

**Maynooth
University**

National University
of Ireland Maynooth

Statistical analysis and mathematical
modelling of lymphocyte population
dynamics

Author:
HoChan Cheon

Supervisor:
Prof. Ken Duffy

*A thesis submitted in fulfilment of the requirements
for the Ph.D. degree in Applied Mathematics*

at the

Hamilton Institute
Maynooth University
Maynooth, Co. Kildare
Ireland

Department head: Prof. Ken Duffy

February 22, 2023

Contents

Summary	v
Acknowledgements	vii
List of Figures	xxv
List of Tables	xxvii
1 Introduction	1
1.1 Abstract	1
1.2 Functional roles of the immune system	1
1.2.1 Adaptive immune system	3
1.2.2 Dysfunction of the immune system	5
1.3 Experiment methods for measuring lymphocyte populations	6
1.4 Models of lymphocyte population dynamics	10
1.5 Thesis outline	13
2 Cyton2: A mathematical model of lymphocyte population dynamics	15
2.1 Abstract	15
2.2 Introduction	16
2.3 Model structure	21
2.4 Analysing data to investigate the model assumptions	22
2.4.1 Time-lapse microscopy of B and T cell families	22
2.4.2 Induced dependency through right censoring of timers	28
2.4.3 Using filming data to determine appropriate distribution classes for the timers	33

2.4.4	Data selection and tree collapse	38
2.4.5	Monte-Carlo simulation	41
2.4.6	Bayesian framework	41
2.5	Equations for the mean population size	44
2.5.1	Optimisation strategy for estimating the model parameters	47
2.6	Equations for the variability around the mean	48
2.7	Modifications for calculating number of cells in generation zero	54
2.7.1	Partial activation, no death	54
2.7.2	Partial activation, death program	55
2.7.3	Slow reprogramming	55
2.8	Discussion	56
3	Cyton2 applications: Model evaluation; Predicting effects of signal addition	59
3.1	Abstract	59
3.2	Precision and accuracy of the model estimates using B cell data	61
3.3	Positive signal integration from multiple stimuli in CD8 ⁺ T cell	65
3.4	Predicting negative signal addition from immunosuppressive drugs	69
3.5	Discussion	74
4	Beyond lymphocyte proliferation and population dynamics: B cell differentiation	76
4.1	Abstract	76
4.2	Exploring B cell differentiation	77
4.3	Simulation of B cell differentiation process and expected data structure	79
4.4	Statistical analysis of clonal dependence of differentiation	82
4.5	Analysis of molecular regulation of Class Switch Recombination	86
4.6	Mathematical modelling	89
4.6.1	IgG1 switching	92
4.6.2	IgG1/IgG2b switching	93
4.7	Model fitting	94
4.8	Discussion	96
5	Cyton2 application to the human system: hyper- and hypo-immune responses	98
5.1	Abstract	98

5.2	Immunodeficiency and autoimmunity	99
5.3	Exploring features of cell population dynamics from CVID and SS patients	100
5.4	Reprogrammed death time and activation fraction	104
5.5	Measuring human B cell health using the Cyton2 model	108
5.5.1	Alternate strategy for estimating Cyton2 division parameters	108
5.5.2	Determining relative contribution of division parameters of aberrant response	111
5.6	Computing likelihood of CVID via Bayesian-like simulation	114
5.7	Discussion	118
6	Conclusion	120
6.1	Future work	123
A	Experiments designed and conducted by collaborators	124
A.1	Chapter 2: Time-lapse microscopy experiment	124
A.1.1	Mice	124
A.1.2	CD8 ⁺ T cell isolation	125
A.1.3	<i>in vitro</i> cell culture	125
A.1.4	Live cell imaging and cell tracking	126
A.2	Chapter 3: Drug experiment	127
A.2.1	Cells	127
A.2.2	Culture conditions and stimulation	127
A.2.3	Drugs	127
A.3	Chapter 4: Differentiation experiment	128
A.3.1	Mice	128
A.3.2	B cell isolation and labelling	128
A.3.3	B cell culture	128
A.3.4	Multiplexed cell labelling, stimulation and sorting	129
A.3.5	Antibody staining and flow cytometry	129
A.3.6	Index sorting of multiplex labelled cells for single-cell qPCR	130
A.3.7	Single-cell quantitative reverse transcription PCR	130
A.3.8	I γ 1-sorting using molecular beacons	130
A.3.9	LCMV infection	132
A.3.10	Statistics	132

CONTENTS

B Supplementary materials for Chapter 4	133
B.1 Figures	133
Bibliography	139

Summary

Lymphocytes, comprising of B and T cells, are important members of the adaptive immune system of vertebrates that play a crucial role in defending against harmful pathogens. They are equipped with receptors capable of recognising specific antigens. After activation, they proliferate to form an exponentially growing clone army. Eventually, those cells cease to divide and then largely die over a period of weeks, but leave a small number of cells, called memory cells, that can rapidly respond to any repeated infection. To study such non-linear population dynamics, experimental systems have been designed that generate data at the level of populations, families and single cells to elucidate underlying mechanisms that regulate expansion, cessation, and contraction of cell numbers.

In this thesis, we report on the development of a novel stochastic model of cellular population dynamics, based on [Hawkins et al. \(2007a\)](#), that accounts for experimentally observed correlation structure within family members. In particular, the inheritance of cell division, cessation, and death times within a stochastic model framework considered, and their impact on cell population dynamics are investigated. Model assumptions are informed by datasets from time-lapse microscopy experiments and statistically tested within the Bayesian framework. Consequences of the dependencies are demonstrated with family trees generated by a Monte-Carlo simulation. To assess the model's ability to extract meaningful inferences from population-level data, we design an optimisation strategy to estimate model parameters and investigate its accuracy and precision for a given dataset from *in vitro* murine system. With the analysis pipeline, the model is applied to both *in vitro* murine and human lymphocyte populations to test hypotheses and draw meaningful biological conclusions. For instance, we demonstrate signal integration for T cells from transgenic mice as a linear sum in a time domain, and as a result,

the model successfully recapitulates the data. Lastly, we extend the remit of the stochastic modelling framework by exploring mechanisms of B cell differentiation to antibody-secreting cells and their class switching to different isotypes. A simple probabilistic model that captures molecular changes within these cells sheds light on the process of determining the types of antibodies to produce and predicting the magnitude associated with them.

Acknowledgements

This thesis would not have been possible without the support and advice of many people. Supervisors, colleagues, collaborators and friends, there are no simple words to show how grateful I am to you all.

First of all, I would like to thank Prof. Ken Duffy for his incredible patience and great guidance. During his supervision, I have learned so much from his remarkable insight and reasoning for solving problems. Often, he broadened my horizons beyond what I could see and influenced me to become a better researcher and person. Also, I would like to thank Prof. Phil Hodgkin for his amazing commitment and determination to science, which, ultimately, has led me to enter the wonder of immunology. Without him, I would not be on this magnificent journey of attempting to make use of physics and mathematics in biology.

I would like to thank my collaborators at Walter and Eliza Hall Institute (WEHI) in Australia. Su, Andrey, Miles, Jessica, Jie, Anthony, David, Melissa and Vanessa, I had a great time working with you all and appreciate your patience in helping me understand biology from the beginning.

I would like to thank past and present members of Quantitative T-cell Immunology and Immunotherapy (QuanTII) for their expertise and for providing a research environment that made it possible to complete this work. I am grateful to all the seniors and fellows that I met during the meetings for giving me the opportunity to share my work and scope other interesting research within the field.

I would like to thank my office mates Alan, Estevão, Gianfelice, Eleni, Alessandra, Maeve, Hannah, Amin and Niloufar for interesting and extremely unusual conversations and discussions that I would not be able to find in other places. A special thank to Giulio for his help in continuing the work. Another special thank to Rosemary and Kate for their massive support behind the scenes to make all arrangements possible throughout the PhD.

CONTENTS

I would like to thank my friends in Singapore, Indonesia, Australia, and South Korea. A special mention goes to my dear friend, Asyraaf Ahmad, for helping me understand biology in general and for going through difficult times together.

Lastly, I would like to thank my mother, my father, and my sisters for their moral support and cheer during times that I found difficult.

List of Figures

1.1	Typical immune response, adapted from Murphy et al. (2008); Punt et al. (2019). When naïve lymphocytes get activated by a specific antigen, they undergo a latency phase approximately for 2 days, then begin dividing to mount a full immune response. A large clone of identical cells is generated during this proliferation phase, which generally replicates two to four times every 24 hours, that significant proportion of cells differentiate into effector cells to eradicate the pathogens. Afterwards, a contraction phase begins during which most of the generated cells die, and leave a small subset of cells that larger in size than the initial number of responding lymphocytes to equip the host with protection against the same antigen in the future. Should a second encounter to occur, the response is faster and larger without the latency phase.	2
1.2	Profiling cell lineage through microscope experiments. (<i>top</i>) [Adapted from Movie S1 and Supplementary Fig.2A of Hawkins et al. (2009)] Bright-field images of B cells in a Terasaki well are taken every 2 minutes for 4-5 days, and each founder cell is manually tracked to ensure accuracy. (<i>mid</i>) [Adapted from Fig.1 of Duffy et al. (2012)] Using B cells from Blimp1 GFP reporter mice to track differentiation to antibody-secreting cells (ASCs). Blimp1 is a transcription factor that is expressed in all ASCs. From this transgenic reporter mice, cells express a GFP when they differentiated into ASCs. (<i>bottom</i>) [Adapted from Movie S1 of Mitchell et al. (2018)] Bright-field images are taken every 3 minutes for 5 days. To generate family tree, phase-contrast images can be analysed with semi-automated software. . .	7

1.3 **Characteristic of the cell staining dyes for measuring division profile.** (*left*) A founder cell in a family is initially stained with fluorescence substance such as CFSE and CTV (Lyons and Parish, 1994). As the cell divides, two daughter cells inherit half the intensity from their mother cell regardless of time at which it divides. (*right*) An example histogram plot of cell population binned based on the levels of fluorescence. Each peak represents the number of cells in particular generation. Grey area is typically a measure of auto-fluorescence in control population, e.g. unstimulated cells, to determine the maximum division number that can be resolved before the cells become indistinguishable from their natural emission of light. 8

1.4 **Example time-series data of lymphocyte population for modelling typical immune response.** (*left*) Total number of cells (*black* solid-line) and generational information (*colour* solid-lines) as a function of time. (*right*) Cell numbers per generation at harvested time points from the experiment. 9

2.1 **Overview of the two Cyton models.** (A) The original Cyton model (Hawkins et al., 2007a) where stochastic times to divide and to die are chosen independently after each cell division. Cells cease their motivation to divide based on division-counting mechanism. (B) The Cyton2 model incorporates significant correlation in division times between siblings, as well as familial inheritance of death and division destiny times. (C) A consequence of the correlation and inheritance is that the resulting family trees are heterogeneous, but highly concordant. By exploiting this property, a family tree can be summarised by substituting the average values of its times and fate at each generation. An example of clonally collapsed family tree and its key variables is shown. 20

2.2 **Example family and clonally collapsed trees of CpG-stimulated B cells in the filming datasets.** Each panel shows one family. A horizontal line represents the lifetime of a cell in dividing (—), dying (—), or lost (—) state. Clonally collapsed tree is shown as a single time-line (—) below the family tree. Time to first division (●), average time to last division (●), average time to death (×) and average subsequent division time ($t_{div}^{k \geq 1}$) are annotated on the collapsed line. These values are shown in the legend. (A) Experiment CpG-stimulated B cells (B-exp1). (B) Repeat of B cell experiment (B-exp2). 22

2.3 **Extracting times to fates from CpG-stimulated B cells (A)** Clonally collapsed family trees of B cells. (B) Rank ordered times to events of families. (C) Correlation coefficient (ρ) estimated using bivariate normal distribution with 95% credible interval is reported for each pair. 90%, 95% and 99% density regions are plotted over the data. For null, $H_0: \rho = 0$, and alternative, $H_1: \rho \neq 0$, hypotheses, Bayes Factors ($BF_{01} = 1/BF_{10}$) were calculated. If the data is more probable under H_0 , then it is BF_{01} times more favoured than H_1 (*blue-scale*), and vice versa (*red-scale*). Distributions of the times are collated into 1 hour time intervals and shown in the diagonal panels. 23

2.4 **Extracting times to fates from CD8⁺ T cells in the presence of 1U, 3U or 10U IL-2.** (A) Clonally collapsed family trees of T cells. (B) Rank ordered times to events of families. (C) Correlation coefficient (ρ) estimated using bivariate normal distribution with 95% credible interval is reported for each pair. 90%, 95% and 99% density regions are plotted over the data. For null, $H_0: \rho = 0$, and alternative, $H_1: \rho \neq 0$, hypotheses, Bayes Factors ($BF_{01} = 1/BF_{10}$) were calculated. If the data is more probable under H_0 , then it is BF_{01} times more favoured than H_1 (*blue-scale*), and vice versa (*red-scale*). Distributions of the times are collated into 1 hour time intervals and shown in the diagonal panels. 24

- 2.5 **Extracting times to fates from the repeat of CpG-stimulated B cells and CD8⁺ T cells in the presence of 1U or 10U IL-2.** For CD8⁺ T cell data, measurements from two independent but identical experiment setup are aggregated. **(A1,B1,C1)** Four key Cyton2 variables, T_{div}^0 , $\{T_{div}^k\}_{k \geq 1}$, T_{ld} (replacing T_{dd} as a proxy measure) and T_{die} for all cells in each family are shown as a series of cascade plots. Average coefficient of variation (CV) for each variable is annotated. The lost cells are not shown. **(A2,B2,C2)** Pair-plot of the Cyton2 variables. Distributions of the times are collated into 1 hour time interval and shown in the diagonal panels. Correlation coefficient (ρ) was estimated using bivariate normal distribution with 95% credible interval for each pair. 90%, 95% and 99% density regions are plotted over the data. Given two hypotheses ($H_0: \rho = 0$ and $H_1: \rho \neq 0$), Bayes Factor ($BF_{01} = 1/BF_{10}$) was calculated. If the data is more probable under H_0 , then it is BF_{01} times more favoured than H_1 (*blue-scale*), otherwise H_1 is BF_{10} (*red-scale*) times more favoured than H_0 . 26
- 2.6 **Extracting times to fates from CD8⁺ T cell costimulation experiments.** Four key Cyton2 variables, T_{div}^0 , $\{T_{div}^k\}_{k \geq 1}$, T_{ld} (replacing T_{dd} as a proxy measure) and T_{die} for all cells in each family are shown as a series of cascade plots. Average coefficient of variation (CV) for each variable is annotated. The lost cells are not shown. **(A1-4)** Experiment T-exp1 consists of N4 only, N4 + α CD28, N4 + IL-2 and N4 + α CD28 + IL-2. **(B1-4)** Experiment T-exp2 consists of N4 only, N4 + α CD28, N4 + IL-2 and N4 + α CD28 + IL-2. . 27

2.7 **Simulation under the independence assumption.** 10^6 Cyton2 families were simulated given fitted lognormal distributions of T_{div}^0 , $T_{div}^{k \geq 1}$, T_{ld} , T_{die} from respective filming datasets. **(A)** Three example families from MC simulation parameterised as CpG-stimulated B cells: dividing (—) and dying (—) states. The realisations of T_{div}^0 (●), $T_{div}^{k \geq 1}$ ($t_{div}^{k \geq 1}$), T_{ld} (●), T_{dd} (★) and T_{die} (×) are annotated on a collapsed line. As a feature of inheritance, the progeny cells double in number synchronously whenever division occurs, and likewise, they reach destiny and death at the same time. **(B)** For all simulated families, each variable was randomly sampled from the fitted Cyton distribution (*inset*), and the samples are labeled as true sample time (●). Their corresponding observable sample times (●) are shown along with the data points (●) from the filming datasets. Distributions of the sampled true and observable times of each variable are shown in the diagonal panels. The observable and unobservable regions are separated by upper and lower sections of $y = x$ line (---), respectively. The Bayes factors are reported for the true and observable pairs. 31

2.8 **Simulation under the independence assumption.** 10^6 Cyton2 families were simulated given fitted lognormal distributions of T_{div}^0 , $T_{div}^{k \geq 1}$, T_{ld} , T_{die} from respective filming datasets. For all simulated families, each variable was randomly sampled from the fitted Cyton distribution (*inset*), and the samples are labeled as true sample time (●). Their corresponding observable sample times (●) are shown along with the data points (●) from the filming datasets. Distributions of the sampled true and observable times of each variable are shown in the diagonal panels. The observable and unobservable regions are separated by upper and lower sections of $y = x$ line (---), respectively. The Bayes factors are reported for the true and observable pairs. **(A)** 1U IL-2. **(B)** 3U IL-2. **(C)** 10U IL-2. 32

2.9	Best parametric distribution class.	(A1) Empirical CDF of the measured times are overlaid with CDFs of Gamma, Lognormal, Normal, Weibull, delayed Exponential and delayed Lognormal distributions. 95% confidence bands are plotted by randomly drawing 10^4 samples from respective posterior distributions. (A2, B1-3) The in-sample deviance (\bullet), and WAIC scores (\circ) with 1 standard deviation error bar are shown. The y-axis is sorted from lowest (<i>top</i>) to highest (<i>bottom</i>) WAIC score. The lower WAIC score indicates better descriptor of the data. Except the top-ranked one, the value of the difference of WAIC (\blacktriangle) between a candidate and the top-ranked are shown with 1 standard error of the difference.	34
2.10	Best parametric distribution classes.	The eCDFs of measurements of times to fates are presented. Six candidate distribution classes are parameterised by: (α_G, β_G) for Gamma; (m, s) median and scale for lognormal; (μ, σ) mean and standard deviation for Normal; (α_W, β_W) for Weibull; (λ, c) rate and shift for delayed Exponential; and, (m_d, s_d, c) median, scale and shift for delayed Lognormal. The CDF of the candidates was plotted by taking mean values of posterior distribution of hyper-parameters. The confidence band was plotted by sampling hyper-parameter values from posterior distribution. (A1-2) Repeat of CpG-stimulated B cell (B-cpg2). Rank-ordered WAIC plot is shown. (B1-3) CD8 ⁺ T cell in the presence of 1U, 3U and 10U of IL-2. For T cell data, measurements from two independent but identical experiment setup are aggregated. Corresponding WAIC plot is shown in Fig. 2.9.	35

2.11 **Best parametric distribution classes.** The eCDFs of measurements of times to fates are presented. Four candidate distribution classes are parameterised by: (α_G, β_G) for Gamma; (m, s) median and scale for lognormal; (μ, σ) mean and standard deviation for Normal; (α_W, β_W) for Weibull; (λ, c) rate and shift for delayed Exponential; and, (m_d, s_d, c) median, scale and shift for delayed Lognormal. The CDF of the candidates was plotted by taking mean values of posterior distribution of hyper-parameters. The confidence band was plotted by sampling hyper-parameter values from posterior distribution. **(A1-4)** Experiment T-exp1 consists of N4 only, N4 + α CD28, N4 + IL2 and N4 + α CD28 + IL-2. **(B1-4)** Experiment T-exp2 consists of N4 only, N4 + α CD28, N4 + IL-12 and N4 + α CD28 + IL-12. Minimum of two observations per variable was required for the calculation. 36

2.12 **Clonally collapsed clones for *all* founder B and CD8⁺ T cells.** For CD8⁺ T cell data with 1U, 3U and 10U of IL-2, measurements from two independent but identical experiment setup are aggregated. The collapsed trees are rank-ordered with respect to the total life time. **(A-F)** Clonally collapsed trees for all families before the filtering. The average division (\bullet) and death (\times) times are marked per family, where the colours represent generation that the event was observed. The lost times are not shown. 39

2.13 **Variability of the reduced Cyton2 model.** **(A)** Example Cyton2 parameter estimates from fitting B cell data presented in Fig. 3.1A3. Initial number of cells (N_0), maximum generation (G) and subsequent division time (m) are annotated as a plot title. Probability density and cumulative distribution functions are shown for T_{div}^0, T_{dd} and T_{die} . **(B)** Predicted population size of 0th (*left*) and 1st through 8th generations (*right*). Each solid-line represents average number of cells. Shade area represents upper and lower bounds around the mean ($\mu \pm 1.96\sigma$). 49

2.14 **Total number of cells vs time and analysis of the evolution of the variance.** (A) Predicted population size. Solid-line represents the sum of average cell numbers through all generations. Shade area represents upper and lower bounds around the mean ($\mu \pm 1.96\sigma$). (B) Coefficient of variation as a function of time. The onsets (more than 1% of population) and ends (more than 99% of population) of division and destiny events are marked as solid- and dashed-lines. 51

2.15 **MC simulation of Cyton1 and Cyton2 processes.** (A) Average population size (solid line) and 95% confidence band (shade area). (B) Coefficient of variations. 53

3.1 **The precision of the parameter estimates with CpG-stimulated $Bim^{-/-}$ murine B cell FACS data.** (A1-3) The best-fit reduced Cyton2 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 1000 synthetic datasets, which were created by randomly sampling the original data with replacement per time point. (B) From the sets of estimated parameter vectors, bi-plot of principle component analysis (PCA) result is shown. (C) The marginal coefficients of variation (CV) was calculated with 95% confidence interval from bootstrapping. 60

3.2 **The accuracy of the model fit with CpG-stimulated $Bim^{-/-}$ B cell FACS data.** The root-mean-square error (RMSE) was evaluated over all available data points after fitting the reduced Cyton2 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. 62

3.3	The accuracy of the model fit with CpG-stimulated $Bim^{-/-}$ B cell FACS data.	The root-mean-squared error (RMSE) was evaluated over all available data points after fitting the reduced Cyton2 model to the synthetic datasets. Similarly, the reference RMSE (<i>purple</i>) was obtained after fitting the model to datasets assuming only three replicates are available while maintaining all time points. (A) All possible combinations of positions of time points for $k = 4$ case. The best (<i>blue</i>) and worst (<i>red</i>) examples of Cyton fits are shown. (B) Best examples for $k = 5, 6$ are shown. The worst cases failed to provide good fits (not shown), resulting in large confidence bands and an order of magnitude difference in RMSE. .	64
3.4	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 (\bullet : mean \pm SEM) overlaid with the model extrapolation and 95% confidence band from bootstrapping. (B1-3) 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, α_{CD27} and α_{CD28} ; 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.	66

3.5 **Linear sum of the signals from simulated trees (correspond to Fig.3 in Marchingo et al. (2014)).** For each harvested time points from FACS data, percentage cohort number (with respect to the first time point) as a function of MDN is shown (●: mean ± SEM). The Agent-Based Model was used to generate family trees. The times to first, to destiny and to death were randomly sampled from the estimated *normal* distributions by fitting the reduced Cyton2 model to N4, αCD27 and αCD28 datasets simultaneously. Note that the subsequent division time was set as a constant to be consistent with the model. To match the data, each round of the simulation was initialised with 6066 (N4), 8252 (αCD27) and 8377 (αCD28) clones and ran for $t \in [0, 140]$ with $\Delta t = 0.5$ in hours. This process was repeated 1000 times to obtain 95% confidence bands around the mean. Increase in MDN is labelled in each panel with arrows. The predicted MDNs for αCD27+αCD28 were calculated either by summing the contribution from each individual stimulation (---) or simulating the trees directly with the summed timers (—). 68

3.6 **FACS data and summary statistics of OT-I/Bim^{-/-} CD8⁺ T cells in the presence of Rapamycin, Dexamethasone or both.** (*top-row*) Dose-dependent reduction in total cell numbers are shown for Rapamycin alone, Dexamethasone alone, two fixed Rapamycin concentrations as bases for varying Dexamethasone. (*mid-row*) Summary statistics of cell survival by cohort numbers, i.e. removal of doubling effect. While Dexamethasone shows earlier cell death as dosage increases, Rapamycin has little to no effect. (*bot-row*) Summary statistics of cell proliferation by mean division number, i.e. average number of division rounds. CD8⁺ T cells undergo similar amount of division rounds irrespective of Dexamethasone concentration, whereas significant reduction in the division counts for higher Rapamycin concentration. Consequently, both drugs suppress cell expansion through reduction in the survival or proliferation, and they achieve so without affecting one another. 70

3.7 **Parameter estimates of Cyton2 model from OT-I/Bim^{-/-} CD8⁺ T cells in the presence of Rapamycin, Dexamethasone or both.** (A) Estimated Log-normal distributions of time to first division, time to division destiny and time to death. The same colour represents an equal concentration of Rapamycin. Different line styles were used to indicate specific concentration of Dexamethasone. Shaded area around each line represent 95% confidence band. (B) Estimated subsequent division time. The same colour scheme was adopted for consistency. A vertical dashed-line represents the mean time for a given Rapamycin concentration. 71

3.8 **Predicting cell population size in combination of Rapamycin and Dexamethasone.** Each panel shows the effect of one or combined drug in cell population size. Top-left panel is the control i.e. without the drugs. The best-fit model (*solid-line*) and predicted cell numbers (*dashed-line*) are shown. Shade area around the curves represent 95% confidence band. 73

4.1 **MC simulation of cell-type change and expected data structure.** Members of a family are simulated with two distinct mechanisms of cell fate: (A1) Autonomous determination of cell fate for each cell; and, (A2) Clonally dependent fate with positive correlation coefficient, $r \in [0, 1]$. Once cells change their type with generation-dependent probabilities, $\{p_g : g = 0, 1, \dots, G\}$ (Hasbold et al., 2004), their offspring inherit the state. For each generation, the events of cell-type change are simulated with binomial and beta-binomial distributions for autonomous and clonally dependent cases, respectively, given number of unchanged cells (n_g). Variable, v_g , denotes number of type-changed cells. (B) The expected number and standard deviation of type-changed cells calculated from 10^7 families. (C) Simulated with 200 founder cells with randomly assigned division times. Offspring that reached their fates are randomly assigned expression levels from a single lognormal or multivariate lognormal distribution, parametrised by median and log-variance, (m, s) , and mean vector and covariance matrix, $\boldsymbol{\mu} \in \mathbb{R}^{v_g}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{v_g \times v_g}$, respectively. . . . 80

4.2 **Flow cytometry-based multiplex division tracking assay and experiment measurements of Blimp and IgG1 expression levels.** (A) Lineage tracing of activated B cells for ASC and CSR by multiplexed cell labelling. (B) Distributions of the expression levels of Blimp (*red*) and IgG1 (*blue*). Cascade plots of rank-ordered Blimp and IgG1 expression levels. Generation information is colour coded. (C) All possible combinations of pairs of the expression levels for cells in the same generation and family. 83

4.3 **Testing null hypotheses of clonal dependence.** Permutation test results of the Blimp (*left*) and IgG1 (*right*) data presented in Fig. 4.2B. The vertical dashed-line refers to the data statistics of the original data. The distribution shows the density of the \log_{10} of the average clonal variance computed for 500,000 permutations of the data. The lower one-sided p-values are shown in legend for both tests described in the main text. These values indicate rejection of the hypotheses with a significance level of 0.05. 86

4.4 **B cell antibody class switching is characterised by division-linked AID expression and division-independent I γ 1 transcription.** [Corresponding to Figures 3 & 4 from Horton et al. (2022)] (A) Experiment measurements of proportions of B cells expressing AID (triplicate) and I γ 1. (B) Proposed molecular regulations of AID and I γ 1 at single-cell level. Once a cell expresses AID, its offspring inherit the state of mother cell and stay AID⁺. In contrast, each cell autonomously determines I γ 1, its state is re-randomised at each generation. (C) experiment measurements (Mean \pm SEM) and model prediction of proportion of IgG1⁺ cells. Percentage switching efficiency of IgG1 given both AID and I γ 1 expressions. 87

4.5 **Model of multiple antibody class switching in the presence of intracellular competition between GLTs.** [Corresponding to Figure 5 from Horton et al. (2022)] (A) Measured proportions of AID⁺, I γ 1⁺ and I γ 2b⁺ cells for varying concentrations of IL-4. (B) Model predictions of joint probability of observing IgG1⁺ and IgG2b⁺ cells that are in generation g (c.f. Eq. 4.8) given empirical estimates of hazard probabilities of AID and of constant probabilities GLTs per IL-4. Shade area represents 95% confidence band. 90

5.1	<p>FACS data and summary statistics of human B cells obtained from healthy donors, CVID and SS patients. Cells were stimulated with combined anti-Ig engagement of the B cell receptor and CpG DNA to induce T cell independent response. (<i>top-row</i>) Total cell numbers, normalised to the initial cell number. Summary statistics of the responses are given by (<i>mid-row</i>) normalised total cohort number for cell survival and (<i>bot-row</i>) mean division number for cell proliferation.</p>	101
5.2	<p>Permutation tests on the measurements and summary statistics between HD (<i>blue</i>), CVID (<i>red</i>) and SS (<i>green</i>). For time points at 72h, 96h and 120h, the tests were performed in two sets: HD vs CVID and HD vs SS on total cell numbers (<i>top-row</i>), total cohort numbers (<i>mid-row</i>) and mean division numbers (<i>bot-row</i>). On each panel, data points are shown in a scatter plot, and differences of the means are shown below it. Solid horizontal line indicates zero difference. The error bar is a 95% confidence interval estimated using bootstrap method, and its distribution is shown as well. The one-sided p values from the permutation tests are annotated with corresponding colours. 10^7 permutations and bootstrap samples were used for the calculation.</p>	103
5.3	<p>Initial cell death program and transition to the new reprogrammed death timer. Activation of human naïve B cells take approximately 25 hours. The time-series total cohort number and the best-fit model for (A) unstimulated cells (<i>orange</i>) and (B) CpG (<i>red</i>), anti-Ig (<i>green</i>) and CpG + anti-Ig (<i>blue</i>) stimulated cells. Solid-line and shade-area represent the mean and 95% confidence band, respectively. Vertical dash-line is the activation time. The parameter estimates are shown in the legend and bottom-left of the plot panel.</p>	107

5.4 **Estimated division parameters of the Cyton2 model from HD, CVID and SS datasets.** (A) Representative best-fit Cyton2 model for one donor from each group: *left*-HD, *mid*-CVID and *right*-SS. (B) Cumulative distribution function (CDF) of the estimated RVs that represent time to first division (T_{div}^0), time to division destiny (T_{dd}) and time to death (T_{die}). Solid, dashed and dash-dotted lines indicate mean CDF curves computed across HD, CVID and SS, respectively. Shade-area is lower and upper 95% percentile of the CDFs, calculated from collection of individual donors per group. (C) Summary of all Cyton2 parameters. Each dot in scatter plot represents one donor, and the error bar is 95% confidence interval from bootstrapping. Marginal distributions of subsequent division time (m) and activation fraction (f_A) are shown. Dashed line represents a mean value. 110

5.5 **Using the estimated Cyton2 parameters to compute relative contributions compare to the HD average.** (A) An example calculation at each stage for a CVID patient. Black solid line and dashed line indicate the HD average and original best-fit model, respectively. Dash-dotted line shows change in the response at each stage, where the colour represents a substituted parameter. Stacked bar plot shows percentage contribution using total cell numbers at t^* relative to that of the HD average. (B) Relative contributions for all CVID and SS individuals. 113

5.6 **Marginal distribution of key fitted Cyton2 parameters and parameteric distribution classes that describe each of the components for simulation.** (A) Bayesian correlation analysis for independence of the parameters between HD individuals ($N = 34$). Bayesian Factor (BF) and correlation coefficient with 95% credible interval are annotated in each panel (adopted method described in Section 2.4.6). Shade areas represent 90, 95 and 99% density regions. (B) Fitted various parametric distribution classes for each component. Truncated Normal ($t\mathcal{N}(a = 0, b, \mu, \sigma)$), Lognormal ($\mathcal{LN}(m, s)$), Normal ($\mathcal{N}(\mu, \sigma)$) and Gamma (Gamma(α_G, β_G)) distributions were used for $m_{div}^0, m_{dd}, m_{die}$ and m . By definition $f_A \in [0, 1]$, so Truncated Normal ($t\mathcal{N}(a = 0, b = 1, \mu, \sigma)$) and Beta (Beta(α_B, β_B)) distributions were tested. The best descriptor was determined by the lowest AIC value marked with an asterisk (*). 115

5.7 **Simulation of HD with parameter sampling and likelihood of measured CVID.** (A1) Total cell numbers and mean division number (MDN) of 500,000 simulated Cyton2 model overlaid with CVID measurements. Each coloured solid line indicates an individual donor. Dashed-line represents average HD by sampling $m_{div}^0, m_{dd}, m_{die}, m$ and f_A parameters. Shade-area is 95% confidence band. (A2) Empirical CDF (eCDF) plot of the normalised total cell numbers for HD and CVID and the simulation at 96h (*top-row*) and 120h (*bot-row*). Location of CVID data points are shown on the eCDF of the simulated total cells. (B) The likelihood of lowest five CVID total cell number measurements given the simulation at 96h for all combination of the parameter components that were sampled from the fitted distributions. 116

B.1 Clonal analysis of B cell fates by multiplexed fluorescent labelling. [Corresponding to Supplementary Figure S1 from Horton et al. (2022) and appear in Section 4.4] (A) An example of the flow cytometry gating layout used to identify clonal populations according to multiplexed labelling status. (B-D) Data relating to Fig. 4.2, showing distributions of the intraclonal correlations calculated using Spearman’s for the indicated cell fate following repeated sampling of cell pairs either from within clones, as experimentally measured, or following randomisation. 134

B.2 Emergence of class-switched B cells is characterised by clonal heterogeneity in division-linked AID expression and division-independent I γ 1 transcription. [Corresponding to Supplementary Figure 3 from Horton et al. (2022) and appear in Section 4.5] Lineage tracing of individual B cell clones analysed for (A) AID expression or (C) I γ 1 transcription. Repeated sampling of cell pairs from within clones or following randomisation for (B) AID expression and (D) I γ 1 transcription. 135

B.3 Clonal and single-cell analysis of hCD2/AID expression and I γ 1 germline transcription. [Corresponding to Supplementary Figure S3 from Horton et al. (2022) and appear in Section 4.5] Data as in Supplementary Fig. B.2, clonal analysis of AID and I γ 1 expression. Distributions of intraclonal correlations by Spearman’s for (A) AID and (B) I γ 1 from measured data or following randomisation. (C) hCD2/AID reporter B cells were activated with LPS and IL-4 for 72 hours prior to staining for surface hCD2, single cell index sorting into a 384-well plate and analysis of I γ 1 expression by sc-qPCR. Per-cell levels of both hCD2/AID and I γ 1 were determined by relating single-cell qPCR results to index sort FACS data. Expression levels are indicated by FACS channel values for hCD2/AID and relative quantity (RQ) for I γ 1. Rho value represents Spearman’s rank correlation coefficient. (D) Single-cell analysis of I γ 1 expression in B cells activated with LPS and IL-4 for 72 hours prior to staining for surface IgG1 and index sorting into a 384-well for qPCR. I γ 1 expression is displayed based on surface expression of IgG1. Representative of 2 independent experiments. 136

B.4 Pre-existing fate potential is identifiable in naïve B cells for ASC differentiation, but not CSR, and is linked to receptor sensitivity. [Corresponding to Figure 2 from Horton et al. (2022) and appear in Section 4.8] **(A)** Sorting of naïve Blimp^{gfp/+} B cells by expression of LPS receptors RP105 and TLR4/MD2, which were subsequently stimulated with LPS for 3 days and **(B)** analysed for differentiation. **(C)** Sort gates for 2 populations of naïve wild-type B cells by IL-4R α expression, which were subsequently stimulated with LPS and varying concentrations of IL-4 for 3 days and **(D)** analysed for CSR to IgG1. Representative of 2 independent experiments. Mean \pm standard error. Statistical analysis using the Holm-Sidak’s multiple comparison test. * $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$ 137

B.5 Diverse clonal mechanisms for controlling fate heterogeneity in B cells. [Corresponding to Figure 7 from Horton et al. (2022) and appear in Section 4.8] **(A)** Individual naïve cells vary in construction, including variation in receptor expression, to generate a population with diverse potential for ASC differentiation. **(B)** ASC fate selection and timing are conserved within clones. **(C)** Isotype switching utilises cell autonomous probability modulation that leads to variegated clones. **(D)** External cytokines determine the likelihood of binary GLT expression in a division-independent manner. This intersects with a separate probability for AID expression to regulate the switching fate decision. 138

List of Tables

2.1	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} (<i>blue-scale</i>) times more favoured than H_1 , otherwise H_1 is BF_{10} (<i>red-scale</i>) times more favoured than H_0 . The variable pairs that marked with $N < 2$ or NA indicate that the analysis could not be performed, for it requires minimum of two clones to be observed in the sample.	29
2.2	Bayesian independence test of the times to fates from simulation. The test was performed with $N_{\text{True}} = 10^6$ 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 ($N_{\text{Obs.}}$) 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 (<i>blue-scale</i>), otherwise it is indicated by BF_{10} (<i>red-scale</i>).	33

<p>2.3 Number of cells observed for division, death and lost events in each generation. For each condition of CD8⁺ T cell with 1U, 3U and 10U IL-2, measurements from two independent but identical experiment setup are pooled. The total number of cells in each generation (i.e. divided + dead + lost cells) is equal to twice of divided cells from previous generation. A cell is considered lost if it becomes indistinguishable to the nearby cells, or survives until the end of given experiment time frame.</p>	38
<p>2.4 Number of clones used in the analysis of the filming 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.</p>	40
<p>2.5 Bayes factor interpretation (Jeffreys, 1961).</p>	42
<p>2.6 List of candidate parametric distribution classes.</p>	43
<p>2.7 Cyton1 and Cyton2 parameter settings for MC simulation. The asterisk denotes the parameter values that were set and fixed from the Cyton2 estimates. Other Cyton1 parameters were obtained by fitting the model to the same dataset presented in Fig. 2.13. For Cyton1 simulation, (T_{div}^0, T_{die}^0) denotes times to <i>first</i> division and death; $(T_{div}^{g>0}, T_{die}^{g>0})$ denotes times to <i>subsequent</i> division and death; and, γ_g is progressor fraction, c.f Eq. 1.1. The times are lognormally distributed with corresponding median and shape parameters.</p>	52
<p>A.1 Oligo sequences</p>	131

1

Introduction

1.1 Abstract

In the following sections, we outline the background of immunology that motivates our work. In particular, we introduce the overall function of the immune system, then narrow it down to specific cell types called lymphocytes (B and T cells) as these cells will be the main topic of this thesis. Lymphocytes are equipped with protective mechanisms that are capable of eliminating foreign pathogens that cause damage to host body, generally in the form of an infection. One of the important lymphocytes' behaviours is creating a large clone of the same type that recognises a specific foreign substance, after which a majority of the population becomes effector cells to actively neutralise the invaders and leave a small proportion of memory cells for repeated infections in the future. To investigate this phenomenon, researchers performed various experiments and derived mathematical models that aim to elucidate the underlying mechanism of cell population dynamics. From the literature, we will report recent findings from experimental methods that were conducted at single-cell and population-level.

1.2 Functional roles of the immune system

The immune system is a protective mechanism that constitutes a wide variety of immune cells to defend against harmful microbes or to prevent growth of malignant cells. Amongst various types of the immune cells in our body, lymphocytes are the essential cells that are adaptively protecting the host from harmful substances such as pathogens (e.g. viruses and bacteria) while recognise the host to avoid

1.2. FUNCTIONAL ROLES OF THE IMMUNE SYSTEM

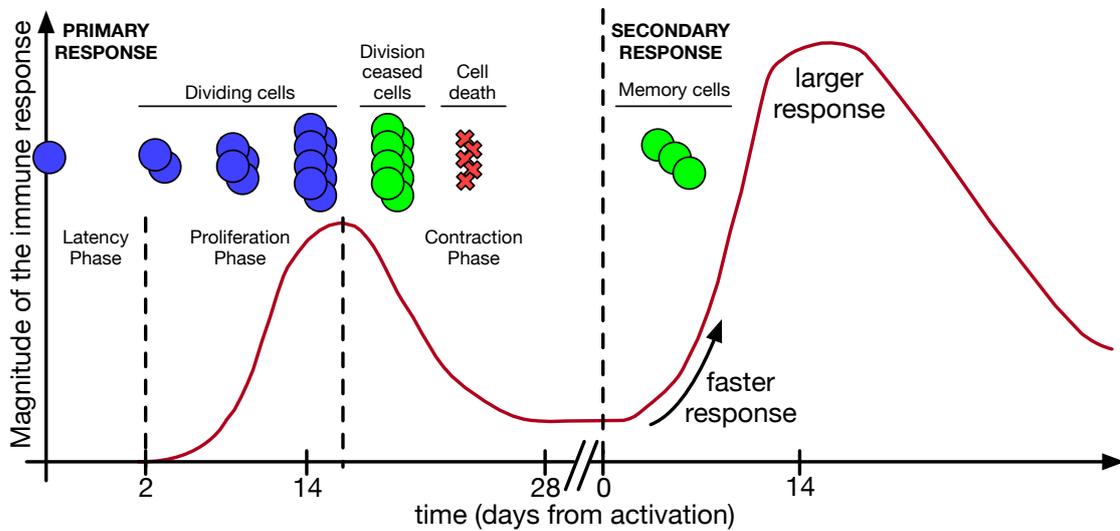


Fig. 1.1: Typical immune response, adapted from [Murphy et al. \(2008\)](#); [Punt et al. \(2019\)](#). When naïve lymphocytes get activated by a specific antigen, they undergo a latency phase approximately for 2 days, then begin dividing to mount a full immune response. A large clone of identical cells is generated during this proliferation phase, which generally replicates two to four times every 24 hours, that significant proportion of cells differentiate into effector cells to eradicate the pathogens. Afterwards, a contraction phase begins during which most of the generated cells die, and leave a small subset of cells that larger in size than the initial number of responding lymphocytes to equip the host with protection against the same antigen in the future. Should a second encounter to occur, the response is faster and larger without the latency phase.

damaging nearby friendly tissue as a form of self-tolerance. These cells are developed in primary lymphoid organs, which includes the bone marrow and the thymus, and mount an immune response to fight off the intruders in secondary lymphoid organs such as the lymph nodes and spleen.

The *innate immune system*, which can be found in all multicellular plants and animals, acts as the first line of defence against pathogens. In our body, this system include not only the epidermis to form a physical barrier between the exterior to the interior, but also white bloods cells such as macrophages, Natural Killer (NK) cells, dendritic cells to actively remove the pathogens by phagocytosis or cytotoxic activity ([Murphy et al., 2008](#)). Most of the immune cells are capable of recognising pathogen-associated molecular patterns (PAMPs) such as *lipopolysaccharide* (LPS) on Gram-negative bacteria (see [Alberts et al., 2002](#), Chapter 25). These components are commonly detected in innate cells through specific pattern recognition receptors

1.2. FUNCTIONAL ROLES OF THE IMMUNE SYSTEM

(PRRs). As PAMPs are not present in healthy bodies, the specificity of PRRs establish extremely well defined self vs non-self discrimination. When these cells recognise PAMPs, they become activated and mediate a rapid response within minutes or hours even if the host has never encountered particular pathogen previously (Murphy et al., 2008). Despite the effective response for eliminating common pathogens, the innate immune cells have limited number of conserved, germline-coded PRRs so that they cannot recognise different variants of molecular structure of microbial or non-microbial molecules. In such cases, the pathogens may have the chance to evade the system, thus, a more flexible system has evolved. As it can detect and adapt to seemingly any foreign material this additional protection mechanism is named the adaptive immune system.

1.2.1 Adaptive immune system

This system is developed in all jawed vertebrates (gnathostomata) as a result of evolutionary process (Cooper and Alder, 2006). Key characteristics of the adaptive immune system that distinguish from the innate system are (i) the ability to prepare and enhance the response for a reoccurring threat (see Secondary Response in Fig. 1.1), and (ii) each of the adaptive immune cells is likely to be specific to different threats.

Two main actors that orchestrate an adaptive immune response are lymphocytes, which constitute B and T cells (NK cell is categorised as a lymphocyte but belong to the innate immune system). These cells are equipped with membrane-bound randomised receptors called B cell receptor (BCR) and T cell receptor (TCR), respectively, that only bind to small repertoire of antigen through lock-and-key mechanism, i.e. specific to the reciprocal shape of each receptor. The strength of this binding depends on its affinity, and stronger it binds, likely to induce a greater immune response from these antigen-specific B and T cells. An important study from Tonegawa (1983) reported stochastic generation of the receptors from somatic mutation called V(D)J recombination, which explains how variety of BCRs and TCRs are achieved in a lymphocyte population.

Pathways of lymphocyte activation can be characterised either directly by contacting antigens presented on the surface of a pathogen or secreted in the blood plasma from harmful cells, or indirectly by help of antigen-presenting cells (APCs). The APCs, such as macrophages or dendritic cells, present antigens sampled from

1.2. FUNCTIONAL ROLES OF THE IMMUNE SYSTEM

parts of a pathogen via phagocytosis. A lymphocyte that recognises its specific antigen and receives the appropriate signals will become activated and undergo rapid cell division to form a large clone of cells with identical specificity. If the cells have never been exposed to the antigen previously, that is in “naïve” state, they undergo a latency phase for approximately 2 days before initiating the extensive mitotic division (see Primary Response [Fig. 1.1](#)). The idea that only certain types of B and T cells are selected in a pre-existing group of lymphocytes, which then produce identical clones, was first postulated by Frank M. Burnet in 1957 and known as *clonal selection theory* ([Burnet, 1957](#)). The theory was revised over the years and is currently a well accepted paradigm of adaptive immunity ([Miller, 1963](#); [Bretscher and Cohn, 1970](#); [Lafferty et al., 1980](#); [Mueller et al., 1989](#); [Janeway, 1989](#); [Matzinger, 1994](#); [Baxter and Hodgkin, 2002](#); [Hodgkin et al., 2007](#)).

A large proportion of clonally expanded B and T cells provides humoral and cell-mediate immunity, respectively, in their effector state ([Punt et al., 2019](#), Chapter 1). Effector B cells secrete antibodies (immunoglobulins) that attach to their cognate antigen to mark a cell for improving recognition and death by phagocytic cells, hindering the cell from binding healthy host cells or biologically inactivating the pathogens ([Alberts et al., 2002](#), Chapter 24). T cells can be further distinguished into two sub-types known as CD4⁺ and CD8⁺ T cell depending on the presence of cell-surface molecule (cluster of differentiation or CD) at the end of its development in the thymus ([Murphy et al., 2008](#), Chapter 8). CD4⁺ and CD8⁺ T cell are also commonly known as *helper* and *killer* T cell, for their functional roles in the immunity. While CD8⁺ T cells are responsible for directly eliminating abnormal cells by secreting serine protease (granzyme B) along with perforin to trigger apoptosis of the targeted cells, CD4⁺ T cells release cell-signalling protein called cytokines to modulate the humoral and cell-mediated responses, or to control the growth of certain cell population (e.g. interleukin-2 or IL-2 for growth of T cell population), or to enhance/inhibit actions carried by other cell types (e.g. promoting cytotoxicity of CD8⁺ T cells).

Once the pathogen is successfully suppressed, the majority of the clonally expanded lymphocytes begin to die, and leave a small subset of long-lived cells known as memory cells that are larger in population size than the initial number of responding cells. These cells preserve the same receptor for a long period of time in case of reappearance of the antigen. Should the body experience a subsequent encounter, the cells generate larger clones in numbers more rapidly than their naïve progenitors,

1.2. FUNCTIONAL ROLES OF THE IMMUNE SYSTEM

thereby providing the host more effective protection against the same antigen (Fig. 1.1). This long-term protection is a fundamental feature of the adaptive immune system and main principle of how vaccines work.

1.2.2 Dysfunction of the immune system

A successful immune response is carefully modulated process that could potentially fail at multiple points. There are three major classes of reactions related to dysfunction of the immune system.

Hypersensitivity. A failure of regulating the immune response and overreaction to an antigen that normally would be harmless to the host body. This results in an allergic reaction (Galli et al., 2008).

Autoimmunity. This occurs when lymphocytes fail to distinguish healthy self from non-self, resulting in attacking ones own body. Studies in human and animal models reveal that genetic and multiple environmental factors contribute to autoimmunity. However, the exact causes are generally unknown as it manifests differently organ to organ: some are limited to particular tissue and others are more system-wide (Rosenblum et al., 2015). Examples of autoimmune disorders are type 1 diabetes, coeliac disease, multiple sclerosis and Sjögren syndrome (Romão et al., 2018; Caio et al., 2019).

Immunodeficiency. This refers to a failure of mounting an appropriate immune response and responding generally in much weaker form compared to healthy individuals. There are two distinct classes of immunodeficiency depending on its cause: primary and secondary (Durandy et al., 2013; Sánchez-Ramón et al., 2019). A primary immunodeficiency (PID) is caused by intrinsic factors such as genetic inheritance or developmental defects. Conversely, a secondary (or acquired) immunodeficiency is caused by external factors such as HIV infection or malnutrition. A famous case of PID is a young boy named David P. Vetter (1971-1983) who had severe combined immunodeficiency. Mild exposure of typically innocuous pathogen could be fatal for him, so he was also known as the “bubble boy” by the media, for his intricate containment system that protected him from immediate surroundings (Tommasini et al., 2018).

From this complex mechanism of the adaptive immune system, we will mainly consider proliferation dynamics of B and T cells in this thesis. Although full

1.3. EXPERIMENT METHODS FOR MEASURING LYMPHOCYTE POPULATIONS

understanding of an immune response requires an inclusion of intricate network of interactions between various types of immune cells, the proliferation dynamics provides important clues in illuminating inner workings of the response. In the next sections, we outline experiment techniques to measure the kinetics at population and single-cell levels, which will be central to our development of a new mathematical model and to test it subsequently.

1.3 Experiment methods for measuring lymphocyte populations

Various experimental techniques have been developed that allow quantitative measurement of the number of cells and division rounds in a population. To have a greater control on the variables of the experiment, these techniques have been developed for *in vitro* studies, i.e. a controlled environment such as a cell culture. Such system minimises unpredictable interactions between other types of cells or environments so that outcomes can be reproducible under the same condition. The other type of study is performed in *in vivo* setting, where cells are allowed to grow within a living organism to which they naturally belong. Here, the main complexity arise as it becomes significantly difficult to keep track of all possible factors that undergo during the growth, thus, the experiment becomes relatively more sensitive to the exact settings at the time of the study.

As our main focus in this thesis is identifying the fundamental mechanism that drives the proliferation of cells rather than results from the complex interactive nature, we will mainly analyse datasets obtained from *in vitro* system unless mentioned otherwise. In the following sections, we will introduce descriptions of frequently used experiment methods in the literature.

Time-lapse microscopy. One popular approach is taking images of a cell sequentially for a fixed period of time *in vitro* setting (Hawkins et al., 2009; Downey et al., 2011; Duffy et al., 2012; Dowling et al., 2014; Chakravorty et al., 2014; Kinjyo et al., 2015; Shokhirev et al., 2015; Mitchell et al., 2018; Zhou et al., 2018; Cornwell and Nordon, 2019). Several techniques have been developed to help visualise various features of lymphocytes behave in these cultures. For example, Sakaue-Sawano et al. (2008) and his colleagues developed a technique called Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) in order to observe cells in a different cell cycle

1.3. EXPERIMENT METHODS FOR MEASURING LYMPHOCYTE POPULATIONS

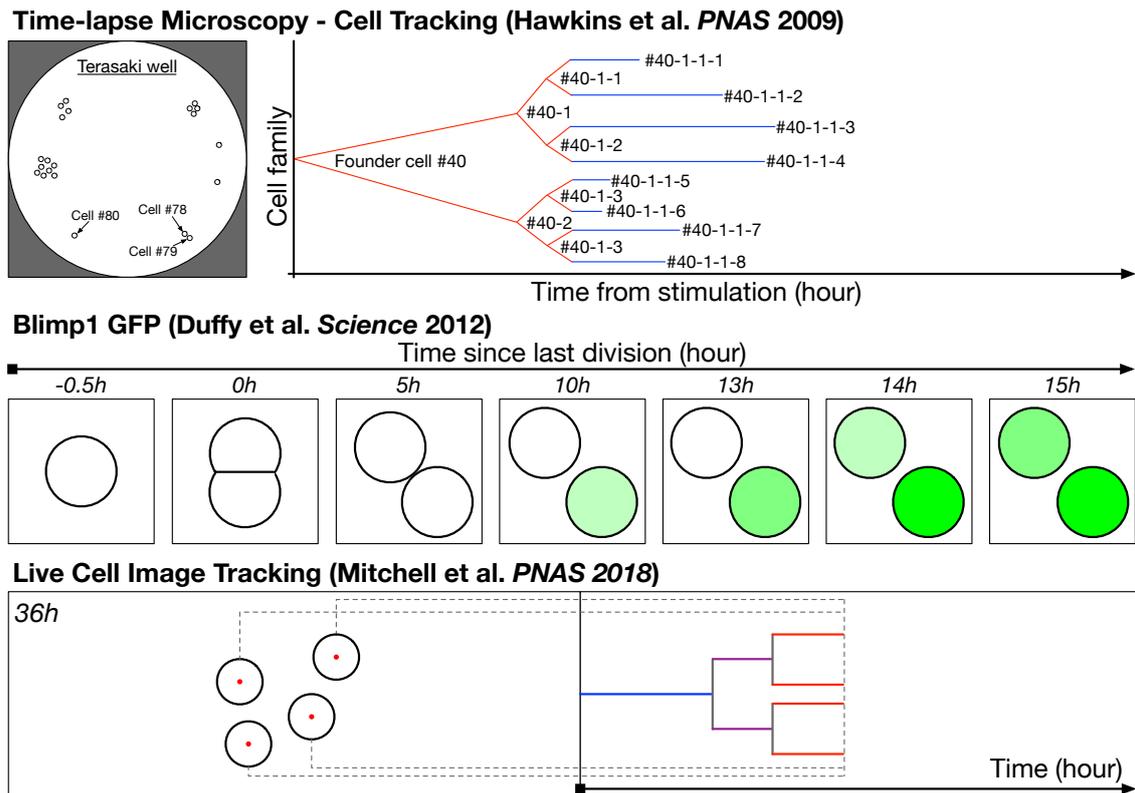


Fig. 1.2: Profiling cell lineage through microscope experiments.

(*top*) [Adapted from Movie S1 and Supplementary Fig.2A of Hawkins et al. (2009)] Bright-field images of B cells in a Terasaki well are taken every 2 minutes for 4-5 days, and each founder cell is manually tracked to ensure accuracy.

(*mid*) [Adapted from Fig.1 of Duffy et al. (2012)] Using B cells from Blimp1 GFP reporter mice to track differentiation to antibody-secreting cells (ASCs). Blimp1 is a transcription factor that is expressed in all ASCs. From this transgenic reporter mice, cells express a GFP when they differentiated into ASCs.

(*bottom*) [Adapted from Movie S1 of Mitchell et al. (2018)] Bright-field images are taken every 3 minutes for 5 days. To generate family tree, phase-contrast images can be analysed with semi-automated software.

1.3. EXPERIMENT METHODS FOR MEASURING LYMPHOCYTE POPULATIONS

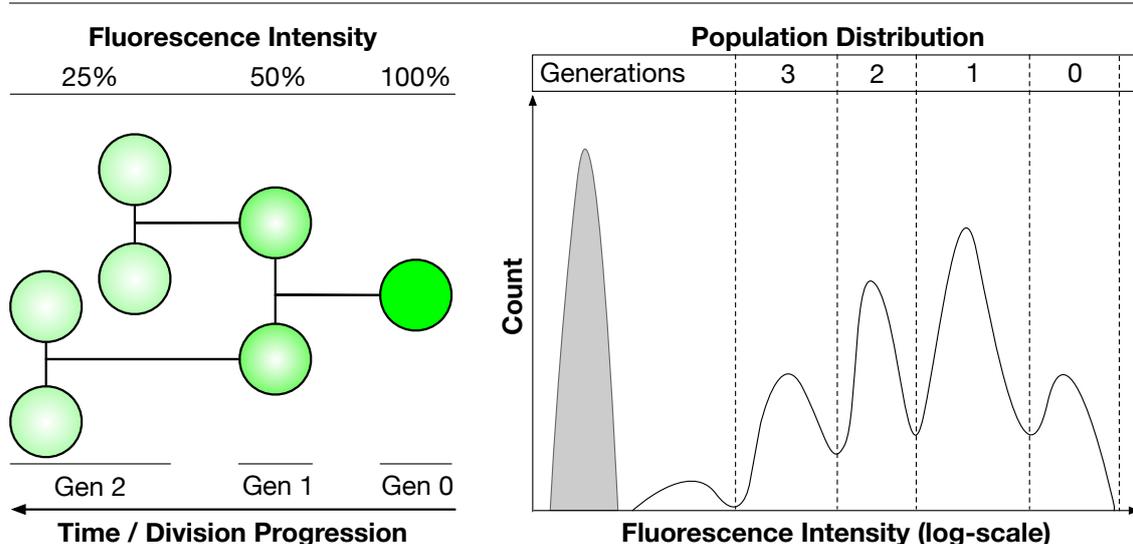


Fig. 1.3: Characteristic of the cell staining dyes for measuring division profile. (*left*) A founder cell in a family is initially stained with fluorescence substance such as CFSE and CTV (Lyons and Parish, 1994). As the cell divides, two daughter cells inherit half the intensity from their mother cell regardless of time at which it divides. (*right*) An example histogram plot of cell population binned based on the levels of fluorescence. Each peak represents the number of cells in particular generation. Grey area is typically a measure of auto-fluorescence in control population, e.g. unstimulated cells, to determine the maximum division number that can be resolved before the cells become indistinguishable from their natural emission of light.

phase. In this system, cells express a Green Fluorescent Protein (GFP; discovered in 1962 by Shimomura et al. (1962)) during S/G₂/M phases, while a red fluorescent protein is expressed during G₁ phase in the cell cycle. Hence, the exact timing of cell division and death can be accurately tracked, e.g. Dowling et al. (2014). Other examples, such as studies in Duffy et al. (2012); Zhou et al. (2018), used B Lymphocyte-Induced Maturation Protein 1 (Blimp1) GFP reporter mice to observe B cell differentiation to antigen-secreting cells (ASCs). Blimp1 is a transcription factor known to be critical for ASC differentiation (Kallies et al., 2004). The cells can be identified as they express GFP during the development (see *mid* Fig. 1.2). Fucci or reporter mice are particularly useful for measuring times at which cells change their state. However, it is also possible to directly observe without any fluorescent property exerted in the system (Hawkins et al., 2009; Mitchell et al., 2018) (see *top & bottom* Fig. 1.2). The images taken from these methods can be stacked to reconstruct entire cellular lineages. Time-lapse microscopy has advantages,

1.3. EXPERIMENT METHODS FOR MEASURING LYMPHOCYTE POPULATIONS

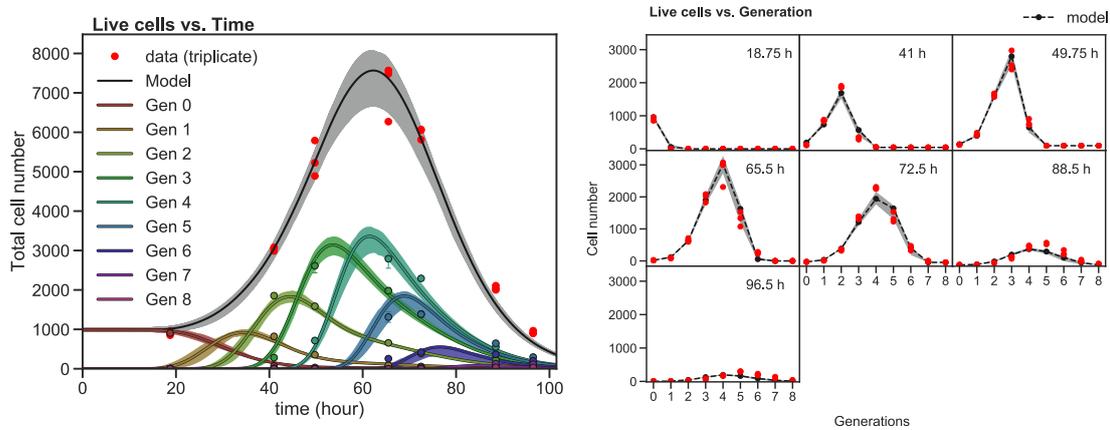


Fig. 1.4: Example time-series data of lymphocyte population for modelling typical immune response. (*left*) Total number of cells (*black solid-line*) and generational information (*colour solid-lines*) as a function of time. (*right*) Cell numbers per generation at harvested time points from the experiment.

including ability to measure the exact timing of cell fates such as division, death or differentiation for each offspring, heterogeneity in cell life times and generation information. Using these measurements and with utilisation of mathematical model, important characteristics of population dynamics have been revealed: (i) the cell fates are determined autonomously and stochastically as a consequence of competition between times to these fates (Duffy et al., 2012; Zhou et al., 2018); (ii) the existence of correlation structure in cell fates so that it is more likely to observe a symmetric family tree of proliferating B and T cell (Hawkins et al., 2009; Dowling et al., 2014; Shokhirev et al., 2015; Mitchell et al., 2018), which was extensively incorporated in mathematical models in Duffy and Subramanian (2009); Wellard et al. (2010); Markham et al. (2010) for predicting cell numbers.

This imaging technique has a few limitations: (i) the cells cannot be accurately traced for a long time as cells become numerous in the field of view. This hinders determining each cell's relationship across frames; (ii) the cells continuously move around the medium and begin to form three dimensional structures (Errington et al., 2005); and (iii) the technique is best optimised for *in vitro* studies, but difficult to implement for *in vivo* system as it is limited by accessibility and depth of the region of interest. It is possible to perform *in vivo* experiment with specialised microscope, for example using intravital microscopy (Celso et al., 2009; Hawkins et al., 2016), but are typically sampled at lower frame rates and they fail to keep track of cells

1.4. MODELS OF LYMPHOCYTE POPULATION DYNAMICS

that leave the immediate field of view. Together, the *in vitro* experiments typically yield small sample size and possible to track cell division up to 10 generations.

Division tracking via cell staining. The method of division tracking was first introduced by [Lyons and Parish \(1994\)](#) using a fluorescent dye called *carboxyfluorescein diacetate succinimidyl ester* (CFSE), and subsequently further developed and assessed by [Quah and Parish \(2012\)](#); [Tempany et al. \(2018\)](#) for other colours including Cell Trace Violet (CTV) or Cell Trace Yellow (CTY). It is one of the most frequently used high-throughput experimental techniques, used particularly in immunology, to estimate the population size while tracking the generation information of proliferating cells. When cells that are initially stained with fluorescent dye divide, their offspring inherit half of the intensity level from their parent cell. This dilution property can be used to infer cells in different generations from their fluorescent intensity through flow cytometry. We illustrate this method in [Fig. 1.3](#). A typical measurement in most applications of division tracking assay is distribution of cell population binned across generation that are usually sampled at a particular harvested time, but this can be extended for multiple time points to form a time-series data (see example [Fig. 1.4](#)). Recent study in [Horton et al. \(2018\)](#) further refined this technique to track the offspring of individual clones using multiple distinct colours of the dyes.

In the following chapters, we will analyse datasets obtained from both time-lapse microscopy and division tracking assay to verify our assumptions that are crucial in developing a new lymphocyte population model, and to test the model for drawing biologically meaningful inferences.

1.4 Models of lymphocyte population dynamics

In the previous section, we introduced two major experimental techniques to study evolution of cells at single- and population-level. However, due to the non-linear nature of cell expansion, it is challenging to extract useful information that captures features of the dynamics. So, many mathematical models that adopt diverse techniques had been proposed to determine and calculate quantitative features. A well-known example is the Smith-Martin (SM) model, where the authors suggested a semi-stochastic two-state model to represent underlying processes governing a cell's lifetime ([Smith and Martin, 1973](#)). Elegant studies in [Pilyugin et al. \(2003\)](#);

1.4. MODELS OF LYMPHOCYTE POPULATION DYNAMICS

Bernard et al. (2003); Ganusov et al. (2005) extended the SM model to estimate biologically important parameters such as birth/death and turn-over rates. Other approaches like system of ordinary or partial differential equations (ODEs or PDEs) were substantially used in Revy et al. (2001); Boer and Perelson (2005); Asquith et al. (2006); Boer et al. (2006); Ganusov et al. (2007); Luzyanina et al. (2007); Banks et al. (2010, 2011, 2012); Hasenauer et al. (2012); Boer and Perelson (2013); Luzyanina et al. (2014); Banks et al. (2015); Luzyanina and Bocharov (2018); and, branching process in Yates et al. (2007); Subramanian et al. (2008); Hyrien and Zand (2008); Duffy and Subramanian (2009); Wellard et al. (2010); Hyrien et al. (2010); Miao et al. (2011); Nordon et al. (2011); Duffy et al. (2012); and, statistical/probabilistic techniques in Hasbold et al. (1999a,b); Deenick et al. (2003); Leon et al. (2004); Markham et al. (2010); Yates et al. (2017); Pandit and Boer (2019); Amo et al. (2020); Belluccini et al. (2022). In addition, a complex network of molecular interactions and signalling pathways (e.g. Beyer et al., 2011) was used to predict population dynamics (Shokhirev et al., 2015; Mitchell et al., 2018). However, the vast majority of these models assume that the cell's fate is independent of its family's history, which is inconsistent with previously published experiment results that trace fates in cell lineages (Hawkins et al., 2009; Duffy et al., 2012; Marchingo et al., 2016; Pham et al., 2018; Mitchell et al., 2018).

In this section we will describe one particular model called Cyton model (Hawkins et al., 2007a; Callard and Hodgkin, 2007; Duffy et al., 2012), built on years of conceptualisation and empirical evidence that cells can control multiple fates internally independently of each other (Gett and Hodgkin, 1998; Deenick et al., 1999; Gett and Hodgkin, 2000; Rush and Hodgkin, 2001; Tangye et al., 2003a,b; Hodgkin et al., 2005; Hodgkin, 2005), that allows capturing the growth of a lymphocyte population. Our intent is to recapitulate the fundamental concept behind this model, and new empirical findings over the decades that motivates our work in this thesis.

In Hawkins et al. (2007a), authors proposed that the cell fates are governed by competition of cellular machineries: times to division and to death of a proliferating cell. They also coined the term *division destiny* (DD) to refer to the number of divisions cells underwent before returning to quiescent, or resting, state. These machineries were assumed to operate independently of each other (Gett and Hodgkin (2000) introduces the “Law of Independence”). Together with division tracking assay and the model, a detailed experimental protocol followed by standardised quantitative framework was introduced in Hawkins et al. (2007b) and implemented

1.4. MODELS OF LYMPHOCYTE POPULATION DYNAMICS

in [Shokhirev and Hoffmann \(2013\)](#). This model was applied extensively to analyse B and T cell populations and deduce influences of one or more external signals in final immune response ([Hawkins et al., 2013](#); [Marchingo et al., 2014](#)).

The Cyton model follows two simple rules for dividing lymphocytes: (i) division and survival operations are repeated through subsequent generation after the first division and, (ii) upon division, individual cells will draw new division and death times from the appropriate distribution (see [Fig. 2.1A](#)). Formally, let random variables (RVs) T_{div}^g and T_{die}^g be the division and death times, respectively, in generation g such that $P(T_{div}^g < t)$ and $P(T_{die}^g < t)$ represent probabilities of division and death times less than time at $t \in [0, \infty)$. These RVs are assumed to be log-normally distributed based on empirical observations, although other right-skewed distribution classes were shown to be well approximating the data, e.g Weibull or Gamma ([Hawkins et al., 2009](#)). A progenitor cell is initially assigned (T_{div}^0, T_{die}^0) independently, and whichever comes first determines the fate of the cell. If $T_{div}^0 < T_{die}^0$, then the cell generates two daughter cells each of which draws new (T_{div}^1, T_{die}^1) values, otherwise the cell is removed from the system. This process repeats so long as the division time outcompetes the death time for each offspring. With this construct and assuming that T_{div}^g and T_{die}^g are independent and identically distributed RVs, we can formalise number of dividing cells in each generation by following set of equations:

$$n_{div}^g(t) = \begin{cases} N_0 \gamma_0 [1 - P(T_{die}^0 < t)] f_{T_{div}^0}(t) & \text{if } g = 0 \\ 2\gamma_g \int_0^t n_{div}^{g-1}(\tau) [1 - P(T_{die}^g < t - \tau)] f_{T_{div}^g}(t - \tau) d\tau & \text{if } g > 0 \end{cases},$$

and, for number of dying cells,

$$n_{die}^g(t) = \begin{cases} N_0 [1 - \gamma_0 P(T_{div}^0 < t)] f_{T_{die}^0}(t) & \text{if } g = 0 \\ 2 \int_0^t n_{div}^{g-1}(\tau) [1 - \gamma_g P(T_{div}^g < t - \tau)] f_{T_{die}^g}(t - \tau) d\tau & \text{if } g > 0 \end{cases},$$

where N_0 is initial total cell numbers, $f(\cdot)$ is probability distribution function (PDF) and $\gamma_g \in [0, 1]$ is progressor fraction (i.e. DD) that captures proportion of the population that will traverse to the next generation in respond to the stimulation.

Generally, the progressor fraction is defined using normal distribution,

$$\gamma_g = \begin{cases} \gamma_0 & \text{if } g = 0 \\ \frac{P(G > g + 1)}{P(G > g)} & \text{if } g > 0 \end{cases}, \quad (1.1)$$

where $G \sim \mathcal{N}(\mu, \sigma)$ is normally distributed RV that describes DD given mean (μ) and standard deviation (σ). The expected total number of cells is

$$N_g(t) = \begin{cases} N_0 - \int_0^t n_{div}^0(\tau) + n_{die}^0(\tau) d\tau & \text{if } g = 0 \\ \int_0^t 2n_{div}^{g-1}(\tau) - n_{div}^g(\tau) - n_{die}^g(\tau) d\tau & \text{if } g > 0 \end{cases}.$$

This equation is known as the Cyton model. The parameters for division, death and progressor fraction can be estimated with a population dataset obtained from division tracking assay using numerical techniques such as in least-squares sense (e.g. [Marquardt, 1963](#)) or via genetic algorithm (e.g. [Storn and Price, 1997](#)). We will use this model as a basis of our work and adapt it by incorporating recent experimental findings for predicting cell numbers of *in vitro* systems.

1.5 Thesis outline

In this thesis we will introduce a mathematical model of lymphocyte proliferation dynamics that, in contrast to earlier versions, include familial correlation structures, which has been reported in experimental studies for both B and T lymphocytes. Despite the non-linear nature of the dynamics and the complexity due to the correlation, we developed a general framework that neither is over-parameterised nor computationally expensive in data fitting in order to draw meaningful biological inferences on the lymphocyte population dynamics.

In [Chapter 2](#), we establish the motivation for revisiting the Cyton model ([Hawkins et al., 2007a](#)), and derive a newer mathematical model called Cyton2 that is more aligned to the recent experimental studies. To do so, we utilise published B cell data from [Hawkins et al. \(2009\)](#) and primary T cell data, obtained from murine *in vitro* single-cell filming experiments, to substantiate two crucial assumptions embedded in the model: that the cellular machineries (e.g. time to first division)

are well approximated as being independent of each other, and to establish the best distribution class that describes the execution time of modules.

In [Chapter 3](#), we fit the Cyton2 model to lymphocyte population datasets that are generally acquired from division tracking assays (e.g. CFSE or CTV). We first assess the robustness of the model by systematically removing the information such as the repeated samples or data points from one or more harvested times, then evaluate the model extrapolation. Also, we draw similar but alternative interpretation of the study published in [Marchingo et al. \(2014\)](#) about the linear additive nature of the co-stimulatory signals in CD8⁺ T cells. The majority of the work presented in [Chapters 2 and 3](#) was published in [Cheon et al. \(2021\)](#). All the analyses, model implementation and datasets are publicly available on a [GitHub repository](#).

In [Chapter 4](#), another important mechanism of the adaptive immunity was explored: differentiation. In particular, we modelled switching of antibody subclass of B cells, which is a feature that changes the functional properties but not the specificity of the antibody, based on the empirical evidence of the underlying molecular regulations. To do so, a simple probabilistic approach was employed. We published this work in [Horton et al. \(2022\)](#).

Finally, in [Chapter 5](#), we tested the model for human systems, in particular, to hypo- and hyper-immune responses from immuno-compromised patients. We identified key characteristics of these patients in cell population dynamics when compared to healthy donors and developed a method to quantify health of naïve human B cells. Also, we questioned the hypothesis of unlikely event of occurring patient-like immune response from healthy individuals. More specifically, we investigated B cells from immunodeficient and autoimmune patients and utilised the model to dissect division modules that led to those aberrant behaviours.

We provided the details of experiments whose output we analysed, as performed by our collaborators from WEHI, in [Appendix A](#) for completeness. The experiment details of time-lapse microscopy for the primary CD8⁺ T cell data presented in [Chapter 2](#) were shown in [Section A.1](#). Methods for the drug experiment, obtained from the division tracking assay, in [Chapter 3](#) were listed in [Section A.3](#). In [Section A.2](#), we reported technical details of the differentiation experiment presented in [Chapter 4](#), performed by Miles B. Horton.

2

Cyton2: A mathematical model of lymphocyte population dynamics

2.1 Abstract

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.

In this chapter, 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 established 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 in [Chapter 3](#), 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.

2.2 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., 2007a](#)). 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 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 Cell Trace 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. (2007a); 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. (2007a) 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.

2.2. INTRODUCTION

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 et al., 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 (Turner et al., 2008).

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., 2007a; 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, 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](#); [Zhou et al., 2018](#); [Horton 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. In the following sections, we propose 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 correlation in the likelihood of each alternative fate and determine a suitable class of parametric

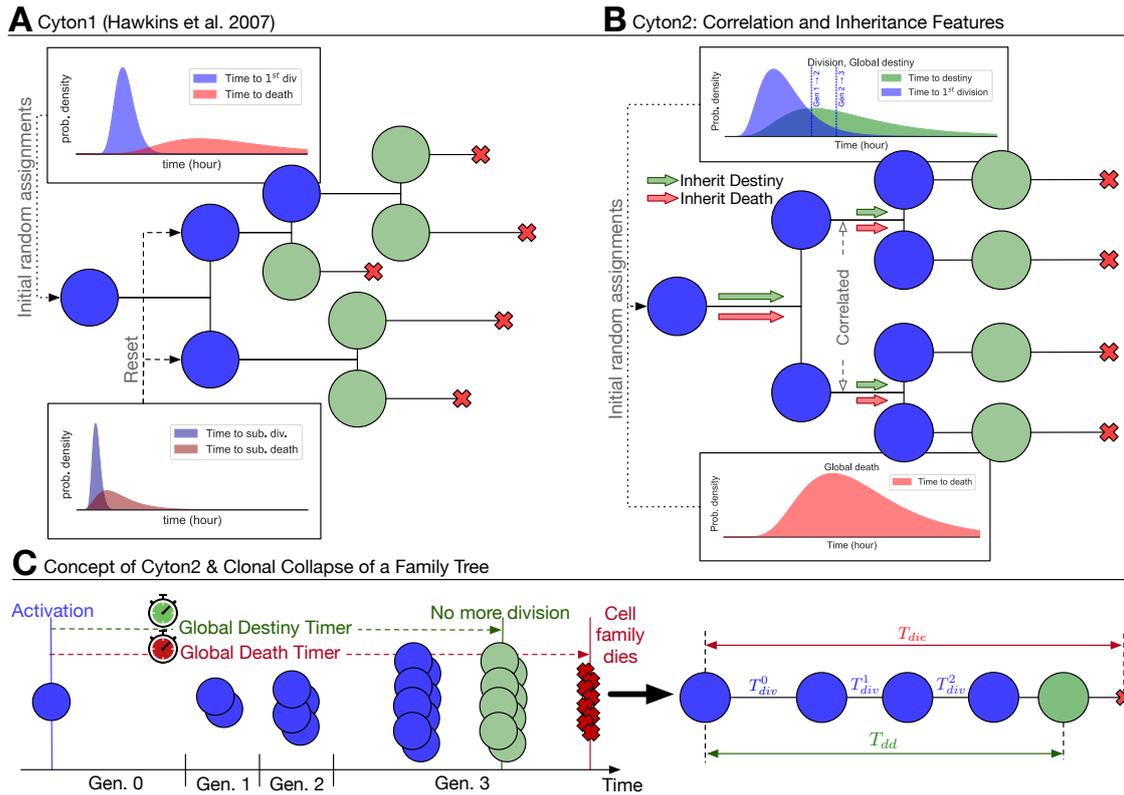


Fig. 2.1: Overview of the two Cyton models. (A) The original Cyton model (Hawkins et al., 2007a) where stochastic times to divide and to die are chosen independently after each cell division. Cells cease their motivation to divide based on division-counting mechanism. (B) The Cyton2 model incorporates significant correlation in division times between siblings, as well as familial inheritance of death and division destiny times. (C) A consequence of the correlation and inheritance is that the resulting family trees are heterogeneous, but highly concordant. By exploiting this property, a family tree can be summarised by substituting the average values of its times and fate at each generation. An example of clonally collapsed family tree and its key variables is shown.

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.

2.3 Model structure

Recent experimental findings suggest three important modifications to the original Cyton model for stimulus-induced proliferation bursts (Fig. 2.1A): 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 (Fig. 2.1B). Thus we posit the new Cyton2 stochastic model using sets of random variables that corresponds to a global death timer, a global destiny timer, and division-time machinery (Fig. 2.1C).

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 $\min(T_{die}, T_{dd})$.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

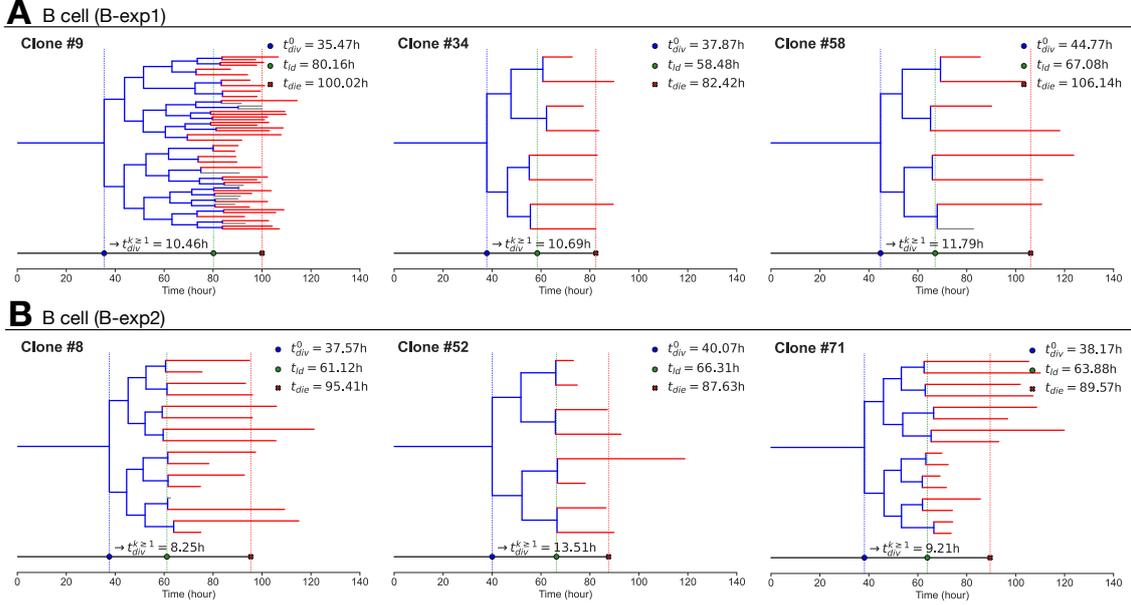


Fig. 2.2: Example family and clonally collapsed trees of CpG-stimulated B cells in the filming datasets. Each panel shows one family. A horizontal line represents the lifetime of a cell in dividing (—), dying (—), or lost (—) state. Clonally collapsed tree is shown as a single time-line (—) below the family tree. Time to first division (●), average time to last division (●), average time to death (×) and average subsequent division time ($t_{div}^{k \geq 1}$) are annotated on the collapsed line. These values are shown in the legend. (A) Experiment CpG-stimulated B cells (B-exp1). (B) Repeat of B cell experiment (B-exp2).

- At time $t < \min(T_{die}, T_{dd})$, cells in the family are in generation $G(t) = \max\{g : \sum_{k=0}^g T_{div}^k < t\}$.

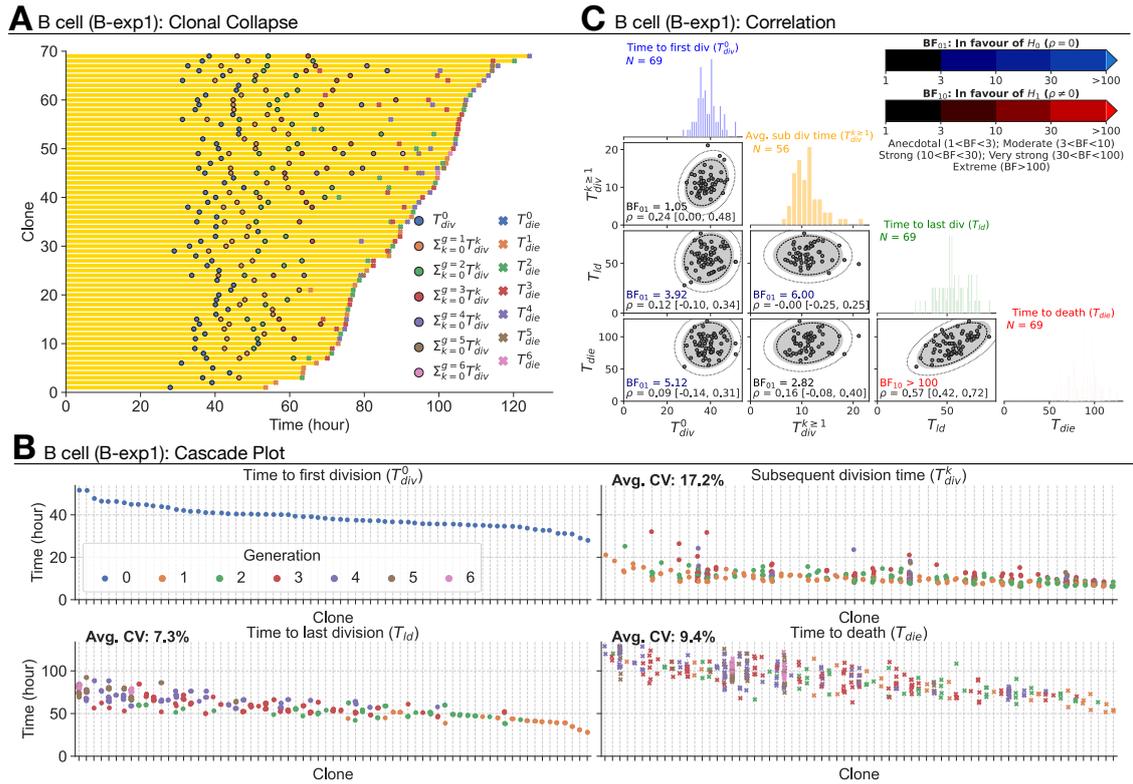
To properly assess the appropriateness of the Cyton2 as a fine-grained description required time-lapse microscopy data. To that end we re-analysed previously published B cell data sets as well as new, primary CD8⁺ T cell datasets.

2.4 Analysing data to investigate the model assumptions

2.4.1 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),

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS



2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

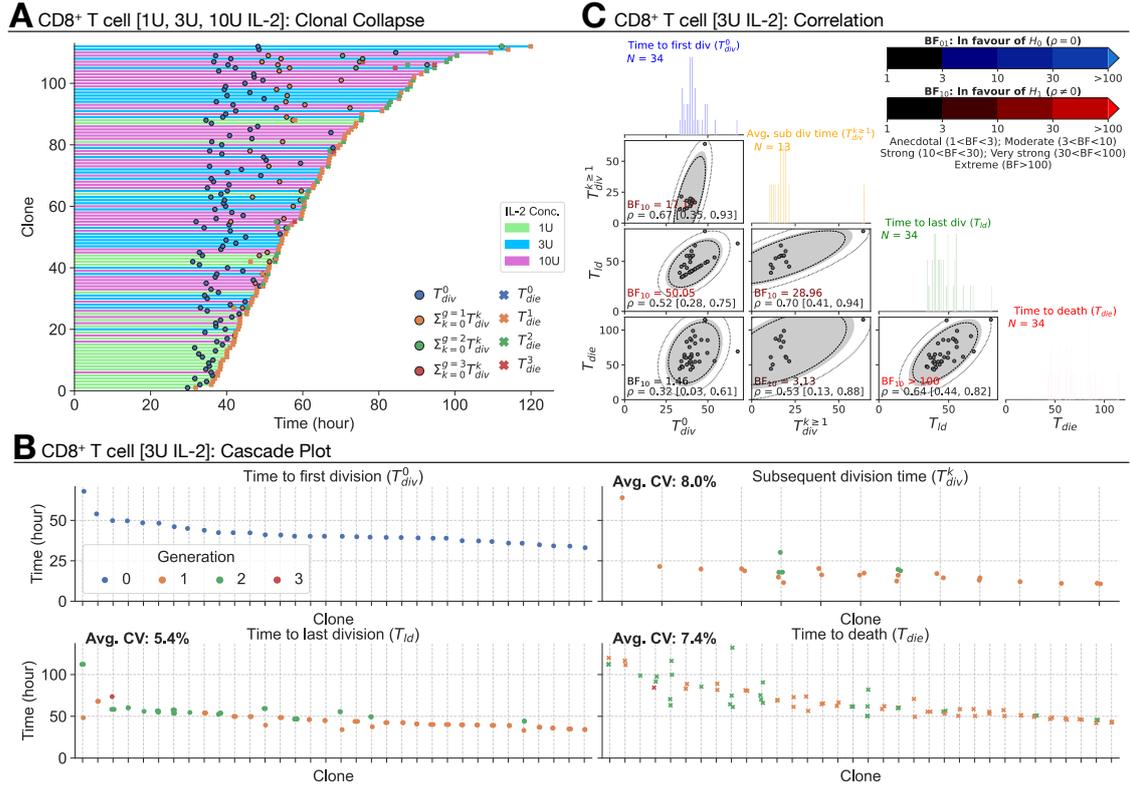


Fig. 2.4: Extracting times to fates from CD8⁺ T cells in the presence of 1U, 3U or 10U IL-2. (A) Clonally collapsed family trees of T cells. (B) Rank ordered times to events of families. (C) Correlation coefficient (ρ) estimated using bivariate normal distribution with 95% credible interval is reported for each pair. 90%, 95% and 99% density regions are plotted over the data. For null, $H_0: \rho = 0$, and alternative, $H_1: \rho \neq 0$, hypotheses, Bayes Factors ($BF_{01} = 1/BF_{10}$) were calculated. If the data is more probable under H_0 , then it is BF_{01} times more favoured than H_1 (*blue-scale*), and vice versa (*red-scale*). Distributions of the times are collated into 1 hour time intervals and shown in the diagonal panels.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

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 Section 2.4.4 for method). This process is illustrated in Fig. 2.1C and was applied to each B cell family as shown in Figs. 2.2 and 2.3A. Measurements corresponding to key Cyton2 variables are further illustrated in the cascade plots Fig. 2.3B 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 hours, are substantially longer than times to subsequent divisions. These data also confirmed the relatively consistent subsequent division times (≈ 10 hours) 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 Fig. 2.5A1). 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 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 Section 2.4.6) (Fig. 2.3C) and tabulated in Table 2.1. 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

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

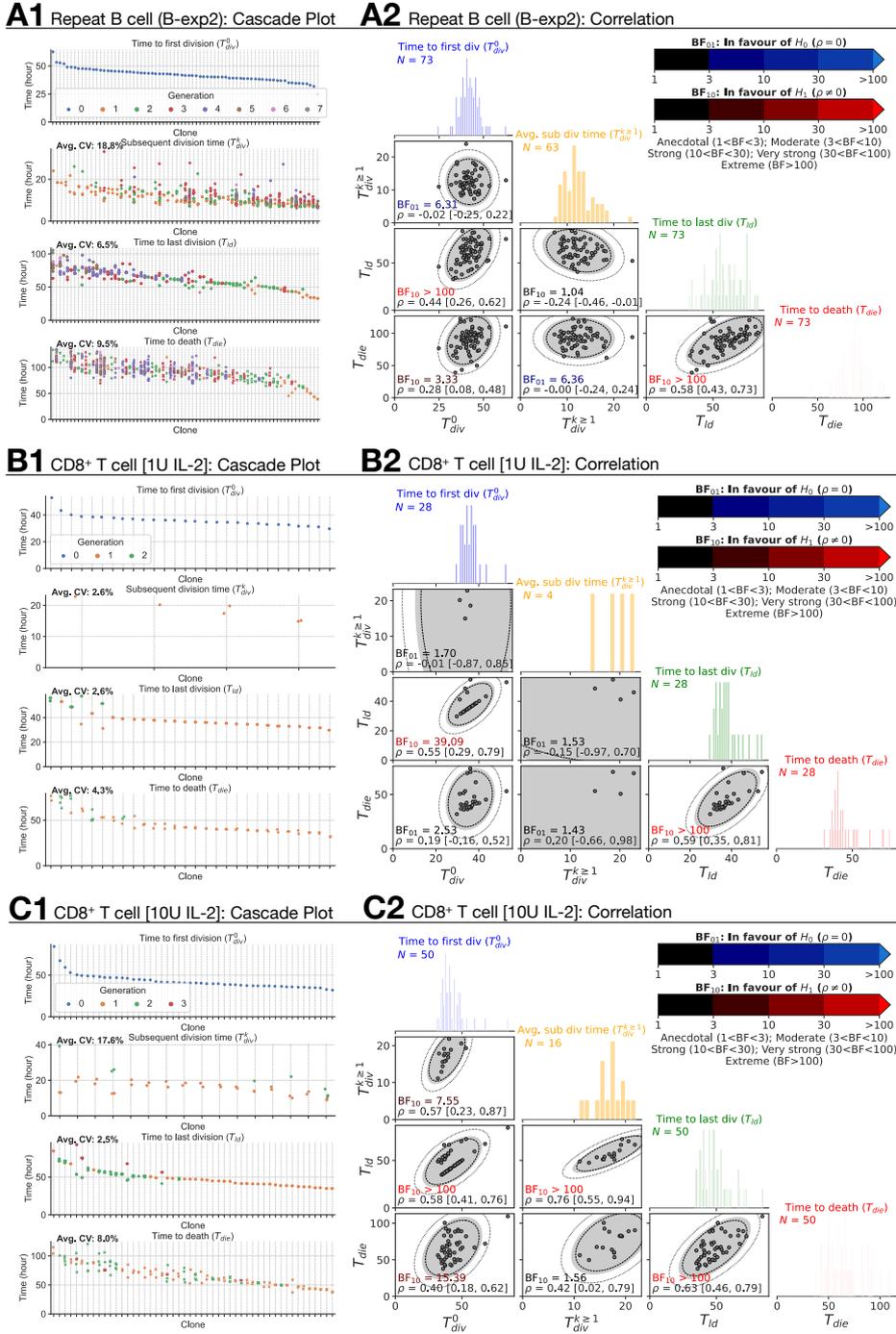


Fig. 2.5: Extracting times to fates from the repeat of CpG-stimulated B cells and CD8⁺ T cells in the presence of 1U or 10U IL-2. For CD8⁺ T cell data, measurements from two independent but identical experiment setup are aggregated. (A1,B1,C1) Four key Cyton2 variables, T_{div}^0 , $\{T_{div}^k\}_{k \geq 1}$, T_{ld} (replacing T_{dd} as a proxy measure) and T_{die} for all cells in each family are shown as a series of cascade plots. Average coefficient of variation (CV) for each variable is annotated. The lost cells are not shown. (A2,B2,C2) Pair-plot of the Cyton2 variables. Distributions of the times are collated into 1 hour time interval and shown in the diagonal panels. Correlation coefficient (ρ) was estimated using bivariate normal distribution with 95% credible interval for each pair. 90%, 95% and 99% density regions are plotted over the data. Given two hypotheses ($H_0: \rho = 0$ and $H_1: \rho \neq 0$), Bayes Factor ($BF_{01} = 1/BF_{10}$) was calculated. If the data is more probable under H_0 , then it is BF_{01} times more favoured than H_1 (blue-scale), otherwise H_1 is BF_{10} (red-scale) times more favoured than H_0 .

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

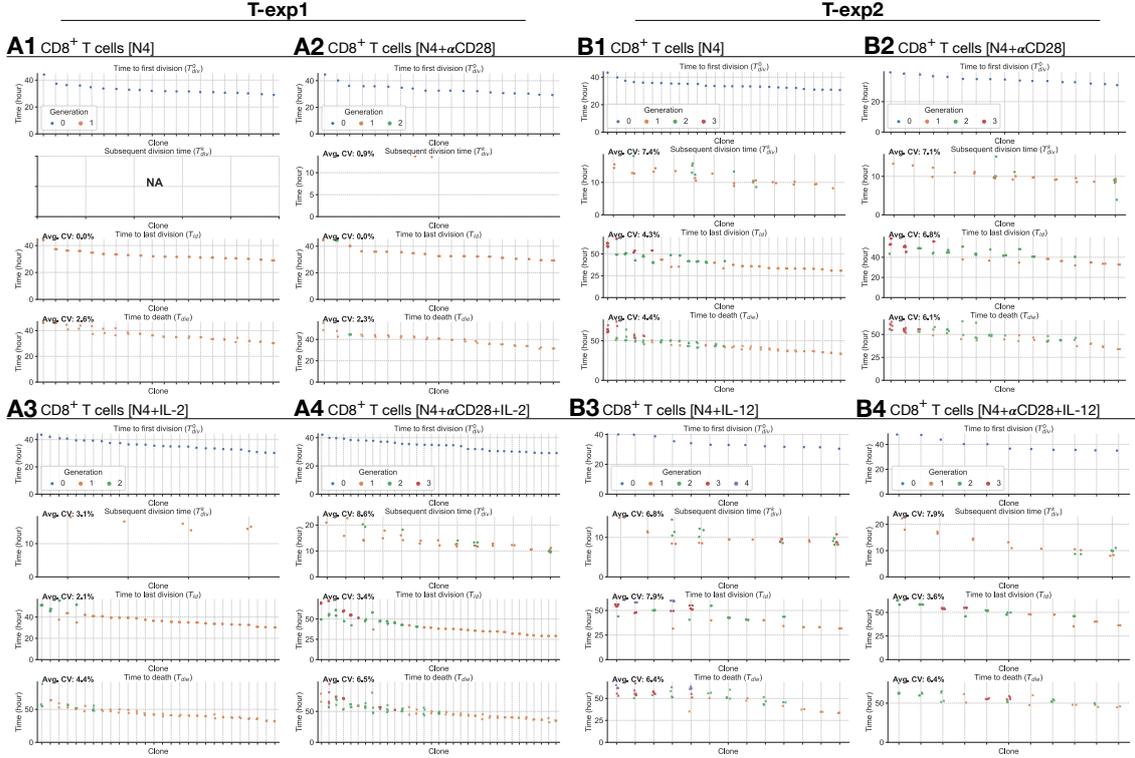


Fig. 2.6: Extracting times to fates from $CD8^+$ T cell costimulation experiments. Four key Cyton2 variables, T_{div}^0 , $\{T_{div}^k\}_{k \geq 1}$, T_{ld} (replacing T_{dd} as a proxy measure) and T_{die} for all cells in each family are shown as a series of cascade plots. Average coefficient of variation (CV) for each variable is annotated. The lost cells are not shown. **(A1-4)** Experiment T-exp1 consists of N4 only, N4 + α CD28, N4 + IL-2 and N4 + α CD28 + IL-2. **(B1-4)** Experiment T-exp2 consists of N4 only, N4 + α CD28, N4 + IL-12 and N4 + α CD28 + IL-12.

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 hours. In the first dataset (i) the cells were stimulated with α CD3 and co-incubated with 1U/mL, 3U/mL or 10U/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

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

IL-12 (T-exp2). Details for live imaging and data extraction are given in Methods. In Fig. 2.4, 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 ($\approx 40\text{h}$) than the subsequent division times ($\approx 18\text{h}$) for 1U, 3U and 10U of IL-2. Also, the spread of the times within a family show similar or lower average CVs than that of B cells (Fig. 2.4B for 3U; see Fig. 2.5B1,C1 for 1U and 10U). We applied the same calculation to (ii) and (iii) datasets and reached the same conclusions (see Fig. 2.6 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 (Fig. 2.4C; see Fig. 2.5B2,C2 for 1U and 10U). These were further supported by BF calculations, which show strong evidence in favour of H_1 . We noticed the same results for T-exp1 and T-exp2 datasets (Table 2.1).

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.

2.4.2 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 2.4.1, 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

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

Cell Type	Stimulation	Number of clones (N) & Bayes Factor ^a ($BF_{01} = 1/BF_{10}$) & Correlation Coefficient (ρ [CI])					
		$(T_{div}^0, T_{div}^{k \geq 1})$	(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})
B	CpG (B-exp1)	$N = 56$	$N = 69$	$N = 69$	$N = 56$	$N = 56$	$N = 69$
		$BF_{01} = 1.05$	$BF_{01} = 3.92$	$BF_{01} = 5.12$	$BF_{01} = 6.00$	$BF_{01} = 2.82$	$BF_{10} > 100$
		0.24 [0.01, 0.48]	0.12 [-0.10, 0.35]	0.09 [-0.14, 0.31]	0.0 [-0.25, 0.25]	0.16 [-0.08, 0.40]	0.57 [0.42, 0.72]
	CpG (B-exp2)	$N = 63$	$N = 73$	$N = 73$	$N = 63$	$N = 63$	$N = 73$
	$BF_{01} = 6.31$	$BF_{10} > 100$	$BF_{10} = 3.33$	$BF_{10} = 1.04$	$BF_{01} = 6.36$	$BF_{10} > 100$	
	-0.02 [-0.25, 0.22]	0.44 [0.26, 0.62]	0.28 [0.08, 0.48]	-0.24 [-0.46, -0.01]	0.00 [-0.24, 0.24]	0.58 [0.43, 0.73]	
CD8 ⁺ T	1U IL-2	$N = 4$	$N = 28$	$N = 28$	$N = 4$	$N = 4$	$N = 28$
		$BF_{01} = 1.70$	$BF_{10} = 39.09$	$BF_{01} = 2.53$	$BF_{01} = 1.53$	$BF_{01} = 1.43$	$BF_{10} > 100$
		-0.01 [-0.86, 0.85]	0.55 [0.29, 0.78]	0.19 [-0.15, 0.53]	-0.15 [-0.96, 0.71]	0.20 [-0.66, 0.98]	0.59 [0.35, 0.81]
	3U IL-2	$N = 13$	$N = 34$	$N = 34$	$N = 13$	$N = 13$	$N = 34$
		$BF_{10} = 17.17$	$BF_{10} = 50.05$	$BF_{10} = 1.46$	$BF_{10} = 28.96$	$BF_{10} = 3.13$	$BF_{10} > 100$
		0.67 [0.36, 0.93]	0.52 [0.28, 0.75]	0.32 [0.03, 0.61]	0.70 [0.41, 0.94]	0.52 [0.12, 0.87]	0.64 [0.44, 0.82]
10U IL-2	$N = 16$	$N = 50$	$N = 50$	$N = 16$	$N = 16$	$N = 50$	
	$BF_{10} = 7.55$	$BF_{10} > 100$	$BF_{10} = 15.39$	$BF_{10} > 100$	$BF_{10} = 1.56$	$BF_{10} > 100$	
	0.57 [0.23, 0.86]	0.58 [0.41, 0.75]	0.40 [0.18, 0.62]	0.76 [0.55, 0.94]	0.42 [0.02, 0.79]	0.63 [0.46, 0.78]	
CD8 ⁺ T (T-exp1)	N4	$N < 2$	$N = 20$	$N = 20$	$N < 2$	$N < 2$	$N = 20$
		NA	$BF_{10} > 100$	$BF_{10} = 8.31$	NA	NA	$BF_{10} = 8.31$
		1.00 [1.00, 1.00]	0.53 [0.22, 0.82]			0.53 [0.22, 0.82]	
	N4 + α CD28	$N < 2$	$N = 19$	$N = 19$	$N < 2$	$N < 2$	$N = 19$
		NA	$BF_{10} = 85.63$	$BF_{10} = 18.83$	NA	NA	$BF_{10} = 60.52$
		0.67 [0.42, 0.89]	0.59 [0.30, 0.86]			0.66 [0.40, 0.89]	
	N4 + IL-2	$N = 4$	$N = 28$	$N = 28$	$N = 4$	$N = 4$	$N = 28$
		$BF_{01} = 1.30$	$BF_{10} = 29.98$	$BF_{10} = 2.09$	$BF_{01} = 1.31$	$BF_{01} = 1.58$	$BF_{10} > 100$
		0.24 [-0.64, 0.99]	0.54 [0.28, 0.78]	0.38 [0.06, 0.67]	-0.24 [-0.99, 0.63]	-0.12 [-0.96, 0.73]	0.69 [0.49, 0.87]
	N4 + α CD28 + IL-2	$N = 13$	$N = 33$	$N = 13$	$N = 13$	$N = 33$	$N = 33$
$BF_{10} = 4.17$		$BF_{10} = 1.03$	$BF_{01} = 3.28$	$BF_{01} = 1.26$	$BF_{01} = 1.77$	$BF_{10} > 100$	
	0.55 [0.17, 0.89]	0.30 [-0.01, 0.59]	0.14 [-0.18, 0.45]	0.33 [-0.14, 0.78]	0.26 [-0.23, 0.72]	0.71 [0.54, 0.87]	
CD8 ⁺ T (T-exp2)	N4	$N = 12$	$N = 27$	$N = 27$	$N = 12$	$N = 12$	$N = 27$
		$BF_{10} = 15.12$	$BF_{01} = 2.04$	$BF_{01} = 2.31$	$BF_{10} = 2.69$	$BF_{10} = 3.53$	$BF_{10} > 100$
		0.68 [0.35, 0.94]	0.22 [-0.13, 0.56]	0.20 [-0.15, 0.54]	0.52 [0.10, 0.89]	0.55 [0.15, 0.90]	0.95 [0.91, 0.98]
	N4 + α CD28	$N = 12$	$N = 17$	$N = 17$	$N = 12$	$N = 12$	$N = 17$
		$BF_{01} = 1.34$	$BF_{01} = 2.84$	$BF_{01} = 2.75$	$BF_{01} = 2.82$	$BF_{01} = 1.48$	$BF_{10} > 100$
		0.33 [-0.18, 0.78]	0.13 [-0.32, 0.57]	0.15 [-0.30, 0.59]	-0.02 [-0.55, 0.52]	0.30 [-0.19, 0.78]	0.83 [0.67, 0.96]
	N4 + IL-12	$N = 9$	$N = 13$	$N = 13$	$N = 9$	$N = 9$	$N = 13$
		$BF_{10} = 14.97$	$BF_{01} = 2.58$	$BF_{01} = 2.39$	$BF_{01} = 2.21$	$BF_{01} = 2.24$	$BF_{10} > 100$
		0.74 [0.41, 0.98]	0.13 [-0.38, 0.63]	0.17 [-0.34, 0.66]	0.14 [-0.48, 0.72]	0.13 [-0.48, 0.73]	0.95 [0.88, 0.99]
	N4 + α CD28 + IL-12	$N = 7$	$N = 11$	$N = 11$	$N = 7$	$N = 7$	$N = 11$
$BF_{01} = 1.33$		$BF_{01} = 2.47$	$BF_{01} = 2.48$	$BF_{01} = 1.00$	$BF_{10} = 1.16$	$BF_{10} = 29.19$	
	0.32 [-0.33, 0.90]	0.12 [-0.43, 0.66]	0.12 [-0.44, 0.66]	0.40 [-0.23, 0.93]	0.44 [-0.17, 0.95]	0.74 [0.45, 0.96]	

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

Table 2.1: 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} (*blue-scale*) times more favoured than H_1 , otherwise H_1 is BF_{10} (*red-scale*) times more favoured than H_0 . The variable pairs that marked with $N < 2$ or NA indicate that the analysis could not be performed, for it requires minimum of two clones to be observed in the sample.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

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 (Figs. 2.3B and 2.4B). 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 (see Section 2.4.5). 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 2.4.3). In Fig. 2.7A, 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 Fig. 2.7B. By construction, the BF's 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 BF's are consistent with the experimental data in favouring the alternative hypothesis ($H_1 : \rho \neq 0$) for some pairs (Table 2.2). 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. We followed a similar protocol for the CD8⁺ T cell data where a higher degree of right-censorship occurs due to the underlying distributions having greater overlap (Fig. 2.8). The BF's 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. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

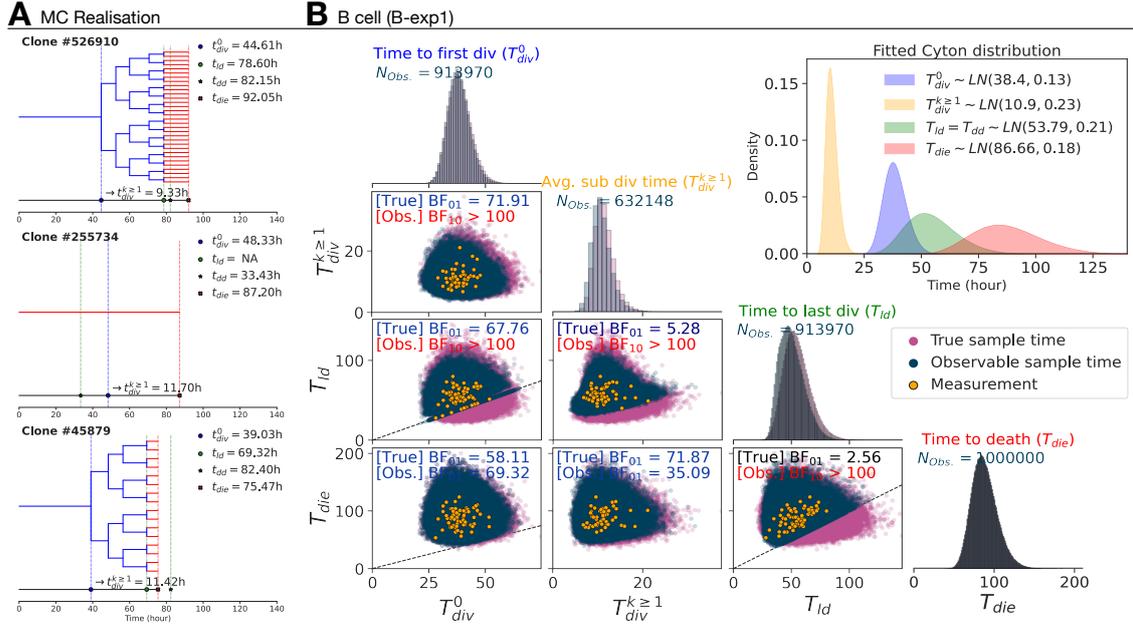


Fig. 2.7: Simulation under the independence assumption. 10^6 Cyton2 families were simulated given fitted lognormal distributions of T_{div}^0 , $T_{div}^{k \geq 1}$, T_{ld} , T_{die} from respective filming datasets. **(A)** Three example families from MC simulation parameterised as CpG-stimulated B cells: dividing (—) and dying (—) states. The realisations of T_{div}^0 (●), $T_{div}^{k \geq 1}$ ($t_{div}^{k \geq 1}$), T_{ld} (●), T_{dd} (★) and T_{die} (×) are annotated on a collapsed line. As a feature of inheritance, the progeny cells double in number synchronously whenever division occurs, and likewise, they reach destiny and death at the same time. **(B)** For all simulated families, each variable was randomly sampled from the fitted Cyton distribution (*inset*), and the samples are labeled as true sample time (●). Their corresponding observable sample times (●) are shown along with the data points (●) from the filming datasets. Distributions of the sampled true and observable times of each variable are shown in the diagonal panels. The observable and unobservable regions are separated by upper and lower sections of $y = x$ line (---), respectively. The Bayes factors are reported for the true and observable pairs.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

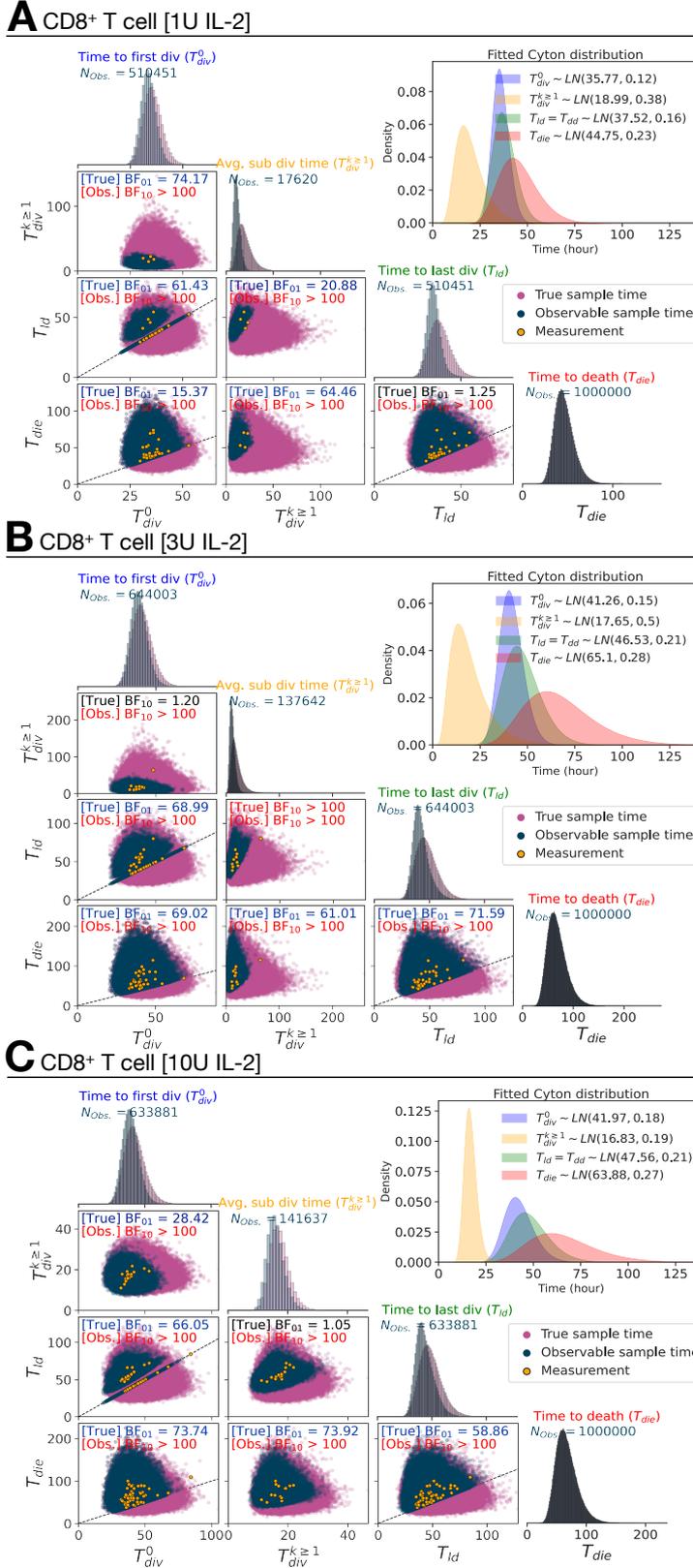


Fig. 2.8: Simulation under the independence assumption. 10^6 Cyton2 families were simulated given fitted lognormal distributions of T_{div}^0 , $T_{div}^{k \geq 1}$, T_{id} , T_{die} from respective filming datasets. For all simulated families, each variable was randomly sampled from the fitted Cyton distribution (*inset*), and the samples are labeled as true sample time (\bullet). Their corresponding observable sample times (\bullet) are shown along with the data points (\bullet) from the filming datasets. Distributions of the sampled true and observable times of each variable are shown in the diagonal panels. The observable and unobservable regions are separated by upper and lower sections of $y = x$ line (---), respectively. The Bayes factors are reported for the true and observable pairs. (A) 1U IL-2. (B) 3U IL-2. (C) 10U IL-2.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

Cell Type	Stim.	Percentage of number of observable times ($N_{\text{Obs.}}/N_{\text{True}}$) & Bayes Factor ^a ($\text{BF}_{01} = 1/\text{BF}_{10}$)											
		$(T_{\text{div}}^0, T_{\text{die}}^{\leq 1})$		$(T_{\text{div}}^0, T_{\text{die}})$		$(T_{\text{div}}^0, T_{\text{die}})$		$(T_{\text{div}}^{\leq 1}, T_{\text{die}})$		$(T_{\text{div}}^{\leq 1}, T_{\text{die}})$		$(T_{\text{die}}, T_{\text{die}})$	
		True	Obs.	True	Obs.	True	Obs.	True	Obs.	True	Obs.	True	Obs.
B (B-exp1)	CpG	$\text{BF}_{01} = 71.91$	63.2%	$\text{BF}_{01} = 67.76$	91.4%	$\text{BF}_{01} = 58.11$	91.4%	$\text{BF}_{01} = 5.28$	63.2%	$\text{BF}_{01} = 71.87$	63.2%	$\text{BF}_{01} = 2.56$	91.3%
			$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} = 69.32$		$\text{BF}_{10} > 100$		$\text{BF}_{10} = 35.09$		$\text{BF}_{10} > 100$
CD8 ⁺ T	1U IL-2	$\text{BF}_{01} = 74.17$	17.6%	$\text{BF}_{01} = 61.43$	51.0%	$\text{BF}_{01} = 15.37$	51.0%	$\text{BF}_{01} = 20.88$	17.6%	$\text{BF}_{01} = 64.46$	17.6%	$\text{BF}_{01} = 1.25$	51.0%
			$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$
	3U IL-2	$\text{BF}_{10} = 1.20$	13.8%	$\text{BF}_{01} = 68.99$	64.4%	$\text{BF}_{01} = 69.02$	64.4%	$\text{BF}_{10} > 100$	13.8%	$\text{BF}_{01} = 61.01$	13.8%	$\text{BF}_{01} = 71.59$	64.4%
			$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$
	10U IL-2	$\text{BF}_{01} = 28.42$	14.1%	$\text{BF}_{01} = 66.05$	63.2%	$\text{BF}_{01} = 73.74$	63.2%	$\text{BF}_{01} = 1.05$	14.1%	$\text{BF}_{01} = 73.92$	14.1%	$\text{BF}_{01} = 58.86$	63.2%
			$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$		$\text{BF}_{10} > 100$

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

Table 2.2: Bayesian independence test of the times to fates from simulation. The test was performed with $N_{\text{True}} = 10^6$ 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 ($N_{\text{Obs.}}$) 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 (*blue-scale*), otherwise it is indicated by BF_{10} (*red-scale*).

2.4.3 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., 2007a).

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

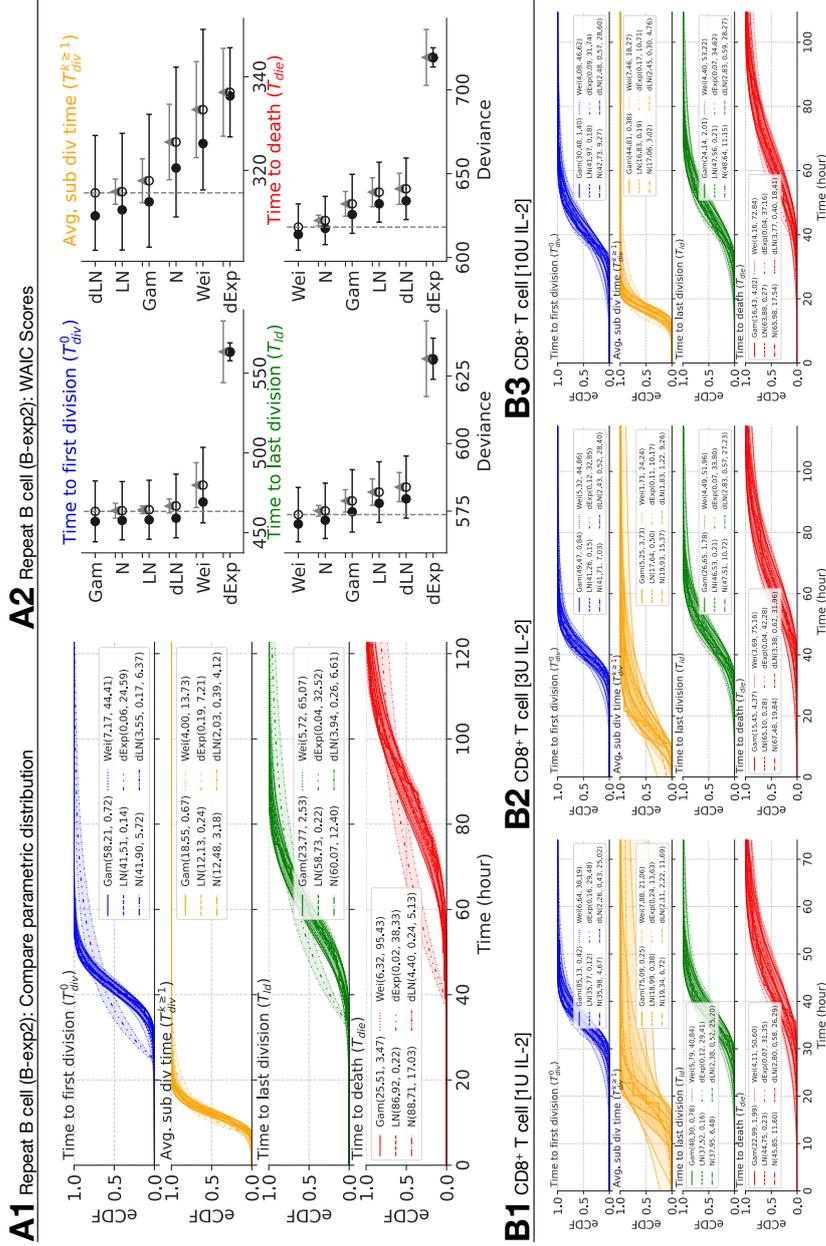


Fig. 2.10: Best parametric distribution classes. The eCDFs of measurements of times to fates are presented. Six candidate distribution classes are parameterised by: (α_G, β_G) for Gamma; (m, s) median and scale for lognormal; (μ, σ) mean and standard deviation for Normal; (α_W, β_W) for Weibull; (λ, c) rate and shift for delayed Exponential; and, (m_d, s_d, c) median, scale and shift for delayed Lognormal. The CDF of the candidates was plotted by taking mean values of posterior distribution of hyper-parameters. The confidence band was plotted by sampling hyper-parameter values from posterior distribution. **(A1-2)** Repeat of CpG-stimulated B cell (B-cpg2). Rank-ordered WAIC plot is shown. **(B1-3)** CD8⁺ T cell in the presence of 1U, 3U and 10U of IL-2. For T cell data, measurements from two independent but identical experiment setup are aggregated. Corresponding WAIC plot is shown in **Fig. 2.9**.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

For B cells, the empirical cumulative distribution function (eCDF) measured times overlaid with CDFs of four candidate distributions with 95% confidence bands in [Fig. 2.9A1](#). 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_d, s_d, 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 [Section 2.4.6](#)) to quantitatively determine the best fit ([Fig. 2.9A2](#)) ([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, $\Delta\text{SE} = \pm 12.3$). The delayed Lognormal was most preferred candidate for $T_{div}^{k \geq 1}$ (263.4), but the Lognormal was a close second (264.2, $\Delta\text{SE} = \pm 2.7$). 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, $\Delta\text{SE} = \pm 4.9$). We observed similar results in the repeat of B cell data (B-exp2), except for T_{die} where the Weibull provided the best fit ([Fig. 2.10A1-2](#)).

For CD8⁺ T cells, we present the rank-ordered WAIC plot for all IL-2 concentrations in [Fig. 2.9B1-3](#) (see [Fig. 2.10B1-3](#) 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 10U IL-2 in which Weibull was top-ranked (80.5), but only marginally so compared to the Normal (81.3, $\Delta\text{SE} = \pm 1.6$), the Gamma (83.0, $\Delta\text{SE} = \pm 3.3$) and the Lognormal (83.3, $\Delta\text{SE} = \pm 2.6$). Note that the estimates of WAIC for $T_{div}^{k \geq 1}$ in 1U 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 ([Fig. 2.11](#)).

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

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

Cell Type	Stimulation	Initial # clones	Observed cells: # divided cells (# dead cells) [# lost cells]							
			Gen 0	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 6	Gen 7
B	CpG (B-exp1)	108	108 (0) [0]	169 (30) [17]	204 (67) [67]	153 (147) [108]	82 (134) [90]	37 (95) [32]	1 (58) [15]	0 (0) [2]
	CpG (B-exp2)	88	88 (0) [0]	145 (24) [7]	150 (99) [41]	175 (101) [24]	75 (203) [72]	18 (93) [39]	2 (18) [16]	0 (4) [0]
CD8 ⁺ T	1U IL-2	109	29 (79) [1]	6 (50) [2]	0 (12) [0]	-	-	-	-	-
	3U IL-2	90	68 (22) [0]	77 (48) [11]	46 (35) [73]	2 (4) [86]	4 (0) [0]	1 (0) [7]	0 (0) [2]	-
	10U IL-2	163	101 (62) [0]	122 (71) [9]	94 (59) [91]	12 (12) [164]	0 (0) [24]	-	-	-
CD8 ⁺ T (T-exp1)	N4	45	20 (25) [0]	0 (40) [0]	-	-	-	-	-	-
	N4 + α CD28	41	22 (19) [0]	8 (36) [0]	0 (4) [12]	-	-	-	-	-
	N4 + IL-2	37	28 (9) [0]	6 (50) [0]	0 (12) [0]	-	-	-	-	-
	N4 + α CD28 + IL-2	47	36 (11) [0]	31 (41) [0]	20 (38) [4]	0 (19) [21]	-	-	-	-
CD8 ⁺ T (T-exp2)	N4	37	30 (7) [0]	27 (33) [0]	12 (34) [8]	0 (16) [8]	-	-	-	-
	N4 + α CD28	38	33 (5) [0]	52 (14) [0]	8 (30) [66]	0 (16) [0]	-	-	-	-
	N4 + IL-12	29	22 (7) [0]	33 (11) [0]	12 (18) [36]	5 (17) [2]	0 (6) [4]	-	-	-
	N4 + α CD28 + IL-12	36	32 (4) [0]	55 (9) [0]	22 (20) [68]	0 (8) [36]	-	-	-	-

Table 2.3: Number of cells observed for division, death and lost events in each generation. For each condition of CD8⁺ T cell with 1U, 3U and 10U IL-2, measurements from two independent but identical experiment setup are pooled. The total number of cells in each generation (i.e. divided + dead + lost cells) is equal to twice of divided cells from previous generation. A cell is considered lost if it becomes indistinguishable to the nearby cells, or survives until the end of given experiment time frame.

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.

2.4.4 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. Total number cells observed for each of these variables are tabulated in [Table 2.3](#). All of these variables were measured from $t = 0h$, 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 $x = \langle x_1, x_2, \dots, x_j \rangle$ with $x_j \in \{1, 2\}$ is a finite and ordered sequence of 1s and 2s. Beginning with a founder cell, defined as $x = \langle 0 \rangle$, we denote its first and second daughter cells in generation 1 by $x = \langle 1 \rangle$ and $x = \langle 2 \rangle$, respectively. In general, $\langle x_1, x_2, \dots, x_j \rangle$ 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, $x = \langle 1, 1, 2 \rangle$ denotes the second daughter of the

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

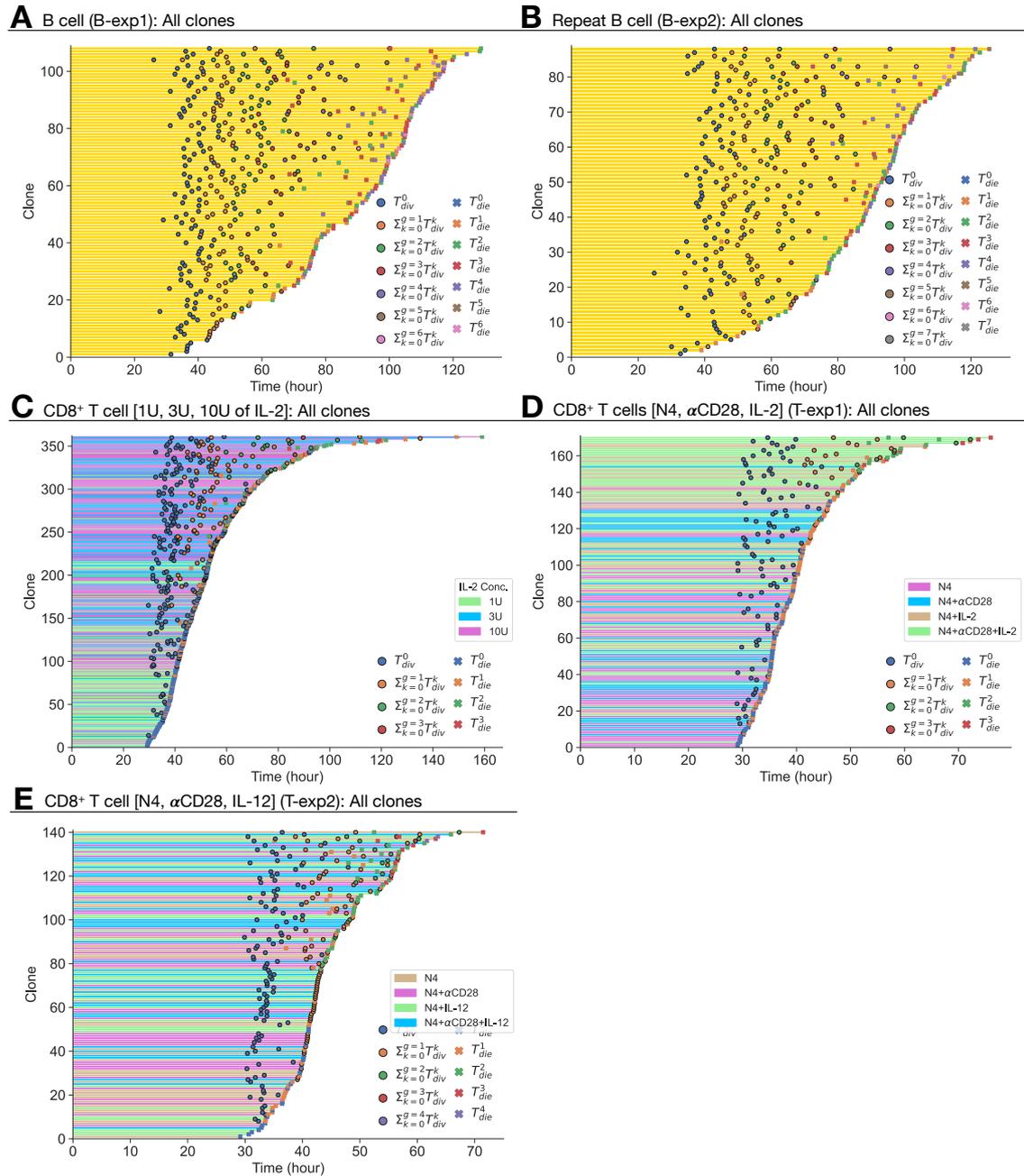


Fig. 2.12: Clonally collapsed clones for *all* founder B and CD8⁺ T cells. For CD8⁺ T cell data with 1U, 3U and 10U of IL-2, measurements from two independent but identical experiment setup are aggregated. The collapsed trees are rank-ordered with respect to the total life time. **(A-F)** Clonally collapsed trees for all families before the filtering. The average division (\bullet) and death (\times) times are marked per family, where the colours represent generation that the event was observed. The lost times are not shown.

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

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	CpG (B-exp1)	69 (63.9%)	56 (51.9%)	69 (63.9%)	69 (63.9%)
	CpG (B-exp2)	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%)

Table 2.4: Number of clones used in the analysis of the filming 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.

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 $g(\langle 0 \rangle) = 0$. 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 2.4.1](#) and [2.4.3](#), 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 have 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

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

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 (Fig. 2.1C). 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 (see Fig. 2.12 for clonally collapsed trees for *all* clones from B and CD8⁺ T cells). Table 2.4 shows the number of retained clones used in all analyses presented in this chapter after applying the filtering rule.

2.4.5 Monte-Carlo simulation

We used Monte-Carlo (MC) method 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).

2.4.6 Bayesian framework

In Section 2.4.1 and 2.4.2, 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 $\mu_x, \mu_y \sim U(0, 1000)$, $\sigma_x, \sigma_y \sim$

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

Bayes Factor: BF_{01} (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)
3 - 10	Moderate evidence for H_0 (H_1)
1 - 3	Anecdotal evidence for H_0 (H_1)
1	No evidence

Table 2.5: Bayes factor interpretation ([Jeffreys, 1961](#)).

$U(0, 1000)$ 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 $\text{BF}_{01} = P(\mathcal{D}|H_0)/P(\mathcal{D}|H_1)$ 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](#)):

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

2.4. ANALYSING DATA TO INVESTIGATE THE MODEL ASSUMPTIONS

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

Table 2.6: List of candidate parametric distribution classes.

where r denotes for the sample correlation defined as $r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$. For the interpretation of Bayes factor, we adopted the discrete categories of evidential strength proposed in [Jeffreys \(1961\)](#) ([Table 2.5](#)).

In [Section 2.4.3](#), 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 2.6](#) 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:

$$\text{WAIC}(z, \Theta) = -2 \left(\sum_{i=1}^n \log \left[\frac{1}{S} \sum_{s=1}^S P(z_i | \Theta_s) \right] - \sum_{i=1}^n \text{Var}_{s=1}^S \log(P(z_i | \Theta_s)) \right), \quad (2.1)$$

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. 2.1](#) 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

2.5. EQUATIONS FOR THE MEAN POPULATION SIZE

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,

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

2.5 Equations for the mean population size

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 2.3](#) 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 $\boldsymbol{\theta} = (T_{div}^0, \{M_g\}_{g \geq 1}, T_{dd}, T_{die})$ denoting time to first division, subsequent division time per generation, time to destiny and time to death, respectively.

Generation zero ($g = 0$)

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

2.5. EQUATIONS FOR THE MEAN POPULATION SIZE

death are useful modifications suited to particular applications. Such modifications are discussed further in [Section 2.7](#).

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

$$Z_0(t) = \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{\min(t, T_{dd}) < T_{div}^0\}}, \quad (2.2)$$

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 2.4.1](#), 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,

$$\mathbb{E}[Z_0(t)] = P(T_{die} > t) \left[P(T_{div}^0 > t)P(T_{dd} > t) + \int_0^t dP(T_{dd} \leq \tau)P(T_{div}^0 > \tau) \right]. \quad (2.3)$$

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.

Subsequent generations ($g > 0$)

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

$$Z_g(t) = 2^g \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{T_{div}^0 + \sum_{k=1}^{g-1} M_k \leq \min(t, T_{dd}) < T_{div}^0 + \sum_{k=1}^g M_k\}}. \quad (2.4)$$

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,

2.5. EQUATIONS FOR THE MEAN POPULATION SIZE

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

$$\begin{aligned} P(X_g \leq \min(t, T_{dd}) < X_g + M_g) &= P(X_g \leq t < X_g + M_g) P(T_{dd} > t) \\ &\quad + P(X_g < T_{dd} < X_g + M_g) P(T_{dd} \leq t). \end{aligned}$$

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

$$\begin{aligned} \mathbb{E}[Z_g(t)] &= 2^g P(T_{die} > t) \times \\ &\left[P(T_{dd} > t) \int_0^t dP(X_g \leq \tau) P(M_g > t - \tau) + \int_0^t dP(T_{dd} \leq \tau) P(X_g < \tau < X_g + M_g) \right]. \end{aligned} \tag{2.5}$$

Together with [Eq. 2.3](#) and [2.5](#), 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

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

where $\boldsymbol{\theta} = (T_{div}^0, \{M_g\}_{g \geq 1}, T_{dd}, T_{die})$ 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 ([Figs. 2.3B, 2.4B, 2.5A1,B1,C1](#) and [2.6A1-4,B1-4](#)). 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

2.5. EQUATIONS FOR THE MEAN POPULATION SIZE

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. 2.3](#) 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. 2.5](#) 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 throughout the thesis.

$$\tilde{y}_g(t; \tilde{\theta}) := N_0 \mathbb{E}[\tilde{Z}_g(t; \tilde{\theta})]. \quad (2.6)$$

2.5.1 Optimisation strategy for estimating the model parameters

Division structured population datasets obtained from FACS were fitted to the reduced Cyton2 model (i.e. [Eq. 2.6](#)). 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,

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

2.6. EQUATIONS FOR THE VARIABILITY AROUND THE MEAN

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 1000 times, which resulted in 1000 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.

2.6 Equations for the variability around the mean

In the previous section, we derived equations for the mean population size at time t and showed that the variability in parameter estimates can be obtained by employing bootstrap method as a part of fitting procedure. There are two sources of variability in the system: measurement variability; and intrinsic variability due to the stochastic nature of the dynamics. The bootstrap method propagates the measurement variability in estimating parameters for a given dataset. Here, we will derive analytic expressions for the variance of the reduced Cyton2 model to examine the intrinsic variability of the system. With Eqs. 2.2 and 2.4, we derive it by substituting to the usual variance formula $\text{Var}(X) = \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ for generation zero and subsequent generations separately. First, let us define the

2.6. EQUATIONS FOR THE VARIABILITY AROUND THE MEAN

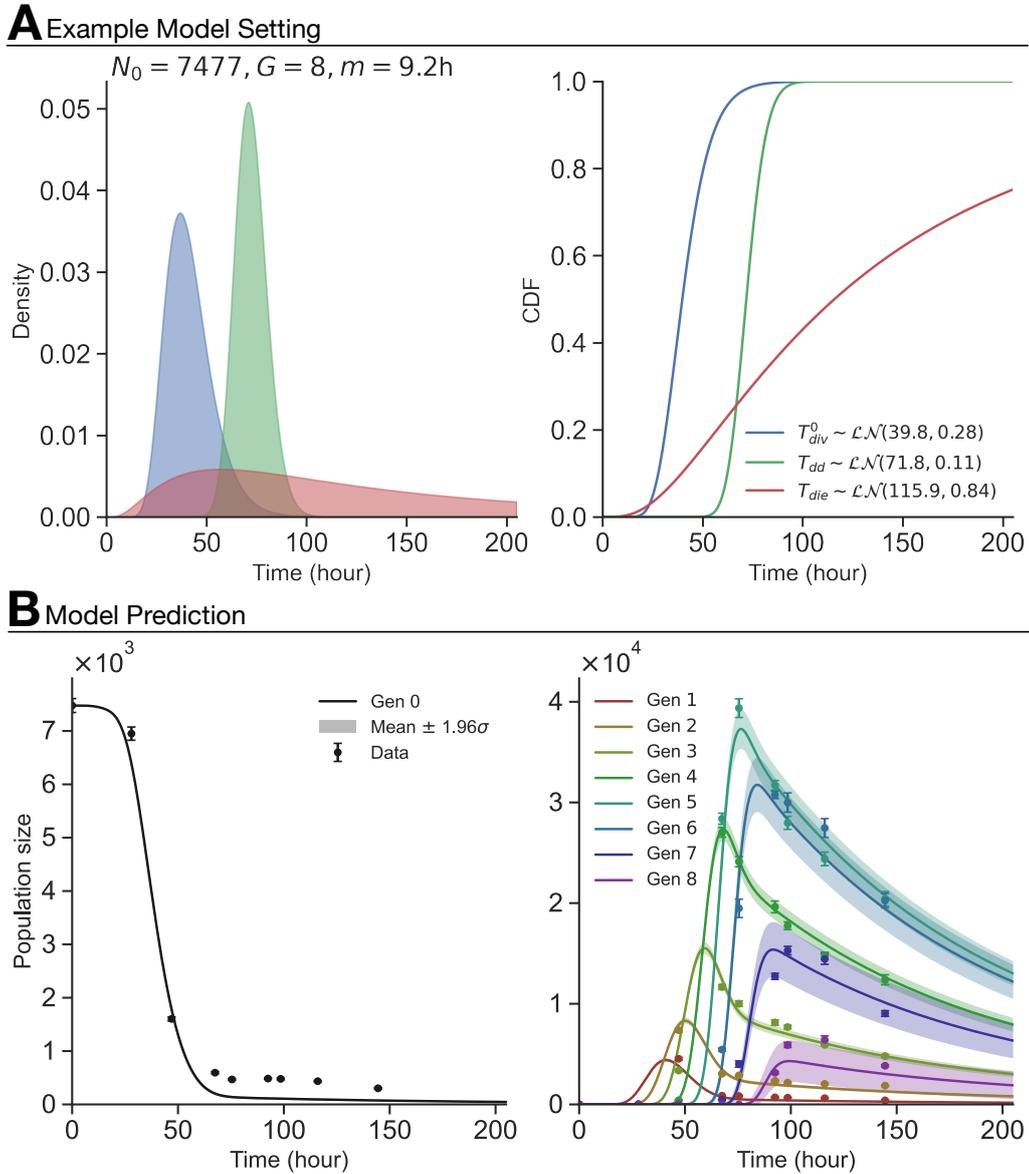


Fig. 2.13: Variability of the reduced Cyton2 model. (A) Example Cyton2 parameter estimates from fitting B cell data presented in Fig. 3.1A3. Initial number of cells (N_0), maximum generation (G) and subsequent division time (m) are annotated as a plot title. Probability density and cumulative distribution functions are shown for T_{div}^0 , T_{dd} and T_{die} . (B) Predicted population size of 0th (left) and 1st through 8th generations (right). Each solid-line represents average number of cells. Shade area represents upper and lower bounds around the mean ($\mu \pm 1.96\sigma$).

2.6. EQUATIONS FOR THE VARIABILITY AROUND THE MEAN

following sets to simplify the equations,

$$\begin{aligned} A &\equiv \{T_{die} > t\}, \\ B &\equiv \{\min(t, T_{dd}) < T_{div}^0\}, \\ C &\equiv \{T_{div}^0 + (g-1)m \leq \min(t, T_{dd}) < T_{div}^0 + gm\}, \end{aligned}$$

such that the primitive form of $\tilde{Z}_g(t)$ becomes,

$$\tilde{Z}_g(t) = \begin{cases} \mathbb{1}_A \mathbb{1}_B & \text{if } g = 0 \\ 2^g \mathbb{1}_A \mathbb{1}_C & \text{if } g > 0 \end{cases}. \quad (2.7)$$

Generation zero. Substituting the $g = 0$ case of Eq. 2.7 into the variance equation yields,

$$\text{Var}(\tilde{Z}_0(t)) = \mathbb{E}[\tilde{Z}_0(t)^2] - \left(\mathbb{E}[\tilde{Z}_0(t)]\right)^2 = \mathbb{E}[\tilde{Z}_0(t)](1 - \mathbb{E}[\tilde{Z}_0(t)]),$$

as $\mathbb{E}[(\mathbb{1}_A \cdot \mathbb{1}_B)^2] = \mathbb{E}[\mathbb{1}_A^2 \cdot \mathbb{1}_B^2] = \mathbb{E}[\mathbb{1}_A \cdot \mathbb{1}_B]$ from the independence assumption. We substitute the expected cell number to obtain the following variance equation for generation zero,

$$\begin{aligned} \text{Var}(\tilde{Z}_0(t)) &= P(T_{die} > t)P(T_{div}^0 > t)P(T_{dd} > t) \\ &\quad \times [1 - P(T_{die} > t)P(T_{div}^0 > t)P(T_{dd} > t)] \\ &+ P(T_{die} > t) \int_0^t f_{T_{dd}}(\tau)P(T_{div}^0 > \tau)d\tau \\ &\quad \times \left[1 - P(T_{die} > t) \int_0^t f_{T_{dd}}(\tau)P(T_{div}^0 > \tau)d\tau\right] \\ &\quad - 2P(T_{die} > t)^2P(T_{div}^0 > t)P(T_{dd} > t) \int_0^t f_{T_{dd}}(\tau)P(T_{div}^0 > \tau)d\tau. \end{aligned}$$

Subsequent generations. Similar to generation zero, we use $g > 0$ of Eq. 2.7 to obtain,

$$\text{Var}(\tilde{Z}_g(t)) = \mathbb{E}[\tilde{Z}_g(t)^2] - \left(\mathbb{E}[\tilde{Z}_g(t)]\right)^2 = \mathbb{E}[\tilde{Z}_g(t)](2^g - \mathbb{E}[\tilde{Z}_g(t)]).$$

2.6. EQUATIONS FOR THE VARIABILITY AROUND THE MEAN

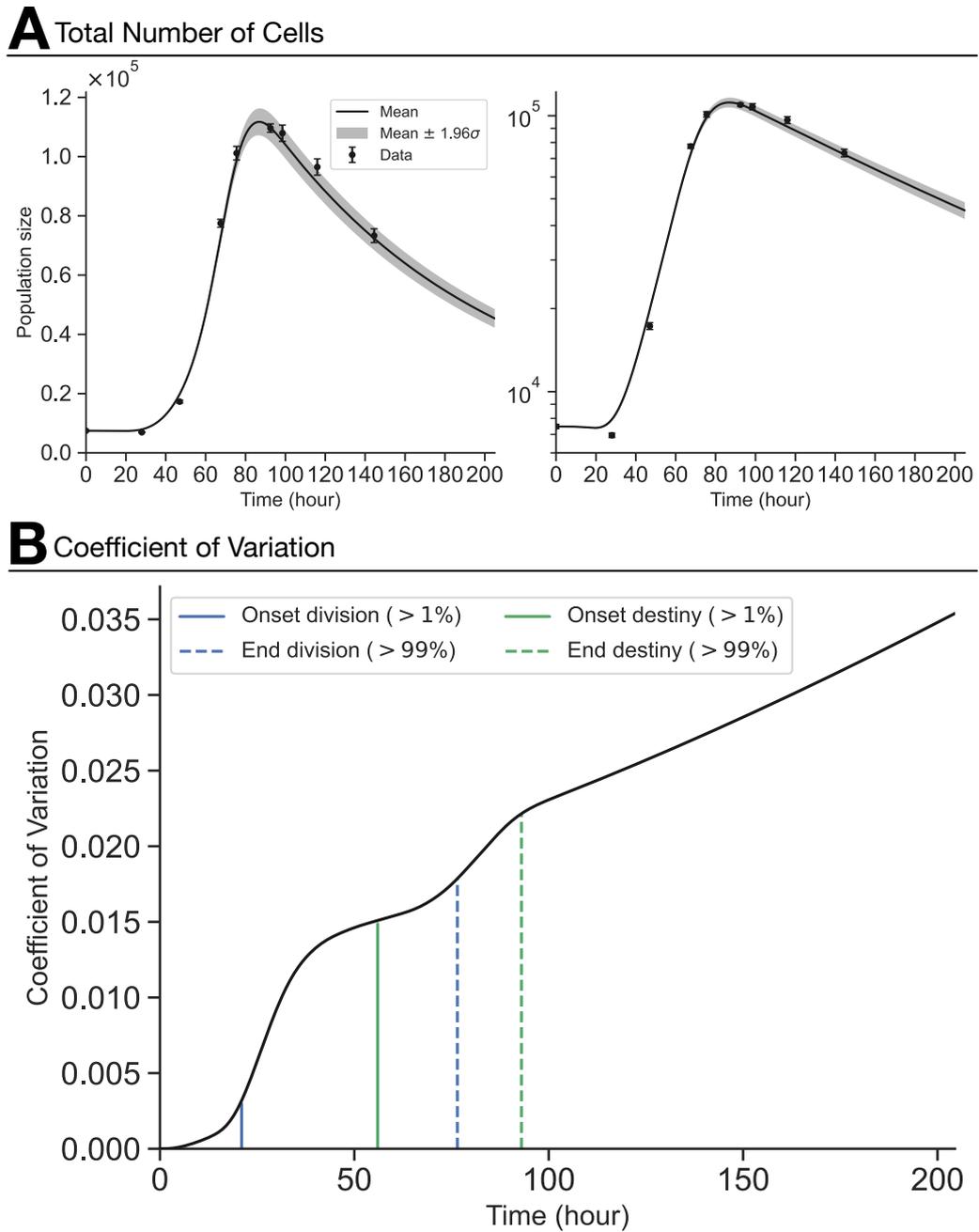


Fig. 2.14: Total number of cells vs time and analysis of the evolution of the variance. (A) Predicted population size. Solid-line represents the sum of average cell numbers through all generations. Shade area represents upper and lower bounds around the mean ($\mu \pm 1.96\sigma$). (B) Coefficient of variation as a function of time. The onsets (more than 1% of population) and ends (more than 99% of population) of division and destiny events are marked as solid- and dashed-lines.

2.6. EQUATIONS FOR THE VARIABILITY AROUND THE MEAN

Cyton1 Parameters										
T_{div}^0		T_{die}^0		$T_{div}^{g>0}$		$T_{die}^{g>0}$		γ_g		
m_{div}^0	s_{div}^0	m_{die}^0	s_{die}^0	$m_{div}^{g>0}$	$s_{div}^{g>0}$	$m_{die}^{g>0}$	$s_{die}^{g>0}$	γ_0	μ	σ
39.81*	0.28*	41.49	0.14	9.2*	0.03	93.34	0.810	1.0	3.72	1.277

Cyton2 Parameters						
T_{div}^0		T_{dd}		T_{die}		m
m_{div}^0	s_{div}^0	m_{dd}	s_{dd}	m_{die}	s_{die}	
39.81*	0.28*	71.82	0.11	115.88	0.84	9.2*

Table 2.7: Cyton1 and Cyton2 parameter settings for MC simulation. The asterisk denotes the parameter values that were set and fixed from the Cyton2 estimates. Other Cyton1 parameters were obtained by fitting the model to the same dataset presented in Fig. 2.13. For Cyton1 simulation, (T_{div}^0, T_{die}^0) denotes times to *first* division and death; $(T_{div}^{g>0}, T_{die}^{g>0})$ denotes times to *subsequent* division and death; and, γ_g is progressor fraction, c.f Eq. 1.1. The times are lognormally distributed with corresponding median and shape parameters.

Thus,

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

Provided initial cell number N_0 , the standard deviation can be computed by $\sigma(\tilde{Z}_g(t)) = \sqrt{N_0 \text{Var}(\tilde{Z}_g(t))}$ for each generation. An example model output for first 8 generations is shown in Fig. 2.13B based on the estimated parameter values from B cell data discussed in Chapter 3 (c.f. Fig. 3.1A3). Given the

2.6. EQUATIONS FOR THE VARIABILITY AROUND THE MEAN

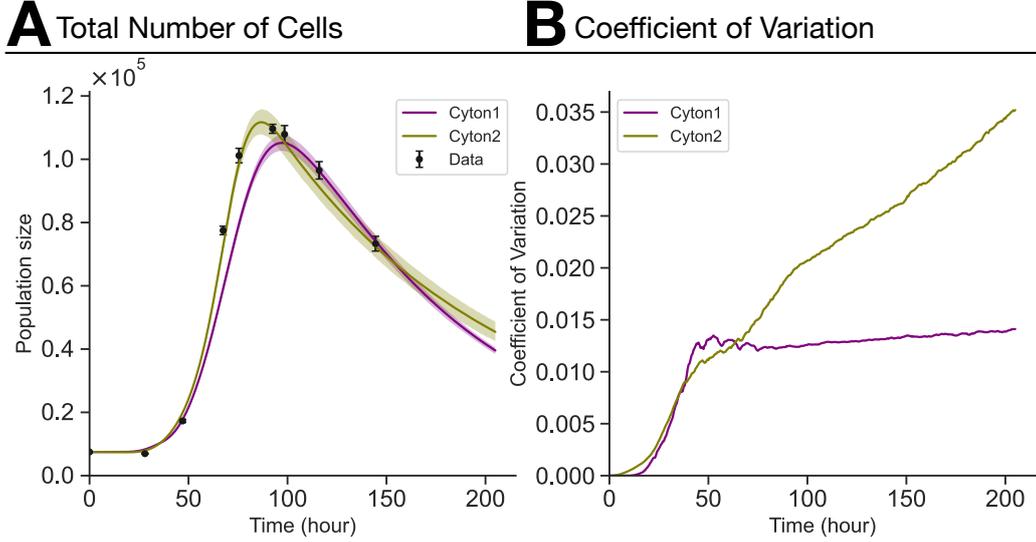


Fig. 2.15: MC simulation of Cyton1 and Cyton2 processes. (A) Average population size (solid line) and 95% confidence band (shade area). (B) Coefficient of variations.

average cell numbers for each generation, the total number of cells is simply the sum $\mu_{\tilde{Z}}(t) := N_0 \sum_{g=0}^G \mathbb{E}[\tilde{Z}_g(t)]$. Similarly, the total variance is given by $\sigma_{\tilde{Z}}^2(t) := N_0 \sum_{g=0}^G \text{Var}(\tilde{Z}_g(t))$ (see Fig. 2.14A). Using the coefficient of variation (CV), defined as

$$\text{CV}(t) = \frac{\sigma_{\tilde{Z}}(t)}{\mu_{\tilde{Z}}(t)} = \frac{\sqrt{N_0 \sum_{g=0}^G \text{Var}(\tilde{Z}_g(t))}}{N_0 \sum_{g=0}^G \mathbb{E}[\tilde{Z}_g(t)]},$$

we evaluated the evolution of the variability in relation to the mean of the cell population. As reported in Subramanian et al. (2008), our results show that the CV increases at two distinct stages: onsets of division and of destiny (Fig. 2.14B). At long time-scales, particularly after the end of destiny stage, the CV continues to rise at a rapid rate, indicating a greater variability due to newly introduced correlation structure. A similar conclusion using branching process formulation was reported in Duffy and Subramanian (2009) that compares different degrees of correlation in a family.

We compared Cyton1 and Cyton2 to investigate the impact of familial correlation in the inherent variability of the population dynamics. To do so, we utilised MC simulation and estimated the variances empirically. As these models have slightly different stochastic process (e.g. generation versus timer mechanism for DD), the

2.7. MODIFICATIONS FOR CALCULATING NUMBER OF CELLS IN GENERATION ZERO

parameters of Cyton1 were matched by fitting the model to the same dataset. However, the time to first division and mean subsequent division time were set and fixed to the values obtained from fitting the Cyton2 model (Fig. 2.15A). The parameter settings for the simulation are summarised in Table 2.7. Here, we ran 1000 iterations and simulated 7477 families per iteration to match the aforementioned model settings. In Fig. 2.15B, the estimated CVs of Cyton1 and Cyton2 as a function of time were plotted. The result indicates that both models follow a similar trend as cells begin to divide and reach DD. However, at long time-scale, the CV from Cyton2 continuously rises, while its value from Cyton1 remains steady despite the average number of cells decreases. Similar conclusion was reported in Subramanian et al. (2008); Duffy and Subramanian (2009).

The variance equations are useful for estimating error in parameter estimation. For example, Milutinović and Boer (2007) considered contribution of the inherent stochasticity of cell dynamics with respect to the total variance of the data fluctuations. Using ordinary differential equations to model an immune response and with Gillespie's simulations, they concluded that a significant portion of the variance originates from the intrinsic noise, thus, suggested an approach to include the noise for parameter estimation. We appreciate the advantage of utilisation of the variance equations, however, the mean equations will be largely used for the following chapters.

2.7 Modifications for calculating number of cells in generation zero

In Section 2.5 we derived a formula for calculating expected number of cells in generation zero, assuming cells are immediately programmed upon activation to adopt RVs for Cyton2 evolution. Here we offer a series of revised formulas for immune activation situations where the calculation of cells in generation zero is more complex.

2.7.1 Partial activation, no death

In many culture arrangements, only a fraction of the cells receive the activation stimulus and adopt the new program governed by Cyton2 parameters. Here the population is first divided into fractions. The number of stimulated cells is given as

2.7. MODIFICATIONS FOR CALCULATING NUMBER OF CELLS IN GENERATION ZERO

$N_0 \cdot p$, where $p \in [0, 1]$ is stimulated cell fraction, and is treated as in [Section 2.5](#) to calculate overall expected number of cells. The unstimulated cells, $N_0(1 - p)$, remain undivided. This introduces one new parameter, p .

In essence, it is equivalent of introducing a Bernoulli RV, $A : \Omega \mapsto \{0, 1\}$, that represents an event of activation for the founder cell in a family tree, where $\mathbb{E}[A] = p \in [0, 1]$ is the probability of activation. Then, the number of cells in a given family $Z_g(t)$ (c.f. [Eq. 2.2](#) and [2.4](#)) needs to be modified to

$$Z_g(t) = \begin{cases} (1 - A) + A \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{\min(t, T_{dd}) < T_{div}^0\}} & \text{for } g = 0 \\ 2^g A \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{T_{div}^0 + \sum_{k=1}^{g-1} M_k \leq \min(t, T_{dd}) < T_{div}^0 + \sum_{k=1}^g M_k\}} & \text{for } g > 0 \end{cases} \quad (2.8)$$

2.7.2 Partial activation, death program

Here the arrangement is as above, but the unstimulated cells will die according to a distinct probability. To do so, we need an extra RV for the unstimulated time to death distribution, T_U . After splitting the population into the fractions by parameter p , the unstimulated cells remain in generation zero, but die according to distribution of T_U . Similar to [Eq. 2.8](#), the modified equation will be

$$Z_g(t) = \begin{cases} (1 - A) \mathbb{1}_{\{T_U > t\}} + A \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{\min(t, T_{dd}) < T_{div}^0\}} & \text{for } g = 0 \\ 2^g A \mathbb{1}_{\{T_{die} > t\}} \mathbb{1}_{\{T_{div}^0 + \sum_{k=1}^{g-1} M_k \leq \min(t, T_{dd}) < T_{div}^0 + \sum_{k=1}^g M_k\}} & \text{for } g > 0 \end{cases} \quad (2.9)$$

2.7.3 Slow reprogramming

In a common situation, cells in culture take some time to integrate the activation signals and reprogram survival from the unstimulated fate. The signals leading to reprogramming are in a race with the initial survival program of the cells. For this we define an activation time for completion of the program and assume that cells behave as if unstimulated to that activation time. If adopting this method, the activation time becomes time zero for the calculation of PDFs/CDFs of RVs. This scenario can also occur when only a fraction of cells are stimulated in culture. Hence, calculation of cells requires values for p , $t_A \in \mathbb{R}_{>0}$ (activation time), and T_U

according to

$$Z_g(t) = \begin{cases} \mathbb{1}_{\{T_U > t\}} & \text{for } t \leq t_A, g = 0 \\ \mathbb{1}_{\{T_U > t_A\}} \left[(1 - A) \mathbb{1}_{\{T_U > t\}} + A \mathbb{1}_{\{X_{die} > t\}} \mathbb{1}_{\{\min(t, X_{dd}) < X_{div}^0\}} \right] & \text{for } t > t_A, g = 0 \\ \mathbb{1}_{\{T_U > t_A\}} \left[2^g A \mathbb{1}_{\{X_{die} > t\}} \mathbb{1}_{\{X_{div}^0 + \sum_{k=1}^{g-1} M_k \leq \min(t, X_{dd}) < X_{div}^0 + \sum_{k=1}^g M_k\}} \right] & \text{for } t > t_A, g > 0 \end{cases} \quad (2.10)$$

where $X_{div}^0, X_{dd}, X_{die} : \Omega \mapsto (t_A, \infty)$ are shifted RVs such that $X = T + t_A$. Note that Eq. 2.10 reduces to Eq. 2.9 when $t_A = 0$.

2.8 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; Hasenauer et al., 2012; Banks et al., 2012; Mazzocco et al., 2017), with a few notable exceptions (Wellard et al., 2010; Zilman et al., 2010; Hyrien 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; Amo et al., 2020). As such, as with the original Cyton model (Hawkins et al., 2007a; 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).

3

Cyton2 applications: Model evaluation; Predicting effects of signal addition

3.1 Abstract

In this chapter, three applications are discussed using B and T cell population datasets obtained from fluorescence-activated cell sorting (FACS) analysis, which is a type of flow cytometry, performed by our WEHI collaborators: (i) We assessed the model using root-mean-squared-error (RMSE) with respect to amount of information in the dataset and reported identifiability of the parameters. Here, the B cell experiment was designed to test robustness and reproducibility of the assay such that larger sample size was collected; (ii) We reanalysed CD8⁺ T cell data published in [Marchingo et al. \(2014\)](#) to show a novel perspective of linear sum of signals from costimulants. This additive process was performed in time domain, that is sum of Cyton2 random variables, and predicted the population dynamics in the presence of multiple costimulatory molecules; and (iii) A primary CD8⁺ T cell dataset was analysed to address whether the additivity of the signal integration can be applied to immunosuppressive drugs, in particular Rapamycin and Dexamethasone which are known for limiting the cell expansion, thereby predicting the negative signal effect when used in combination.

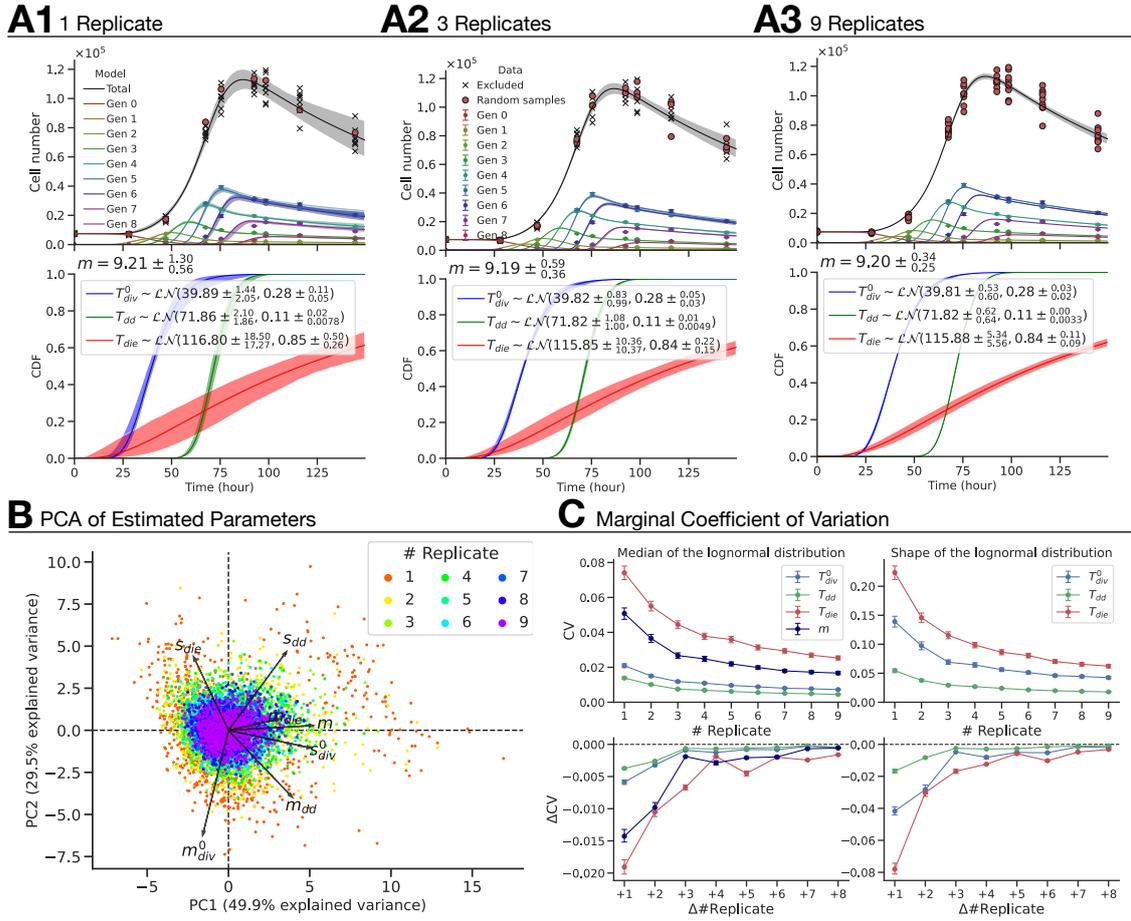


Fig. 3.1: The precision of the parameter estimates with CpG-stimulated $Bim^{-/-}$ murine B cell FACS data. (A1-3) The best-fit reduced Cyton2 model (top), which has seven fitted parameters (bottom): $T_{div}^0 \sim \mathcal{LN}(m_{div}^0, s_{div}^0)$, $T_{dd} \sim \mathcal{LN}(m_{dd}, s_{dd})$, $T_{die} \sim \mathcal{LN}(m_{die}, s_{die})$ and, subsequent division time, m . For a given replicate number, the model was fitted to 1000 synthetic datasets, which were created by randomly sampling the original data with replacement per time point. (B) From the sets of estimated parameter vectors, biplot of principle component analysis (PCA) result is shown. (C) The marginal coefficients of variation (CV) was calculated with 95% confidence interval from bootstrapping.

3.2 Precision and accuracy of the model estimates using B cell data

While information from time-lapse microscopy in [Chapter 2](#) 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 [Section 2.5](#), for fitting to such data to challenge Cyton2's applicability. We challenged the model with both B and CD8⁺ T cell datasets.

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.

3.2. PRECISION AND ACCURACY OF THE MODEL ESTIMATES USING B CELL DATA

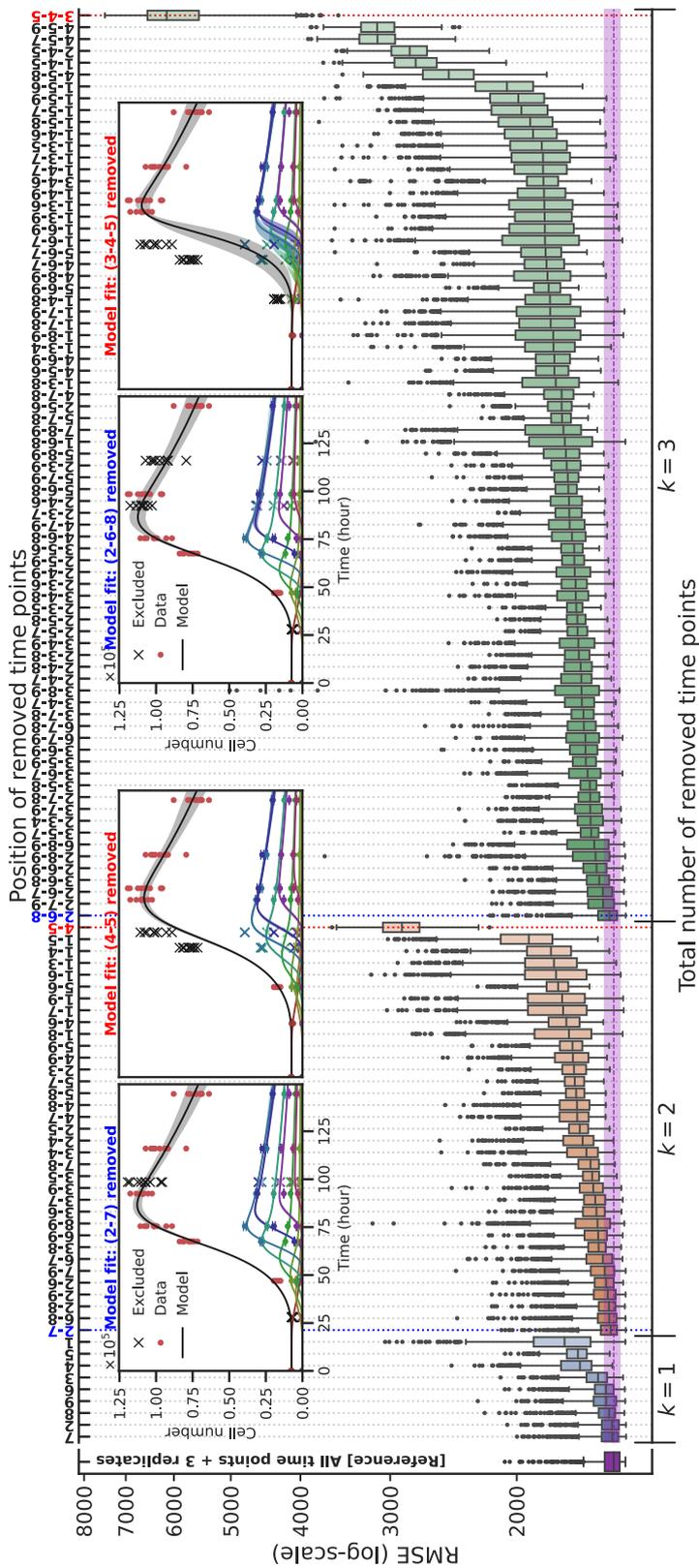


Fig. 3.2: The accuracy of the model fit with CpG-stimulated $Bim^{-/-}$ B cell FACS data. The root-mean-square error (RMSE) was evaluated over all available data points after fitting the reduced Cyton2 model to the synthetic datasets. The reference (*purple*) was obtained after fitting the model to datasets 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.

3.2. PRECISION AND ACCURACY OF THE MODEL ESTIMATES USING B CELL DATA

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 [Fig. 3.1A1-3](#), the best-fit model and the estimated parameters with 95% confidence intervals are shown (see [Section 2.5.1](#) in [Chapter 2](#)). 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 component analysis (PCA) ([Fig. 3.1B](#)). 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 ([Fig. 3.1C](#)). 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 ([Fig. 3.2](#)). 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 1000 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 [Fig. 3.2](#), we present rank ordered values of the RMSE when one, two or three time-points are removed (see [Fig. 3.3](#) for greater than three). The RMSE of model fits using all time points with three replicates is also shown as a

3.2. PRECISION AND ACCURACY OF THE MODEL ESTIMATES USING B CELL DATA

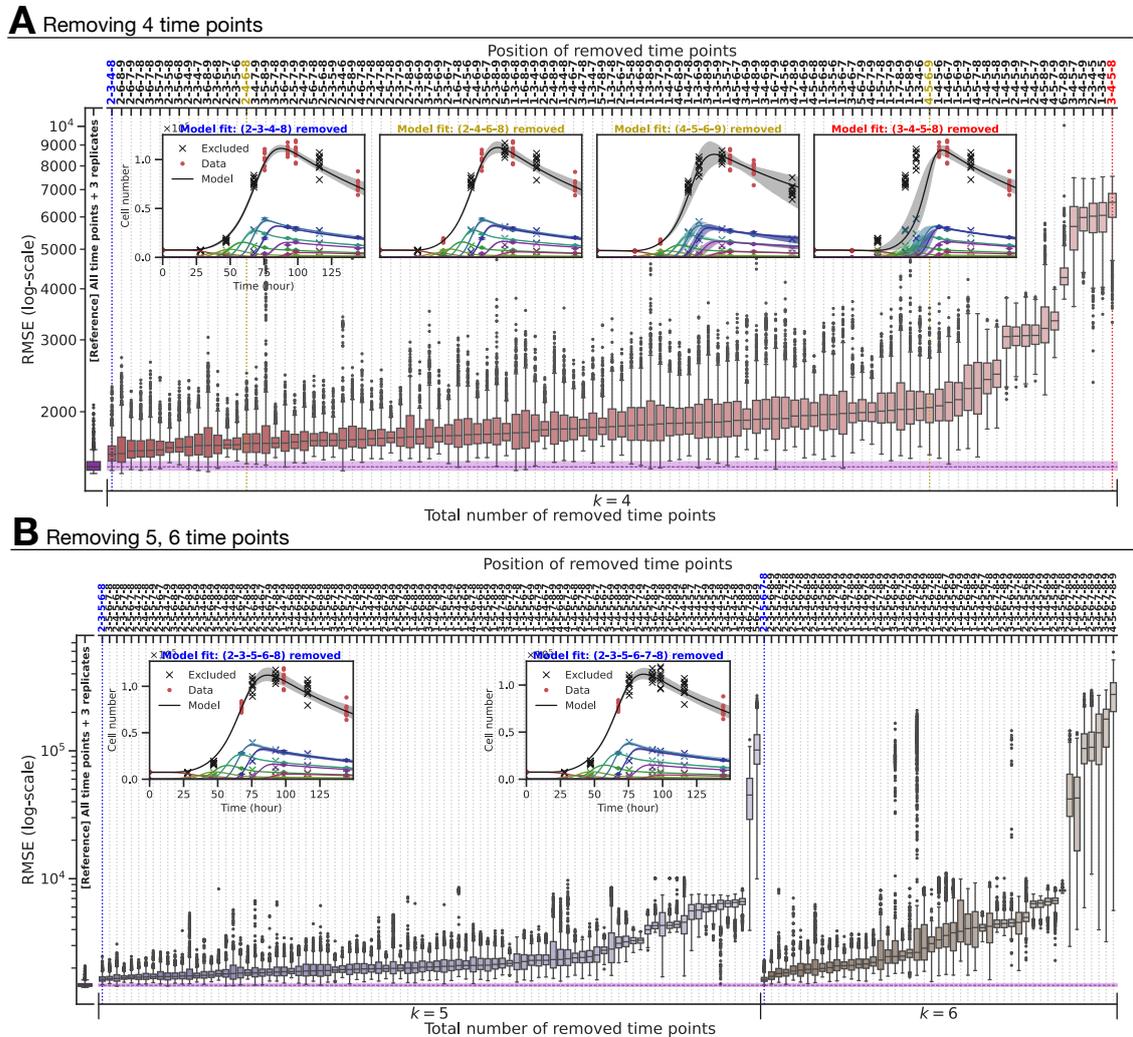


Fig. 3.3: The accuracy of the model fit with CpG-stimulated $Bim^{-/-}$ B cell FACS data. The root-mean-squared error (RMSE) was evaluated over all available data points after fitting the reduced Cyton2 model to the synthetic datasets. Similarly, the reference RMSE (purple) was obtained after fitting the model to datasets assuming only three replicates are available while maintaining all time points. **(A)** All possible combinations of positions of time points for $k = 4$ case. The best (blue) and worst (red) examples of Cyton fits are shown. **(B)** Best examples for $k = 5, 6$ are shown. The worst cases failed to provide good fits (not shown), resulting in large confidence bands and an order of magnitude difference in RMSE.

3.3. POSITIVE SIGNAL INTEGRATION FROM MULTIPLE STIMULI IN CD8⁺ T CELL

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 Fig. 3.3B). 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.

3.3 Positive signal integration from multiple stimuli in CD8⁺ T cell

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 (5µg/mL) and CD28 (2µg/mL), both alone and in combination. Cells were harvested at 27h, 44h, 52.5h, 66.5h, 69h, 76.5h, 90h, 101h and 115.5h 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

3.3. POSITIVE SIGNAL INTEGRATION FROM MULTIPLE STIMULI IN CD8⁺ T CELL

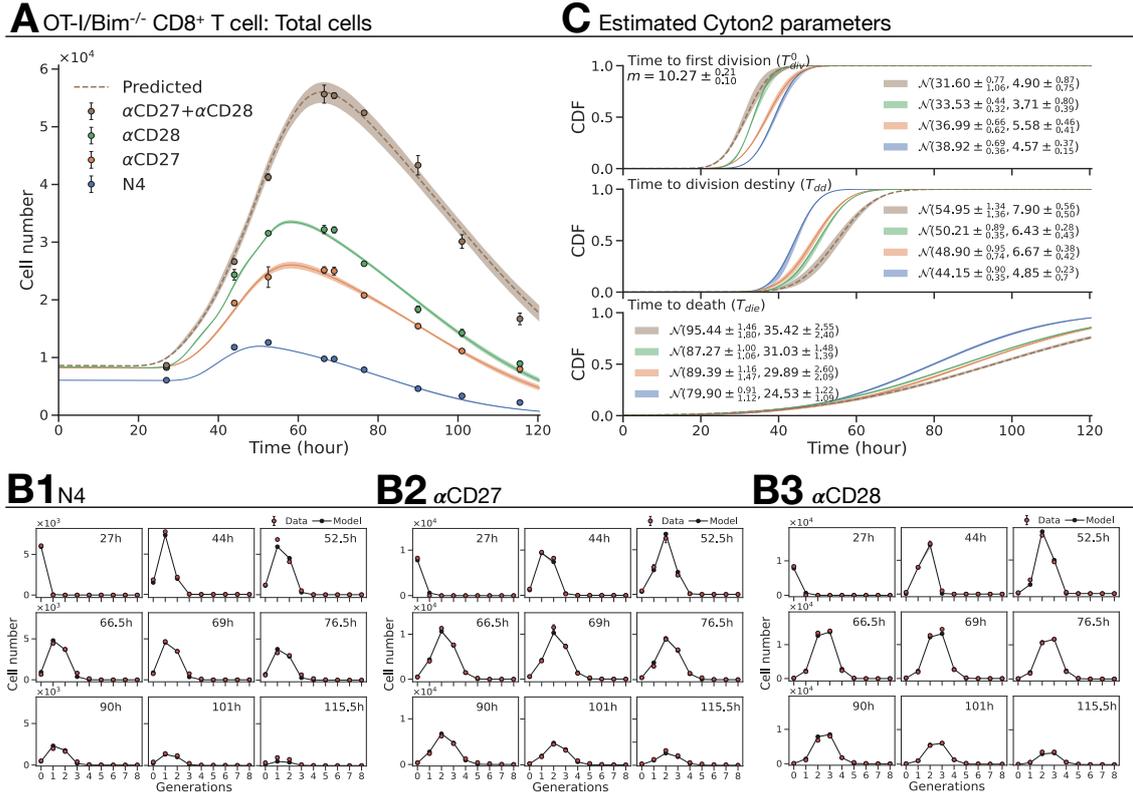


Fig. 3.4: Fitting the Cyton2 model to OT-I/Bim^{-/-} CD8⁺ T cell FACS data (Marchingo et al., 2014). The cells were stimulated with N4 as basis for all other conditions. **(A)** Harvested total cell numbers (\bullet : mean \pm SEM) overlaid with the model extrapolation and 95% confidence band from bootstrapping. **(B1-3)** 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, α CD27 and α CD28; 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.

3.3. POSITIVE SIGNAL INTEGRATION FROM MULTIPLE STIMULI IN CD8⁺ T CELL

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

$$T_{dd}^{\alpha CD27+\alpha CD28} \sim \mathcal{N}\left(\mu_{N4} + \Delta\mu_{\alpha CD27} + \Delta\mu_{\alpha CD28}, \sqrt{\sigma_{N4}^2 + \Delta\sigma_{\alpha CD27}^2 + \Delta\sigma_{\alpha CD28}^2}\right), \quad (3.1)$$

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 [Fig. 3.4A](#), 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 [Section 2.5.1 in Chapter 2](#)) ([Fig. 3.4B1-3](#)), 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 [Fig. 3.4C](#). 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. 3.1](#). Strikingly, this successfully recreated the expansion kinetics of OT-I/Bim^{-/-} CD8⁺ T cells in the presence of both $\alpha CD27$ and $\alpha CD28$ ([Fig. 3.4A](#)), 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 Monte-Carlo simulation given the fitted and predicted parameter estimates ([Fig. 3.5](#)). These results illustrate the merit of Cyton2 in uncovering how simple operations can underlie highly non-linear population dynamics.

3.3. POSITIVE SIGNAL INTEGRATION FROM MULTIPLE STIMULI IN CD8⁺ T CELL

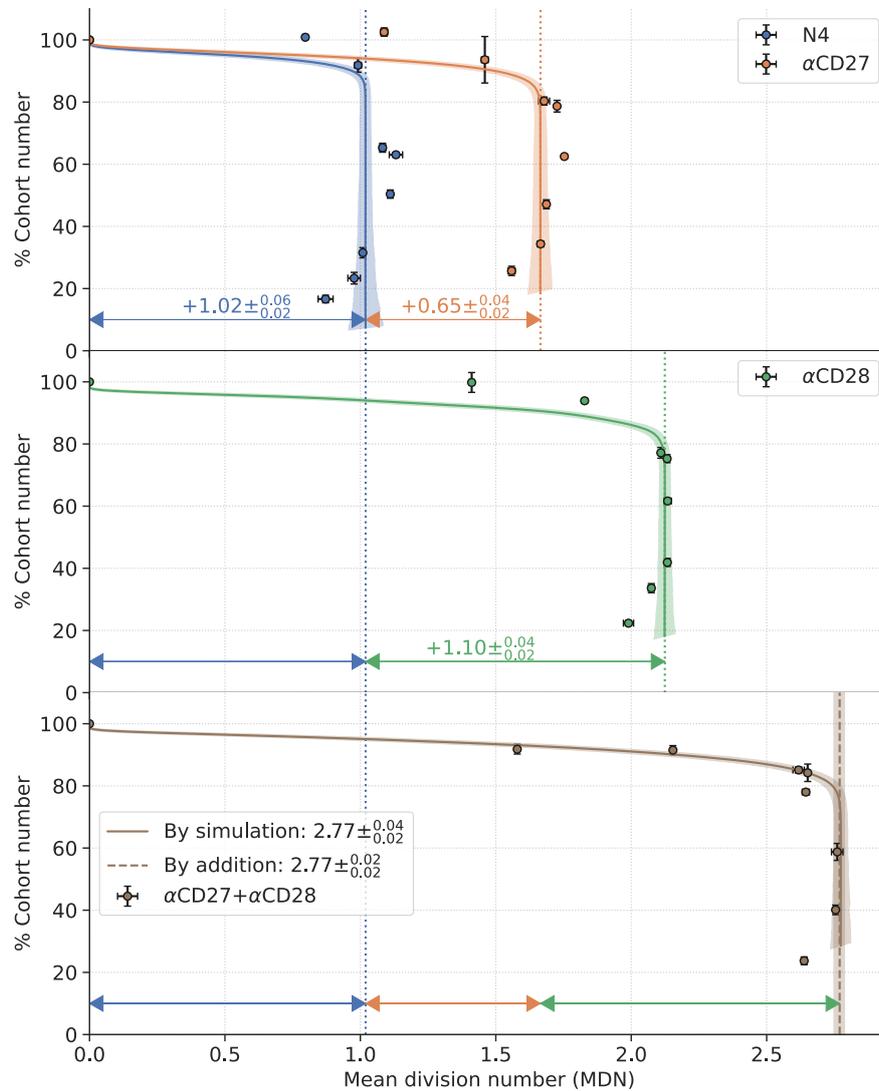


Fig. 3.5: Linear sum of the signals from simulated trees (correspond to Fig.3 in Marchingo et al. (2014)). For each harvested time points from FACS data, percentage cohort number (with respect to the first time point) as a function of MDN is shown (●: mean \pm SEM). The Agent-Based Model was used to generate family trees. The times to first, to destiny and to death were randomly sampled from the estimated *normal* distributions by fitting the reduced Cyton2 model to N4, α CD27 and α CD28 datasets simultaneously. Note that the subsequent division time was set as a constant to be consistent with the model. To match the data, each round of the simulation was initialised with 6066 (N4), 8252 (α CD27) and 8377 (α CD28) clones and ran for $t \in [0, 140]$ with $\Delta t = 0.5$ in hours. This process was repeated 1000 times to obtain 95% confidence bands around the mean. Increase in MDN is labelled in each panel with arrows. The predicted MDNs for α CD27+ α CD28 were calculated either by summing the contribution from each individual stimulation (---) or simulating the trees directly with the summed timers (—).

3.4 Predicting negative signal addition from immunosuppressive drugs

In the previous section, we found that the combined effect of multiple costimulatory signals manifests as a linear sum of contributions from each stimulus with respect to the timers. This additivity recapitulated the larger expansion of the cell population that was induced from the combination of two “*positive*” milder stimuli. The importance of understanding and predicting such dynamics can be extended to clinical application such as personalised chemotherapy for immunocompromised patients who suffer from lack of responding cells to fight off infections. However, it is also possible for the immune cells to overrespond, as seen from autoimmunity, allergy or organ transplantation, in which case it is often desired to suppress the reaction to reduce the damage to nearby healthy cells. In this section, we present an illustration of the model to shed light on cell population dynamics under the influences that work in combination. To do so, we examined FACS data collected from our WEHI collaborators, where they investigated the effect of two immunosuppressive compounds known as Rapamycin and Dexamethasone on antigen-induced proliferation of CD8⁺ T cells from OT-I/Bim^{-/-} transgenic mice. The experimental details are shown in [Appendix A.2](#).

Rapamycin (Sirolimus) was first isolated in 1972 from *Streptomyces hygroscopicus* bacterium found in Easter Island (Rapa Nui) ([Vézina et al., 1975](#)) and is now commonly used in preventing kidney organ transplant rejection ([Ballou and Lin, 2008](#)). It achieves its immunosuppressant function by inhibiting cell cycle progression that is usually motivated by the protein *mammalian Target of Rapamycin* (mTOR) ([Fingar et al., 2004](#)), which results in reduced T cell activation and proliferation ([Zheng et al., 2007](#); [Finlay et al., 2012](#)).

Dexamethasone (a corticosteroid) was first synthesised in 1950s, and has been used for treating autoimmune diseases such as rheumatoid arthritis, multiple sclerosis and psoriasis ([Coutinho and Chapman, 2011](#)). Although it is one of the most widely used drugs, the underlying mechanism behind immunosuppression is not well understood. Studies in [Galili et al. \(1980\)](#); [Purton et al. \(2004\)](#); [Giles et al. \(2018\)](#); [Chen et al. \(2018\)](#) reported a reduction in T cell survival through apoptosis, however, it still remains unclear if there exists any effect on cell proliferation. An analysis from our collaborator revealed that the decrease in total cell population

3.4. PREDICTING NEGATIVE SIGNAL ADDITION FROM IMMUNOSUPPRESSIVE DRUGS

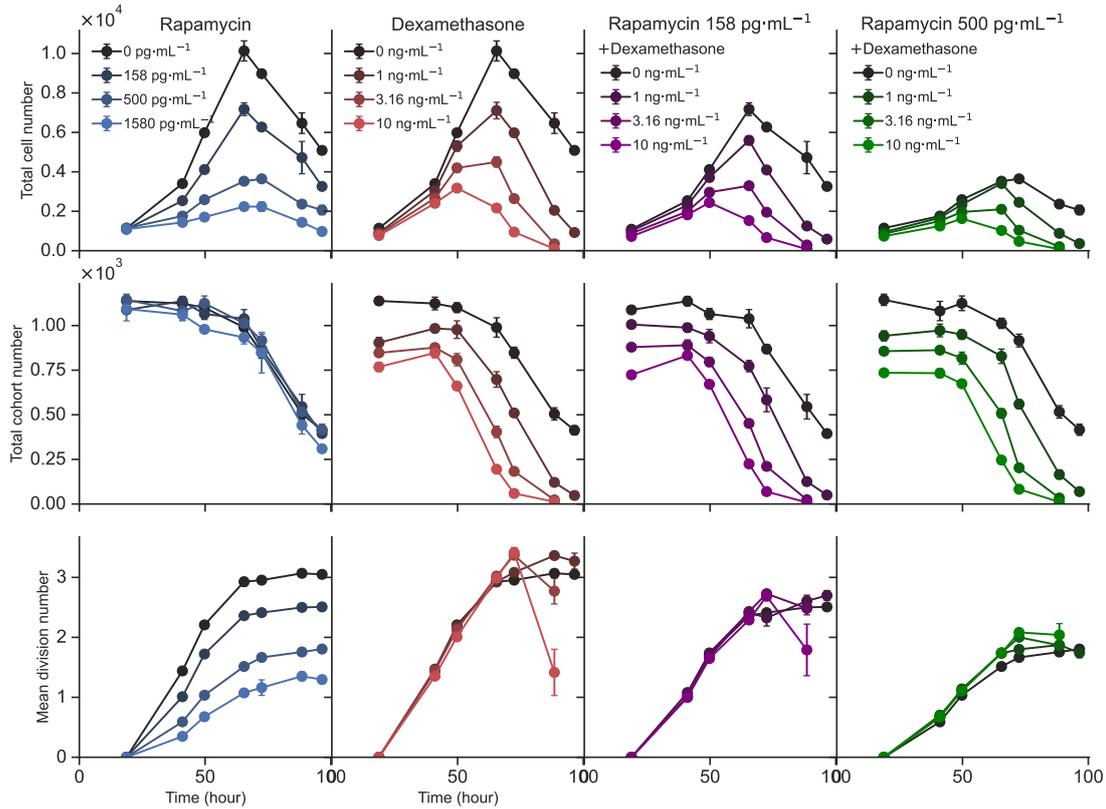


Fig. 3.6: FACS data and summary statistics of OT-I/Bim^{-/-} CD8⁺ T cells in the presence of Rapamycin, Dexamethasone or both. (*top-row*) Dose-dependent reduction in total cell numbers are shown for Rapamycin alone, Dexamethasone alone, two fixed Rapamycin concentrations as bases for varying Dexamethasone. (*mid-row*) Summary statistics of cell survival by cohort numbers, i.e. removal of doubling effect. While Dexamethasone shows earlier cell death as dosage increases, Rapamycin has little to no effect. (*bot-row*) Summary statistics of cell proliferation by mean division number, i.e. average number of division rounds. CD8⁺ T cells undergo similar amount of division rounds irrespective of Dexamethasone concentration, whereas significant reduction in the division counts for higher Rapamycin concentration. Consequently, both drugs suppress cell expansion through reduction in the survival or proliferation, and they achieve so without affecting one another.

3.4. PREDICTING NEGATIVE SIGNAL ADDITION FROM IMMUNOSUPPRESSIVE DRUGS

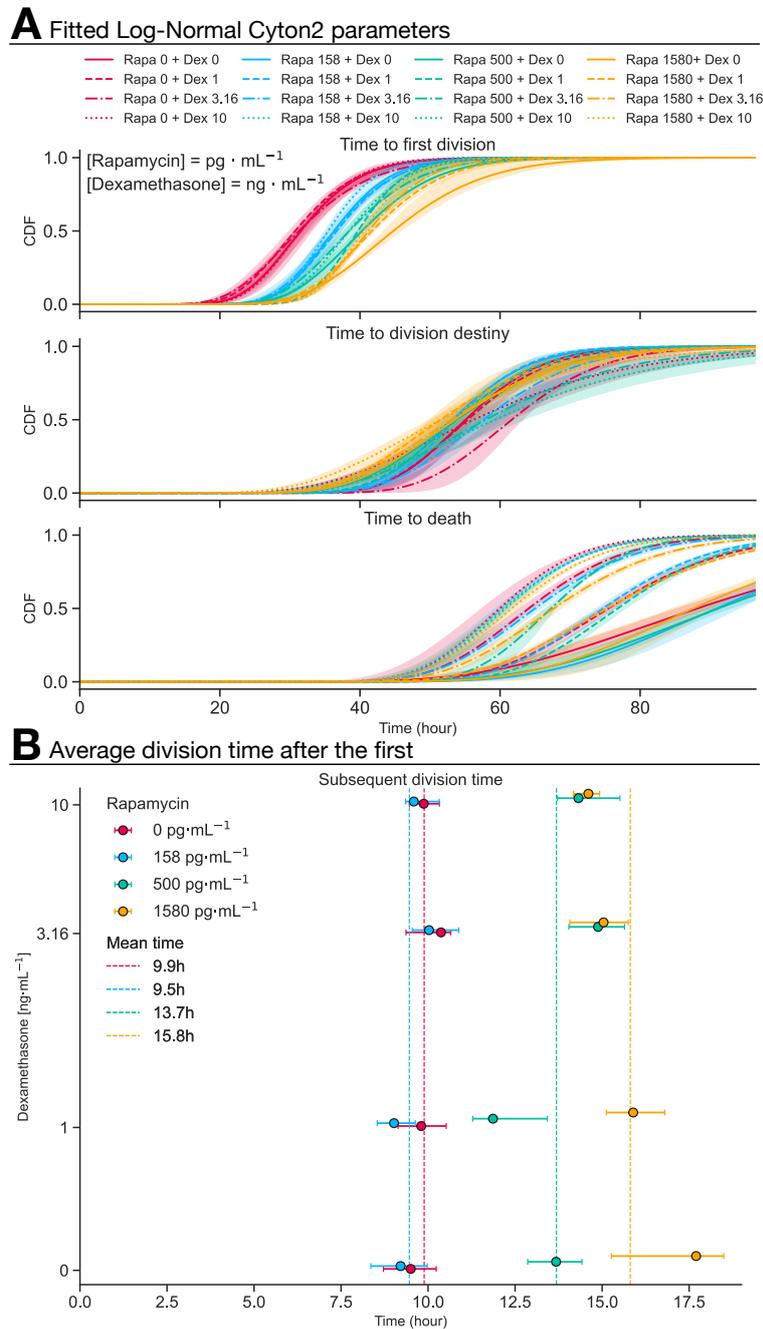


Fig. 3.7: Parameter estimates of Cyton2 model from OT-I/*Bim*^{-/-} CD8⁺ T cells in the presence of Rapamycin, Dexamethasone or both. (A) Estimated Log-normal distributions of time to first division, time to division destiny and time to death. The same colour represents an equal concentration of Rapamycin. Different line styles were used to indicate specific concentration of Dexamethasone. Shaded area around each line represent 95% confidence band. (B) Estimated subsequent division time. The same colour scheme was adopted for consistency. A vertical dashed-line represents the mean time for a given Rapamycin concentration.

3.4. PREDICTING NEGATIVE SIGNAL ADDITION FROM IMMUNOSUPPRESSIVE DRUGS

numbers of stimulated T cells was not mediated by changes in the proliferation, but rather by shortening of their life time (Lye, 2016, Chapter 4).

To use Cyton2 to assess the effect of each of these drugs, alone and in combination, on proliferation and survival, time series experiments were performed and analysed by our collaborator at WEHI. CTV-labeled OT-I/Bim^{-/-} CD8⁺ T cells were initially stimulated with N4 peptide and α CD28 (6.32 μ g/mL) as base positive stimuli to induce the cell expansion. These cells were further stimulated with or without various concentrations of Rapamycin (0, 15, 500, 1580 pg/mL) and Dexamethasone (0, 1, 3.16, 10 ng/mL) after which they were harvested at 18.75h, 41h, 49.75h, 65.5h, 72.5h, 88.5h, 96.5h for all three replicates per time point (see *top-row* of Fig. 3.6). We observed a decrease in cell population for both drugs. To further investigate the effects, we computed total cohort and mean division numbers (MDN) at each harvested time point, t , given by,

$$C(t) = \sum_{g=0}^G c_g(t) = \sum_{g=0}^G \frac{n_g(t)}{2^g},$$

$$\text{MDN}(t) = \frac{\sum_{g=0}^G g \cdot c_g(t)}{\sum_{g=0}^G c_g(t)},$$

where $n_g(t)$ and $c_g(t)$ are the cell number and cohort number, respectively, in generation g at time t ; and G is the maximum observed generation from the FACS dataset. As $C(t)$ removes the expansion effect contributed from each cell family in the population, this is a monotonically decreasing function that stays constant if no cell death has occurred. In *mid-* and *bottom-*rows of Fig. 3.6, we observed that Rapamycin significantly reduces MDN while having little to no effect on cell death. Conversely, Dexamethasone did not affect MDN and achieved cell reduction via inducing early cell death as indicated by a rapid drop in $C(t)$. More importantly, these drugs act independently when used in combination, that is the magnitude of the reduction is proportional to the contributions from Rapamycin on MDN and Dexamethasone on the cell death.

To assess the effects quantitatively, we fitted the Cyton2 model and extracted the model parameters as described in Section 2.5.1. Here, the best fit model was obtained per drug condition prescribed with three log-normal distributions for times to first division, to division destiny and to death, and a constant subsequent division time. We summarised the parameter estimates in Fig. 3.7. The results indicate

3.4. PREDICTING NEGATIVE SIGNAL ADDITION FROM IMMUNOSUPPRESSIVE DRUGS

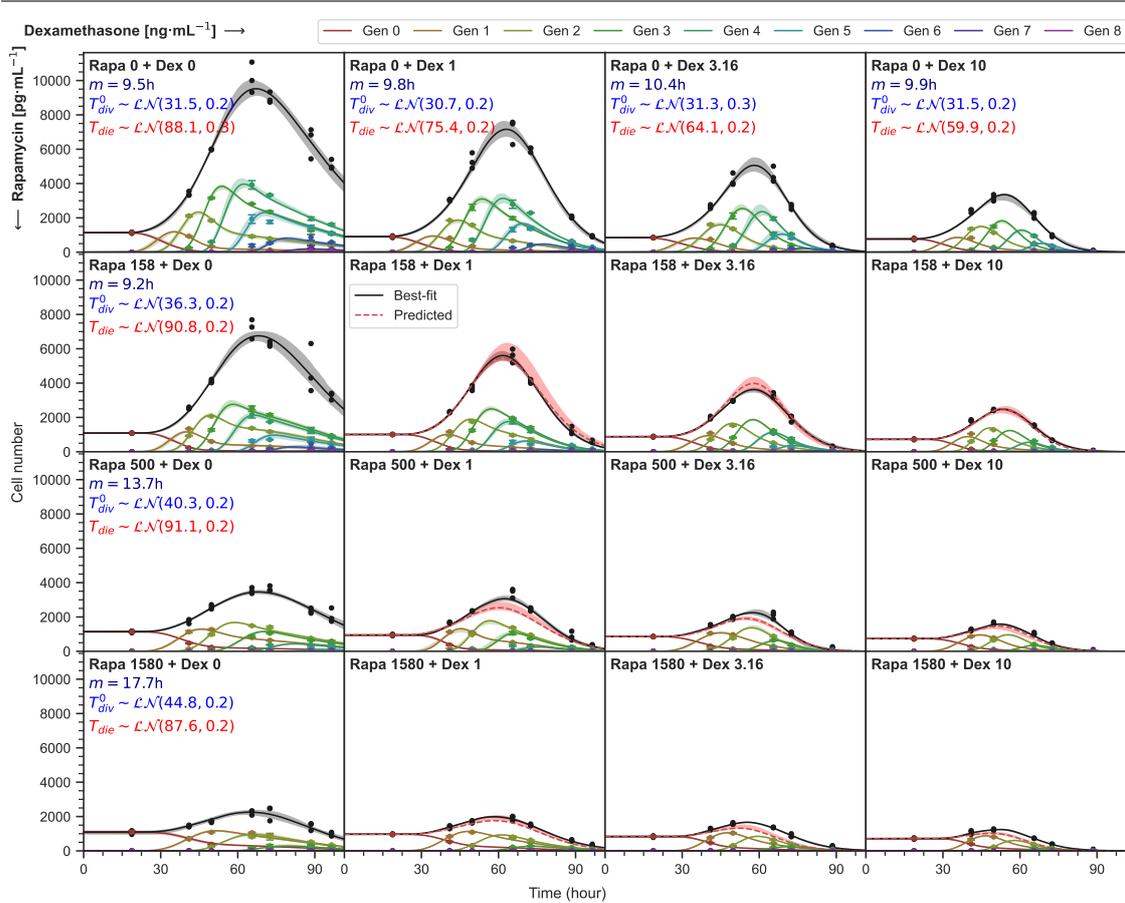


Fig. 3.8: Predicting cell population size in combination of Rapamycin and Dexamethasone. Each panel shows the effect of one or combined drug in cell population size. Top-left panel is the control i.e. without the drugs. The best-fit model (*solid-line*) and predicted cell numbers (*dashed-line*) are shown. Shade area around the curves represent 95% confidence band.

that the division times did not change with respect to Dexamethasone but mean division times slow down due to Rapamycin, resulting in lower number of cells in the population. Conversely, an earlier time to death was observed for higher Dexamethasone concentration, while division times were unaffected with Rapamycin. However, the time to division destiny appears to have no apparent alteration with either of the drugs. This is consistent with the observation in Fig. 3.6 that the drugs operate independently and that all three timers, for division, death and destiny can be regulated independently in the cell. This result also strongly supports the translation of our interpretation into the time domain, as seen similarly in the previous section for costimuli. To predict cell numbers for the combined effect, we

first took parameter estimates from fitting Rapamycin or Dexamethasone alone (see *top-row* and *first-column* of [Fig. 3.8](#)), then replaced the division times according to respective Rapamycin concentration, and similarly, replaced the death time according to corresponding Dexamethasone dosage. For example, Rapamycin 158 pg/mL plus Dexamethasone 1 ng/mL would require estimates from Rapamycin 158 pg/mL alone for division times and from Dexamethasone 1 ng/mL alone for death time. Average time to division destiny was used for the prediction. Strikingly, this simple operation captured OT-I/Bim^{-/-} CD8⁺ T cells under the influence of two immunosuppressive drugs in combination. These results further illustrate how Cyton2 unveils complex non-linear effect of negative signal addition with a simple operation.

3.5 Discussion

In summary, we extracted significant biological meaning from time series data from different stimulation conditions using the Cyton2 model for *in vitro* B and T cell populations. The model was quantitatively evaluated by fitting artificial datasets randomly and conditionally sampled from the observed dataset. The precision of the parameter estimate increases as we introduce more replicates into the procedure. This was qualitatively manifested as a small cluster in size from PCA analysis and tightened lower and upper bounds of 95% confidence intervals in cell number extrapolation. We also found that the model can accurately capture the B cell population dynamics even with limited information. It was shown that contributing factors are availability for data points at early, expansion and contraction phases of the dynamics. So long as these points are present (as low as three time points, see [Fig. 3.3](#)), the model is capable of estimating parameters with reasonable confidence.

In [Marchingo et al. \(2014\)](#), the authors deduced that CD8⁺ T cell populations follows a simple additive rule in handling multiple signals from costimulatory receptors. To do so, they calculated a summary statistics, mean division number, and showed that the final outcome (i.e. α CD27 plus α CD28) can be predicted by algebraic linear sum of those statistics. We reanalysed the same dataset and successfully captured the dynamics, but with an alternate methodology. Here, we estimated parameters of the RVs from each stimulation (N4, α CD27, α CD28) by simultaneously fitting the model, then add the contributions linearly in time to predict CD8⁺ T cells stimulated with α CD27 plus α CD28.

A natural question arises from the signal addition is whether it can be applied

to suppressive signals. An experiment utilising CD8⁺ T cell proliferation was used to address this question. Here, the model was capable of predicting the effect of two immunosuppressive drugs, Rapamycin and Dexamethasone, in combination given the effects of individual drug.

All analyses performed in this chapter were on the datasets obtained from *in vitro* murine experiments. In biology, it is common to extend this type of experiment to human or human-like systems for its applicability in mind such as a new drug development. As such, we will discuss further merit of Cyton2 applied to human B and T cells in [Chapter 5](#).

4

Beyond lymphocyte proliferation and population dynamics: B cell differentiation

4.1 Abstract

In this chapter, we present work, some of which appeared in [Horton et al. \(2022\)](#), that was accomplished in collaboration with partners from Prof. Philip Hodgkin's lab at Walter and Eliza Hall Institute (WEHI). The publication concentrated on the biological insights that can be drawn from the quantitative analysis of clonal B cell stimulation data. Here, we expand upon and extend the development, and application, of the mathematical models and techniques that were required and that we principally developed. Up to this chapter, we considered population dynamics that only involve clonal expansion and contraction of activated lymphocytes. However, a successful immune response involves differentiation to effector cells, for example antibody-secreting cells (ASCs) for B cells, in order to eradicate foreign pathogens. Thus, if experimental methods can be designed that can record measurements on differentiation, it is natural to incorporate the differentiation process into modelling population dynamics. Furthermore, it is of interest for modelling purposes to know how patterns of differentiation arise within families. To acquire such data requires some form of lineage tracing. With collaborators at WEHI having developed a clonal multiplex assay for lymphocytes ([Marchingo et al., 2016](#); [Horton et al., 2018](#)), and has since been developed for hematopoietic stem cells ([Tak et al., 2021](#)), that has the potential to record information on molecular components as well as division numbers within clonal families, the development of a mathematical and statistical framework to assess the resulting information was needed. We analysed *in vitro* clonal samples

4.2. EXPLORING B CELL DIFFERENTIATION

from murine B cells using multiplex cell-labelling technique and single-cell sorting to dissect the characteristics of ASC differentiation and antibody isotype switching at the clonal and molecular level. The regulation of some of the key molecules that determine the cell's fate were incorporated in a simple probabilistic model to predict single or concurrent switching frequency under varying in time conditions.

4.2 Exploring B cell differentiation

A proportion of naïve B cells during the proliferation initiates an isotype class switch in antibody type (e.g. IgM to IgG), and differentiate into replicating (plasmablast), non-replicating (plasma cell) antibody-secreting cells (ASCs) and long-lived memory cells (Stavnezer et al., 2008), thereby creating broad heterogeneous populations of effector B cells (Kallies et al., 2004; Hasbold et al., 2004). A similar process is observed for CD4⁺ and CD8⁺ T cells, leading to functionally different role in their effector states such as releasing cytokines and initiating cytotoxic activity (Seder and Ahmed, 2003; Yates et al., 2004; Kaech and Wherry, 2007; Zhu et al., 2010; Gerlach et al., 2010, 2013; Verdon et al., 2020).

Various experimental techniques have been employed for observing B cell differentiation. One commonly used method is to stain for markers usually in combinations, such as Syndecan-1, B220, CD19, CD38, CD43 and CD63L, as indirect indicators of a cell having differentiated to become an ASC and measure their expression levels with flow cytometry (Sanderson et al., 1989; Tangye et al., 2003c; Taylor et al., 2015; Barwick et al., 2016; Krishnamurty et al., 2016; Boonyaratanakornkit and Taylor, 2019). Alternatively, Kallies et al. (2004) introduced *B lymphocyte induced maturation protein-1* (Blimp-1) GFP reporter mice by utilising the unique characteristic that Blimp-1 is rapidly induced during the B cell differentiation into ASCs (Turner et al., 1994; Shapiro-Shelef et al., 2003, 2005; Wang et al., 2019). The latter technique, together with time-lapse microscopy, was extensively utilised in studies, for example, Fairfax et al. (2007); Duffy et al. (2012); Zhou et al. (2018) to shed light on the mechanism of B cell fate determination.

To examine differentiation, B cells need to become activated and undergo clonal expansion. Here, we present two methods that are commonly used in the lab environment to induce such reactions in the absence of BCR stimulation by antigen. One way to achieve this is to provide strong signals from via Toll-like receptors (TLRs) using *lipopolysaccharides* (LPS). Unlike CpG DNA, another TLR agonist

4.2. EXPLORING B CELL DIFFERENTIATION

that induces proliferation without differentiation, many of LPS-stimulated cells differentiate to ASCs while dividing and accumulate and increase in frequency with successive generations (Hasbold et al., 2004). Furthermore, LPS-stimulated B cells undergo isotype switching and alter the isotype chosen in response to different cytokines (Coffman et al., 1988). The other way is to mimic T-cell dependent activation through CD40, for example, with soluble α CD40 antibody. This system is typically accompanied with “switching” cytokines such as IL-4, IL-5, Interferon Gamma (IFN- γ) and Transforming growth factor Beta (TGF- β) to initiate class switching of immunoglobulin isotype (Coffman et al., 1988; Hodgkin et al., 1994; Hasbold et al., 1999b; Deenick et al., 1999, 2005; Oracki et al., 2010).

As with studying proliferation dynamics, diverse mathematical models have been proposed to deduce underlying mechanisms and to estimate biologically relevant parameters. Quantitative analyses of population datasets from division tracking assays provided insights on the characteristic of B cell fates: isotype switching requires cell division and is division-linked (Hodgkin et al., 1996; Tangye et al., 2003c). Furthermore, it was reported that isotype switching is stochastic and independent of the decision to become an ASC (Hasbold et al., 2004). Measurements of exact timings of individual cell’s fate from imaging platform in controlled, *in vitro* settings, and using probabilistic model, Duffy et al. (2012); Zhou et al. (2018) deduced that the cells reached their fates autonomously and accurately predicted cell numbers subjected to stochastic competition. Another *in vitro* study from Hawkins et al. (2013) used the original Cyton model to infer that the rate of differentiation to ASC depends on the type of stimulation in the system. In particular, LPS induces a constant rate regardless of its concentration so long as it is above a certain threshold. However, this rate changes with α CD40 stimulation strength: a higher proportion of differentiation per generation is seen with lower α CD40 concentrations. Other mathematical models in the literature cover mechanism of B cell fate in a broader context. Examples include: (i) modelling a B cell regulatory network in germinal centre (Martínez et al., 2012); (ii) stochastic compartmental model to describe inter- and extra-cellular dynamics of B cell maturation and exit from germinal centre (Thomas et al., 2019); and, (iii) hybrid ODE formulation with a stochastic component to model the effect of joint regulation of IL-2 and IL-4 in B cell proliferation and differentiation (Atitey and Anchang, 2022).

In the following sections, we will introduce an analysis of B cell differentiation in an attempt to unravel clonal characteristics of the process at molecular level based

4.3. SIMULATION OF B CELL DIFFERENTIATION PROCESS AND EXPECTED DATA STRUCTURE

on a dataset from an experimental system that was performed *in vitro* with murine B cells. As the multiplex lineage tracing assay provides distinctive information than that of time-lapse microscopy experiments, or other similar methods, development of a model that is distinct from previous ones was necessary. To do so, we first performed a gedanken experiment with two simple distinct rules that anticipate the core characteristics to be expected in the resulting data. The two alternative generative patterns are: (i) each offspring autonomously determines its cell fate independently of other cells; and, (ii) cells within a family determine their fates in a clonally dependent manner. Here, we generated family trees via Monte-Carlo (MC) simulation to obtain a simulated dataset that can be compared to features observed within experimental data. In addition to the theoretical work, we analysed flow cytometry-based lineage tracing (Marchingo et al., 2016; Horton et al., 2018) datasets and derived a number of simple probabilistic models, based on those two rules, to predict the proportion of differentiated B cells, particularly, IgG1⁺ and IgG2b⁺ cells.

4.3 Simulation of B cell differentiation process and expected data structure

In this section, we present MC simulation of the B cell differentiation that generates clonally-independent and -dependent family trees. Here, clonal independence indicates that each cell's fate is determined autonomously irrespective of other sibling cells, whereas clonal dependence means close association of fate change between related cells. For both cases, we assume that no cell death occurs in the system, and cell fate is an irreversible process. One distinguishing characteristic between the two systems is that the offspring in the clonally-dependent case inherit the state from parent cells. Overall description of the simulation is shown in Fig. 4.1A1,A2. Ultimately, we aim to extract and compare statistical characteristics in the number of type-changed cells that are regenerated by those two rules and obtain theoretical expectations for the data structure that one could expect from experiments.

First we considered generating symmetric family trees to identify unique statistical features that manifest as a distinguishing factor between the two systems. Let n_g be the number of unchanged cells in a given family for generation, $g = 0, 1, \dots, G$, where each simulated family begins with $n_0 = 1$ founder cell. Given division-

4.3. SIMULATION OF B CELL DIFFERENTIATION PROCESS AND EXPECTED DATA STRUCTURE

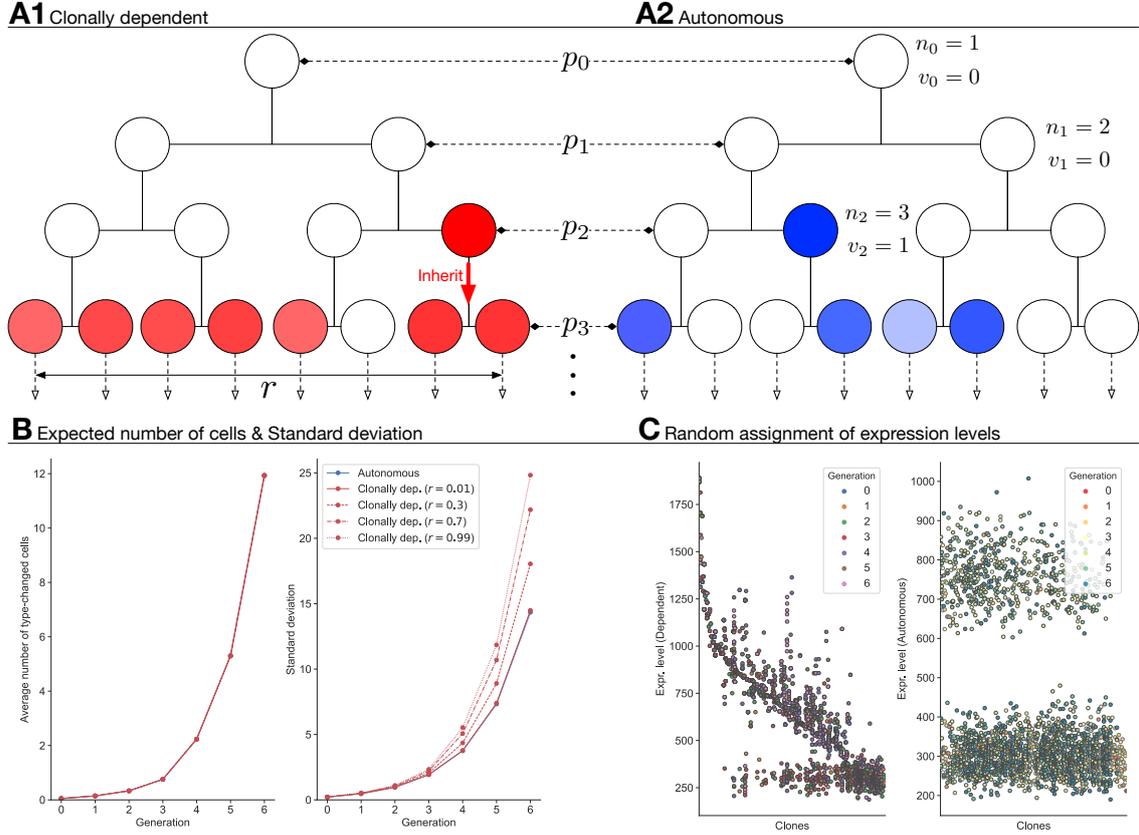


Fig. 4.1: MC simulation of cell-type change and expected data structure. Members of a family are simulated with two distinct mechanisms of cell fate: **(A1)** Autonomous determination of cell fate for each cell; and, **(A2)** Clonally dependent fate with positive correlation coefficient, $r \in [0, 1]$. Once cells change their type with generation-dependent probabilities, $\{p_g : g = 0, 1, \dots, G\}$ (Hasbold et al., 2004), their offspring inherit the state. For each generation, the events of cell-type change are simulated with binomial and beta-binomial distributions for autonomous and clonally dependent cases, respectively, given number of unchanged cells (n_g). Variable, v_g , denotes number of type-changed cells. **(B)** The expected number and standard deviation of type-changed cells calculated from 10^7 families. **(C)** Simulated with 200 founder cells with randomly assigned division times. Offspring that reached their fates are randomly assigned expression levels from a single lognormal or multivariate lognormal distribution, parametrised by median and log-variance, (m, s) , and mean vector and covariance matrix, $\boldsymbol{\mu} \in \mathbb{R}^{v_g}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{v_g \times v_g}$, respectively.

4.3. SIMULATION OF B CELL DIFFERENTIATION PROCESS AND EXPECTED DATA STRUCTURE

dependent differentiation probabilities $\{p_g\}$ (Hasbold et al., 2004), the number of type-changed cells is randomly decided using a binomial distribution, $v_g = B(n_g, p_g)$, for the clonally-independent case as n_g numbers of cells independently determine their fate. To model the clonally-dependent case, we used a beta-binomial distribution, $v_g = BB(n_g, \alpha_g, \beta_g)$, to include degrees of correlation between sibling cells in a given generation such that v_g objects were picked out of n_g cells with probability p_g and intraclass correlation $r \in [0, 1]$. More specifically, the beta-binomial distribution is parameterised by $\alpha_g = p_g(1 - r)/r$ and $\beta_g = (1 - p_g)(1 - r)/r$ (Marchingo et al., 2016). With this formulation, when $r = 0$ each cell determines the fate independent of others, whereas for $r = 1$ all cells in the same generation either maintain or change their states. We ran simulations for 10^7 families and recorded $n_g^{(i)}$ and $v_g^{(i)}$ for every i^{th} family. Each realisation of a family tree was set to terminate at maximum generation $G = 6$, thus, we obtain sets $\mathcal{N} = \{\mathbf{n}^{(i)}\}$ and $\mathcal{V} = \{\mathbf{v}^{(i)}\}$ where $\mathbf{n}^{(i)}, \mathbf{v}^{(i)} \in \mathbb{Z}_+^G$. The expected number and standard deviation of \mathcal{V} are shown in Fig. 4.1B. For both hypothesised scenarios, the average numbers of type-changed cells are approximately the same, however, the variance increases as the correlation coefficient r increases. This result indicates that the consequence of clonal dependency can be statistically identified by comparing the variances.

Next, we extended the simulation above and extracted information from simulated trees to establish expected data structure. To do so, we randomly assigned expression levels that correspond to experimentally observable quantities directly after determining cells' fates at each generation per family. In a typical experiment setup, these levels are measured using one or more biological markers (e.g. cell surface staining) to identify cell fates in a given population. To reflect correlation structure and to ensure positive expression values, the assignment was performed using a multivariate lognormal distribution parameterised by a mean vector $\boldsymbol{\mu} \in \mathbb{R}^{v_g}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{v_g \times v_g}$ with correlation coefficient r . The sizes of mean vector and covariance matrix depend on the number of type-changed cells in each generation. For the clonally-independent case we set $r = 0$ in the matrix such that each cell gets assigned autonomously and independently of the other cells. For the clonally-dependent case, we set $r = 0.9$ to remark structural differences in possible experimental observation subjected to a strong correlation. As all cells exhibit auto-fluorescence, we assigned these levels separately using a lognormal distribution. Furthermore, we allocated random division times for each cell to create frayed family trees, which are often observed in real experiments. At the beginning of each simu-

4.4. STATISTICAL ANALYSIS OF CLONAL DEPENDENCE OF DIFFERENTIATION

lation, a founder cell was randomly assigned time to first division from a lognormal distribution with median and log-variance ($T_{div}^0 \sim \mathcal{LN}(24, 0.3)$) and its offspring picked a random inter-division times from a exponential distribution with mean of 10 hours ($T_{div}^{k>0} \sim \text{Exp}(10)$). We recorded these times and examined expression levels of cells at 48 hours. In Fig. 4.1C, rank-ordered plot of the expression levels is shown. As expected, the correlation induces tight clusters of cell expression levels in a given family and cascade pattern throughout the clones. However, two clouds of the expression levels are shown for the autonomous case, indicating small portion of cells has type-changed at random and the rest remains in the initial state. These structural differences can be utilised to illuminate underlying mechanisms of cell fate determination.

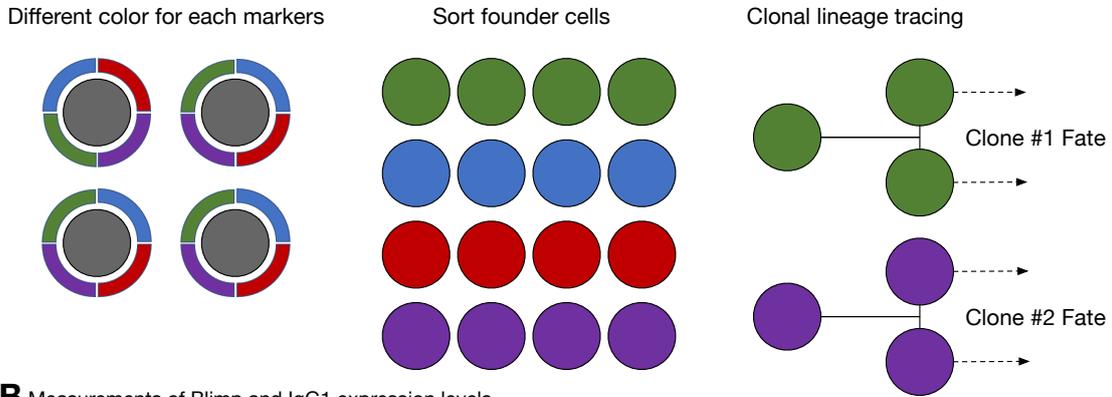
4.4 Statistical analysis of clonal dependence of differentiation

Lineage tracing

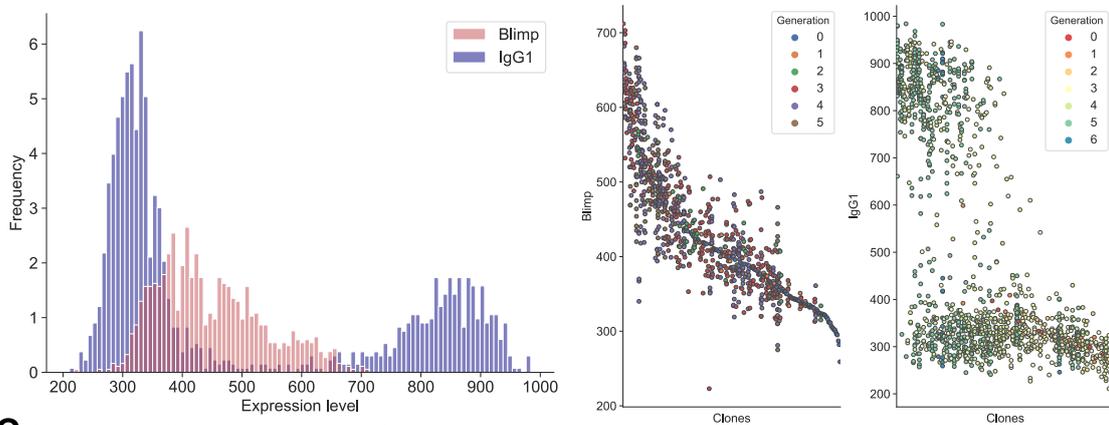
Conventional lineage tracing and fate mapping approaches such as time lapse microscopy have been widely used to investigate emergence of heterogeneous cell types in a population. However, scientific discovery and progress have been impeded by some practical problems such as aggregation of cells *in vitro* and difficulties in cell tracing over long period of time *in vivo*. Recently developed techniques such as barcoding via retrovirus-tagged progenitors (Gerlach et al., 2010; Naik et al., 2013, 2014) and heritable genome editing via CRISPR technology (McKenna et al., 2016) have contributed to design higher throughput experiments and led to important discoveries in the development of diverse cell types across multiple biological systems. As the detail of division progression is important for understanding the inner workings of cell fate determination, Horton et al. (2018) introduced utilisation of multiplexed division tracking dyes used in *in vitro* settings to simultaneously measure division progression, typically up to 8 divisions before cells are indistinguishable from autofluorescence, as well as cell fates with flow cytometry-based phenotyping (Fig. 4.2A). This method extends well established fluorescence-based assay, for example with *carboxyfluorescein succinimidyl ester*, and provides information of clonal lineage in a high-throughput manner.

4.4. STATISTICAL ANALYSIS OF CLONAL DEPENDENCE OF DIFFERENTIATION

A Multiplex division tracking assay



B Measurements of Blimp and IgG1 expression levels



C Relationship of the expression levels between offspring

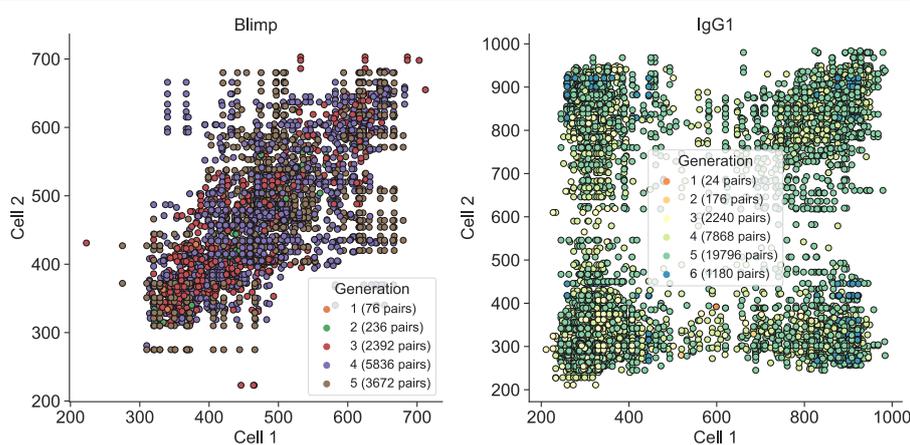


Fig. 4.2: Flow cytometry-based multiplex division tracking assay and experiment measurements of Blimp and IgG1 expression levels. (A) Lineage tracing of activated B cells for ASC and CSR by multiplexed cell labelling. (B) Distributions of the expression levels of Blimp (*red*) and IgG1 (*blue*). Cascade plots of rank-ordered Blimp and IgG1 expression levels. Generation information is colour coded. (C) All possible combinations of pairs of the expression levels for cells in the same generation and family.

4.4. STATISTICAL ANALYSIS OF CLONAL DEPENDENCE OF DIFFERENTIATION

We obtained datasets that contain information on differentiation of B cells into ASCs and IgG1 lymphoblasts, indicated by Blimp and IgG1 expression, respectively (see [Supplementary Fig. B.1A](#)). Division information and rank-ordered cascade plots of expression levels are shown in [Fig. 4.2B](#), including the marginal distributions. We noticed that the expression level of Blimp exhibits a similar pattern with the simulation result of clonally dependent families. On the other hand, in the culture conditions used, large proportion of cells remain unswitched ($\lesssim 500$) and those who switched to IgG1⁺ appear to reach fates independently of other cells in the family. Furthermore, the association of fate determination between sibling cells in a family revealed that Blimp has relatively strong correlation, whereas this association disappears with IgG1 ([Fig. 4.2C](#)). To augment our point, quantitative assessments of the association is presented in [Supplementary Fig. B.1B-D](#) by calculating Spearman's correlation coefficient for repeated sampling of cell pairs either within clones or selected randomly irrespective of family.

Statistical inference

To quantitatively assess the clonal dependence of cell fate determination, we performed permutation tests with the following information from the datasets: (i) expression level; (ii) generation; and, (iii) clonal membership. Here, we define multiplex data as a sequence of three elements for every i^{th} cell,

$$\mathcal{D} = \{(e_i, g_i, c_i)\}_{i=1}^N,$$

where N is the total number of cells, $e_i \in \mathbb{R}$ is the expression level, $g_i \in \{0, 1, \dots, G\}$ is generation and $c_i \in \mathcal{C} = \{1, 2, \dots, C\}$ is label for clonal membership with $G, C \in \mathbb{N}$. For our tests, we posit two null hypotheses,

- *Test 1: Every cell's expression level is equal in distribution irrespective of generation and clonal membership.*
- *Test 2: Every cell's expression level is equal in distribution irrespective of clonal membership only.*

4.4. STATISTICAL ANALYSIS OF CLONAL DEPENDENCE OF DIFFERENTIATION

Given null hypotheses, we define the variance of the expression level averaged across all clones as our test statistics,

$$T(\mathcal{D}) = \frac{1}{C} \sum_{x \in \mathcal{C}} \sigma^2((e_i : c_i = x, i = 1, \dots, N)),$$

where $\sigma^2(A) = \frac{1}{|A|-1} \sum_{a \in A} \left(a - \frac{1}{|A|} \sum_{a' \in A} a' \right)^2$. A collection of data permutations, denoted Q , is all possible reassignments of the labels of the individual observation, $i \rightarrow \tilde{\pi}(i)$, so that each of a reordered data set is constructed by

$$\pi : \mathcal{D} \rightarrow \mathcal{D}_\pi = \{(e_{\tilde{\pi}(i)}, g_i, c_i)\}_{i=1}^N.$$

If we have 1,000 cells, then there are 1,000 factorial possible permutations, requiring MC method to draw $B \in \mathbb{N}$ number of samples from Q uniformly at random to evaluate $T(\mathcal{D}_\pi)$ and estimate the following one-sided test p-values:

$$\hat{p} = \frac{1 + \sum_{i=1}^B \mathbb{1}_{\{T(\mathcal{D}_\pi) \leq T(\mathcal{D})\}}}{1 + B}.$$

With this recipe, we can establish allowable permutations depending on the tests we wish to proceed. In particular, *Test 1* has no restriction, whereas we limit the permutation of the expression levels between the cells in the same generation for *Test 2*. [Fig. 4.3](#) presents the results of the tests encoded with p-values estimated from MC with $B = 500,000$, resulting in rejections of the null hypotheses for both *Test 1* and *Test 2*. These rejections indicate that the data have lower average variance than that of expected variance under the the null hypothesis, thus have greater clone relatedness. However, the measurement of IgG1 expression level has a peculiar pattern, where more than half of the cells remain in initial state ($\approx 67\%$ in IgG1⁻) and exhibit similar data structure as shown in [Section 4.3](#) (c.f. [Fig. 4.1C](#)) even though we assumed clonal independence in the simulation. In the following section, we introduce measurements of molecules, in particular germ-line transcripts (GLTs), that are essential part of the mechanism leading to immunoglobulin class switch recombination (CSR), and analyse them to elucidate molecular regulation of CSRs to IgG1 (and later IgG2b) that could lead to the observation of IgG1 expression level.

4.5. ANALYSIS OF MOLECULAR REGULATION OF CLASS SWITCH RECOMBINATION

Permutation test: clonal dependence

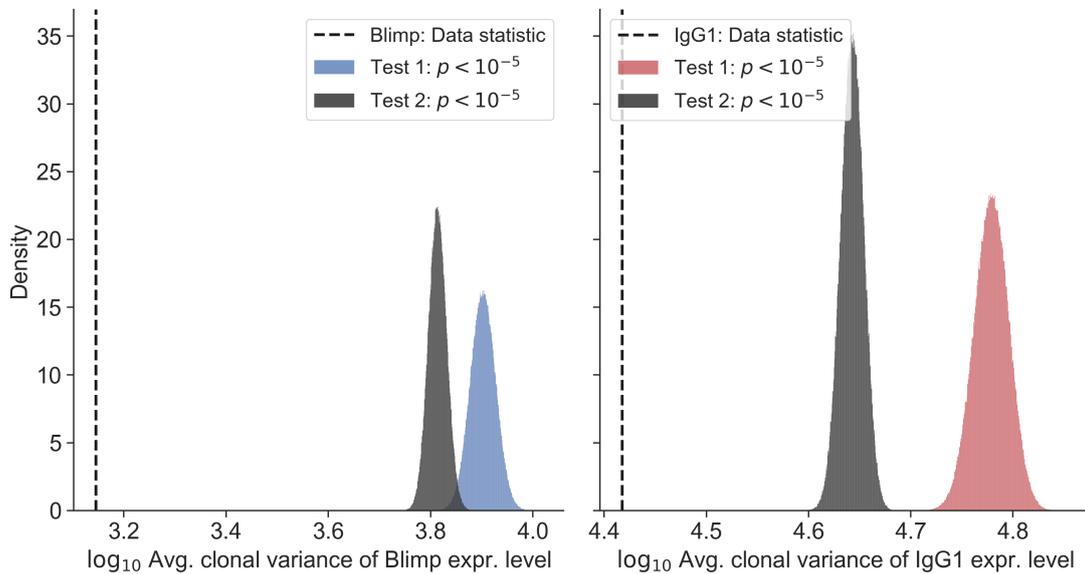


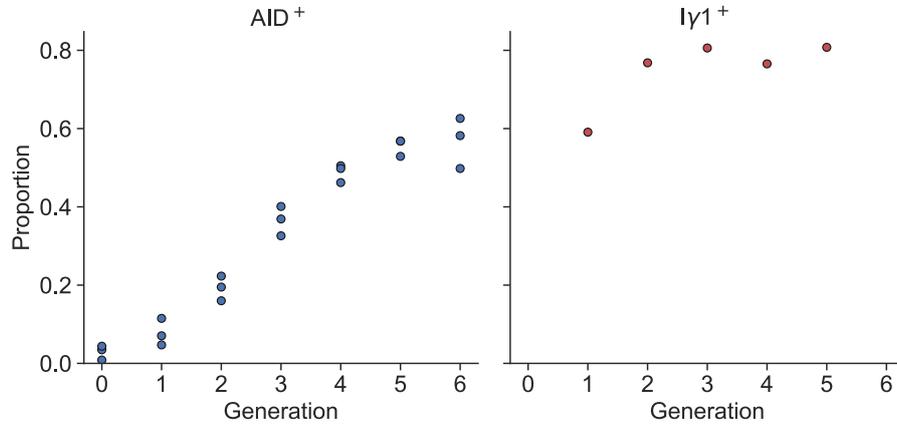
Fig. 4.3: Testing null hypotheses of clonal dependence. Permutation test results of the Blimp (*left*) and IgG1 (*right*) data presented in Fig. 4.2B. The vertical dashed-line refers to the data statistics of the original data. The distribution shows the density of the \log_{10} of the average clonal variance computed for 500,000 permutations of the data. The lower one-sided p-values are shown in legend for both tests described in the main text. These values indicate rejection of the hypotheses with a significance level of 0.05.

4.5 Analysis of molecular regulation of Class Switch Recombination

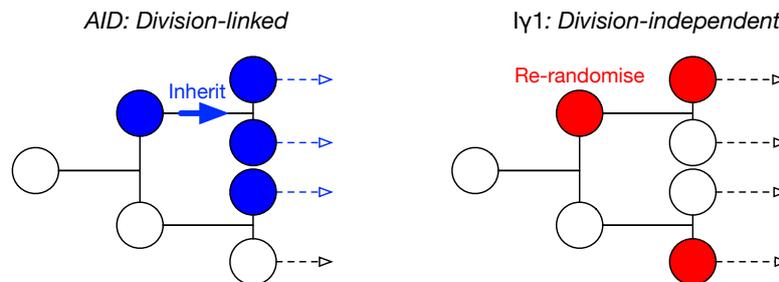
To understand further the development of the characteristic familial isotype switching patterns, we examined the familial expression of other key mediators of switching. There are two main molecular components that are key to the process of isotype class switching (e.g. from IgM to IgG1) in B cell differentiation. The first is the DNA-editing enzyme *Activation-Induced cytidine Deaminase* (AID) (Muramatsu et al., 2000) and the other is $\text{I}\gamma 1$ GLT. These components are not expressed in naïve B cells and both are induced by external stimuli during B cell activation (Stavnezer-Nordgren and Sirlin, 1986; Zhang et al., 2019). To investigate molecular regulation of CSR, our collaborators at WEHI obtained measurements of AID expression and $\text{I}\gamma 1$ transcription (Supplementary Fig. B.2A,C). AID expression was

4.5. ANALYSIS OF MOLECULAR REGULATION OF CLASS SWITCH RECOMBINATION

A Proportion of AID⁺ and I γ 1⁺ B cells



B Proposed AID and I γ 1 regulation



C Model prediction of IgG1 cells & residual-sum-of-squares of model fit

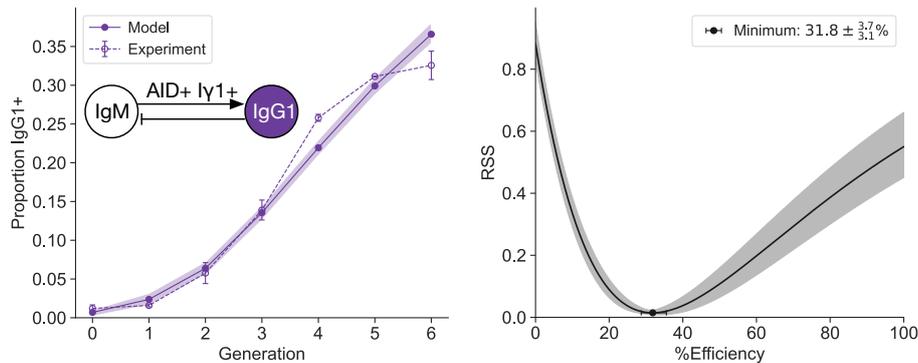


Fig. 4.4: B cell antibody class switching is characterised by division-linked AID expression and division-independent I γ 1 transcription. [Corresponding to Figures 3 & 4 from [Horton et al. \(2022\)](#)] (A) Experiment measurements of proportions of B cells expressing AID (triplicate) and I γ 1. (B) Proposed molecular regulations of AID and I γ 1 at single-cell level. Once a cell expresses AID, its offspring inherit the state of mother cell and stay AID⁺. In contrast, each cell autonomously determines I γ 1, its state is re-randomised at each generation. (C) experiment measurements (Mean \pm SEM) and model prediction of proportion of IgG1⁺ cells. Percentage switching efficiency of IgG1 given both AID and I γ 1 expressions.

4.5. ANALYSIS OF MOLECULAR REGULATION OF CLASS SWITCH RECOMBINATION

measured in clonal families using multiplex fluorescent dyes with flow cytometry together with a human-CD2 (hCD2) reporter, and I γ 1 transcripts were measured by single-cell quantitative polymerase chain reaction (sc-qPCR) combined with index-sorting (see experiment details in [Appendix A.3](#)). These measurements are shown in [Supplementary Fig. B.2A,C](#). Here, repeated random cell pairs and quantification of clonal correlations indicate noticeable discordance in expression of both molecules. Furthermore, co-expression analysis of hCD2/AID and I γ 1 exhibits no correlation between the two molecules ([Supplementary Fig. B.3](#)). These results demonstrate that the molecules are each expressed stochastically and independently of each other. Next, we sought to develop a model of CSR based on the intersection of expression of both GLT and AID, by chance, in the same cells, and present results of the model predictions. Mathematical derivation and reasoning of the model are shown separately in [Section 4.6](#) to clearly distinguish biological findings inferred from the data and model.

It was reported that antibody class switching is a division-linked process, that is, the frequency of switched cells during B cell activation increases with each successive round of the cell cycle ([Hodgkin et al., 1996](#); [Deenick et al., 1999](#)). To dissect molecular regulation of CSR fates in relation to generation, we inspected expression of AID and I γ 1 of dividing B cells. Qualitatively, it is evident from inspection of [Fig. 4.4A](#) that these two molecules behave distinctively: AID is division-linked as the proportion of AID⁺ B cells increases cumulatively with generation number, creating non-symmetric families, whereas I γ 1 transcription frequency is constant and unaffected by cell division ([Fig. 4.4A](#)). From these measurements, we proposed two separate mechanisms to model stochastic emergence of AID and I γ 1 molecules: (i) A cell can express AID at any generation independent of other sibling cells, and once it expresses, the subsequent offspring inherit the state and stay AID⁺. This is irreversible process as to accommodate increase in frequency of AID⁺ cells; (ii) Each cell autonomously expresses I γ 1, and it is re-randomised upon each cell division irrespective of clonal lineage or generation ([Fig. 4.4B](#)). Based on these two rules, we constructed a simple probabilistic model and asked if it was consistent with, and could predict, the number of observed IgG1 switched cells in each generation (see [Section 4.6.1](#) for the equations). In particular, we assumed that the switching occurs when both AID and I γ 1 are expressed, by chance, in the same cell with some efficiency which were modelled by a binomial distribution given empirical probabilities of the two molecules. In [Fig. 4.4C](#), we present the model of predicted proportion of IgG1

switched cells with estimated switching efficiency of approximately 32%, revealing excellent accord with the data.

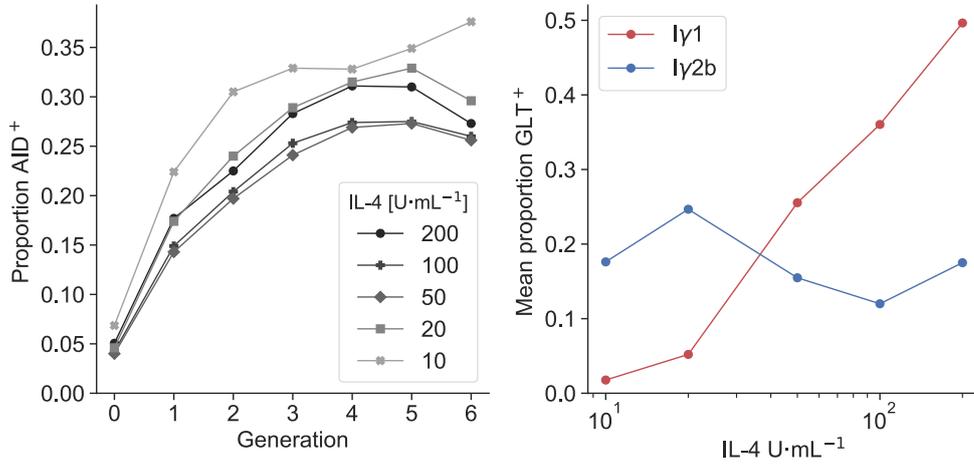
B cells are restricted to expressing a single isotype at a time, however, they are capable of switching to different isotypes, e.g. IgG2b, directly from IgM and sequentially by first expressing another switched isotype (Stavnezer et al., 2008). Here, we investigated dynamics of multiple isotype switching of B cells, particularly from IgM to IgG1 and/or to IgG2b cells. To do so, we obtained measurements of another GLT, i.e. I γ 2b expression along with AID and I γ 1, that responsible for isotype class switch to IgG2b. Induction of concurrent switching to two isotypes was achieved by stimulating B cells with LPS, IL-4 and TGF- β , and the proportions of IgG1 and IgG2b switched cells within families were determined with the multiplex clonal assay. In Fig. 4.5A, we present experimentally measured proportions of AID⁺, I γ 1⁺ and I γ 2b⁺ cells for five different concentrations of IL-4. While we can observe a division-dependent accumulation of AID⁺ cells, as previously described, for all IL-4, I γ 1 expression frequency increases in a dose dependent manner, consistent with its role in promoting switching to IgG1 cells. However, I γ 2b remains constant irrespective of the IL-4 titration. Here, we postulated that the fate decisions during concurrent antibody class switching are likely resolved by intracellular competition between GLTs, with more strongly expressed transcripts outcompete the others to determine outcomes of a switching cell. We included this rule in the probabilistic model as shown in Fig. 4.5B to predict IgG1 and IgG2b switching, provided empirical estimates of AID, I γ 1 and I γ 2b probabilities with a constant switching efficiencies (see Section 4.6.2 for the equations). This model closely recapitulated the dynamics of multiple isotype switching under varying external IL-4 signal strength, supporting the hypothesis that conflicting CSR fate decisions are determined by the competition at molecular level.

4.6 Mathematical modelling

So far, we have presented datasets and model results of isotype class switching in the previous section. Here, we show derivation of the model equations and parameter estimation method to predict proportion of IgG1 and IgG2b switched cells. More specifically, all models of CSR used in Section 4.5 assume that cells express two main components of the isotype switching stochastically and independently: AID and GLTs (I γ 1 and I γ 2b), with these molecules exhibiting distinct mechanisms during

4.6. MATHEMATICAL MODELLING

A Proportion of AID⁺, I γ 1⁺ and I γ 2b⁺ B cells



B Model of direct and sequential isotype class switching

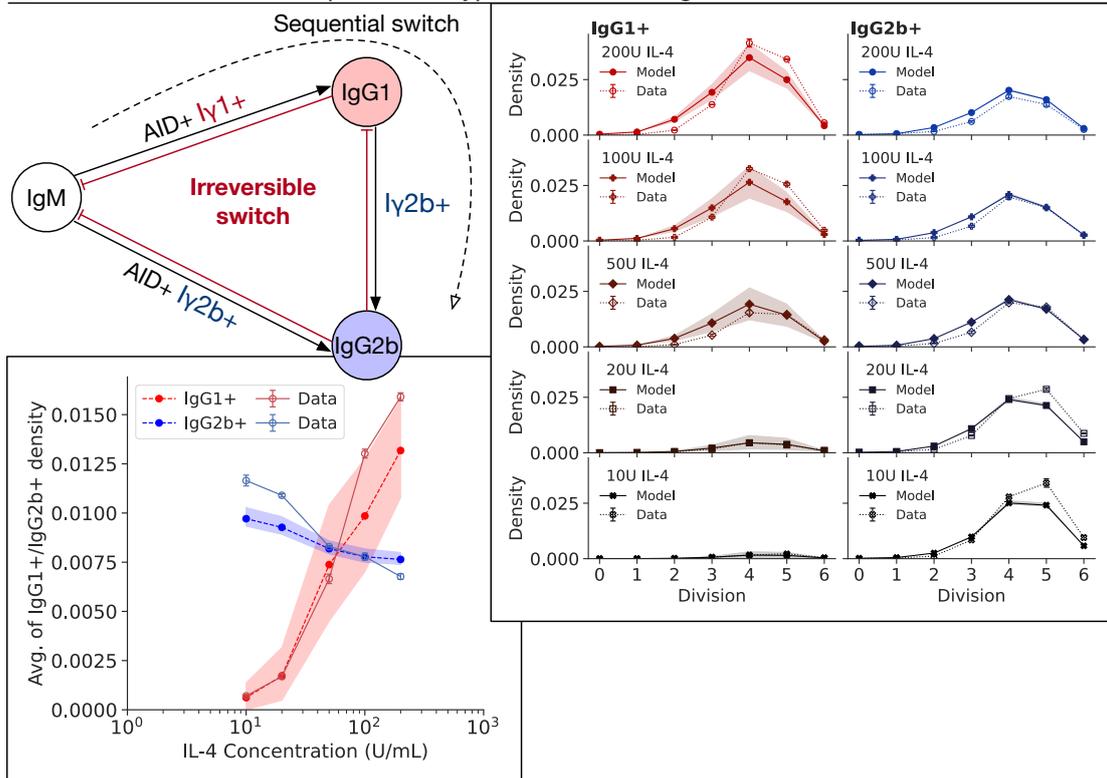


Fig. 4.5: Model of multiple antibody class switching in the presence of intracellular competition between GLTs. [Corresponding to Figure 5 from Horton et al. (2022)] (A) Measured proportions of AID⁺, I γ 1⁺ and I γ 2b⁺ cells for varying concentrations of IL-4. (B) Model predictions of joint probability of observing IgG1⁺ and IgG2b⁺ cells that are in generation g (c.f. Eq. 4.8) given empirical estimates of hazard probabilities of AID and of constant probabilities GLTs per IL-4. Shade area represents 95% confidence band.

4.6. MATHEMATICAL MODELLING

the cell proliferation phase. While the frequency of AID positive cells increases with successive generation as offspring inherit the AID state, the frequencies of I γ 1 and I γ 2b positive cells remain unaffected by cell division. Consequently, when cells autonomously express both components in each generation, a fraction of them switches to either IgG1 or IgG2b depending on the type of the expressed GLT that corresponds to its final fate. In this section, we construct a probabilistic model based on the probabilities of expressing AID and GLTs, as well as its switching efficiency, to calculate proportions of IgG1 and IgG2b.

Let $\{\rho_g^A \in [0, 1] : g = 0, 1, \dots, G\}$ be the expected proportion of cells that are AID⁺ in generation g , and $\{p_k^I\}$ be the probability that a cell is I γk positive for $k = 1, 2$, representing I γ 1 and I γ 2b, respectively. In addition, if a cell is both AID⁺ and I γk ⁺, then it can switch to IgG1 and IgG2b with probability $\epsilon_k \in [0, 1]$, encoded by $k = 1, 2$ respectively (recombination efficiency). To begin the model building, we define A_g^+ (A_g^-) as number of AID⁺ (AID⁻) cells and $G_g^{(k)}$ as number of IgG k ⁺ cells. Here, the model assumes that all cells are initially in the IgM state and that switching occurs once per-generation and is an irreversible process, i.e., once a cell switches to IgG k ⁺, it is inherited by its offspring and cannot switch back to a lower k . Furthermore, we assume there is no cell death occurs during the entire process and all cells in a family stop dividing at D generation (division destiny).

Prior to building the model, we first need to convert $\{\rho_g^A\}$, which are accumulating hazard as once a cell is AID⁺ its offspring stay positive, into the per-generation hazard probabilities $\{p_g^A\}$:

$$p_g^A = \begin{cases} \rho_0^A & \text{if } g = 0, \\ \frac{\rho_g^A - \rho_{g-1}^A}{1 - \rho_{g-1}^A} & \text{if } g > 0. \end{cases}$$

Hence, p_g^A is the probability that an AID⁻ cell becomes AID⁺ in generation g , while ρ_g^A is the overall proportion of AID⁺ cells in generation g . In summary, the model is parameterised by $\{p_g^A\}$, $\{p_k^I\}$, ϵ_k in which we used empirical estimates of $\{p_g^A\}$ and $\{p_k^I\}$ as model inputs.

In the following sections, we derive the equations separately for generation zero and subsequent generation to then calculate proportions of IgG1 and IgG2b for two scenarios: (i) IgM \rightarrow IgG1; (ii) IgM \rightarrow IgG1/IgG2b.

4.6.1 IgG1 switching

Generation zero

The number of remaining AID⁻ cells satisfies the binomial distribution with probability $1 - p_0^A$, that is, $A_0^- = \text{Bin}(N_0, 1 - p_0^A)$. Assuming that $N_0 = 1$ is a starting cell in a family, then the expected number of A_0^+ is $\mathbb{E}[A_0^+] = 1 - \mathbb{E}[A_0^-] = p_0^A$. Thus, the expected number of cells that become IgG1⁺ in generation zero is

$$\mathbb{E}[G_0^{(1)}] = \mathbb{E}[A_0^+] \epsilon_1 p_1^I = p_0^A \epsilon_1 p_1^I. \quad (4.1)$$

Subsequent generation

For $0 < g \leq D$, similar to generation zero, the remaining AID⁻ cells in generation g conforms to a recurrence relation $A_g^- = \text{Bin}(2A_{g-1}^-, 1 - p_g^A)$, where the factor two is required to account for the doubling effect from cell division. Then, the expected number of A_g^+ is $\mathbb{E}[A_g^+] = 2^g - \mathbb{E}[A_g^-] = 2^g (1 - \prod_{i=0}^{g-1} (1 - p_i^A))$. For IgG1 cells, A_g^+ cells switch to IgG1⁺ with probability $\epsilon_1 p_1^I$, but all offspring of cells that have switched in previous generations remain switched. Thus, the expected number of $G_g^{(1)}$ is given by the following recursive formula,

$$\begin{aligned} \mathbb{E}[G_g^{(1)}] &= 2\mathbb{E}[G_{g-1}^{(1)}] + (\mathbb{E}[A_g^+] - 2\mathbb{E}[G_{g-1}^{(1)}]) \epsilon_1 p_1^I \\ &= 2^g \left[\mathbb{E}[G_0^{(1)}] (1 - \epsilon_1 p_1^I)^g + \epsilon_1 p_1^I \sum_{i=0}^{g-1} (1 - \epsilon_1 p_1^I)^i \left(1 - \prod_{j=0}^{g-i} (1 - p_j^A) \right) \right]. \end{aligned} \quad (4.2)$$

Together with Eq. 4.1 and 4.2, we can calculate the proportion of IgG1 switched cells by multiplying 2^{-g} ,

$$\mathbb{E}[\tilde{G}_g^{(1)}; \boldsymbol{\theta}] = \begin{cases} \mathbb{E}[G_0^{(1)}] = p_0^A \epsilon_1 p_1^I & \text{if } g = 0, \\ \mathbb{E}[G_0^{(1)}] (1 - \epsilon_1 p_1^I)^g + \epsilon_1 p_1^I \sum_{i=0}^{g-1} (1 - \epsilon_1 p_1^I)^i \left(1 - \prod_{j=0}^{g-i} (1 - p_j^A) \right) & \text{if } g > 0, \end{cases} \quad (4.3)$$

where $\boldsymbol{\theta} = (p_g^A, \epsilon_1, p_1^I)$ is the model parameter.

4.6.2 IgG1/IgG2b switching

In this section, we model the system with multiple isotypes by adding an extra switching state, IgG2b, which can occur in two ways: (i) directly from the IgM state; or (ii) sequentially from the IgG1 state. Note that the reverse switching is strictly prohibited. Here, we assume that both $I\gamma 1$ and $I\gamma 2b$ operate independently of each other, and should both $I\gamma 1$ and $I\gamma 2b$ be expressed, and both recombinations be efficient, we assume that IgG1 switching occurs. Furthermore, the efficiencies are assumed to be constant irrespective of IL-4 concentration.

Generation zero

We adopt the same construct as shown in [Section 4.6.1](#) to calculate the expected number of A_0^+ such that $\mathbb{E}[A_0^+] = p_0^A$. For IgG1 and IgG2b calculation, by addressing the conflict where A_0^+ cells can become both $I\gamma 1^+$ and $I\gamma 2b^+$ throughout the generation, the number of $IgGk^+$ cells at the end of the division must be determined. We can resolve such conflict by assigning A_0^+ cells into three categories: (i) $M_0^{(0)}$, number of cells failed to switch with probability $(1 - \epsilon_1 p_1^I)(1 - \epsilon_2 p_2^I)$; (ii) $M_0^{(1)}$, number of IgG1 switching cells with probability $\epsilon_1 p_1^I \epsilon_2 p_2^I + \epsilon_1 p_1^I (1 - \epsilon_2 p_2^I)$; and, $M_0^{(2)}$, number of IgG2b switching cells with probability $(1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I$. More specifically,

$$(M_0^{(0)}, M_0^{(1)}, M_0^{(2)}) = \text{Mult}(A_0^+, (1 - \epsilon_1 p_1^I)(1 - \epsilon_2 p_2^I), \\ \epsilon_1 p_1^I \epsilon_2 p_2^I + \epsilon_1 p_1^I (1 - \epsilon_2 p_2^I), (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I).$$

This entails that IgG2b switching is allowed only if $I\gamma 2b$ is expressed exclusively. Hence, the expected numbers of $G_0^{(1)}$ and $G_0^{(2)}$ are

$$\begin{aligned} \mathbb{E}[G_0^{(1)}] &= \mathbb{E}[M_0^{(1)}] = p_0^A (\epsilon_1 p_1^I \epsilon_2 p_2^I + \epsilon_1 p_1^I (1 - \epsilon_2 p_2^I)), \\ \mathbb{E}[G_0^{(2)}] &= \mathbb{E}[M_0^{(2)}] = p_0^A (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I. \end{aligned} \tag{4.4}$$

Subsequent generation

As both $I\gamma 1$ and $I\gamma 2b$ operate independent of AID, adding IgG2b state in the model does not alter the AID calculation shown in [Section 4.6.1](#). Thus, for $0 < g \leq D$, the expected value of A_g^+ remains the same, i.e., $\mathbb{E}[A_g^+] = 2^g (1 - \prod_{i=0}^g (1 - p_i^A))$. However, the recursive formula for expected number of $G_g^{(1)}$ and $G_g^{(2)}$ needs an extra

term to accommodate IgG1 to IgG2b transition, that is,

$$\begin{aligned}\mathbb{E}[G_g^{(1)}] &= 2\mathbb{E}[G_{g-1}^{(1)}] + \mathbb{E}[M_g^{(1)}] - \mathbb{E}[X_g], \\ \mathbb{E}[G_g^{(2)}] &= 2\mathbb{E}[G_{g-1}^{(2)}] + \mathbb{E}[M_g^{(2)}] + \mathbb{E}[X_g],\end{aligned}$$

where $X_g = \text{Bin}(2G_{g-1}^{(1)}, (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I)$ is sequentially switched cells to IgG2b. Re-expressing $\mathbb{E}[G_g^{(2)}]$ yields,

$$\begin{aligned}\mathbb{E}[G_g^{(2)}] &= 2^g \left[\mathbb{E}[G_0^{(2)}] (1 - (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I)^g + \right. \\ &\quad \left. (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I \sum_{i=0}^{g-1} (1 - (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I)^i \left(1 - \prod_{j=0}^{g-i} (1 - p_j^A) \right) \right],\end{aligned}\tag{4.5}$$

and for $\mathbb{E}[G_g^{(1)}]$,

$$\begin{aligned}\mathbb{E}[G_g^{(1)}] &= 2^g \mathbb{E}[G_0^{(1)}] s^g - (\epsilon_1 p_1^I \epsilon_2 p_2^I + \epsilon_1 p_1^I (1 - \epsilon_2 p_2^I)) \times \\ &\quad \left[\sum_{i=0}^{g-1} 2^{i+1} s^i \mathbb{E}[G_{g-i-1}^{(2)}] - 2^g \sum_{i=0}^{g-1} s^i \left(1 - \prod_{j=0}^{g-i} (1 - p_j^A) \right) \right],\end{aligned}\tag{4.6}$$

where $s = 1 - \epsilon_1 p_1^I \epsilon_2 p_2^I - \epsilon_1 p_1^I (1 - \epsilon_2 p_2^I) - (1 - \epsilon_1 p_1^I) \epsilon_2 p_2^I$. Thus, the proportion of IgG1 and IgG2b can be obtained by multiplying 2^{-g} to [Eqs. 4.4, 4.5](#) and [4.6](#) with the model parameter $\beta = (p_g^A, \epsilon_1, p_1^I, \epsilon_2, p_2^I)$:

$$\begin{aligned}\mathbb{E}[\tilde{G}_g^{(1)}; \beta] &= \mathbb{E}[G_g^{(1)}] 2^{-g} \\ \mathbb{E}[\tilde{G}_g^{(2)}; \beta] &= \mathbb{E}[G_g^{(2)}] 2^{-g}.\end{aligned}\tag{4.7}$$

4.7 Model fitting

Given empirical estimates of $\{p_g^A\}$, $\{p_1^I\}$ and $\{p_2^I\}$ from experiments, we found optimal efficiencies, ϵ_1^* and ϵ_2^* , by fitting the models to the IgG1 and IgG2b datasets using least-squares with Levenberg-Marquardt (LM) optimisation algorithm ([Marquardt, 1963](#)) implemented in LMFIT ([Newville et al., 2014](#)). Let $\{\mathcal{D}_g^{(1)}\}$ and $\{\mathcal{D}_g^{(2)}\}$ denote the sets of measured IgG1 and IgG2b proportions, respectively. The cost function

for the model in [Section 4.6.1](#) (c.f. [Eq. 4.3](#)) is

$$C(\boldsymbol{\theta}) = \sum_{g=0}^G \left(\mathcal{D}_g^{(1)} - \mathbb{E}[\tilde{G}_g^{(1)}] \right)^2,$$

such that we find the minimum,

$$\epsilon_1^* = \arg \min_{\epsilon_1} C(\boldsymbol{\theta}).$$

We performed grid search on $\epsilon_1 \in [0, 1]$ with $\Delta\epsilon_1 = 0.001$ and stored the best-ranking 100 solutions as candidates according to the residual sum-of-squares (RSS). Then, these candidates were used as initial values for LM algorithm and recorded RSS for each candidate to identify the best-fit parameter by the lowest RSS. To obtain the 95% confidence interval, we performed the bootstrap method ([Efron, 1979](#)) with artificial datasets that were resampled with replacement from the original measured data. In particular, we resampled $\{p_g^A\}$, $\{p_1^I\}$ and $\{\mathcal{D}_g^{(1)}\}$ simultaneously at each iteration to propagate measurement errors from the experiment. We repeated this process 10^4 times, which resulted in 10^4 additional estimates of the parameter. We identified lower 2.5th and upper 97.5th percentiles as lower and upper limits of confidence interval.

For the model fit in [Section 4.6.2](#) (c.f. [Eq. 4.7](#)), we simultaneously fit across all IL-4 concentrations, $i \in \mathcal{I} = \{1, 2, \dots, I\}$, for both IgG1 and IgG2b datasets. The cost function is a weighted sum,

$$C(\boldsymbol{\beta}) = \sum_{i \in \mathcal{I}} \sum_{g=0}^G \mathcal{F}_{i,g} \left[\left(\mathcal{D}_{i,g}^{(1)} - \mathbb{E}[\tilde{G}_{i,g}^{(1)}] \right)^2 + \left(\mathcal{D}_{i,g}^{(2)} - \mathbb{E}[\tilde{G}_{i,g}^{(2)}] \right)^2 \right], \quad (4.8)$$

where $\mathcal{F}_{i,g}$ is the measured frequencies of the cells in generation g at i^{th} IL-4 concentration. Here, the term $\mathcal{F}_{i,g} \cdot \mathbb{E}[\tilde{G}_{i,g}^{(k)}]$ (and $\mathcal{F}_{i,g} \cdot \mathcal{D}_{i,g}^{(k)}$) can be interpreted as a density of joint probability of observing IgG k^+ cells and that the cells are in generation g . We find the minimum of ϵ_2 while fixing $\epsilon_1 = \epsilon_1^*$ obtained above:

$$\epsilon_2^* = \arg \min_{\epsilon_2} C(\boldsymbol{\beta}).$$

Similarly, we applied grid search on $\epsilon_2 \in [0, 1]$ with $\Delta\epsilon_2 = 0.001$, but we kept top 200 solutions as candidates to widen the search space. Then, we performed LM

algorithm on each candidate to obtain 200 RSS values and identified the best-fit by the lowest RSS. For 95% confidence interval, similar bootstrapping technique was applied: resampling of $\{p_1^I\}$, $\{p_2^I\}$, $\{\mathcal{D}_{i,g}^{(1)}\}$ and $\{\mathcal{D}_{i,g}^{(2)}\}$ per IL-4 concentration. However, we excluded $\{p_g^A\}$ from the resampling as we only have 1 replicate in this particular instance and included resampling of ϵ_1 from reported 95% confidence interval above. The resampling process was repeated 10^4 times to acquire additional 10^4 estimates from which we calculate 95% confidence interval of ϵ_2^* .

4.8 Discussion

In an adaptive immune response, clonally selected naïve lymphocytes expand and differentiate to multiple alternative fates. It is evident that delivery of signals from other cell types in complex microenvironments play a role in this regulation. However, the emergence of heterogeneous populations, even under controlled *in vitro* conditions, implies that variation in the initial state of cells, or a utilisation of autonomous randomisation process, is an inherent characteristic of the system that is orchestrated with the external influences. In this chapter, we explored potential sources of inherent variation in murine B cells by tracing fate differences within clonal families at the population level using multiplex assay and single-cell qPCR.

Single-cell lineage tracing revealed that ASC differentiation fates are heritable through generations during proliferation. In particular, we demonstrated that the characteristic structure of Blimp expression level was quantitatively consistent with the hypothesis that cells inherit states from their parent cells. The quantifications of the associativity between siblings or within clones were achieved through calculations of Spearman's correlation coefficient and hypothesis testing via permutation tests, respectively. Our colleagues at WEHI further identified that the variable predisposition toward ASC differentiation is in part due to broadly distributed TLR expression within naïve B cell population (Supplementary Fig. B.4 and Supplementary Fig. B.5A,B).

In contrast to ASC differentiation, antibody class switching was instead more consistent with the hypothesis and operated independently of clonal lineage and was attributable to cell-intrinsic stochastic processes (Supplementary Fig. B.5C). This was identified through observations of AID expression and I γ 1 GLT transcription as they progressed through generations. While the frequency of AID expressed cells increased as division progressed, the proportion of I γ 1⁺ cells was effectively constant.

This data informed us for building a simple probabilistic model. By assuming that two independent stochastic events, AID and GLT, must occur in the same cell, we predicted the rate of isotype switching per generation observed in the experiment. We extended this model to sequential switching from IgM to IgG1/IgG2b as a function of IL-4 concentration. One of the key observations is that the probability of IgG1 and IgG2b switching directly changes with respect to corresponding I γ 1 and I γ 2b GLT transcripts, which in turn altered the resulting frequencies of switched cells in a predictable manner (Supplementary Fig. B.5D). Knowing the frequency of AID and GLTs within a population, we account for the proportions of switched cells, thereby predicting the effect of external signals in B cell differentiation.

Together these results illustrate how stochastic processes based on the simple rules of molecular regulation recapitulate the generation of diverse antibody types during the clonal expansion phase that was carefully moderated with external signals in a controlled *in vitro* environment.

5

Cyton2 application to the human system: hyper- and hypo-immune responses

5.1 Abstract

In this chapter, we present an application of the Cyton2 model for the lymphocytes that are found from immuno-compromised human patients, when compared with healthy donors (HDs). In collaboration with the partners from the immunology division at Walter and Eliza Hall Institute (WEHI) led by Prof. Philip Hodgkin, we interrogated human B cell data that were obtained from two distinct patient groups with the following primary diseases: (i) Common Variable Immunodeficiency (CVID); and (ii) Sjögren Syndrome (SS). CVID is the most common clinically significant immune defects where the majority of the cases are known to be caused by the combination of multiple small genetic defects ([Bogaert et al., 2016](#); [Ameratunga et al., 2018](#)). SS is an autoimmune rheumatic disease where hyperactivity of B cell is considered a hallmark of the disorder. As these diseases exhibit two opposite antibody responses, hypo- and hyper-response respectively, we questioned the compounding impact of one or more defects in the proliferation modules of naïve B cells in comparison to HDs. Utilising the Cyton2 model first introduced in [Chapter 2](#) to dissect each division components, we developed a method to measure human B cell health of CVID and SS patients when compared with HD. Furthermore, by assuming that components of the Cyton2 model are selected at random in the population at large, we extrapolated to estimate the likelihood of observing these disorders amongst HD by a Monte-Carlo simulation of the model that resembles Bayesian inference.

5.2 Immunodeficiency and autoimmunity

In all the work prior to this chapter, we have explored *in vitro* stimulated lymphocytes from the murine system, typically transgenic mice for better control of environmental factors or for reporting change in states of cells, e.g. differentiation to antibody-secreting cells. A natural question was whether or not our mathematical model and analysis techniques developed so far could be useful when applied to data from human lymphocytes, which have greater variance person-to-person in terms of immune response due to various conditions such as genetic factors or living conditions. Nevertheless, an effective immune response must (i) lead to a response of an appropriate magnitude to protect against pathogens, (ii) prevent uncontrolled expansion of lymphocytes and (iii) be capable of discriminating between self and non-self antigens. Given such complexity in regulation, mathematical models and techniques may be useful in interrogating and dissecting a dysregulated immune response in quantitative manner.

Primary antibody deficiencies occurring because of defective B cells are the most common primary immunodeficiency (PID) in the human population (Durandy et al., 2013). Genetic mutations, such as mutation in *Bruton's Tyrosine Kinase* affects early B cell development (Conley et al., 2009), or malfunctions in any component in the innate and/or adaptive immune systems, such as T- and B-cell interaction, can cause PIDs. To diagnose and treat patients of these conditions, it is essential to characterise their genetic features. The PID was once thought to be caused by a single-gene abnormality, however, studies (Bykowsky et al., 1996; Kornfeld et al., 1996) revealed that significant variability can be manifested amongst patients even with the same gene defect. Amongst many types of PIDs, Common Variable Immunodeficiency (CVID) is the most prevalent type that typically has a complex genetic basis (Salzer et al., 2009) with still many unknown genetic causes (Bogaert et al., 2016). CVID patients are diagnosed to make weak, or deficient antibody responses following infections. Here, we will focus on the population dynamics of naïve B cells collected from CVID patients and placed in controlled *in vitro* conditions in an attempt to dissect components of the affected division parameters that led to deficient response. It is our hypothesis that many combinations of small genetic effects can alter the B cell response and lead to antibody deficiency. However, the functional assays of proliferation and survival will provide an assessment of the net health of each of the Cyton elements.

5.3. EXPLORING FEATURES OF CELL POPULATION DYNAMICS FROM CVID AND SS PATIENTS

Autoimmunity, which lies at the opposite spectrum of immune deficiencies, has specific characteristic features: individuals overproduce antibodies and fail to distinguish self from non-self, resulting in harm to the host. To date, there are over 80 different autoimmune diseases that have been characterised for humans (Vojdani et al., 2014). Of those types, we investigated a particular disease, called Sjögren Syndrome (SS). Similar to CVID, The exact cause of the disease remain unknown, but hormonal, genetic and environmental factors (e.g. infectious agents) are involved in the initiation of SS (Voulgarelis and Tzioufas, 2010; Kroese et al., 2014). The hallmark of SS is hyperactivity of B cells, mainly IgG plasma cells, that leads to irreversible destruction of the glandular tissue. These hyperactive B cells secrete auto-antibodies and other various cytokines. In our application, by fitting a variant of the Cyton2 model to proliferation data, we aim to see if the differences in CVID and SS responses are reflected in differences in particular timers within the models.

One of key insights to our analysis presented in this chapter is that the defective components in B cell can be inferred and extracted from the behaviour of cell expansion. Although identifying the exact genetic cause or other elements that are responsible for CVID and SS require extensive research at the molecular level, we aim to obtain high-level overview of the characteristic features and differences between the disease types when compared with HDs using existing experimental techniques (e.g. flow cytometry) and the Cyton2 model.

5.3 Exploring features of cell population dynamics from CVID and SS patients

To identify characteristic differences between the healthy controls and patients, we interrogated B cell proliferation *in vitro* from a time-series division assay performed by WEHI members. Human naïve B cells were stimulated with CpG DNA and anti-Ig to generate a T cell independent proliferative response. In Fig. 5.1 top-row, we present cell number counts for 34 HDs, 18 CVID patients and 22 SS donors. We normalised these measurements with respect to the initial cell numbers seeded for each donor to remove effect of observing larger cell expansion due to differences in starting cell numbers, that is $\tilde{n}(t_i) = n(t_i)/n(t_0)$ where $n(t_i)$ is the total cell numbers measured at t_i hour with typically $t_0 = 0$ hour. Summary statistics were calculated, in particular total cohorts and mean division counts (i.e. cell survival and

5.3. EXPLORING FEATURES OF CELL POPULATION DYNAMICS FROM CVID AND SS PATIENTS

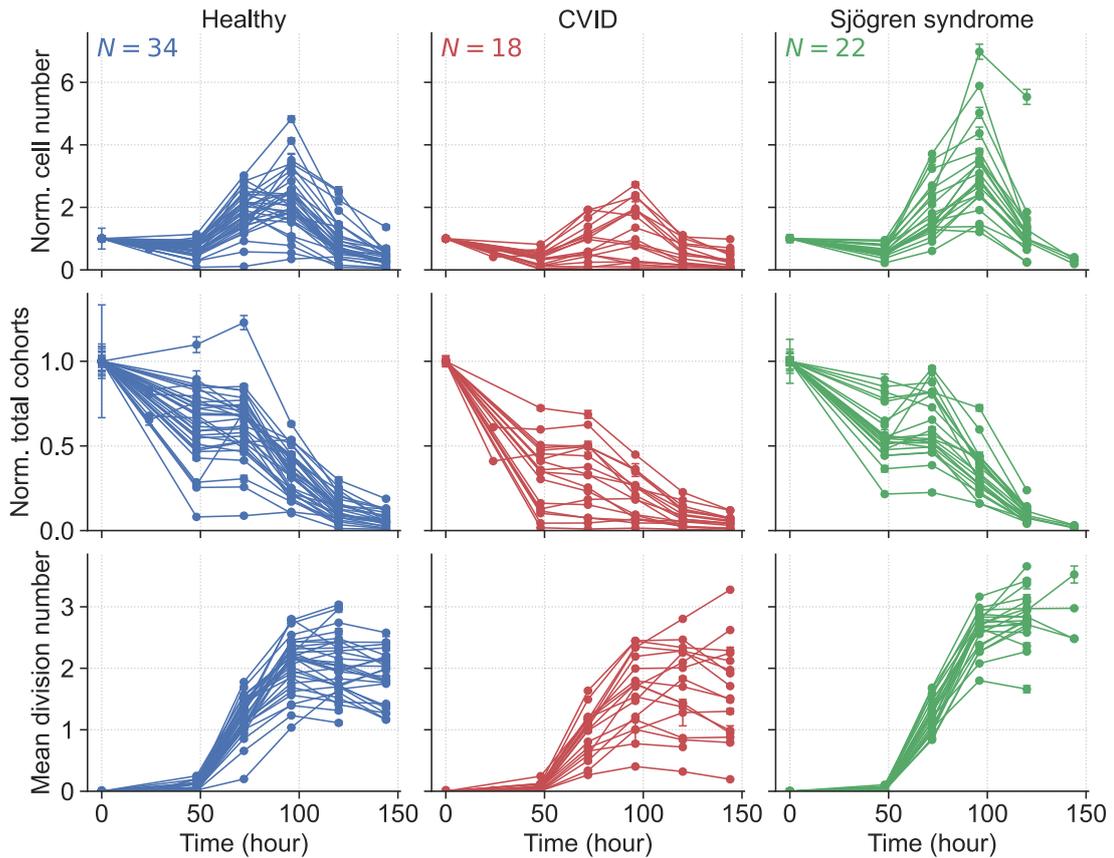


Fig. 5.1: FACS data and summary statistics of human B cells obtained from healthy donors, CVID and SS patients. Cells were stimulated with combined anti-Ig engagement of the B cell receptor and CpG DNA to induce T cell independent response. (*top-row*) Total cell numbers, normalised to the initial cell number. Summary statistics of the responses are given by (*mid-row*) normalised total cohort number for cell survival and (*bot-row*) mean division number for cell proliferation.

5.3. EXPLORING FEATURES OF CELL POPULATION DYNAMICS FROM CVID AND SS PATIENTS

expansion), to facilitate qualitative comparison between the groups (mid- and bottom of Fig. 5.1). With the responses from healthy donors as a reference, we observed that both patient groups follow similar pattern in decreasing in cell numbers up to approximately 50 hours, then exhibit aberrant behaviours afterwards. Qualitatively, it was difficult to infer whether the CVID patients have lower mean division numbers as the variability between the donors was larger, however, cells tend to die earlier than those from the HD, leading to deficient cell counts. On the other hand, it was clear that the most SS cells underwent one more round of division on average than the HD cells, while the cell survival profile was largely the same.

Next, we applied multiple permutation tests with data points collected at 72h, 96h and 120h to obtain quantitative assessments of those qualitative observations. The mean numbers of cells, cohorts and division were used as test statistics for the selected time points, and the null hypothesis was that the difference of means between HD and patient groups is zero. In Fig. 5.2, a series of Gardner-Altman plots (Gardner and Altman, 1986; Ho et al., 2019), are shown to report statistical significance and to capture effect size with estimated 95% confidence interval via bootstrap method. Here, 10^7 permutations and bootstrap samples were gathered to compute p value and confidence interval for each case. We found that rejection occurred for all tests between CVID and HD, which entails, together with the negative effect sizes, CVID patients have lower cell counts, earlier cell death and undergo less cell division compared to cells from the healthy group. Interestingly, no rejections were found between SS and HD for the cohort number tests, which suggests the survival of cells from the SS group behaves similarly to those from HD. However, we observed statistically significant results for the mean division numbers after 96h. Positive effect sizes are consistent with the known characteristic of SS (Kroese et al., 2014), where B cells divide more and lead to a hyperactivity.

In summary, the key characteristics of CVID cells are identified as faster cell death and lower average number of divisions, hence, overall less cell expansion. Meanwhile, SS cells experience more divisions, likely due to faster division rate, but the death rate is largely unaltered compared to the HD cells, which results in larger cell expansion.

5.3. EXPLORING FEATURES OF CELL POPULATION DYNAMICS FROM CVID AND SS PATIENTS

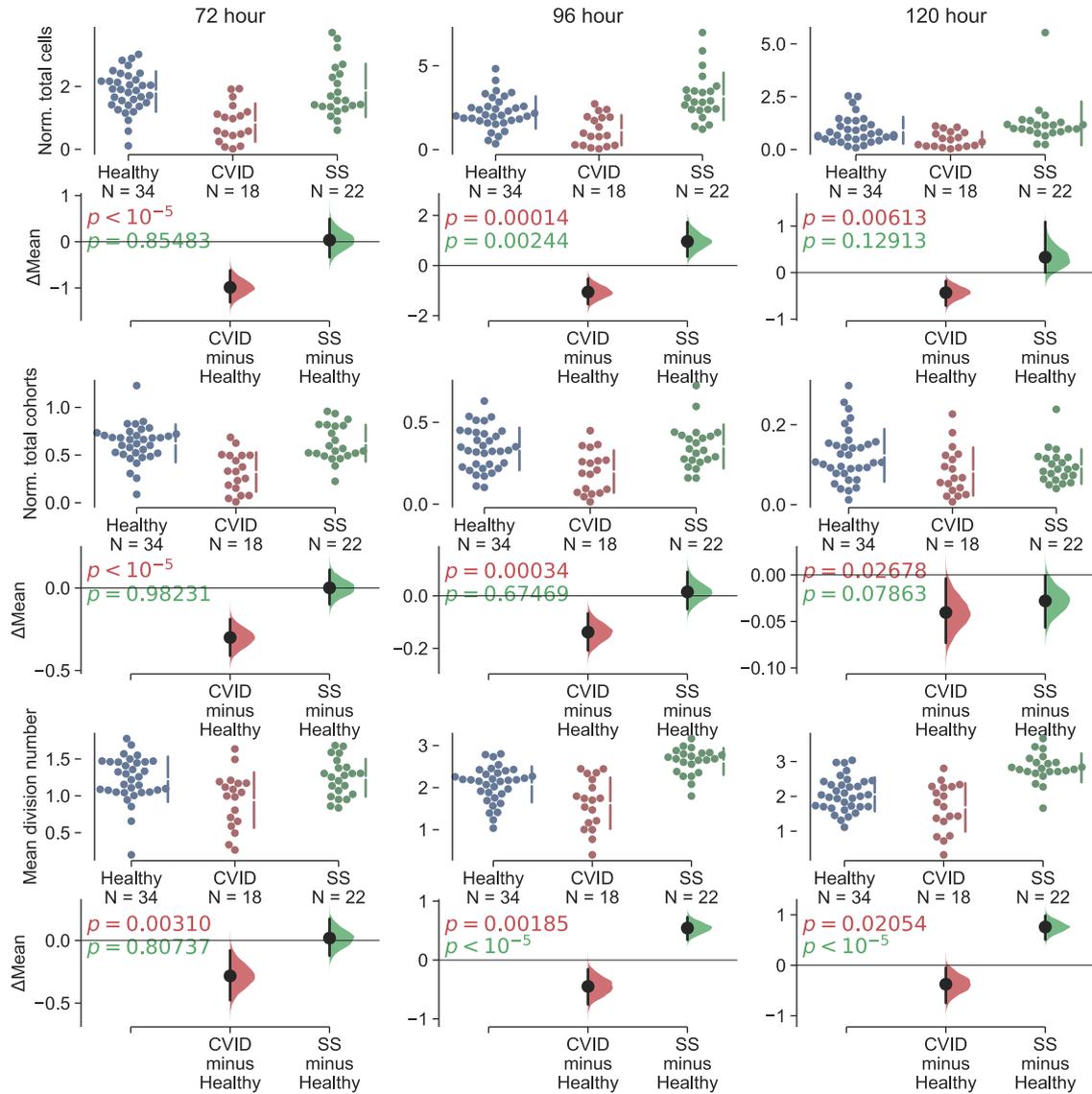


Fig. 5.2: Permutation tests on the measurements and summary statistics between HD (blue), CVID (red) and SS (green). For time points at 72h, 96h and 120h, the tests were performed in two sets: HD vs CVID and HD vs SS on total cell numbers (*top-row*), total cohort numbers (*mid-row*) and mean division numbers (*bot-row*). On each panel, data points are shown in a scatter plot, and differences of the means are shown below it. Solid horizontal line indicates zero difference. The error bar is a 95% confidence interval estimated using bootstrap method, and its distribution is shown as well. The one-sided p values from the permutation tests are annotated with corresponding colours. 10^7 permutations and bootstrap samples were used for the calculation.

5.4 Reprogrammed death time and activation fraction

In this section, we introduce a new parameter that is necessarily required to explain a specific phenomenon, observed distinctly from murine cells. Our collaborators at WEHI performed additional experiments at earlier time points, particularly at 24h and 48h, for the activation of naïve human B cells stimulated *in vitro* with CpG, polyclonal anti-Ig or both. Both stimulants mimic T cell independent activation in which CpG and anti-Ig signal through Toll-like receptor 9 (TLR9) and B cell receptor (BCR), respectively. As shown in Fig. 5.1, cell numbers tend to decrease for all conditions prior to the 50h mark, which raises questions about what proportion of cells had perceived stimulatory signals and whether they follow an alternate cell death program before commencing division. Common activation markers of naïve B cells, CD69, CD86 and Human Leukocyte Antigen-DR (HLA-DR), were measured after 24h and 48h along with media alone (data not shown). It was found that expression of CD69, CD86 and HLA-DR increased after 24h in cultures with CpG, anti-Ig and CpG plus anti-Ig. More specifically, CD69 was upregulated for approximately 92% of CpG stimulated cells and 86% of anti-Ig stimulated cells at 24h. Furthermore, CD86 was upregulated for 82% of CpG stimulated cells and 64% of anti-Ig stimulated cells at 48h. Similarly, a high proportion of cells were expressing CD69 and CD86 when stimulated with both CpG and anti-Ig. This result indicates that the majority of cells are activated in response to CpG and/or anti-Ig, irrespective of whether they divide. Hence, we postulated that B cells in culture undergo two distinct death programs and switch to a different survival profile approximately between 24h and 48h.

Next, we formalised a new method to estimate the reprogrammed death time, which we denote T_{die} , that will be utilised in the Cyton2 model. To do so, our collaborators conducted experiments to observe behaviour of cells without any stimulation. As we posit that the death time was reprogrammed only for the cells that commenced division, we assumed that these unstimulated cells follow a death program that was initially set in place throughout the course (Hawkins et al., 2007a). So, using datasets from the experiments, we first defined a new random variable (RV) that captures the time to death for unstimulated cells, T_U , and we assumed that it is lognormally distributed.

Let $n_g(t_i)$ denote the set of time-course cell number data, where $i \in \{0, 1, \dots, I\}$ and $g \in \{0, 1, \dots, G\}$ are time and generation index, respectively. We converted

5.4. REPROGRAMMED DEATH TIME AND ACTIVATION FRACTION

the cell numbers to a precursor cohort number according to the equation: $c_g(t_i) = n_g(t_i)2^{-g}$ (Gett and Hodgkin, 2000) and summed over all generations at each harvested time-points to obtain the total cohort number given by:

$$S(t_i) = \sum_{g=0}^G c_g(t_i).$$

This cohort conversion removes an effect of cell expansion from the total cell numbers by incorporating the doubling factor 2^{-g} . As a result, $S(t_i)$ is a monotonically decreasing function, where in particular, if one could follow a time-series of a single culture, $S(t_i)$ would be $S(t_0) \geq S(t_1) \geq \dots \geq S(t_I)$ for $t_0 < t_1 < \dots < t_I$. Biologically, this means that the total cohorts decreases only if the culture undergoes cell death, otherwise it would remain constant. We utilised this property to estimate the death time distribution with complementary cumulative distribution function (CCDF) of T_U : $f_{T_U}(t) = 1 - F_{T_U}(t; m_U, s_U)$, where F is the CDF of T_U with median, m_U , and log-variance, s_U . In our application, we used a normalised quantity for all time points, denoted $\tilde{S}(t_i) = S(t_i)/S(t_0)$.

Given estimated T_U , we then utilised datasets from experiments, where cells were stimulated with CpG, anti-Ig or both to estimate $T_{die} \sim \mathcal{LN}(m_{die}, s_{die})$, which is essentially the death time in the Cyton2 model. Here, we defined a piece-wise function to incorporate information of T_U ,

$$s(t; \boldsymbol{\alpha}) = \begin{cases} 1 - F_{T_U}(t; m_U, s_U) & \text{if } t \leq t_a \\ f_{T_U}(t_a)(1 - F_{T_{die}}(t; m_{die}, s_{die})) & \text{if } t > t_a \end{cases}, \quad (5.1)$$

where $\boldsymbol{\alpha} = (t_a, m_{die}, s_{die})$ and $t_a \in \mathbb{R}_{\geq 0}$ is the activation time and $f_{T_U}(t_a)$ is effectively total cohort numbers that represent number of activated cells, entering division and switching to a new death program. We performed a least-squares fit with the following objective function,

$$L(\boldsymbol{\alpha}) = \sum_{i=0}^I (\tilde{S}(t_i) - s(t_i; \boldsymbol{\alpha}))^2, \quad (5.2)$$

5.4. REPROGRAMMED DEATH TIME AND ACTIVATION FRACTION

which we employed numerical methods to find an approximate minimum,

$$\begin{aligned} \{\boldsymbol{\alpha}^*\} &\in \arg \min_{\boldsymbol{\alpha}} L(\boldsymbol{\alpha}); \\ \text{s.t. } t_a &\in [0, 100]; m_{die} \in (0, 500]; s_{die} \in (0, 5]. \end{aligned}$$

With the estimated values, we defined a post-hoc parameter *activation fraction*, $f_A = f_{T_U}(t_a)$. This quantity is useful to determine the proportion of cells that mount the immune response and to correctly extract the division parameters.

We utilised a gradient-based Levenberg-Marquardt (LM) algorithm to minimise Eq. 5.2 and for estimating unstimulated death time. The algorithm is implemented in Python package LMFIT (Newville et al., 2014), a high-level wrapper of SciPy-optimize library (Virtanen et al., 2020). The LM method requires a starting guess for each of the parameter values, so we used 100 random initial sets of values uniformly drawn within the identified parameter ranges. The best-fit was identified by comparing the lowest residual sum of squares (RSS). We used bootstrap method (Efron, 1979) to evaluate variability of the point estimate, $\boldsymbol{\alpha}^*$, by creating 1,000 artificial samples of the original data with replacement. These data were fitted with the identical optimisation routine to generate 1,000 additional estimates from which 95% confidence interval was computed.

We present a summary of the model fits in Fig. 5.3 using four healthy donors. In particular, we showed the estimates of the initial cell death program, T_U , and the reprogrammed death time, T_{die} , in Fig. 5.3A and 5.3B, respectively. We noticed that transition to the new death program take approximately 26 hours since the stimulation. The estimates on the new death time indicate that cells typically survived for a longer period of time compared to the unstimulated cells. Moreover, CpG has the largest effect on the cell death time with larger variance of T_{die} , while anti-Ig stimulated cells died at relatively earlier time. For CpG plus anti-Ig, the death time was estimated between those of CpG and anti-Ig alone. Also, we observed that the estimated f_A varied from donor to donor, but the values were consistent between different conditions within the same donor, and the mean proportion of activated cells was estimated to be 75.3%, indicating the majority of healthy cells commenced division and switched to a new survival program.

5.4. REPROGRAMMED DEATH TIME AND ACTIVATION FRACTION

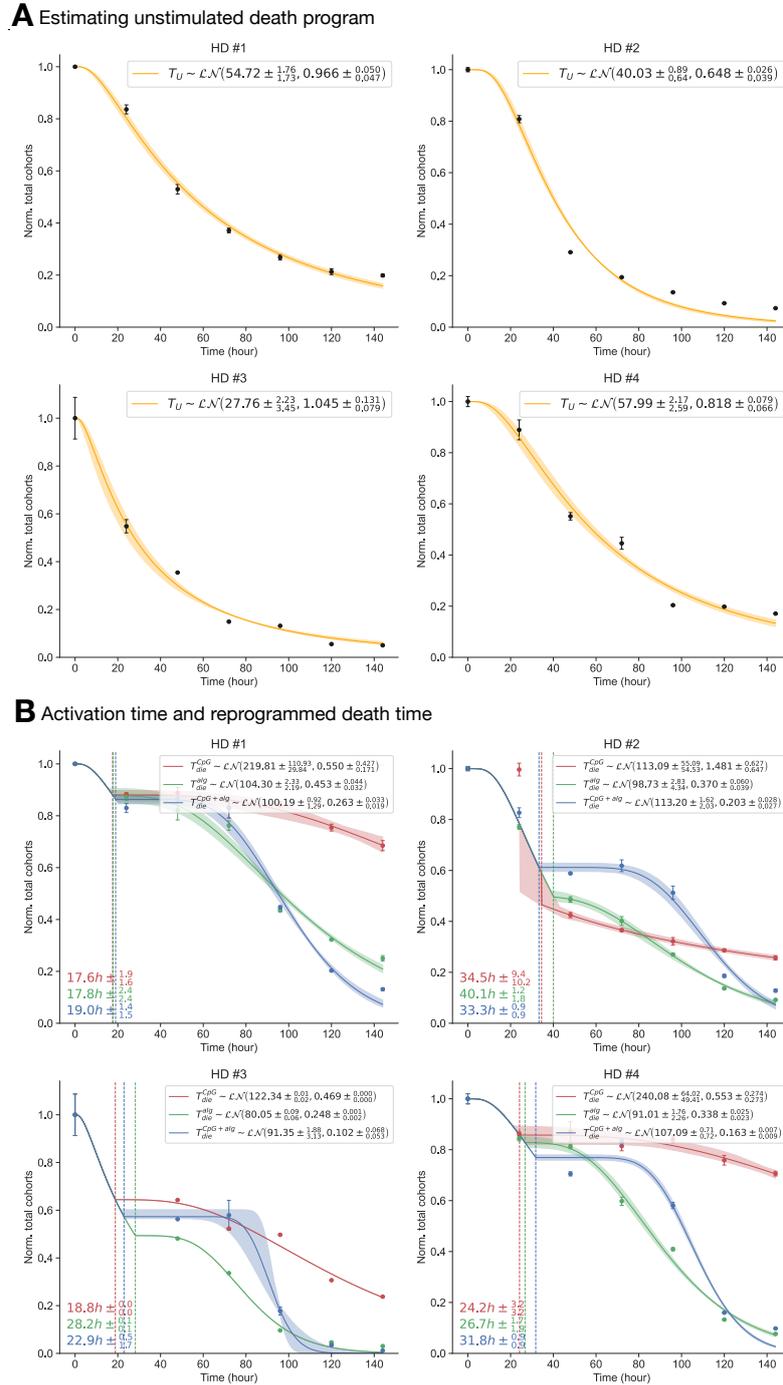


Fig. 5.3: Initial cell death timer and transition to the new reprogrammed death timer. Activation of human naïve B cells take approximately 25 hours. The time-series total cohort number and the best-fit model for (A) unstimulated cells (orange) and (B) CpG (red), anti-Ig (green) and CpG + anti-Ig (blue) stimulated cells. Solid-line and shade-area represent the mean and 95% confidence band, respectively. Vertical dash-line is the activation time. The parameter estimates are shown in the legend and bottom-left of the plot panel.

5.5 Measuring human B cell health using the Cyton2 model

In this section, we developed a method to quantify characteristic features of HD, CVID and SS B cell population dynamics with the Cyton2 model discussed in [Chapter 2](#). One of our objectives was to assign a meaningful score to indicate state of health of B cells from CVID and SS relative to healthy cohorts. In this process, we dissected components that constitute a healthy response in terms of division parameters, estimated by the model. This enabled us to compute relative contribution of aberrant parameters and identified compounding effect of multiple small defects that ultimately led to either a hypo- or hyper-immune response. To do so, we used datasets in [Fig. 5.1](#) and integrated the method to estimate reprogrammed death time discussed in [Section 5.4](#) to the Cyton2 estimation routine.

5.5.1 Alternate strategy for estimating Cyton2 division parameters

Since our measurements contain a proportion of cells that die before changing their survival program and entering the proliferative phase, we needed to isolate these cells for accurate estimation of the parameters from dividing cells. To do so, we introduced an extra step prior to the existing Cyton2 fitting routine. First, we filtered datasets for $t_i > 24$ hour as we observed that the majority of cells became activated and altered their survival program after approximately one day. Then, we slightly modified [Eq. 5.1](#) to estimate f_A and T_{die} using the following equation,

$$s(t; \gamma) = C(1 - F_{T_{die}}(t; m_{die}, s_{die})),$$

where $\gamma = (C, m_{die}, s_{die})$ and $C \in [0, \infty)$ is the total cohort number of activated cells that was further utilised to compute $f_A = C/S(t_0)$. Essentially, this modification allowed us to estimate without the information of unstimulated cell death. The estimated f_A and T_{die} for the reprogrammed death time were used as a prior knowledge of the system and fixed during the Cyton2 fitting process.

As the objective was to estimate the Cyton2 parameters of the dividing cells, the model was provided with an initial cell number, $N_0 = f_A \sum_g n_g(t_0)$, to remove the effect of unactivated cells. This process reduces the number of free parameters from seven to five in the fitting procedure.

5.5. MEASURING HUMAN B CELL HEALTH USING THE CYTON2 MODEL

In summary, we estimated the following set of division parameters: the time to first division, $T_{div}^0 \sim \mathcal{LN}(m_{div}^0, s_{div}^0)$; the time to division destiny, $T_{dd} \sim \mathcal{LN}(m_{dd}, s_{dd})$; and the subsequent division time, $m \in \mathbb{R}_{\geq 0}$. Thus, there are total five parameters to estimate, $\boldsymbol{\beta} = (m_{div}^0, s_{div}^0, m_{dd}, s_{dd}, m)$, such that the objective function is in the form,

$$L(\boldsymbol{\beta}) = \sum_{g=0}^G \sum_{\{t_i\}_{>24}} (n_g(t_i) - y_g(t_i; \boldsymbol{\beta}))^2, \quad (5.3)$$

where $y_g(t_i; \boldsymbol{\beta})$ is the model prediction in unit of cell numbers (c.f. Eq. 2.6). Similarly, we performed a least-squares fit to find an approximate minimum,

$$\begin{aligned} \{\boldsymbol{\beta}^*\} &\in \arg \min_{\boldsymbol{\beta}} L(\boldsymbol{\beta}); \\ \text{s.t. } m_{div}^0 &\in (0, 500]; s_{div}^0 \in (0, 5]; m_{dd} \in (0, 500]; s_{dd} \in (0, 5]; m \in [0, 50], \end{aligned}$$

where we applied the same optimisation and bootstrap method for $\boldsymbol{\beta}^*$ as described previously. With estimated parameter values, $\boldsymbol{\gamma}^*$ and $\boldsymbol{\beta}^*$, we recapitulated immune responses measured from HD, CVID and SS cells.

We present the model fits and summary of the parameter estimates in Fig. 5.4. Examples of the model extrapolation for each of the group are shown in Fig. 5.4A, along with the cell number measurements, to indicate fit quality. Summary CDF plots of the key division parameters (i.e. RVs) are presented in Fig. 5.4B to signify the differences between the groups. Here, we noticed that the estimated medians of T_{div}^0 were similar for all HD, CVID and SS group, however, the variability was slightly larger for the CVID donors. We observed similar results for T_{die} and the subsequent division time (Fig. 5.4C). The most prominent difference was observed for T_{dd} , where cells from CVID and SS group reached division destiny at earlier and later times, respectively, compared to the HD average. This result is consistent with the permutation tests shown in Fig. 5.2. As cells had entered the first division and died at similar times, this result indicates that CVID cells had less time to grow in cell numbers while SS cells had more time to divide, thus recapitulating the key characteristics of immunodeficiency and autoimmune disorder. Moreover, the average activation fractions for the HD and SS cells were estimated approximately 66% and 61%, however, it was significantly lower for the CVID cells ($\approx 34\%$), suggesting the majority of them did not undergo clonal expansion and follow the initial cell death program.

5.5. MEASURING HUMAN B CELL HEALTH USING THE CYTON2 MODEL

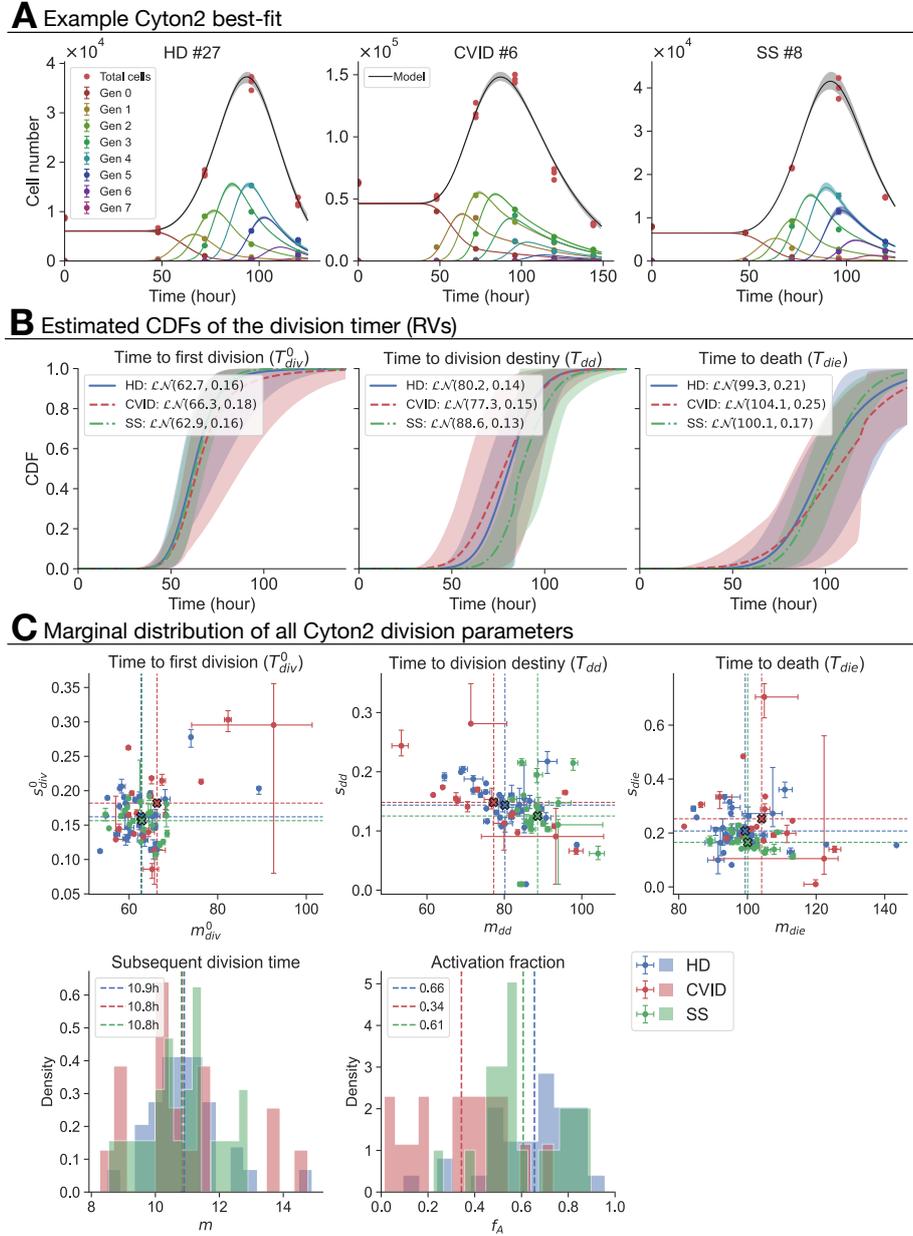


Fig. 5.4: Estimated division parameters of the Cyton2 model from HD, CVID and SS datasets. (A) Representative best-fit Cyton2 model for one donor from each group: *left*-HD, *mid*-CVID and *right*-SS. (B) Cumulative distribution function (CDF) of the estimated RVs that represent time to first division (T_{div}^0), time to division destiny (T_{dd}) and time to death (T_{die}). Solid, dashed and dash-dotted lines indicate mean CDF curves computed across HD, CVID and SS, respectively. Shade-area is lower and upper 95% percentile of the CDFs, calculated from collection of individual donors per group. (C) Summary of all Cyton2 parameters. Each dot in scatter plot represents one donor, and the error bar is 95% confidence interval from bootstrapping. Marginal distributions of subsequent division time (m) and activation fraction (f_A) are shown. Dashed line represents a mean value.

5.5. MEASURING HUMAN B CELL HEALTH USING THE CYTON2 MODEL

5.5.2 Determining relative contribution of division parameters of aberrant response

In this section, we reported how each of the parameter deviates from the average HD response with the parameter estimates using the Cyton2 model. It has been studied that apart from rare, monogenic forms, the majority of CVID cases are caused by mutations in multiple genes, where approximately 80% of them have an unknown cause and their prevalence and significance are currently unknown (Bogaert et al., 2016). We assessed and determined the relative contribution of each parameter to quantify B cell healthy and identified the diverse effects that led to an aberrant immune response.

Define $\theta_k^{(h)}$, $\theta_k^{(c)}$ and $\theta_k^{(s)}$ be the Cyton2 parameters estimated from HD, CVID and SS datasets, respectively, where $k = 1, 2, \dots, 8$ for $\theta = (f_A, m_{div}^0, s_{div}^0, m_{dd}, s_{dd}, m_{die}, s_{die}, m)$ with healthy and patient indices, $h = \{1, 2, \dots, 34\}$, $c = \{1, 2, \dots, 18\}$ and $s = \{1, 2, \dots, 22\}$. Then, we computed the mean healthy donor parameter vector by marginally averaging the components of θ ,

$$\bar{\theta}_k^{\text{HD}} = \frac{1}{34} \sum_{h=1}^{34} \theta_k^{(h)} \quad \forall k = 1, 2, \dots, 8.$$

The Cyton2 model was evaluated with $\bar{\theta}^{\text{HD}}$ such that we obtained a time at which the total cell number of the average healthy donor response was maximised,

$$t^* = \arg \max_t \left\{ \sum_{g=0}^G y_g(t; \bar{\theta}^{\text{HD}}) \right\}$$

which leads to $\bar{N}_{max}^{\text{HD}} = \sum_{g=0}^G y_g(t^*; \bar{\theta}^{\text{HD}})$,

and the corresponding total cell numbers at t^* for CVID and SS responses,

$$N_{base}^{(q)} = \sum_{g=0}^G y_g(t^*; \theta^{(q)}), \tag{5.4}$$

where q indicates the patient index, i.e. c or s . We then proceeded by sequentially substituting the components from $\bar{\theta}^{\text{HD}}$ to $\theta^{(c)}$ and $\theta^{(s)}$ to acquire the total cell numbers at each substitution stage via Eq. 5.4. A new vector at each stage was

5.5. MEASURING HUMAN B CELL HEALTH USING THE CYTON2 MODEL

given by,

$$\boldsymbol{\theta}_i = \boldsymbol{\theta}^{(q)} \mathbb{1}_{(k>i)} + \bar{\boldsymbol{\theta}}^{\text{HD}} \mathbb{1}_{(k\leq i)} \text{ for } i = 0, 1, 3, 5, 7, 8,$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function, and i denotes the number of ordered substituted vector components. For example, $i = 0$ would mean no substitution, $i = 1$ substitutes f_A , $i = 3$ substitutes $f_A, m_{div}^0, s_{div}^0$, and so on. Finally, we enumerated the following evaluations,

$$N_i^{(q)} = \sum_{g=0}^G y_g(t^*; \boldsymbol{\theta}_i) \quad \forall i, q,$$

to compute fold changes of the total cell numbers, $\Delta N_i^{(q)} = N_i^{(q)} / N_0^{(q)}$. Evidently, $N_0^{(q)} = N_{base}^{(q)}$ and $N_8^{(q)} = \bar{N}_{max}^{\text{HD}}$. With these fold changes, we computed relative percentage contribution by,

$$\% \Delta N_i^{(q)} = 100 \left(\frac{\Delta N_i^{(q)} - \Delta N_{i-1}^{(q)}}{\Delta N_8^{(q)}} \right) = 100 \left(\frac{N_i^{(q)} - N_{i-1}^{(q)}}{\bar{N}_{max}^{\text{HD}}} \right) \text{ for } i = 1, 3, 5, 7, 8,$$

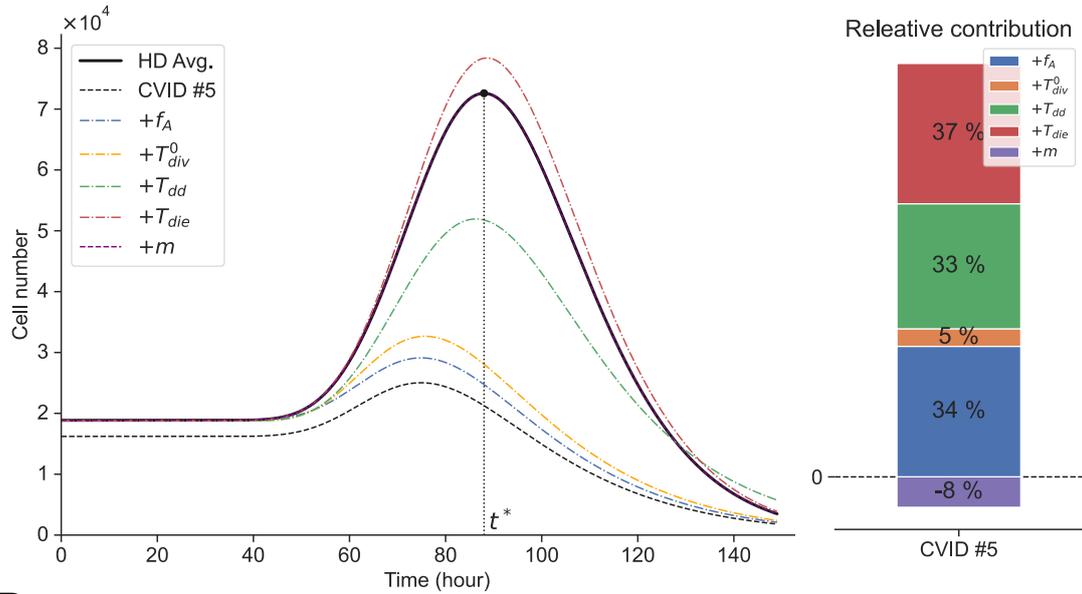
with $\Delta N_0^{(q)} = 0$ for no substitution. We showed an example calculation for one CVID patient in [Fig. 5.5A](#). Essentially, the percentage quantified deviation from the HD average for each of the cellular machinery. We applied and repeated this analysis to all of the CVID and SS donors in [Fig. 5.5B](#). As we picked a fixed time point for the computation, some of stages well exceeded $\bar{N}_{max}^{\text{HD}}$, and follow-up substitution subsequently reduced the total cell numbers to match the HD average. The results indicated that the relative contributions varied greatly between individual CVID patients, which highlights that CVID is a multifactorial disorder with collection of defects in different component lead to immunodeficiency. For example, CVID #1 has approximately equal contribution for all components, whereas CVID #3 suffer significant deviation for f_A and T_{dd} .

In contrast to the CVID patients, we observed consistent relative contributions across the SS donors, where f_A appeared to be the largest factor and followed up by the T_{div}^0 and T_{dd} .

In summary, we developed a method to quantitatively assess B cell health given the average HD response. By comparing the estimated Cyton2 parameters between HD and patients, we evaluated how each component altered cell population dynamics and showed that the accumulation of these changes led to a hyper- or hypo-immune response.

5.5. MEASURING HUMAN B CELL HEALTH USING THE CYTON2 MODEL

A Relative contribution of the Cyton2 parameters compare to the HD avg.



B Relative contribution for all CVID and SS donors

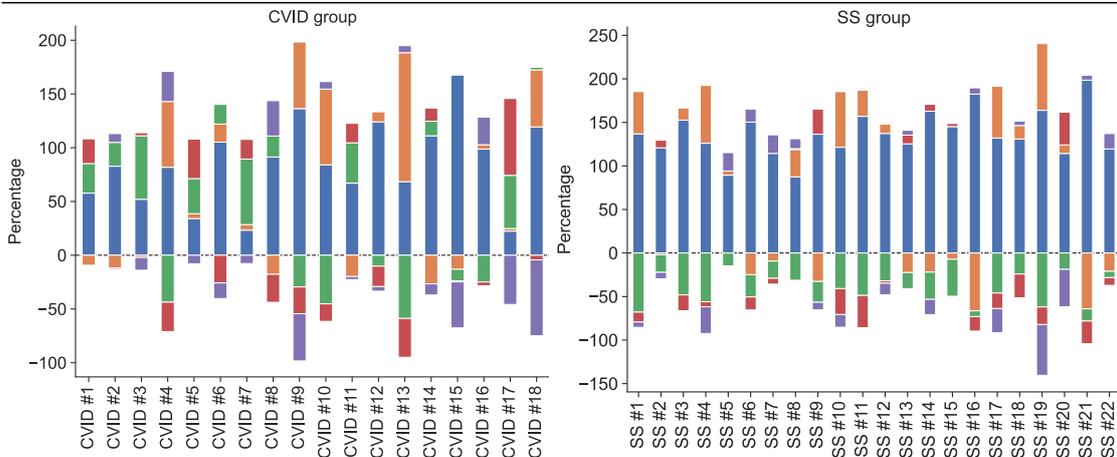


Fig. 5.5: Using the estimated Cyton2 parameters to compute relative contributions compare to the HD average. (A) An example calculation at each stage for a CVID patient. Black solid line and dashed line indicate the HD average and original best-fit model, respectively. Dash-dotted line shows change in the response at each stage, where the colour represents a substituted parameter. Stacked bar plot shows percentage contribution using total cell numbers at t^* relative to that of the HD average. (B) Relative contributions for all CVID and SS individuals.

5.6 Computing likelihood of CVID via Bayesian-like simulation

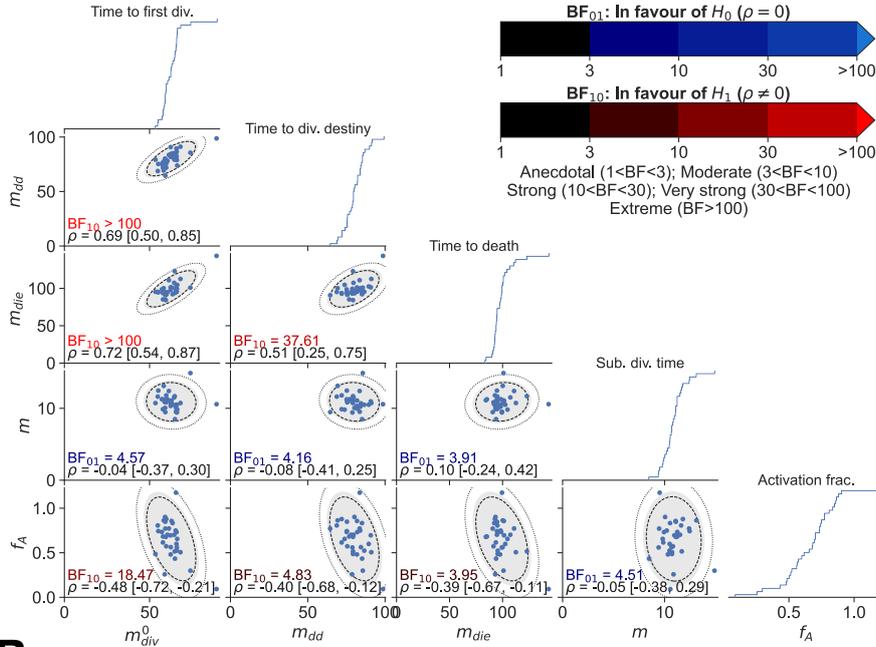
In the previous section, we found evidence supporting our hypothesis that accumulation of small deviations of the components from average HD could lead to CVID disorder. Having created a stochastic model, one feature that we could exploit is to make extrapolations beyond the data. So we questioned whether or not we could further apply the model to estimate likelihood of these diseases given some degrees of variability in the HD parameters. That is, we assume that the genetic makeup of the individual dictates the value in each of the independent modular components of the Cyton2 model under these controlled conditions. We performed a Monte-Carlo (MC) simulation by assuming that components of the Cyton2 model are chosen at random. Essentially, we marginally sampled parameter values to construct a new “healthy” individual that was compared with the CVID measurements. A key ingredient of this method is to determine parametric distribution classes that accurately capture the collection of HD parameter estimates such that a new set of values could be sampled. To first approximation, we limited our sampling to the following parameter set, $(m_{div}^0, m_{dd}, m_{die}, m, f_A)$, and used average log-variances for clear and intuitive demonstration of the method.

First, we investigated correlation structure of the HD parameter estimates to explore relatedness of the parameters within HD group. In [Fig. 5.6A](#), all possible pairs of the parameters are shown. Correlation coefficients were estimated using the Bayesian framework described in [Section 2.4.6](#) along with Bayes Factor (BF) to indicate if the data is more probable under the null hypothesis ($H_0 : \rho = 0$). Here, all pairs that consist of m and f_A were reported to have little to no correlation. Interestingly, the medians of T_{div}^0, T_{dd} and T_{die} had moderate to strong correlation, which indicates cells from a individual HD tends to have later cell destiny and death as they entered the first division at relatively later time. One could raise a question that consideration of these correlations might be crucial during the sampling process, however, we assumed the independence for simplicity.

To randomly sample the components, we marginally fitted candidate parametric distribution classes given as follows: Truncated Normal with left truncation at zero (i.e. $a = 0$), Lognormal, Normal and Gamma distributions for $m_{div}^0, m_{dd}, m_{die}$ and m ; and, Truncated Normal with left and right truncation at zero and one (i.e.

5.6. COMPUTING LIKELIHOOD OF CVID VIA BAYESIAN-LIKE SIMULATION

A Bayesian correlation analysis



B Determining parametric distribution class

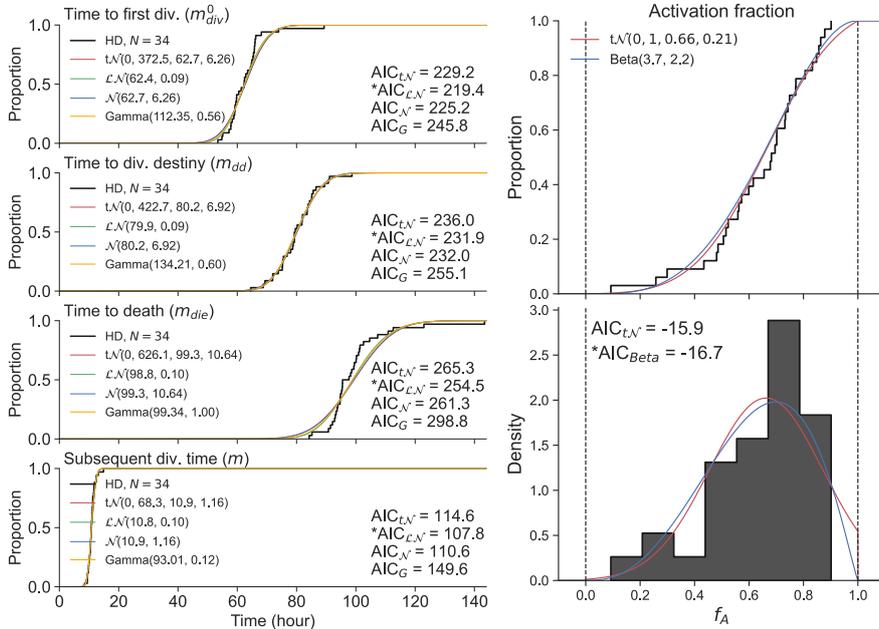


Fig. 5.6: Marginal distribution of key fitted Cyton2 parameters and parametric distribution classes that describe each of the components for simulation. (A) Bayesian correlation analysis for independence of the parameters between HD individuals ($N = 34$). Bayesian Factor (BF) and correlation coefficient with 95% credible interval are annotated in each panel (adopted method described in Section 2.4.6). Shade areas represent 90, 95 and 99% density regions. (B) Fitted various parametric distribution classes for each component. Truncated Normal ($t\mathcal{N}(a = 0, b, \mu, \sigma)$), Lognormal ($\mathcal{LN}(m, s)$), Normal ($\mathcal{N}(\mu, \sigma)$) and Gamma ($\text{Gamma}(\alpha_G, \beta_G)$) distributions were used for $m_{div}^0, m_{dd}, m_{die}$ and m . By definition $f_A \in [0, 1]$, so Truncated Normal ($t\mathcal{N}(a = 0, b = 1, \mu, \sigma)$) and Beta ($\text{Beta}(\alpha_B, \beta_B)$) distributions were tested. The best descriptor was determined by the lowest AIC value marked with an asterisk (*).

5.6. COMPUTING LIKELIHOOD OF CVID VIA BAYESIAN-LIKE SIMULATION

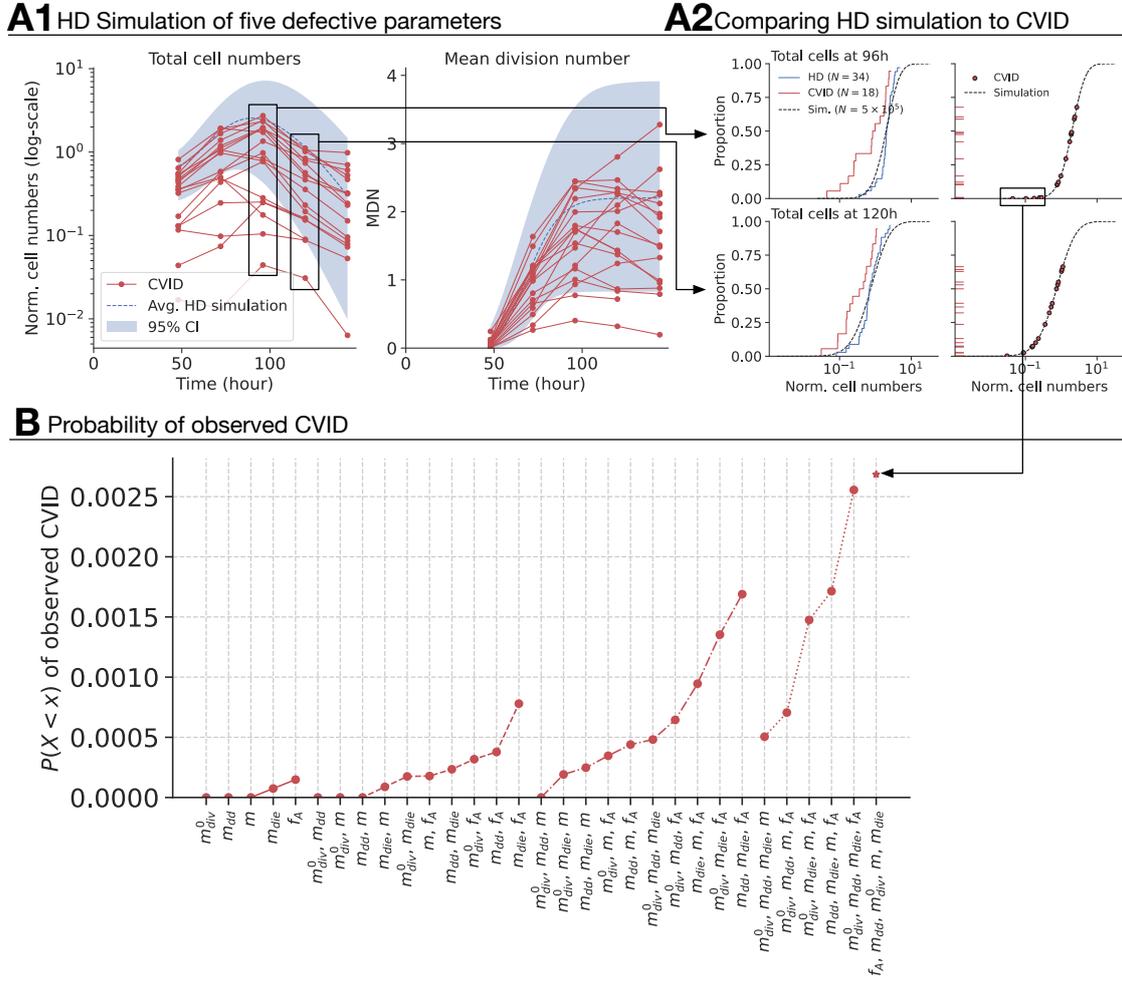


Fig. 5.7: Simulation of HD with parameter sampling and likelihood of measured CVID. (A1) Total cell numbers and mean division number (MDN) of 500,000 simulated Cyton2 model overlaid with CVID measurements. Each coloured solid line indicates an individual donor. Dashed-line represents average HD by sampling $m_{div}^0, m_{dd}, m_{die}, m$ and f_A parameters. Shade-area is 95% confidence band. (A2) Empirical CDF (eCDF) plot of the normalised total cell numbers for HD and CVID and the simulation at 96h (*top-row*) and 120h (*bot-row*). Location of CVID data points are shown on the eCDF of the simulated total cells. (B) The likelihood of lowest five CVID total cell number measurements given the simulation at 96h for all combination of the parameter components that were sampled from the fitted distributions.

5.6. COMPUTING LIKELIHOOD OF CVID VIA BAYESIAN-LIKE SIMULATION

$a = 0, b = 0$), respectively, and Beta distributions for f_A . We present the fit results in Fig. 5.6B. Qualitatively, all candidates appeared to well describe the data. So, we made decision with the model selection criteria, Akaike Information Criterion (AIC) in particular, so that whichever class has the lowest AIC value was picked for further downstream analysis. In our case, Lognormal distribution was the best descriptor for all $m_{div}^0, m_{dd}, m_{die}$ and m and Beta distribution was chosen for f_A . These distributions are specifically helpful as they naturally limit the sampling to positive real values (and $f_A \in [0, 1]$ for a percentage activation), which is suitable for our biological application.

Given these distributions, we constructed a new parameter vector by sampling each component independently of each other for $Q = 500,000$ times to create Q numbers of simulated immune response using the Cyton2 model. Here, we varied the number of components that were simultaneously sampled at each iteration to investigate the level of deviation, which ultimately affected the probability of observing CVID measurements. A summary of sampling all five components overlaid with experimental measurements from CVID patients is plotted in Fig. 5.7A1. To compare with the data, we picked 96h and 120h time points and plotted the locations of CVID data points using empirical CDFs (Fig. 5.7A2). The total cell numbers were normalised to compensate different starting cell numbers. With these information, we calculated the average likelihood, denoted $P(X < x)$ where x is the data, for each combination of the parameter components (Fig. 5.7B). Here, we chose the lowest five CVID samples at 96h, which are located at the left tail of the simulated HD distribution, to reflect that not all CVID patients exhibit low responses in cell counts. Because CVID can be caused by a variety of factors, it is possible that some people have been diagnosed with a low antibody production rate but have a normal rate of cell proliferation. The results show that the likelihood of collectively falling into the very weak overall response range slowly increased as the number of parameter components was added in the sampling process, which indicates combined multiple defects play a role in immunodeficiency. In particular, The defects in f_A , and subsequently with m_{die} and m_{dd} , appeared to dramatically increase the probability, implying that low cell activation and inadequate control of cell death and destiny are potentially sufficient to drag cell counts to levels comparable to CVID patients.

5.7 Discussion

In this chapter, we explored further application of the Cyton2 model in human system to investigate B cell population dynamics. In particular, we summarised key characteristics of CVID and SS cells given the HD cells as the reference point. The variability in the total cell numbers during cell expansion phase for the CVID patients was larger than that of SS patients. Moreover, CVID cells tend to die earlier and undergo less cell expansion, while SS cells on average divided more and died at later time points. These behaviours captured hallmarks of immunodeficiency and autoimmune disorders, respectively.

We used the Cyton2 model and designed a new parameter estimation strategy to capture the proportion of cells that switched to a reprogrammed cell death after activation. This was particularly useful method as it was observed that the majority of human B cells perceived signals but only some fractions of cells actually commenced division. Therefore, we introduced a new parameter called activation fraction that summarised behaviour of cells for the first 24 hours since stimulation and was further utilised in estimating Cyton2 parameters. Ultimately, the model was able to accurately recapitulate cell population dynamics for all CVID and SS patients.

With the new fitting strategy and parameter estimates, we quantitatively assessed B cell health by computing contribution of each parameter components relative to the average HD response. For our application, we used the maximum total cell numbers as a reference metric to measure the B cell health. Here, we found that each CVID patient has diverse defects and their compounding effect led to reduced cell numbers. In contrast, SS patients had comparatively consistent relative contributions across all parameters, where the activation fraction, time to first division and time to division destiny appeared to be the largest contributors to changes associated with this form of autoimmunity.

We then further utilised the Cyton2 model to simulate the immune response of a individual HD by allowing variability around the parameter estimates for the average HD. To do so, each component of estimated Cyton2 parameters was marginally fitted to candidate parametric distribution classes. These distributions were used to sample new parameter values such that a new parameter vector was constructed to feed into the model. This process ultimately allowed us to compute likelihood of observing CVID measurements using normalised total cell number as our comparison point.

For CVID measurements, the likelihood was increased as we introduced number of parameter components that could be sampled from. Currently, there are no precise data on the prevalence of CVID in the literature, however, it is estimated at between 1 in 100,000 and 1 in 10,000 of the population ([Bonilla et al., 2016](#); [Bogaert et al., 2016](#)), which is 25 times smaller than our estimation. Despite with the discrepancy, our result is consistent with the previous finding that CVID is a multifactorial disorder, and the model opens a new possibility to decompose such complex disorder into smaller components to identify governing factors.

Taken together, these results demonstrated valuable applications of the Cyton2 model not only in murine but also in human systems. Despite the complex and different behaviour of cells that normally were not observed in murine cells at early time points, the model was able to dissect non-linear cell population dynamics and could be used as a tool to quantitatively assess and standardise human system.

6

Conclusion

In this thesis, we have proposed a novel stochastic model of lymphocyte population dynamics and demonstrated its various applications in *in vitro* experiments of murine and human B and T cells. Our model extends the original Cyton model (Hawkins et al., 2007a) by incorporating recent findings on lymphocyte behaviours, such as correlated cell fates of progeny cells within a family (Hawkins et al., 2009; Markham et al., 2010; Duffy et al., 2012; Dowling et al., 2014; Mitchell et al., 2018). Using this model, we have developed an analysis framework that can be readily applied to population datasets, particularly those obtained from cell division tracking assays, such as cell staining with fluorescent dyes and Fluorescence-activated cell sorting (FACS).

In Chapter 2, we developed the Cyton2 model, which incorporates the inheritance of death and division stopping times (i.e. division destiny) (Heinzel et al., 2017), and provided a derivation of the mean and variance equations of lymphocyte population dynamics. In this model, cells in a family were assumed to follow internal timers that govern division, cessation, and death. We analysed the stochastic competition of these cellular machineries in detail using datasets from time-lapse microscopy, which provided accurate recordings of the times to each cell's fate (Hawkins et al., 2009). We adopted Bayesian hypothesis testing and Monte-Carlo (MC) simulation to infer the independent operation of the timers. Additionally, we determined the parametric distribution class that best describes the measurements, namely the Lognormal distribution. These results support the crucial assumptions that constitute the model. Using the least-squares method and an appropriate optimisation algorithm, such as Levenberg-Marquardt or genetic algorithms, the model parameters can be estimated for lymphocyte populations under various conditions.

In [Chapter 3](#), we presented three model applications for *in vitro* experiments of murine B and T cells. Firstly, a dataset from CpG-stimulated B cells was used to assess the accuracy and precision of the model estimates. This was a particularly suitable set for testing as it contains larger samples than a typical time-series lymphocyte population dataset. The model was capable of recapitulating the immune response with as little as three time points in the data. Moreover, we demonstrated three replicates per time point, which is a standard sample size in the field, was sufficient to obtain accurate parameter estimates.

We then applied the model to analyse previously published CD8⁺ T cell data from [Marchingo et al. \(2014\)](#) to provide alternate, but functionally consistent, perception of the signal integration. Here, we showed that multiple *positive* signals, which amplify T cell expansion when combined, add linearly and operate independently in the time domain. In particular, we introduced a simple sum of normally distributed random variables, i.e. the timers, to predict the amplified cell expansion.

Lastly, we investigated the effects of combination of *negative* signals on CD8⁺ T cells, carried by two immunosuppressants, Rapamycin and Dexamethasone, at various concentrations. These drugs act independently of each other by either inhibiting cell cycle progression, thus slowing down cell division, or reducing cell survival, respectively. Utilising this characteristic and the model, we predicted the cell population sizes at different degrees of drug concentrations by a simple rule of replacing the affected Cyton2 parameters.

In [Chapter 4](#), we introduced probabilistic models of B cell differentiation into antibody-secreting cells (ASCs) and isotype class switching recombination (CSR). The development of the models and associated analyses were based on the expression measurements, such as *Activation-Induced cytidine Deaminase* (AID) and Germline Transcription (GLT), obtained from multiplexed division tracking assays ([Horton et al., 2018](#)) and single-cell quantitative polymerase chain reaction (sc-qPCR) to record cell fates for each generation. Here, we tested clonal dependencies of these molecules via MC simulation as well as hypothesis testing, and deduced the heritability of AID and the re-randomisation of GLT expression. Moreover, the regulation of these molecules operates independently of each other. Using these distinct familial features, we formulated a stochastic model that captures differentiation of IgM⁺ to IgG1⁺ and/or IgG2b⁺ B cells depending on the AID and GLT states and predicted proportion of antibody switching outcomes under varied conditions.

In [Chapter 5](#), we presented the model application to human lymphocytes to interrogate datasets obtained from immuno-compromised patient groups. Specifically, B cells from Common Variable Immunodeficiency (CVID) and Sjögren Syndrome (SS) patients were analysed to compare with immune responses from healthy donors (HDs) for identifying distinguishing features. The summary statistics and non-parametric hypothesis testing via permutation test revealed that the CVID cells tend to undergo fewer cell divisions and to die earlier than HD cells, leading to an overall small population size. However, the SS cells divide, on average, one additional round of division than HD cells while they both follow a similar death profile, resulting in larger cell numbers over a given period of time. These observations were further dissected using the Cyton2 model to identify specific modules that may be defective and responsible for hypo- and hyper-immune responses. Here, the marginal distributions of each Cyton2 parameter estimates indicated that a low activation fraction, which is defined as the proportion of cells that commence the division program, for the CVID patients was clearly one of the key factors, and that slow time to division destiny for the SS patients was what allowed cells to traverse further generation. More importantly, comparing the results between individuals revealed that the CVID patients exhibited defects in different components, indicating that CVID is a multifactorial disorder. In comparison, the SS patients displayed relatively consistent defects across individuals.

As we found evidence that the compounding effect of small defects could lead to CVID disorder, we utilised the Cyton2 model to simulate an immune response by sampling the parameters at random based on average HD estimates. This enabled us to extrapolate beyond the data and collect instances that, by chance, have lower cell counts, resembling the responses from the CVID cells. For example, a random parameter set that sampled low activation fraction and earlier cell death would result in limited cell expansion. This procedure was repeated to estimate likelihood of CVID. Here, we observed an increase in the probability as we allowed more parameters to be sampled from.

Finally, the implementation of Cyton2 and isotype switching models presented in this thesis are publicly available at <https://github.com/hodgkinlab> in the repositories called `cyton2-paper` for [Chapters 2](#) and [3](#) and `ism-paper` for [Chapter 4](#).

6.1 Future work

Our research focused on modelling the kinetics of lymphocyte population based on the stochastic nature of the cellular machineries and accurately captured immune responses in diverse circumstances. However, there are numerous potential improvements can be made and limitations to overcome.

One of major aspects we can improve the Cyton2 model is to incorporate the full stochastic feature of the timers, that is, introducing a random variable, instead of a constant, for the subsequent division times to allow for variability around the mean. Although we have exploited concordant cell fates in a single family, number of recent experiments in the literature suggest a proportion of frayed families due to slight discrepancies in times and quasi-symmetric fates between progeny cells. In terms of model fitting, adopting a Bayesian framework to obtain posterior distributions of the parameters, rather than their point estimates and confidence intervals via bootstrap method, would provide a more robust and intuitive statistical interpretation of a given system. For example, Approximate Bayesian Computation ([Beaumont, 2019](#)) can be a great approach for complex systems where the likelihood function cannot be easily derived. [Belluccini et al. \(2022\)](#) presented such implementation for similar research questions as those shown in our work.

From a more overarching perspective, incorporating the differentiation process of B and T cells into the model would provide a more realistic depiction of a complete immune response. In [Chapter 4](#), we explored B cell differentiation and isotype class switching at the molecular level to deduce the inner workings and stochastic regulation of the process. However, this work requires further investigation as B cells can differentiate to different classes (e.g. IgM secreting cells), and the stochastic competition involving multiple alternate paths would exhibit non-trivial outcomes. Ultimately, a model that includes processes of cell proliferation, cessation, death, and differentiation concurrently would have potential to unravel complex interactions and elucidate how our immune system protects against harmful pathogens from activation to eradication in a predictable manner.

A

Experiments designed and conducted by collaborators

In this appendix, we present sources of data studied in this thesis, particularly the experimental systems, as implemented by our collaborators, for completeness so that the reader has the full context for the data used in the mathematical models developed in this thesis. [Section A.1](#) provides additional information of the time-lapse microscope experiment for CD8⁺ T cells from [Chapter 2](#). This experiment was planned, analysed and performed by Simone C. Oostindie, Edwin D. Hawkins, Julia M. Marchingo, Susanne Heinzl and Philip D. Hodgkin in [Cheon et al. \(2021\)](#). In [Section A.3](#), the details for drug experiments presented in [Chapter 3](#) are shown. In [Section A.2](#), we report the details for the experiments used in [Chapter 5](#), performed by Miles B. Horton.

A.1 Chapter 2: Time-lapse microscopy experiment

A.1.1 Mice

All mice were maintained under specific pathogen-free conditions in the WEHI animal facilities (Parkville, Victoria, Australia) and used at 5 to 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-

A.1. CHAPTER 2: TIME-LAPSE MICROSCOPY EXPERIMENT

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.

A.1.2 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 Cell Trace Violet (CTV, Invitrogen) to track and monitor cell division in parallel bulk cultured by flow cytometry. Labelling was performed for 20 minutes at 37°C in PBS + 0.1% BSA.

A.1.3 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 40000 cells in 1mL per well in the presence of 1, 3.16 or 10U/mL recombinant human IL-2 (PeproTech) and 25 μ g/mL S4B6. After 24 hours of culture cells were harvested, washed twice with filming medium, counted and resuspended at 5000 to 10000 cell/mL in filming medium supplemented with 25 μ g/mL S4B6 and 1, 3.16 or 10U/mL recombinant human IL-2.

In experiments using N4 peptide CD8⁺ T cells were cultured with 0.01 μ g/mL N4 at 2×10^4 cells per mL in 200 μ L of a 96-well U-bottom plate in the presence of 25 μ g/mL S4B6. 2 μ g/mL α CD28 (clone 37.51, WEHI antibody facility) or 1ng/mL

A.1. CHAPTER 2: TIME-LAPSE MICROSCOPY EXPERIMENT

IL-12 (Peprotec) were added as indicated. Cells were cultured for 24 hours, washed and resuspended at 5000 to 10000 cells/mL in filming medium containing 25 μ g/mL S4B6.

For one experiment CD8⁺ T cells cultured for 24 hours with N4 in presence or absence of α CD28 were split and supplemented or not with 1U/mL rhIL-2 before replating for filming.

In one experiment CD8⁺ T cells were cultured with N4 alone and addition of either α CD28, IL-12 or both for 24 hours, then washed, resuspended and replated for filming without any further stimuli added.

For filming, 250 μ L cell suspension was added per chamber of an 8 well μ -Slide chamber (Ibidi) containing 125 μ m (MGA-125-01) or 70 μ m (MGA-7-01) microgrids (Daniel Day, Microsurfaces). These conditions resulted in a significant portion of microwells containing one cell per well. Before the start of filming, cells were incubated for \approx 2 hours 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.

A.1.4 Live cell imaging and cell tracking

For single cell filming, microgrids (70/125 μ m, Daniel Day, Microsurfaces) were placed into an 8 well chamberslide (Ibidi μ -slide). Chambers were UV sterilised with 40 μ L 100% ethanol in a laminar-flow cabinet for at least 30 minutes until dry. Another 40 μ L ethanol was added and rinsed 10x with filming medium (advanced RPMI 1640 without phenol red). 250 μ L 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 hours using a 470nm LED, just prior to adding the cells.

Live cell imaging was conducted on an environmentally controlled 37°C + 5% CO₂ humidified Zeiss Axiovert 200M 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 470/555nm set at 25%/100%, respectively, with an exposure time of 200 μ s) was used for detecting green and red fluorescence. Red, green, bright-field and out-of-focus images were taken at 165 second intervals for 5 or 6 days. Bright-field and fluorescent raw images of single cells in microgrids were digitally processed

A.2. CHAPTER 3: DRUG EXPERIMENT

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.

A.2 Chapter 3: Drug experiment

A.2.1 Cells

OT-I Bim^{-/-} CD8⁺ T cells isolated from lymph nodes using EasySepTM Mouse CD8⁺ T Cell Isolation Kit (Stemcell Technologies). Cells were labelled with CTV (final concentration 5 μ M) before being cultured.

A.2.2 Culture conditions and stimulation

Cells were placed in 96 well plates at 8000 cells/well in 200 μ L of T cell medium. All cells stimulated with SIINFEKL peptide (N4) 10ng/mL and α CD28 0.2 μ g/mL; anti mouse IL-2 Ab clone S4B6 (25 μ g/mL) added to neutralise endogenously produced IL-2.

A.2.3 Drugs

Dexamethasone (Calbiochem) final concentrations in culture: 10ng/mL, 3.16ng/mL, 1ng/mL. Rapamycin (LC laboratories) final concentrations in culture: 1580pg/mL, 500pg/mL, 158 pg/mL. Cells were cultured at 37°C, 5% CO₂, humidified atmosphere and were harvested at 7 distinct time points: 18.75h, 41h, 49.75h, 65.6h, 72.5h, 88.5h and 96.5h.

A.3 Chapter 4: Differentiation experiment

A.3.1 Mice

C57BL/6, *Blimp-1^{gfp/+}*, *Aicda*-Cre and Ly5.1 mouse strains were maintained under specific pathogen-free conditions in the Walter and Eliza Hall Institute (WEHI) animal facilities and were used at 6-12 weeks of age. All procedures were performed with approval by the WEHI animal ethics committee. The *Blimp-1^{gfp/+}* and *Aicda*-Cre mice have been previously described ([Kallies et al., 2004](#); [Buchholz et al., 2015](#)).

A.3.2 B cell isolation and labelling

Small resting B cells were isolated from spleens by Percoll gradient of single-cell suspensions and subsequent purification by negative isolation using the EasySep Mouse B cell Isolation Kit according to the manufacturer's protocol (STEMCELL technologies). For experiments that measure cell division in bulk B cell populations, isolated B cells were labelled with 5 μ M CellTraceViolet (CTV, Invitrogen) in sterile PBS containing 0.1% Bovine Serum Albumin (PBS 0.1% BSA).

A.3.3 B cell culture

For all cell cultures medium consisted of RPMI1640, 10% fetal calf serum (FCS), 1mM sodium pyruvate, 10mM HEPES, 100 U/ml Penicillin, 100 μ g/ml Streptomycin, non-essential amino acids, 2mM GlutaMAX (all Invitrogen), and 50 μ M 2-ME (Sigma-Aldrich). Purified B cells were stimulated with 15 μ g/ml lipopolysaccharide (LPS, Sigma-Aldrich) with or without indicated concentrations of recombinant mouse IL-4 (purified from baculovirus-transfected Sf21 insect cells) and/or recombinant mouse Transforming growth factor beta (TGF- β , Invitrogen), as described per experiment in the main text. For cultures containing TGF- β , the cytokine was added to cells after 24-hours of culture in order to minimise its suppression of cell proliferation as previously described ([Deenick et al., 1999, 2005](#)). For bulk cultures cells were either stimulated in 6-well plates at a density of 10⁶ cell/ml or in 96-well plates at 10⁴ cell/well where indicated.

A.3. CHAPTER 4: DIFFERENTIATION EXPERIMENT

A.3.4 Multiplexed cell labelling, stimulation and sorting

For clonal lineage tracing experiments purified B cells were sequentially labelled with three division-tracking dyes (all Invitrogen) as previously described (Marchingo et al., 2016; Horton et al., 2018). Briefly, cells were labelled with either 10 μ M, 5 μ M or 0 μ M CellTrace Yellow (CTY), followed by 5 μ M or 0 μ M CTV and then 5 μ M or 0 μ M Cell proliferation dye eFluor670 (CPD). This generates 8 uniquely labelled cell populations. Labelled cells were stimulated in bulk at a density of 10⁶ cell/ml for 18-22 hours under indicated conditions before harvesting and single-cell sorting into 96-well plates according to the unique fluorescent label combinations to generate clonal samples. Bulk population controls for the uniquely labelled populations were also sorted to provide references for gating of clonal lineage and division number. Samples were returned to culture under the same conditions as indicated, clonal samples included approximately 2000 Ly5.1 filler B cells to assist with cell pelleting. All sorting was performed on either a FACSAria III or an Aria Fusion (BD Biosciences).

A.3.5 Antibody staining and flow cytometry

Cells were harvested at indicated time points and following surface Fc γ RII/III blocking with 24G2 antibody, stained on ice for 30 minutes in FACS buffer (PBS with 1% BSA 1mM EDTA and 0.05% sodium azide) with the following antibodies as described; anti-IgG1 FITC (1:400), anti-IgG1 APC (1:1000), anti-IgG1 PE-Cy7 (1:400), anti-IgG1 BV786 (1:200), anti-IgG2b PE-Cy7 (1:400), anti-IgD APC (1:200), anti-TLR4/MD2 PE-Cy7 (1:100), anti-RP105 PE (1:200), anti-CD45.1 BV650 (1:100), anti-CD45.2 PE-Cy7 (1:100), anti-CD45.2 BUV737 (1:100), anti-B220 BV510 (1:800), anti-CD138 BV711 (1:800), anti-CD38 PerCP-eFluor710 (1:600), anti-CD95 PE-Cy7 (1:600) and anti-human-CD2 BV786 (1:100). For IL-4R staining, cells were first stained with purified anti-IL-4R unconjugated (1:100), followed by anti-rat IgG A647 (1:500). Dead cell exclusion was performed using 0.2 μ g/ml propidium iodide (PI) for *in vitro* experiments or using fixable viability dye Alexa 700 (1:100) for analysis of *ex vivo* samples. Analysis of FACS samples was performed on Fortessa X20 or a FACSCanto (BD Biosciences).

A.3.6 Index sorting of multiplex labelled cells for single-cell qPCR

For experiments involving index sorting of multiplex samples, cells were labelled as described previously, however all cells were incubated with either $10\mu\text{M}$ or $5\mu\text{M}$ CTY but not $0\mu\text{M}$, to generate 6 uniquely labelled populations. Samples were harvested at indicated time points and stained with antibodies described in the main text and figure legends. Individual clonal samples were then index sorted into either a 384-well plate for further analysis, labelled and stained bulk controls were also acquired for assistance with gating during analysis. Index sort files were analysed using the FlowJo IndexSort plugin v1.1 and fluorescence data was used to identify cells by their clonal lineage, division number and surface marker expression in relation to subsequent molecular analysis. Single cell index sorting was performed using a FACSAria Fusion.

A.3.7 Single-cell quantitative reverse transcription PCR

For analysis of germline transcript expression in single cells, cell samples were harvested at indicated time points, stained with the described antibodies and index sorted into 384-well plates containing $1\mu\text{l}$ nuclease-free water with 0.2% Triton-X (Sigma-Aldrich) and 5% SupraseIN (Invitrogen). Cell lysate was subjected to one-step RT-PCR using the Qiagen QuantiTect multiplex RT-PCR kit with indicated custom primers and probes from Integrated DNA Technologies (IDT). Primer and probe sequences are described in [Table A.1](#). RT-qPCR was performed using an Applied Biosystems Quantstudio 12k system for 40 cycles. Single-cell qPCR data analysis was performed using LinRegPCR software and expression values are represented using relative quantities (RQ) as previously described, with a detection limit of 40 cycles ([Ståhlberg et al., 2013](#)).

A.3.8 I γ 1-sorting using molecular beacons

Molecular beacons were utilised to identify RNA expression in living cells as described ([Wile et al., 2014](#); [Zhao et al., 2016](#)). Appropriate beacons were designed targeting the I γ 1 germline transcript or β -actin using the *PinMol* algorithm ([Bayer et al., 2019](#)), which utilises the folded secondary structure of a targeted RNA molecule as input to identify accessible regions and optimise beacon properties. Beacons were designed to contain a stem comprised of 2' O-methyl RNA bases in order to maximise cytosolic localisation and minimise non-specific stem opening, thereby

A.3. CHAPTER 4: DIFFERENTIATION EXPERIMENT

Primer/probe	Sequence
I γ 1 Fwd	5'-ATATCGAGAAGCCTGAGGAATG-3'
I γ 1 Rev	5'-GGCCAGTGGATAGACAGATG-3'
I γ 1 Probe	5'-ACAGGTTGAGAGAACCAAGGAAGCTG-3'
I γ 2b Fwd	5'-CAGAGTTCTCACACACAGAAGAA-3'
I γ 2b Rev	5'-ATCCCAGAGTCACAGAGGAA-3'
I γ 2b Probe	5'-CCCTGGGTGTGGAGATACTGG-3'
ActB Fwd	5'-GACTCATCGTACTCCTGCTTG-3'
ActB Rev	5'-GATTACTGCTCTGGCTCCTAG-3'
ActB Probe	5'-CTGGCCTCACTGTCCACCTTCC-3'
I γ 1 molecular beacon 1	5'-mCmGmAmCmGmC*mU*mU*mC*mU*mC*mG*mA*mU*mA*mU*mA*mU*mU*mC*mC*mA*mU*mG*mA*mCmGmUmCmG-3'
I γ 1 molecular beacon 2	5'-mCmGmCmGmAmG*mU*mC*mU*mC*mU*mG*mU*mU*mC*mC*mU*mG*mU*mU*mU*mU*mG*mG*mU*mUmCmGmCmG-3'
Negative control beacon	5'-mCmCmGmCmGmCmG*mU*mU*mG*mA*mG*mA*mG*mA*mA*mC*mC*mA*mA*mG*mG*mA*mA*mG*mC*mU*mG*mGmCmGmCmGmG-3'
Transfer control beacon	5'-mAmCmGmAmCmGmC*mG*mA*mC*mA*mA*mG*mC*mG*mC*mA*mC*mC*mG*mA*mU*mA*mCmGmUmCmGmU-3'

'm_' – denotes 2' O-methyl RNA bases

'm_*' – denotes phosphorothioated 2' O-methyl RNA bases

Table A.1: Oligo sequences

limiting false-positive signals. Molecular beacons were transfected into cells by electroporation using the Lonza 4D-nucleiofactor protocol for stimulated mouse B cells. 6×10^5 stimulated cells per reaction were resuspended in nucleofactor solution and 500nM of the indicated molecular beacon, either target-specific, negative control or transfection control, and electroporated using code DI-100 with a Lonza Amaxa 4D nucleofactor system. After electroporation, cells were resuspended in 200 μ L pre-warmed 37°C medium followed by subsequent culture and analysis at indicated time points by flow cytometry.

A.3. CHAPTER 4: DIFFERENTIATION EXPERIMENT

A.3.9 LCMV infection

Mice were infected with 3×10^3 plaque-forming units (PFU) of LCMV Armstrong by intravenous injection. After 5 days spleens were harvested, and B cells were isolated from single cell suspensions by negative isolation. Purified B cells were then stained with indicated antibodies for 20 minutes, followed by viability staining for 10 minutes at 4°C. Samples were then either sorted for single-cell qPCR, or analysed by flow cytometry.

A.3.10 Statistics

Statistical analysis presented in the [Appendix B](#) was performed by Miles B. Horton using Prism 7. Correlations were analysed using the Pearson's or Spearman's rank correlation coefficient where stated, and bar graph comparisons were performed using the Holm-Sidak's multiple comparisons test.

B

Supplementary materials for **Chapter 4**

In this appendix, we present some key results of experiments and analyses, which appear in [Horton et al. \(2022\)](#), performed by our main collaborators at Walter and Eliza Hall Institute (WEHI). These findings provide important biological foundations that motivated theoretical work of process of B cell differentiation and the development of the mathematical model, both of which are described in **Chapter 4**.

B.1 Figures

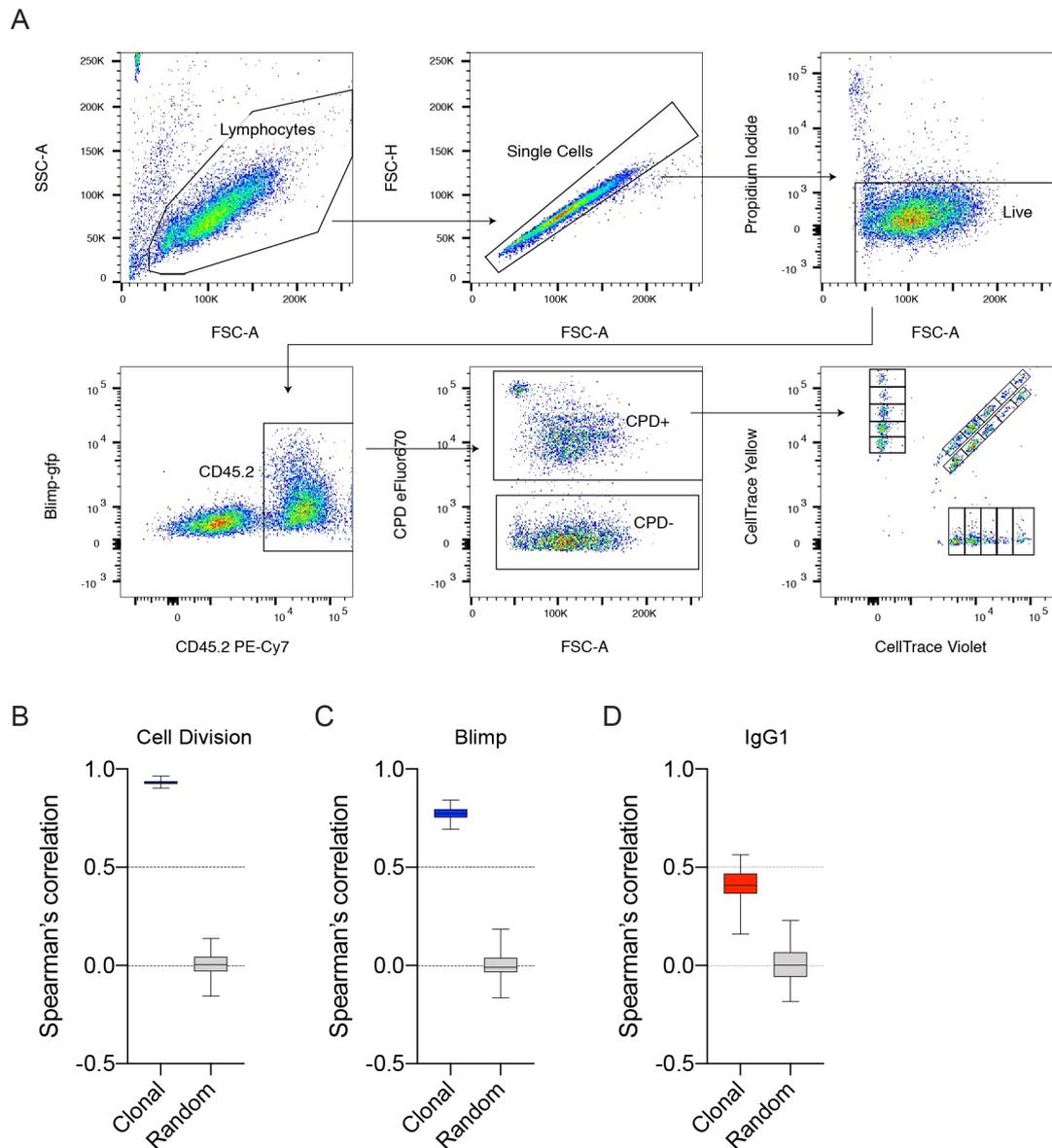


Fig. B.1: Clonal analysis of B cell fates by multiplexed fluorescent labelling. [Corresponding to Supplementary Figure S1 from [Horton et al. \(2022\)](#) and appear in [Section 4.4](#)] **(A)** An example of the flow cytometry gating layout used to identify clonal populations according to multiplexed labelling status. **(B-D)** Data relating to [Fig. 4.2](#), showing distributions of the intraclonal correlations calculated using Spearman's for the indicated cell fate following repeated sampling of cell pairs either from within clones, as experimentally measured, or following randomisation.

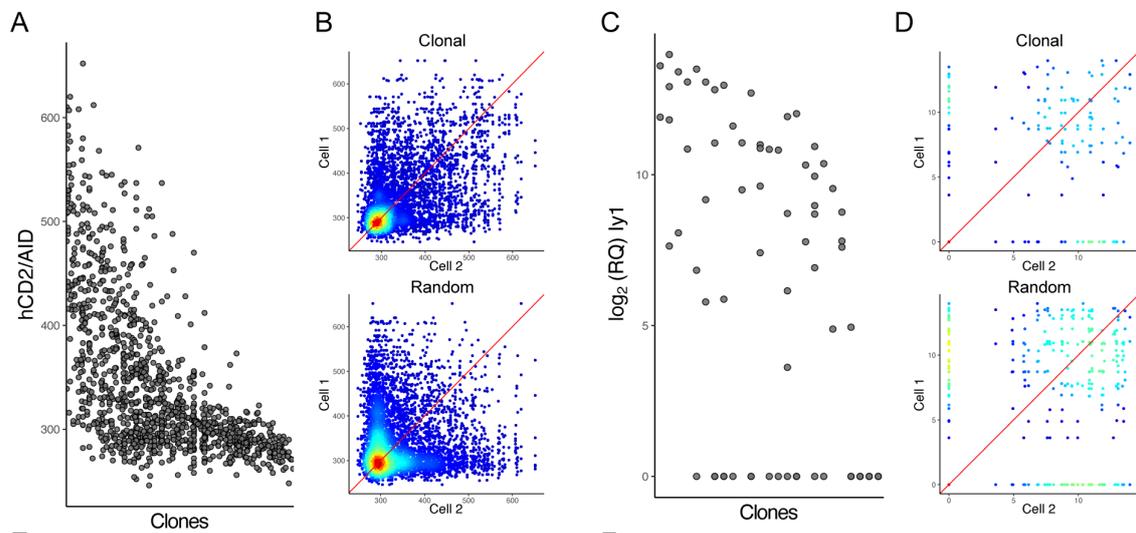


Fig. B.2: Emergence of class-switched B cells is characterised by clonal heterogeneity in division-linked AID expression and division-independent $\text{I}\gamma 1$ transcription. [Corresponding to Supplementary Figure 3 from Horton et al. (2022) and appear in Section 4.5] Lineage tracing of individual B cell clones analysed for (A) AID expression or (C) $\text{I}\gamma 1$ transcription. Repeated sampling of cell pairs from within clones or following randomisation for (B) AID expression and (D) $\text{I}\gamma 1$ transcription.

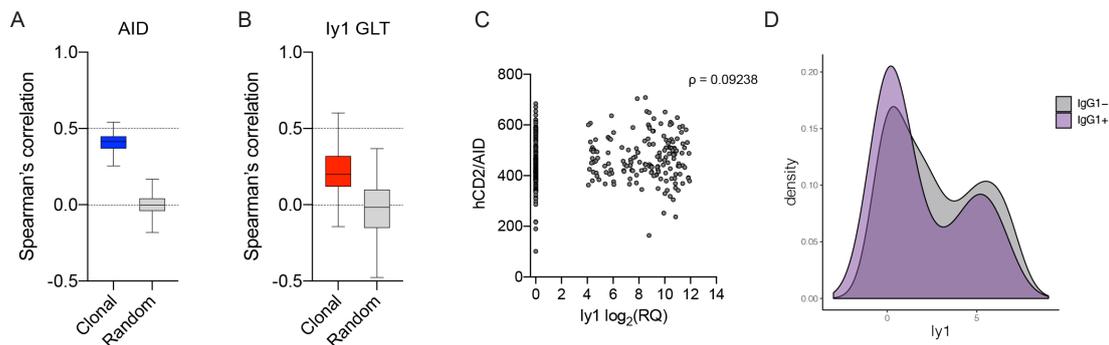


Fig. B.3: Clonal and single-cell analysis of hCD2/AID expression and I γ 1 germline transcription. [Corresponding to Supplementary Figure S3 from Horton et al. (2022) and appear in Section 4.5] Data as in Supplementary Fig. B.2, clonal analysis of AID and I γ 1 expression. Distributions of intraclonal correlations by Spearman's for (A) AID and (B) I γ 1 from measured data or following randomisation. (C) hCD2/AID reporter B cells were activated with LPS and IL-4 for 72 hours prior to staining for surface hCD2, single cell index sorting into a 384-well plate and analysis of I γ 1 expression by sc-qPCR. Per-cell levels of both hCD2/AID and I γ 1 were determined by relating single-cell qPCR results to index sort FACS data. Expression levels are indicated by FACS channel values for hCD2/AID and relative quantity (RQ) for I γ 1. Rho value represents Spearman's rank correlation coefficient. (D) Single-cell analysis of I γ 1 expression in B cells activated with LPS and IL-4 for 72 hours prior to staining for surface IgG1 and index sorting into a 384-well for qPCR. I γ 1 expression is displayed based on surface expression of IgG1. Representative of 2 independent experiments.

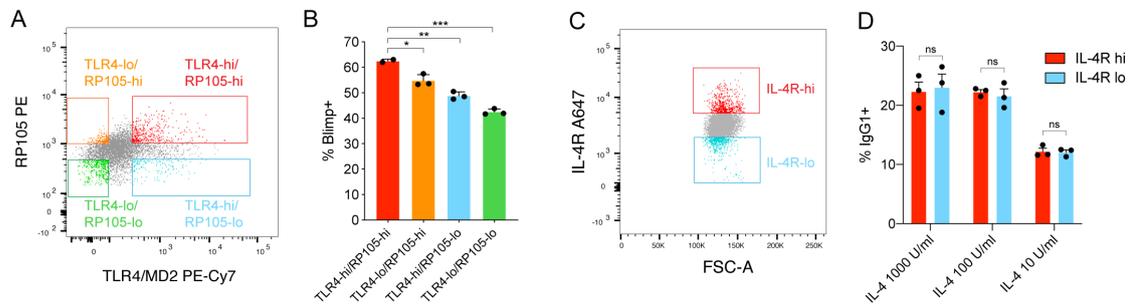


Fig. B.4: Pre-existing fate potential is identifiable in naïve B cells for ASC differentiation, but not CSR, and is linked to receptor sensitivity. [Corresponding to Figure 2 from Horton et al. (2022) and appear in Section 4.8] (A) Sorting of naïve $\text{Blimp}^{\text{gfp}/+}$ B cells by expression of LPS receptors RP105 and TLR4/MD2, which were subsequently stimulated with LPS for 3 days and (B) analysed for differentiation. (C) Sort gates for 2 populations of naïve wild-type B cells by IL-4R α expression, which were subsequently stimulated with LPS and varying concentrations of IL-4 for 3 days and (D) analysed for CSR to IgG1. Representative of 2 independent experiments. Mean \pm standard error. Statistical analysis using the Holm-Sidak's multiple comparison test. $*p < 0.05$, $**p < 0.005$, $***p < 0.0005$.

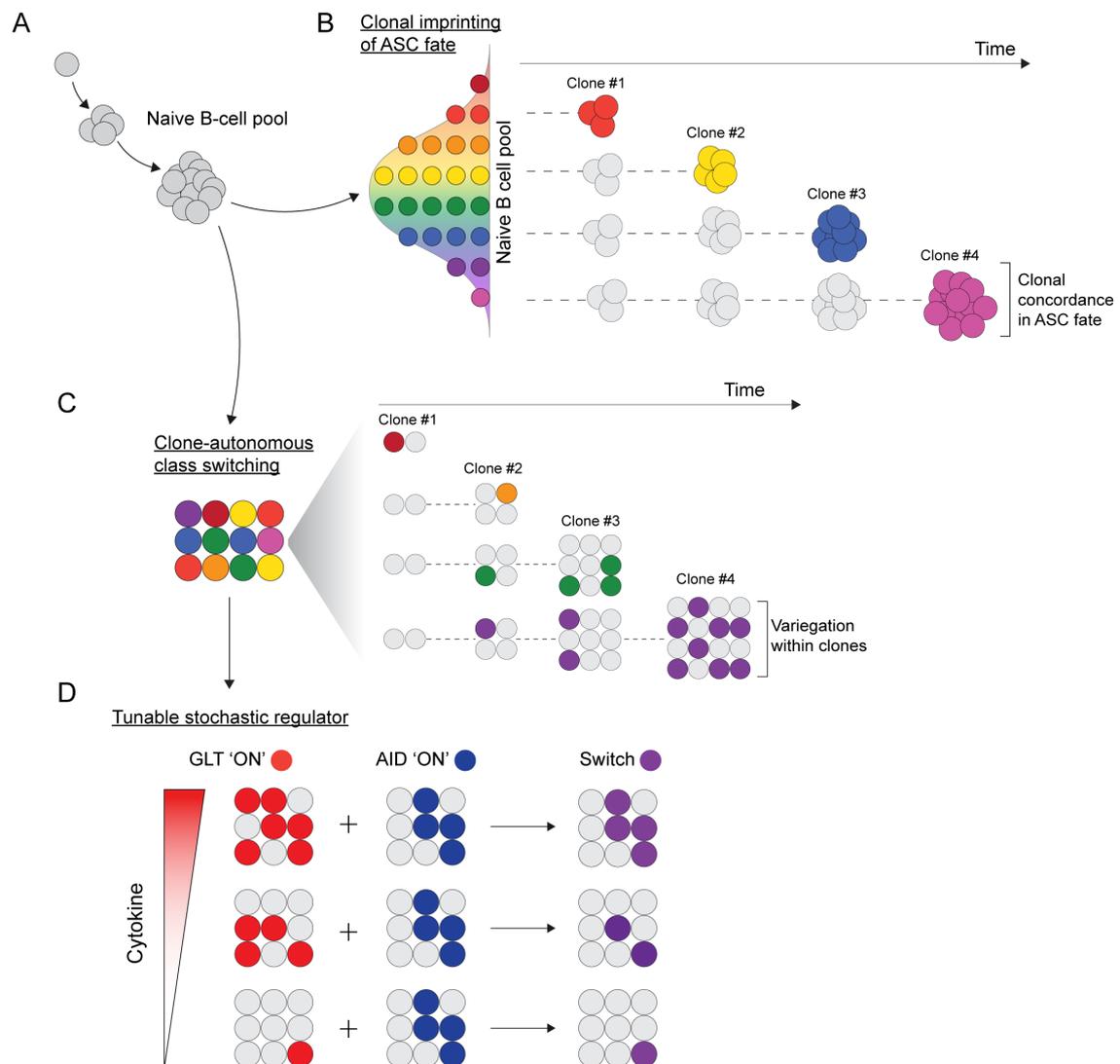


Fig. B.5: Diverse clonal mechanisms for controlling fate heterogeneity in B cells. [Corresponding to Figure 7 from Horton et al. (2022) and appear in Section 4.8] (A) Individual naïve cells vary in construction, including variation in receptor expression, to generate a population with diverse potential for ASC differentiation. (B) ASC fate selection and timing are conserved within clones. (C) Isotype switching utilises cell autonomous probability modulation that leads to variegated clones. (D) External cytokines determine the likelihood of binary GLT expression in a division-independent manner. This intersects with a separate probability for AID expression to regulate the switching fate decision.

Bibliography

- Alberts, Bruce, Johnson, Alexander, Lewis, Julian, Raff, Martin, Roberts, Keith, and Walter, Peter. *Molecular Biology of the Cell*. New York: Garland Science, 2002, 4 ed.
- Ameratunga, Rohan, Woon, See-Tarn, Bryant, Vanessa L., Steele, Richard, Slade, Charlotte, Leung, Euphemia Yee, and Lehnert, Klaus. “Clinical Implications of Digenic Inheritance and Epistasis in Primary Immunodeficiency Disorders.” *Frontiers in Immunology* 8: 1965.
- Amo, Pedro Costa del, Debebe, Bisrat, Razavi-Mohseni, Milad, Nakaoka, Shinji, Worth, Andrew, Wallace, Diana, Beverley, Peter, Macallan, Derek, and Asquith, Becca. “The Rules of Human T Cell Fate in vivo.” *Frontiers in Immunology* 11: 573.
- Antia, Rustom, Bergstrom, Carl T., Pilyugin, Sergei S., Kaech, Susan M., and Ahmed, Rafi. “Models of CD8+ Responses: 1. What is the Antigen-independent Proliferation Program.” *Journal of Theoretical Biology* 221.4 (2003): 585–598.
- Asquith, Becca, Debacq, Christophe, Florins, Arnaud, Gillet, Nicolas, Sanchez-Alcaraz, Teresa, Mosley, Angelina, and Willems, Luc. “Quantifying lymphocyte kinetics in vivo using carboxyfluorescein diacetate succinimidyl ester.” *Proceedings of the Royal Society B: Biological Sciences* 273.1590 (2006): 1165–1171.
- Atitey, Komlan and Anchang, Benedict. “Mathematical Modeling of Proliferative Immune Response Initiated by Interactions Between Classical Antigen-Presenting Cells Under Joint Antagonistic IL-2 and IL-4 Signaling.” *Frontiers in Molecular Biosciences* 9: 777390.
- Ballou, Lisa M. and Lin, Richard Z. “Rapamycin and mTOR kinase inhibitors.” *Journal of Chemical Biology* 1.1-4 (2008): 27–36.
- Banks, H. T., Sutton, Karyn L., Thompson, W. Clayton, Bocharov, Gennady, Roose, Dirk, Schenkel, Tim, and Meyerhans, Andreas. “Estimation of Cell Proliferation Dynamics Using CFSE Data.” *Bulletin of Mathematical Biology* 73.1 (2011): 116–150.

- Banks, H T, Thompson, W Clayton, Peligero, Cristina, Giest, Sandra, Argilaguët, Jordi, and Meyerhans, Andreas. “A division-dependent compartmental model for computing cell numbers in CFSE-based lymphocyte proliferation assays.” *Mathematical Biosciences and Engineering* 9.4 (2012): 699–736.
- Banks, H. Thomas, Kapraun, Dustin F., Link, Kathryn G., Thompson, W. Clayton, Peligero, Cristina, Argilaguët, Jordi, and Meyerhans, Andreas. “Analysis of variability in estimates of cell proliferation parameters for cyton-based models using CFSE-based flow cytometry data.” *Journal of Inverse and Ill-posed Problems* 23.2 (2015): 135–171.
- Banks, H.T., Charles, Frédérique, Jauffret, Marie Doumic, Sutton, Karyn L., and Thompson, W. Clayton. “Label structured cell proliferation models.” *Applied Mathematics Letters* 23.12 (2010): 1412–1415. This paper is tied with section 3.2 in Banks et al. 2011 ”Estimation of...” regarding the change of variable analogous to change of reference frame in Physics.
- Barwick, Benjamin G., Scharer, Christopher D., Bally, Alexander P.R., and Boss, Jeremy M. “Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation.” *Nature immunology* 17.10 (2016): 1216–1225.
- Baxter, Alan G. and Hodgkin, Philip D. “Activation rules: the two-signal theories of immune activation.” *Nature Reviews Immunology* 2.6 (2002): 439–446.
- Bayer, Livia V., Omar, Omar S., Bratu, Diana P., and Catrina, Irina E. “Pin-Mol: Python application for designing molecular beacons for live cell imaging of endogenous mRNAs.” *RNA* 25.3 (2019): 305–318.
- Beaumont, Mark A. “Approximate Bayesian Computation.” *Annual Review of Statistics and Its Application* 6.1 (2019): 379–403.
- Belluccini, Giulia, López-García, Martín, Lythe, Grant, and Molina-París, Carmen. “Counting generations in birth and death processes with competing Erlang and exponential waiting times.” *Scientific Reports* 12.1 (2022): 11289.
- Bernard, Samuel, Pujo-Menjouet, Laurent, and Mackey, Michael C. “Analysis of Cell Kinetics Using a Cell Division Marker: Mathematical Modeling of Experimental

- Data.” *Biophysical Journal* 84.5 (2003): 3414–3424. Age-Maturity model applied to blood cell proliferation.
- Beyer, Tilo, Busse, Mandy, Hristov, Kroum, Gurbiel, Slavyana, Smida, Michal, Haus, Utz-Uwe, Ballerstein, Kathrin, Pfeuffer, Frank, Weismantel, Robert, Schraven, Burkhardt, and Lindquist, Jonathan A. “Integrating Signals from the T-Cell Receptor and the Interleukin-2 Receptor.” *PLoS Computational Biology* 7.8 (2011): e1002121.
- Boer, Rob J De, Ganusov, Vitaly V, Milutinović, Dejan, Hodgkin, Philip D, and Perelson, Alan S. “Estimating Lymphocyte Division and Death Rates from CFSE Data.” *Bulletin of Mathematical Biology* 68.5 (2006): 1011–1031.
- Boer, Rob J. De and Perelson, Alan S. “Estimating division and death rates from CFSE data.” *Journal of Computational and Applied Mathematics* 184.1 (2005): 140–164.
- Boer, Rob J De and Perelson, Alan S. “Quantifying T lymphocyte turnover.” *Journal of theoretical biology* 327: 45–87.
- Bogaert, Delfien J A, Dullaers, Melissa, Lambrecht, Bart N, Vermaelen, Karim Y, Baere, Elfride De, and Haerynck, Filomeen. “Genes associated with common variable immunodeficiency: one diagnosis to rule them all?” *Journal of Medical Genetics* 53.9 (2016): 575.
- Bonilla, Francisco A., Barlan, Isil, Chapel, Helen, Costa-Carvalho, Beatriz T., Cunningham-Rundles, Charlotte, Morena, M. Teresa de la, Espinosa-Rosales, Francisco J., Hammarström, Lennart, Nonoyama, Shigeaki, Quinti, Isabella, Routes, John M., Tang, Mimi L.K., and Warnatz, Klaus. “International Consensus Document (ICON): Common Variable Immunodeficiency Disorders.” *The Journal of Allergy and Clinical Immunology: In Practice* 4.1 (2016): 38–59.
- Boonyaratanakornkit, Jim and Taylor, Justin J. “Techniques to Study Antigen-Specific B Cell Responses.” *Frontiers in Immunology* 10: 1694.
- Bretscher, Peter and Cohn, Melvin. “A Theory of Self-Nonself Discrimination.” *Science* 169.3950 (1970): 1042–1049.

- Buchholz, Veit R, Flossdorf, Michael, Hensel, Inge, Kretschmer, Lorenz, Weissbrich, Bianca, Gräf, Patricia, Verschoor, Admar, Schiemann, Matthias, Höfer, Thomas, and Busch, Dirk H. “Disparate individual fates compose robust CD8+ T cell immunity.” *Science* 340.6132 (2013): 630–5.
- Buchholz, Veit R, Schumacher, Ton N M, and Busch, Dirk H. “T Cell Fate at the Single-Cell Level.” *Annual review of immunology* 34.1 (2015): 65–92.
- Burnet, F. M. “A modification of Jerne’s theory of antibody production using the concept of clonal selection.” *Australian Journal of Science* 20.3 (1957): 67–69.
- Bykowsky, M J, Haire, R N, Ohta, Y, Tang, H, Sung, S S, Veksler, E S, Greene, J M, Fu, S M, Litman, G W, and Sullivan, K E. “Discordant phenotype in siblings with X-linked agammaglobulinemia.” *American journal of human genetics* 58.3 (1996): 477–83.
- Caio, Giacomo, Volta, Umberto, Sapone, Anna, Leffler, Daniel A., Giorgio, Roberto De, Catassi, Carlo, and Fasano, Alessio. “Celiac disease: a comprehensive current review.” *BMC Medicine* 17.1 (2019): 142.
- Callard, Robin and Hodgkin, Phil. “Modeling T- and B-cell growth and differentiation.” *Immunological Reviews* 216.1 (2007): 119–129.
- Celso, Cristina Lo, Fleming, Heather E., Wu, Juwell W., Zhao, Cher X., Miake-Lye, Sam, Fujisaki, Joji, Côté, Daniel, Rowe, David W., Lin, Charles P., and Scadden, David T. “Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche.” *Nature* 457.7225 (2009): 92–96.
- Chakravorty, Rajib, Rawlinson, David, Zhang, Alan, Markham, John, Dowling, Mark R, Wellard, Cameron, Zhou, Jie H S, and Hodgkin, Philip D. “Labour-efficient in vitro lymphocyte population tracking and fate prediction using automation and manual review.” *PloS one* 9.1 (2014): e83251.
- Chen, Liying, Jondal, Mikael, and Yakimchuk, Konstantin. “Regulatory effects of dexamethasone on NK and T cell immunity.” *Inflammopharmacology* 26.5 (2018): 1331–1338.
- Cheon, HoChan, Kan, Andrey, Prevedello, Giulio, Oostindie, Simone C., Dovedi, Simon J., Hawkins, Edwin D., Marchingo, Julia M., Heinzl, Susanne, Duffy,

- Ken R., and Hodgkin, Philip D. "Cyton2: A Model of Immune Cell Population Dynamics That Includes Familial Instructional Inheritance." *Frontiers in Bioinformatics* 1: 723337.
- Coffman, Robert L., Seymour, Brian W. P., Lebman, Deborah A., Hiraki, Debra D., Christiansen, Judy A., Shrader, Barbara, Cherwinski, Holly M., Savelkoul, Huub F. J., Finkelman, Fred D., Bond, Martha W., and Mosmann, Tim R. "The Role of Helper T Cell Products in Mouse B Cell Differentiation and Isotype Regulation." *Immunological Reviews* 102.1 (1988): 5–28.
- Conley, Mary Ellen, Dobbs, A. Kerry, Farmer, Dana M., Kilic, Sebnem, Paris, Kenneth, Grigoriadou, Sofia, Coustan-Smith, Elaine, Howard, Vanessa, and Campana, Dario. "Primary B Cell Immunodeficiencies: Comparisons and Contrasts." *Annual Review of Immunology* 27.1 (2009): 199–227.
- Cooper, Max D. and Alder, Matthew N. "The Evolution of Adaptive Immune Systems." *Cell* 124.4 (2006): 815–822.
- Cornwell, J. A. and Nordon, R. E. "Computational Stem Cell Biology, Methods and Protocols." *Methods in Molecular Biology* 1975: 131–156.
- Coutinho, Agnes E. and Chapman, Karen E. "The anti-inflammatory and immunosuppressive effects of glucocorticoids, recent developments and mechanistic insights." *Molecular and Cellular Endocrinology* 335.1 (2011): 2–13.
- Deenick, E K, Hasbold, J, and Hodgkin, P D. "Switching to IgG3, IgG2b, and IgA is division linked and independent, revealing a stochastic framework for describing differentiation." *Journal of immunology (Baltimore, Md. : 1950)* 163.9 (1999): 4707–14.
- Deenick, Elissa K, Gett, Amanda V, and Hodgkin, Philip D. "Stochastic Model of T Cell Proliferation: A Calculus Revealing IL-2 Regulation of Precursor Frequencies, Cell Cycle Time, and Survival." *The Journal of Immunology* 170.10 (2003): 4963–4972.
- Deenick, Elissa K., Hasbold, Jhagvaral, and Hodgkin, Philip D. "Decision criteria for resolving isotype switching conflicts by B cells." *European Journal of Immunology* 35.10 (2005): 2949–2955.

- Dowling, Mark R, Kan, Andrey, Heinzl, Susanne, Zhou, Jie H S, Marchingo, Julia M, Wellard, Cameron J, Markham, John F, and Hodgkin, Philip D. “Stretched cell cycle model for proliferating lymphocytes.” *Proceedings of the National Academy of Sciences of the United States of America* 111.17 (2014): 6377–82.
- Dowling, Mark R, Milutinović, Dejan, and Hodgkin, Philip D. “Modelling cell lifespan and proliferation: is likelihood to die or to divide independent of age?” *Journal of The Royal Society Interface* 2.5 (2005): 517–526.
- Downey, Mike J., Jeziorska, Danuta M., Ott, Sascha, Tamai, T. Katherine, Koentges, Georgy, Vance, Keith W., and Bretschneider, Till. “Extracting Fluorescent Reporter Time Courses of Cell Lineages from High-Throughput Microscopy at Low Temporal Resolution.” *PLoS ONE* 6.12 (2011): e27886.
- Duffy, K R, Wellard, C J, Markham, J F, Zhou, J H S, Holmberg, R, Hawkins, E D, Hasbold, J, Dowling, M R, and Hodgkin, P D. “Activation-Induced B Cell Fates Are Selected by Intracellular Stochastic Competition.” *Science* 335.6066 (2012): 338–341.
- Duffy, Ken R and Hodgkin, Philip D. “Intracellular competition for fates in the immune system.” *Trends in Cell Biology* 22.9 (2012).
- Duffy, Ken R and Subramanian, Vijay G. “On the impact of correlation between collaterally consanguineous cells on lymphocyte population dynamics.” *Journal of Mathematical Biology* 59.2 (2009): 255–285.
- Durandy, Anne, Kracker, Sven, and Fischer, Alain. “Primary antibody deficiencies.” *Nature Reviews Immunology* 13.7 (2013): 519–533.
- Efron, B. “Bootstrap Methods: Another Look at the Jackknife.” *The Annals of Statistics* 7.1 (1979): 1–26.
- Errington, Rachel J., Marquez, Nuria, Chappell, Sally C., Wiltshire, Marie, and Smith, Paul J. “Time-Lapse Microscopy Approaches to Track Cell Cycle Progression at the Single-Cell Level.” *Current Protocols in Cytometry* 31.1 (2005): 12.4.1–12.4.11.
- Fairfax, Kirsten A., Corcoran, Lynn M., Pridans, Clare, Huntington, Nicholas D., Kallies, Axel, Nutt, Stephen L., and Tarlinton, David M. “Different Kinetics of

- Blimp-1 Induction in B Cell Subsets Revealed by Reporter Gene.” *The Journal of Immunology* 178.7 (2007): 4104–4111.
- Fingar, Diane C, Richardson, Celeste J, Tee, Andrew R, Cheatham, Lynn, Tsou, Christina, and Blenis, John. “mTOR controls cell cycle progression through its cell growth effectors S6K1 and 4E-BP1/eukaryotic translation initiation factor 4E.” *Molecular and cellular biology* 24.1 (2004): 200–16.
- Finlay, David K., Rosenzweig, Ella, Sinclair, Linda V., Feijoo-Carnero, Carmen, Hukelmann, Jens L., Rolf, Julia, Panteleyev, Andrey A., Okkenhaug, Klaus, and Cantrell, Doreen A. “PDK1 regulation of mTOR and hypoxia-inducible factor 1 integrate metabolism and migration of CD8+ T cells.” *The Journal of Experimental Medicine* 209.13 (2012): 2441–2453.
- Galili, Naomi, Galili, Uri, Klein, Eva, Rosenthal, Lili, and Nordenskjöld, Bo. “Human T lymphocytes become glucocorticoid-sensitive upon immune activation.” *Cellular Immunology* 50.2 (1980): 440–444.
- Galli, Stephen J., Tsai, Mindy, and Piliponsky, Adrian M. “The development of allergic inflammation.” *Nature* 454.7203 (2008): 445–454.
- Ganusov, Vitaly V, Milutinović, Dejan, and Boer, Rob J De. “IL-2 Regulates Expansion of CD4 + T Cell Populations by Affecting Cell Death: Insights from Modeling CFSE Data.” *The Journal of Immunology* 179.2 (2007): 950–957.
- Ganusov, Vitaly V., Pilyugin, Sergei S., De Boer, Rob J., Murali-Krishna, Kaja, Ahmed, Rafi, and Antia, Rustom. “Quantifying cell turnover using CFSE data.” *Journal of Immunological Methods* 298.1-2 (2005): 183–200.
- Gardner, M J and Altman, D G. “Confidence intervals rather than P values: estimation rather than hypothesis testing.” *British Medical Journal (Clinical research ed.)* 292.6522 (1986): 746.
- Gerlach, Carmen, Heijst, Jeroen W.J. van, Swart, Erwin, Sie, Daoud, Armstrong, Nicola, Kerkhoven, Ron M., Zehn, Dietmar, Bevan, Michael J., Schepers, Koen, and Schumacher, Ton N.M. “One naive T cell, multiple fates in CD8+ T cell differentiation.” *Journal of Experimental Medicine* 207.6 (2010): 1235–1246.

- Gerlach, Carmen, Rohr, Jan C., Perié, Leila, van Rooij, Nienke, van Heijst, Jeroen W. J., Velds, Arno, Urbanus, Jos, Naik, Shalin H., Jacobs, Heinz, Beltman, Joost B., de Boer, Rob J., and Schumacher, Ton N. M. “Heterogeneous Differentiation Patterns of Individual CD8+ T Cells.” *Science* 340.6132 (2013): 635–639.
- Gett, Amanda V and Hodgkin, Philip D. “Cell division regulates the T cell cytokine repertoire, revealing a mechanism underlying immune class regulation.” *Proceedings of the National Academy of Sciences* 95.16 (1998): 9488–9493.
- . “A cellular calculus for signal integration by T cells.” *Nature Immunology* 1.3 (2000): 239–244.
- Giles, Amber J., Hutchinson, Marsha-Kay N. D., Sonnemann, Heather M., Jung, Jinkyu, Fecci, Peter E., Ratnam, Nivedita M., Zhang, Wei, Song, Hua, Bailey, Rolanda, Davis, Dionne, Reid, Caitlin M., Park, Deric M., and Gilbert, Mark R. “Dexamethasone-induced immunosuppression: mechanisms and implications for immunotherapy.” *Journal for Immunotherapy of Cancer* 6.1 (2018): 51.
- Harris, Theodore E. *The Theory of Branching Processes*. Berlin: Springer-Verlag, 1963.
- Hasbold, J, Gett, AV, Rush, JS, Deenick, E, Avery, D, Jun, J, and Hodgkin, PD. “Quantitative analysis of lymphocyte differentiation and proliferation in vitro using carboxyfluorescein diacetate succinimidyl ester.” *Immunology and Cell Biology* 77.6 (1999a): 516–522.
- Hasbold, J, Hong, J S, Kehry, M R, and Hodgkin, P D. “Integrating signals from IFN-gamma and IL-4 by B cells: positive and negative effects on CD40 ligand-induced proliferation, survival, and division-linked isotype switching to IgG1, IgE, and IgG2a.” *Journal of immunology (Baltimore, Md. : 1950)* 163.8 (1999b): 4175–81.
- Hasbold, Jhagvaral, Corcoran, Lynn M, Tarlinton, David M, Tangye, Stuart G, and Hodgkin, Philip D. “Evidence from the generation of immunoglobulin G-secreting cells that stochastic mechanisms regulate lymphocyte differentiation.” *Nature Immunology* 5.1 (2004): 55–63.

- Hasenauer, J., Schittler, D., and Allgöwer, F. “Analysis and Simulation of Division- and Label-Structured Population Models.” *Bulletin of Mathematical Biology* 74.11 (2012): 2692–2732.
- Hawkins, E D, Markham, J F, McGuinness, L P, and Hodgkin, P D. “A single-cell pedigree analysis of alternative stochastic lymphocyte fates.” *Proceedings of the National Academy of Sciences* 106.32 (2009): 13457–13462.
- Hawkins, E D, Turner, M L, Dowling, M R, van Gend, C, and Hodgkin, P D. “A model of immune regulation as a consequence of randomized lymphocyte division and death times.” *Proceedings of the National Academy of Sciences* 104.12 (2007a): 5032–5037.
- Hawkins, E D, Turner, M L, Wellard, C J, Zhou, J H S, Dowling, M R, and Hodgkin, P D. “Quantal and graded stimulation of B lymphocytes as alternative strategies for regulating adaptive immune responses.” *Nature Communications* 4.1 (2013): 2406.
- Hawkins, Edwin D., Duarte, Delfim, Akinduro, Olufolake, Khorshed, Reema A., Passaro, Diana, Nowicka, Malgorzata, Straszkowski, Lenny, Scott, Mark K., Rothery, Steve, Ruivo, Nicola, Foster, Katie, Waibel, Michaela, Johnstone, Ricky W., Harrison, Simon J., Westerman, David A., Quach, Hang, Gribben, John, Robinson, Mark D., Purton, Louise E., Bonnet, Dominique, and Celso, Cristina Lo. “T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments.” *Nature* 538.7626 (2016): 518–522.
- Hawkins, Edwin D, Hommel, Mirja, Turner, Marian L, Battye, Francis L, Markham, John F, and Hodgkin, Philip D. “Measuring lymphocyte proliferation, survival and differentiation using CFSE time-series data.” *Nature Protocols* 2.9 (2007b): 2057–2067.
- Heinzel, Susanne, Giang, Tran Binh, Kan, Andrey, Marchingo, Julia M, Lye, Bryan K, Corcoran, Lynn M, and Hodgkin, Philip D. “A Myc-dependent division timer complements a cell-death timer to regulate T cell and B cell responses.” *Nature Immunology* 18.1 (2017): 96–103.
- Ho, Joses, Tumkaya, Tayfun, Aryal, Sameer, Choi, Hyungwon, and Claridge-Chang, Adam. “Moving beyond P values: data analysis with estimation graphics.” *Nature Methods* 16.7 (2019): 565–566.

- Hodgkin, P. D. “Quantitative rules for lymphocyte regulation: the cellular calculus and decisions between tolerance and activation.” *Tissue Antigens* 66.4 (2005): 259–266.
- Hodgkin, P D, Lee, J H, and Lyons, A B. “B cell differentiation and isotype switching is related to division cycle number.” *The Journal of Experimental Medicine* 184.1 (1996): 277–281.
- Hodgkin, Philip D. “Modifying clonal selection theory with a probabilistic cell.” *Immunological Reviews* 285.1 (2018): 249–262.
- Hodgkin, Philip D., Castle, Brian E., and Kehry, Marilyn R. “B cell differentiation induced by helper T cell membranes: evidence for sequential isotype switching and a requirement for lymphokines during proliferation.” *European Journal of Immunology* 24.1 (1994): 239–246.
- Hodgkin, Philip D., Hawkins, Edwin D., Hasbold, Jhaguaral, Gett, Amanda V., K., Deenick Elissa, Todd, Hilary F., and Hommel, Mirja. “Monitoring T Cell Proliferation.” *Analyzing T Cell Responses: How to Analyze Cellular Immune Responses against Tumor Associated Antigens*. eds. Dirk Nagorsen and F.M. Marincola, chap. 6. Dordrecht: Springer Netherlands, 2005. 123–141.
- Hodgkin, Philip D, Heath, William R, and Baxter, Alan G. “The clonal selection theory: 50 years since the revolution.” *Nature Immunology* 8.10 (2007): 1019–1026.
- Hoffman, Matthew D. and Gelman, Andrew. “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo.” *The Journal of Machine Learning Research* 15.1 (2014): 1593–1623.
- Hogquist, Kristin A., Jameson, Stephen C., Heath, William R., Howard, Jane L., Bevan, Michael J., and Carbone, Francis R. “T cell receptor antagonist peptides induce positive selection.” *Cell* 76.1 (1994): 17–27.
- Horton, Miles B., Cheon, HoChan, Duffy, Ken R., Brown, Daniel, Naik, Shalin H., Alvarado, Carolina, Groom, Joanna R., Heinzl, Susanne, and Hodgkin, Philip D. “Lineage tracing reveals B cell antibody class switching is stochastic, cell-autonomous, and tuneable.” *Immunity* .

- Horton, Miles B, Prevedello, Giulio, Marchingo, Julia M, Zhou, Jie H S, Duffy, Ken R, Heinzl, Susanne, and Hodgkin, Philip D. “Multiplexed Division Tracking Dyes for Proliferation-Based Clonal Lineage Tracing.” *Journal of Immunology* 201.3 (2018): 1097–1103.
- Hyrien, Ollivier, Chen, Rui, and Zand, Martin S. “An age-dependent branching process model for the analysis of CFSE-labeling experiments.” *Biology Direct* 5.1 (2010): 41.
- Hyrien, Ollivier and Zand, Martin S. “A Mixture Model With Dependent Observations for the Analysis of CSFE–Labeling Experiments.” *Journal of the American Statistical Association* 103.481 (2008): 222–239.
- Janeway, C.A. “Approaching the Asymptote? Evolution and Revolution in Immunology.” *Cold Spring Harbor Symposia on Quantitative Biology* 54.0 (1989): 1–13.
- Jeffreys, Harold. *Theory of Probability*. Oxford: Oxford University Press, 1961, 3 ed.
- Kaech, Susan M. and Wherry, E. John. “Heterogeneity and Cell-Fate Decisions in Effector and Memory CD8+ T Cell Differentiation during Viral Infection.” *Immunity* 27.3 (2007): 393–405.
- Kaech, Susan M., Wherry, E. John, and Ahmed, Rafi. “Effector and memory T-cell differentiation: implications for vaccine development.” *Nature Reviews Immunology* 2.4 (2002): 251–262.
- Kallies, Axel, Hasbold, Jhagvaral, Tarlinton, David M., Dietrich, Wendy, Corcoran, Lynn M., Hodgkin, Philip D., and Nutt, Stephen L. “Plasma Cell Ontogeny Defined by Quantitative Changes in Blimp-1 Expression.” *The Journal of Experimental Medicine* 200.8 (2004): 967–977.
- Kinjyo, Ichiko, Qin, Jim, Tan, Sioh-Yang, Wellard, Cameron J, Mrass, Paulus, Ritchie, William, Doi, Atsushi, Cavanagh, Lois L, Tomura, Michio, Sakaue-Sawano, Asako, Kanagawa, Osami, Miyawaki, Atsushi, Hodgkin, Philip D, and Weninger, Wolfgang. “Real-time tracking of cell cycle progression during CD8+ effector and memory T-cell differentiation.” *Nature Communications* 6.1 (2015): 6301.

- Kornfeld, Stephen J., Haire, Robert N., Strong, Scott J., Tang, Huayang, Sung, Sun-Sang J., Fu, Shu Man, and Litman, Gary W. “A Novel Mutation (Cys145→Stop) in Bruton’s Tyrosine Kinase Is Associated with Newly Diagnosed X-Linked Agammaglobulinemia in a 51-Year-Old Male.” *Molecular Medicine* 2.5 (1996): 619–623.
- Krishnamurty, Akshay T., Thouvenel, Christopher D., Portugal, Silvia, Keitany, Gladys J., Kim, Karen S., Holder, Anthony, Crompton, Peter D., Rawlings, David J., and Pepper, Marion. “Somatically Hypermutated Plasmodium-Specific IgM+ Memory B Cells Are Rapid, Plastic, Early Responders upon Malaria Rechallenge.” *Immunity* 45.2 (2016): 402–414.
- Kroese, Frans GM, Abdulahad, Wayel H, Haacke, Erlin, Bos, Nicolaas A, Vissink, Arjan, and Bootsma, Hendrika. “B-cell hyperactivity in primary Sjögren’s syndrome.” *Expert Review of Clinical Immunology* 10.4 (2014): 483–499.
- Lafferty, Kevin J., Andrus, Linda, and Prowse, Stephen J. “Role of Lymphokine and Antigen in the Control of Specific T Cell Responses.” *Immunological Reviews* 51.1 (1980): 279–314.
- Lee, Ha Youn, Hawkins, Edwin, Zand, Martin S., Mosmann, Tim, Wu, Hulin, Hodgkin, Philip D., and Perelson, Alan S. “Interpreting CFSE Obtained Division Histories of B Cells in Vitro with Smith–Martin and Cyton Type Models.” *Bulletin of Mathematical Biology* 71.7 (2009): 1649–1670.
- Leon, Kalet, Faro, Jose, and Carneiro, Jorge. “A general mathematical framework to model generation structure in a population of asynchronously dividing cells.” *Journal of Theoretical Biology* 229.4 (2004): 455–476.
- Luzyanina, T and Bocharov, G. “Markov Chain Monte Carlo Parameter Estimation of the ODE Compartmental Cell Growth Model.” *Mathematical Biology and Bioinformatics* 13.2 (2018): 376–391.
- Luzyanina, Tatyana, Cupovic, Jovana, Ludewig, Burkhard, and Bocharov, Gennady. “Mathematical models for CFSE labelled lymphocyte dynamics: asymmetry and time-lag in division.” *Journal of Mathematical Biology* 69.6-7 (2014): 1547–1583.
- Luzyanina, Tatyana, Roose, Dirk, Schenkel, Tim, Sester, Martina, Ehl, Stephan, Meyerhans, Andreas, and Bocharov, Gennady. “Numerical modelling of label-

- structured cell population growth using CFSE distribution data.” *Theoretical Biology and Medical Modelling* 4.1 (2007): 26.
- Lye, Bryan. *Developing a Quantitative Framework for Immunotherapy*. Master’s thesis, University of Melbourne, 2016.
- Lyons, A.Bruce and Parish, Christopher R. “Determination of lymphocyte division by flow cytometry.” *Journal of Immunological Methods* 171.1 (1994): 131–137.
- Marchingo, J M, Prevedello, G, Kan, A, Heinzl, S, Hodgkin, P D, and Duffy, K R. “T-cell stimuli independently sum to regulate an inherited clonal division fate.” *Nature Communications* 7.1 (2016): 13540.
- Marchingo, Julia M, Kan, Andrey, Sutherland, Robyn M, Duffy, Ken R, Wellard, Cameron J, Belz, Gabrielle T, Lew, Andrew M, Dowling, Mark R, Heinzl, Susanne, and Hodgkin, Philip D. “Antigen affinity, costimulation, and cytokine inputs sum linearly to amplify T cell expansion.” *Science* 346.6213 (2014): 1123–1127.
- Markham, John F, Wellard, Cameron J, Hawkins, Edwin D, Duffy, Ken R, and Hodgkin, Philip D. “A minimum of two distinct heritable factors are required to explain correlation structures in proliferating lymphocytes.” *Journal of the Royal Society, Interface* 7.48 (2010): 1049–59.
- Marquardt, Donald W. “An Algorithm for Least-Squares Estimation of Nonlinear Parameters.” *Journal of the Society for Industrial and Applied Mathematics* 11.2 (1963): 431–441.
- Martínez, María Rodríguez, Corradin, Alberto, Klein, Ulf, Álvarez, Mariano Javier, Toffolo, Gianna M., Camillo, Barbara di, Califano, Andrea, and Stolovitzky, Gustavo A. “Quantitative modeling of the terminal differentiation of B cells and mechanisms of lymphomagenesis.” *Proceedings of the National Academy of Sciences* 109.7 (2012): 2672–2677.
- Matzinger, P. “Tolerance, Danger, and the Extended Family.” *Annual Review of Immunology* 12.1 (1994): 991–1045.
- Mazzocco, Pauline, Bernard, Samuel, and Pujol-Menjouet, Laurent. “Estimates and impact of lymphocyte division parameters from CFSE data using mathematical modelling.” *PLOS ONE* 12.6 (2017): e0179768.

- McElreath, Richard. *Statistical Rethinking*. CRC Press/Taylor & Francis Group, 2020, 2 ed.
- McKenna, Aaron, Findlay, Gregory M., Gagnon, James A., Horwitz, Marshall S., Schier, Alexander F., and Shendure, Jay. “Whole-organism lineage tracing by combinatorial and cumulative genome editing.” *Science* 353.6298 (2016): aaf7907.
- Miao, Hongyu, Jin, Xia, Perelson, Alan S, and Wu, Hulin. “Evaluation of multitype mathematical models for CFSE-labeling experiment data.” *Bulletin of Mathematical Biology* 74.2 (2011): 300–26.
- Miller, J. F. A. P. “Role of the Thymus in Immunity*.” *British Medical Journal* 2.5355 (1963): 459.
- Milutinović, D. and Boer, R.J. De. “Process Noise: An Explanation for the Fluctuations in the Immune Response during Acute Viral Infection.” *Biophysical Journal* 92.10 (2007): 3358–3367.
- Mitchell, Simon, Roy, Koushik, Zangle, Thomas A, and Hoffmann, Alexander. “Non-genetic origins of cell-to-cell variability in B lymphocyte proliferation.” *Proceedings of the National Academy of Sciences of the United States of America* 115.12 (2018): E2888–E2897.
- Mueller, D L, Jenkins, M K, and Schwartz, R H. “Clonal Expansion Versus Functional Clonal Inactivation: A Costimulatory Signalling Pathway Determines the Outcome of T Cell Antigen Receptor Occupancy.” *Annual Review of Immunology* 7.1 (1989): 445–480.
- Muramatsu, Masamichi, Kinoshita, Kazuo, Fagarasan, Sidonia, Yamada, Shuichi, Shinkai, Yoichi, and Honjo, Tasuku. “Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme.” *Cell* 102.5 (2000): 553–563.
- Murphy, Kenneth, Travers, Paul, and Walport, Mark. *Janeway’s Immunobiology*. New York: Garland Science, 2008, 7 ed.
- Naik, Shalin H., Perié, Leïla, Swart, Erwin, Gerlach, Carmen, Rooij, Nienke van, Boer, Rob J. de, and Schumacher, Ton N. “Diverse and heritable lineage imprinting of early haematopoietic progenitors.” *Nature* 496.7444 (2013): 229–232.

- Naik, Shalin H., Schumacher, Ton N., and Perié, Leila. “Cellular barcoding: A technical appraisal.” *Experimental Hematology* 42.8 (2014): 598–608.
- Newville, Matthew, Stensitzki, Till, Allen, Daniel B., and Ingargiola, Antonino. “LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python.” 2014.
- Nordon, Robert E., Ko, Kap-Hyoun, Odell, Ross, and Schroeder, Timm. “Multi-type branching models to describe cell differentiation programs.” *Journal of Theoretical Biology* 277.1 (2011): 7–18.
- Nordon, Robert E, Nakamura, Masanori, Ramirez, Carole, and Odell, Ross. “Analysis of growth kinetics by division tracking.” *Immunology and Cell Biology* 77.6 (1999): 523–529.
- Oracki, Sarah A, Walker, Jennifer A, Hibbs, Margaret L, Corcoran, Lynn M, and Tarlinton, David M. “Plasma cell development and survival: Plasma cell development and survival.” *Immunological Reviews* 237.1 (2010): 140–159.
- Pandit, Aridaman and Boer, Rob J. De. “Stochastic Inheritance of Division and Death Times Determines the Size and Phenotype of CD8+ T Cell Families.” *Frontiers in Immunology* 10: 436.
- Pham, K., Kan, A., Whitehead, L., Hennessy, R. J., Rogers, K., and Hodgkin, P. D. “Converse Smith-Martin cell cycle kinetics by transformed B lymphocytes.” *Cell Cycle* 17.16 (2018): 1–11.
- Pilyugin, Sergei S., Ganusov, Vitaly V., Murali-Krishna, Kaja, Ahmed, Rafi, and Antia, Rustom. “The rescaling method for quantifying the turnover of cell populations.” *Journal of Theoretical Biology* 225.2 (2003): 275–283.
- Punt, Jenni, Stranford, Sharon A., Jones, Patricia P., and Owen, Judith A. *Kuby Immunology*. New York: Macmillan Education, 2019, 8 ed.
- Purton, Jared F., Monk, Julie A., Liddicoat, Douglas R., Kyparissoudis, Konstantinos, Sakkal, Samy, Richardson, Samantha J., Godfrey, Dale I., and Cole, Timothy J. “Expression of the Glucocorticoid Receptor from the 1A Promoter Correlates with T Lymphocyte Sensitivity to Glucocorticoid-Induced Cell Death.” *The Journal of Immunology* 173.6 (2004): 3816–3824.

- Quah, Benjamin J.C. and Parish, Christopher R. “New and improved methods for measuring lymphocyte proliferation in vitro and in vivo using CFSE-like fluorescent dyes.” *Journal of Immunological Methods* 379.1-2 (2012): 1–14.
- Revy, Patrick, Sospedra, Mireia, Barbour, Boris, and Trautmann, Alain. “Functional antigen-independent synapses formed between T cells and dendritic cells.” *Nature Immunology* 2.10 (2001): 925–931.
- Romão, Vasco C, Talarico, Rosaria, Scirè, Carlo Alberto, Vieira, Ana, Alexander, Tobias, Baldini, Chiara, Gottenberg, Jacques-Eric, Gruner, Heidi, Hachulla, Eric, Mouthon, Luc, Orlandi, Martina, Pamfil, Cristina, Chambrun, Marc Pineton de, Taglietti, Marco, Toplak, Natasa, Daele, Paul van, Laar, Jacob M van, Bombardieri, Stefano, Schneider, Matthias, Smith, Vanessa, Cutolo, Maurizio, Mosca, Marta, and Mariette, Xavier. “Sjögren’s syndrome: state of the art on clinical practice guidelines.” *RMD Open* 4.Suppl 1 (2018): e000789.
- Rosenblum, Michael D., Remedios, Kelly A., and Abbas, Abul K. “Mechanisms of human autoimmunity.” *Journal of Clinical Investigation* 125.6 (2015): 2228–2233.
- Rush, James S and Hodgkin, Philip D. “B cells activated via CD40 and IL-4 undergo a division burst but require continued stimulation to maintain division, survival and differentiation.” *European Journal of Immunology* 31.4 (2001): 1150–1159.
- Sakaue-Sawano, Asako, Kurokawa, Hiroshi, Morimura, Toshifumi, Hanyu, Aki, Hama, Hiroshi, Osawa, Hatsuki, Kashiwagi, Saori, Fukami, Kiyoko, Miyata, Takaki, Miyoshi, Hiroyuki, Imamura, Takeshi, Ogawa, Masaharu, Masai, Hisao, and Miyawaki, Atsushi. “Visualizing spatiotemporal dynamics of multicellular cell-cycle progression.” *Cell* 132.3 (2008): 487–98.
- Salvatier, John, Wiecki, Thomas V., and Fonnesbeck, Christopher. “Probabilistic programming in Python using PyMC3.” *PeerJ Computer Science* 2: e55.
- Salzer, Ulrich, Bacchelli, Chiara, Buckridge, Sylvie, Pan-Hammarström, Qiang, Jennings, Stephanie, Lougaris, Vassilis, Bergbreiter, Astrid, Hagen, Tina, Birmelin, Jennifer, Plebani, Alessandro, Webster, A. David B., Peter, Hans-Hartmut, Suez, Daniel, Chapel, Helen, McLean-Tooke, Andrew, Spickett, Gavin P., Anover-Sombke, Stephanie, Ochs, Hans D., Urschel, Simon, Belohradsky, Bernd H., Ugrinovic, Sanja, Kumararatne, Dinakantha S., Lawrence, Tatiana C., Holm,

- Are M., Franco, Jose L., Schulze, Ilka, Schneider, Pascal, Gertz, E. Michael, Schäffer, Alejandro A., Hammarström, Lennart, Thrasher, Adrian J., Gaspar, H. Bobby, and Grimbacher, Bodo. “Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes.” *Blood* 113.9 (2009): 1967–1976.
- Sanderson, R D, Lalor, P, and Bernfield, M. “B lymphocytes express and lose syndecan at specific stages of differentiation.” *Cell Regulation* 1.1 (1989): 27–35.
- Schindelin, Johannes, Arganda-Carreras, Ignacio, Frise, Erwin, Kaynig, Verena, Longair, Mark, Pietzsch, Tobias, Preibisch, Stephan, Rueden, Curtis, Saalfeld, Stephan, Schmid, Benjamin, Tinevez, Jean-Yves, White, Daniel James, Hartenstein, Volker, Eliceiri, Kevin, Tomancak, Pavel, and Cardona, Albert. “Fiji: an open-source platform for biological-image analysis.” *Nature Methods* 9.7 (2012): 676–682.
- Seder, Robert A and Ahmed, Rafi. “Similarities and differences in CD4+ and CD8+ effector and memory T cell generation.” *Nature Immunology* 4.9 (2003): 835–842.
- Shapiro-Shelef, Miriam, Lin, Kuo-I, McHeyzer-Williams, Louise J, Liao, Jerry, McHeyzer-Williams, Michael G, and Calame, Kathryn. “Blimp-1 Is Required for the Formation of Immunoglobulin Secreting Plasma Cells and Pre-Plasma Memory B Cells.” *Immunity* 19.4 (2003): 607–620.
- Shapiro-Shelef, Miriam, Lin, Kuo-I, Savitsky, David, Liao, Jerry, and Calame, Kathryn. “Blimp-1 is required for maintenance of long-lived plasma cells in the bone marrow.” *The Journal of Experimental Medicine* 202.11 (2005): 1471–1476.
- Shimomura, Osamu, Johnson, Frank H., and Saiga, Yo. “Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea.” *Journal of Cellular Physiology* 59.3 (1962): 223–239.
- Shokhirev, Maxim N, Almaden, Jonathan, Davis-Turak, Jeremy, Birnbaum, Harry A, Russell, Theresa M, Vargas, Jesse A D, and Hoffmann, Alexander. “A multi-scale approach reveals that NF- κ B cRel enforces a B-cell decision to divide.” *Molecular Systems Biology* 11.1 (2015): 783.

- Shokhirev, Maxim Nikolaievich and Hoffmann, Alexander. “FlowMax: A Computational Tool for Maximum Likelihood Deconvolution of CFSE Time Courses.” *PLOS ONE* 8.6 (2013): e67620.
- Smith, J A and Martin, L. “Do Cells Cycle?” *Proceedings of the National Academy of Sciences* 70.4 (1973): 1263–1267.
- Ståhlberg, Anders, Rusnakova, Vendula, Forootan, Amin, Anderova, Miroslava, and Kubista, Mikael. “RT-qPCR work-flow for single-cell data analysis.” *Methods* 59.1 (2013): 80–88.
- Stavnezer, Janet, Guikema, Jeroen E.J., and Schrader, Carol E. “Mechanism and Regulation of Class Switch Recombination.” *Annual Review of Immunology* 26.1 (2008): 261–292.
- Stavnezer-Nordgren, J and Sirlin, S. “Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching.” *The EMBO Journal* 5.1 (1986): 95–102.
- Storn, Rainer and Price, Kenneth. “Differential Evolution - A Simple and Efficient Heuristic for global Optimization over Continuous Spaces.” *Journal of Global Optimization* 11.4 (1997): 341–359.
- Subramanian, Vijay G, Duffy, Ken R, Turner, Marian L, and Hodgkin, Philip D. “Determining the expected variability of immune responses using the cyton model.” *Journal of Mathematical Biology* 56.6 (2008): 861–892.
- Sánchez-Ramón, Silvia, Bermúdez, Arancha, González-Granado, Luis Ignacio, Rodríguez-Gallego, Carlos, Sastre, Ana, Soler-Palacín, Pere, , the ID-Signal Onco-Haematology Group, Allende, Luis, Alsina, Laia, Bielsa, Ana María, Calleja-Antolín, Sara, Cámara, Carmen, Carbone, Javier, Carreras, Carmen, Martín, Ana De Andrés, Deyá, Angela, Díaz de Heredia, Cristina, Dieli-Crimi, Romina, Luis Díez, José, Domínguez-Pinilla, Nerea, Fernández-Pereira, Luis, García, José M, Gil-Herrera, Juana, Gutiérrez, Antonio, Jarque, Isidro, Juan, Manel, Lendínez, Francisco, Llobet, Pilar, López, Mónica, Guía, Ana López de la, López-Hoyos, Marcos, Martín-Nalda, Andrea, Martínez, Mónica, Melero, Josefa, Méndez-Echevarría, Ana, Moral, Pedro, Neth, Olaf, Núñez, María, Ocejo-Vinyals, Gonzalo, Ochoa-Grullón, Juliana, Olbrich, Peter, Oña, Raquel, Pérez-Encinas, Manuel, Pons, Jaime, Rodríguez, Carmen, Sánchez, Berta, Santos-Pérez,

- Juan Luis, Elena Seoane, M, and Vlagea, Alexandru. “Primary and Secondary Immunodeficiency Diseases in Oncohaematology: Warning Signs, Diagnosis, and Management.” *Frontiers in Immunology* 10: 586.
- Tak, Tamar, Prevedello, Giulio, Simon, Gaël, Paillon, Noémie, Benlabiod, Camélia, Marty, Caroline, Plo, Isabelle, Duffy, Ken R, and Perié, Leïla. “HSPCs display within-family homogeneity in differentiation and proliferation despite population heterogeneity.” *eLife* 10: e60624.
- Tangye, Stuart G, Avery, Danielle T, Deenick, Elissa K, and Hodgkin, Philip D. “Intrinsic Differences in the Proliferation of Naive and Memory Human B Cells as a Mechanism for Enhanced Secondary Immune Responses.” *The Journal of Immunology* 170.2 (2003a): 686–694.
- Tangye, Stuart G., Avery, Danielle T., and Hodgkin, Philip D. “A Division-Linked Mechanism for the Rapid Generation of Ig-Secreting Cells from Human Memory B Cells.” *The Journal of Immunology* 170.1 (2003b): 261–269.
- . “A Division-Linked Mechanism for the Rapid Generation of Ig-Secreting Cells from Human Memory B Cells.” *The Journal of Immunology* 170.1 (2003c): 261–269.
- Taylor, Justin J, Pape, Kathryn A, Steach, Holly R, and Jenkins, Marc K. “Humoral immunity. Apoptosis and antigen affinity limit effector cell differentiation of a single naïve B cell.” *Science (New York, N.Y.)* 347.6223 (2015): 784–7.
- Tempany, Jessica C, Zhou, Jie HS, Hodgkin, Philip D, and Bryant, Vanessa L. “Superior properties of CellTrace Yellow™ as a division tracking dye for human and murine lymphocytes.” *Immunology and Cell Biology* 96.2 (2018): 149–159.
- Thomas, Marcel Jan, Klein, Ulf, Lygeros, John, and Martínez, María Rodríguez. “A Probabilistic Model of the Germinal Center Reaction.” *Frontiers in Immunology* 10: 689.
- Tommasini, Alberto, Magnolato, Andrea, and Bruno, Irene. “Innovation for rare diseases and bioethical concerns: A thin thread between medical progress and suffering.” *World Journal of Clinical Pediatrics* 7.3 (2018): 75–82.

- Tonegawa, Susumu. “Somatic generation of antibody diversity.” *Nature* 302.5909 (1983): 575–581.
- Turner, C. Alexander, Mack, David H., and Davis, Mark M. “Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells.” *Cell* 77.2 (1994): 297–306.
- Turner, Marian L., Hawkins, Edwin D., and Hodgkin, Philip D. “Quantitative Regulation of B Cell Division Destiny by Signal Strength.” *The Journal of Immunology* 181.1 (2008): 374–382.
- Vehtari, Aki, Gelman, Andrew, and Gabry, Jonah. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.” *Statistics and Computing* 27.5 (2017): 1413–1432.
- Veiga-Fernandes, Henrique, Walter, Ulrich, Bourgeois, Christine, McLean, Angela, and Rocha, Benedita. “Response of naïve and memory CD8+ T cells to antigen stimulation in vivo.” *Nature Immunology* 1.1 (2000): 47–53.
- Verdon, Daniel J., Mulazzani, Matthias, and Jenkins, Misty R. “Cellular and Molecular Mechanisms of CD8+ T Cell Differentiation, Dysfunction and Exhaustion.” *International Journal of Molecular Sciences* 21.19 (2020): 7357.
- Vézina, C, Kudelski, A, and Sehgal, S N. “Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle.” *The Journal of antibiotics* 28.10 (1975): 721–6.
- Virtanen, Pauli, Gommers, Ralf, Oliphant, Travis E, Haberland, Matt, Reddy, Tyler, Cournapeau, David, Burovski, Evgeni, Peterson, Pearu, Weckesser, Warren, Bright, Jonathan, Walt, Stéfan J van der, Brett, Matthew, Wilson, Joshua, Millman, K Jarrod, Mayorov, Nikolay, Nelson, Andrew R J, Jones, Eric, Kern, Robert, Larson, Eric, Carey, C J, Polat, İlhan, Feng, Yu, Moore, Eric W, VanderPlas, Jake, Laxalde, Denis, Perktold, Josef, Cimrman, Robert, Henriksen, Ian, Quintero, E A, Harris, Charles R, Archibald, Anne M, Ribeiro, Antônio H, Pedregosa, Fabian, Mulbregt, Paul van, Contributors, SciPy 1 0, Vijaykumar, Aditya, Bardelli, Alessandro Pietro, Rothberg, Alex, Hilboll, Andreas, Kloeckner, Andreas, Scopatz, Anthony, Lee, Antony, Rokem, Ariel, Woods, C Nathan, Fulton, Chad, Masson, Charles, Häggström, Christian, Fitzgerald, Clark, Nicholson,

- David A, Hagen, David R, Pasechnik, Dmitrii V, Olivetti, Emanuele, Martin, Eric, Wieser, Eric, Silva, Fabrice, Lenders, Felix, Wilhelm, Florian, Young, G, Price, Gavin A, Ingold, Gert-Ludwig, Allen, Gregory E, Lee, Gregory R, Audren, Hervé, Probst, Irvin, Dietrich, Jörg P, Silterra, Jacob, Webber, James T, Slavič, Janko, Nothman, Joel, Buchner, Johannes, Kulick, Johannes, Schönberger, Johannes L, Cardoso, José Vinícius de Miranda, Reimer, Joscha, Harrington, Joseph, Rodríguez, Juan Luis Cano, Nunez-Iglesias, Juan, Kuczynski, Justin, Tritz, Kevin, Thoma, Martin, Newville, Matthew, Kümmerer, Matthias, Bolingbroke, Maximilian, Tartre, Michael, Pak, Mikhail, Smith, Nathaniel J, Nowaczyk, Nikolai, Shebanov, Nikolay, Pavlyk, Oleksandr, Brodtkorb, Per A, Lee, Perry, McGibbon, Robert T, Feldbauer, Roman, Lewis, Sam, Tygier, Sam, Sievert, Scott, Vigna, Sebastiano, Peterson, Stefan, More, Surhud, Pudlik, Tadeusz, Oshima, Takuya, Pingel, Thomas J, Robitaille, Thomas P, Spura, Thomas, Jones, Thouis R, Cera, Tim, Leslie, Tim, Zito, Tiziano, Krauss, Tom, Upadhyay, Utkarsh, Halchenko, Yaroslav O, and Vázquez-Baeza, Yoshiki. “SciPy 1.0: fundamental algorithms for scientific computing in Python.” *Nature Methods* 17.3 (2020): 261–272.
- Vojdani, Aristo, Pollard, K. Michael, and Campbell, Andrew W. “Environmental Triggers and Autoimmunity.” *Autoimmune Diseases* 2014: 798029.
- Voulgarelis, Michael and Tzioufas, Athanasios G. “Pathogenetic mechanisms in the initiation and perpetuation of Sjögren’s syndrome.” *Nature Reviews Rheumatology* 6.9 (2010): 529–537.
- Wagenmakers, Eric-Jan, Verhagen, Josine, and Ly, Alexander. “How to quantify the evidence for the absence of a correlation.” *Behavior Research Methods* 48.2 (2016): 413–426.
- Wang, Ying-Hsiu, Tsai, Dong-Yan, Ko, Yi-An, Yang, Tsan-Tzu, Lin, I-Ying, Hung, Kuo-Hsuan, and Lin, Kuo-I. “Blimp-1 Contributes to the Development and Function of Regulatory B Cells.” *Frontiers in Immunology* 10: 1909.
- Watanabe, Sumio and Opper, Manfred. “Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory.” *Journal of Machine Learning Research* 11.12 (2010).
- Wellard, C, Markham, J, Hawkins, E D, and Hodgkin, P D. “The effect of correlations

- on the population dynamics of lymphocytes.” *Journal of Theoretical Biology* 264.2 (2010): 443–449.
- Wellard, Cameron, Markham, John F, Hawkins, Edwin D, and Hodgkin, Phillip D. “The Cyton Model for Lymphocyte Proliferation and Differentiation.” *Mathematical Models and Immune Cell Biology*. eds. Carmen Molina-París and Grant Lythe, chap. 5. New York: Springer, 2011. 107–120.
- Wile, Brian M, Ban, Kiwon, Yoon, Young-Sup, and Bao, Gang. “Molecular beacon-enabled purification of living cells by targeting cell type-specific mRNAs.” *Nature Protocols* 9.10 (2014): 2411–2424.
- Yates, Andrew, Callard, Robin, and Stark, Jaroslav. “Combining cytokine signalling with T-bet and GATA-3 regulation in Th1 and Th2 differentiation: a model for cellular decision-making.” *Journal of Theoretical Biology* 231.2 (2004): 181–196.
- Yates, Andrew, Chan, Cliburn, Strid, Jessica, Moon, Simon, Callard, Robin, George, Andrew JT, and Stark, Jaroslav. “Reconstruction of cell population dynamics using CFSE.” *BMC Bioinformatics* 8.1 (2007): 196.
- Yates, Christian A., Ford, Matthew J., and Mort, Richard L. “A Multi-stage Representation of Cell Proliferation as a Markov Process.” *Bulletin of Mathematical Biology* 79.12 (2017): 2905–2928.
- Zhang, Xuefei, Zhang, Yu, Ba, Zhaoqing, Kyritsis, Nia, Casellas, Rafael, and Alt, Frederick W. “Fundamental roles of chromatin loop extrusion in antibody class switching.” *Nature* 575.7782 (2019): 385–389.
- Zhao, Dan, Yang, Yantao, Qu, Na, Chen, Mingming, Ma, Zhao, Krueger, Christopher J., Behlke, Mark A., and Chen, Antony K. “Single-molecule detection and tracking of RNA transcripts in living cells using phosphorothioate-optimized 2'-O-methyl RNA molecular beacons.” *Biomaterials* 100: 172–183.
- Zheng, Yan, Collins, Samuel L., Lutz, Michael A., Allen, Amy N., Kole, Thomas P., Zarek, Paul E., and Powell, Jonathan D. “A Role for Mammalian Target of Rapamycin in Regulating T Cell Activation versus Anergy.” *The Journal of Immunology* 178.4 (2007): 2163–2170.

BIBLIOGRAPHY

Zhou, Jie H S, Markham, John F, Duffy, Ken R, and Hodgkin, Philip D. “Stochastically Timed Competition Between Division and Differentiation Fates Regulates the Transition From B Lymphoblast to Plasma Cell.” *Frontiers in Immunology* 9: 2053.

Zhu, Jinfang, Yamane, Hidehiro, and Paul, William E. “Differentiation of Effector CD4 T Cell Populations*.” *Annual Review of Immunology* 28.1 (2010): 445–489.

Zilman, Anton, Ganusov, Vitaly V., and Perelson, Alan S. “Stochastic Models of Lymphocyte Proliferation and Death.” *PLOS ONE* 5.9 (2010): e12775.