Individual-based Stochastic Modelling of Group 1 Innate Lymphoid Cell Plasticity

Carl Whelan

Born 16th June 1995 in Dublin, Ireland

29th September 2022

Master's Thesis Mathematics

Advisor: Prof. Dr. Anton Bovier

Second Advisor: Prof. Dr. Jan Hasenauer

Institut für Angewandte Mathematik

Mathematisch-Naturwissenschaftliche Fakultät der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Introduction

Natural killer (NK) cells play a critical role in the innate immune system of humans. In particular, they appear to be effective at killing cancer cells, thus controlling tumour initiation and growth. Such properties have led to growing interest in the use of NK cells in cancer immunotherapy.
However, clearly NK cells are not 100% effective in suppressing cancer, and the mechanism by which cancer escapes control by NK cells is therefore of immense biological and medical interest. One contributing factor to this evasion may be phenotypic plasticity among group 1 innate lymphoid cells (to which NK cells belong). In [1] the authors provide evidence that NK cells can switch phenotype to other group 1 innate lymphoid cells, namely type 1 innate lymphoid cells ($ILC1$s) and so-called intermediate type 1 innate lymphoid cells ($int\text{-}ILC1$s). While the exact mechanism behind this phenotypic plasticity remains unknown, it is thought to be stimulated, at least in part, by the cytokine TGF-β.
In this thesis, I attempt to explain the phenotypic plasticity among group 1 innate lymphoid cells using a stochastic individual-based model. This type of model has been well-studied (cf. [2], [3]), are well-suited to modelling cell-populations (where deaths, births and mutations can be thought of as occurring at random times), and have already been used to describe a cancer immunotherapy using other kinds of immune cells, namely T-cells ([4]). One key property of this type of model, which we will exploit repeatedly in this thesis, is that, under relatively weak conditions, the stochastic model obeys a law of large numbers. This means that the system can be approximated via a deterministic model rather than relying on computationally expensive stochastic simulations. This was particularly useful for parameter-fitting, since the system needs to be re-simulated with each new set of parameters considered.
In Section 1 we introduce the biological problem addressed in some more detail and offer some relevant biological background to the model. We also describe the available data which we use for parameter-fitting, and discuss some shortcomings of this data in the context of the model-fitting and validation techniques that could be employed for the rest of the thesis.
Section 2 provides a brief theoretical background to individual-based stochastic models. Here, we construct the stochastic process used in the later sections explicitly, and state the law of large numbers for this kind of process.
In Section 3, we introduce the model fitted for the biological system we wish to describe, i.e. the dynamics of group 1 innate lymphoid cells with phenotypic plasticity). Here, we discuss and compare two models: a first *equal growth rates model* where the growth-rates of all three group 1 innate lymphoid cell phenotypes are assumed to be equal, and a second *varied growth rates model* where we vary the growth-rates within constraints determined by other modelling assumptions we have imposed.

In Section 4 we describe the fitting method used to fit the parameters for the models discussed in Section 3. We also discuss alternative an alternative model-fitting approach that could be employed with more available data, and try out these methods on mock data simulated from the model fit Section 3.
Section 5 concludes and offers an outlook for how this model could be improved to take into account more biologically relevant features (in particular the effect of the cytokine, TGF-β), and how it might be implemented into a full tumour model to model cancer immunotherapy using NK-cells.

Acknowledgements

I would like to express my gratitude to my primary supervisor Prof. Dr. Anton Bovier for his guidance throughout this thesis, to Dr. Anna Kraut for introducing me to the possibilities of individual-based stochastic models through the engaging lecture course she delivered last summer semester, and to Prof. Dr. Tobias Bald for introducing me to the biological problem of innate lymphoid cell plasticity and for sharing his expertise throughout the project. I would also like to thank Dr. Dillon Corvino and Ananthi Kumar for their work in trying to provide additional experimental data within a tight time-frame. Lastly, I would like to thank Prof. Dr. Thomas Unger for encouraging me to study a master's degree in mathematics in Germany, and to thank the DAAD and the University of Bonn for making this ambition possible.

1 Biological Background

The purpose of this section is to introduce the biological problem considered in this thesis. Here, we give some high-level biological context for the models fitted in Sections 3 and 4. Due to lack of available data, not all of the information presented here corresponds to a feature in the final model. For example, the cytokine, TGF-β, is not considered as an explicit model input in the final model although it is a potentially significant factor in driving the plasticity among tumour group 1 ILC phenotypes. This is because there was not sufficient data to credibly fit separate natural switch-rate parameters and TGF-β-induced switch-rates in the final model (see Section 2). However, the biological context is still relevant for considering potential model improvements and designing future experiments, as discussed in Section 5. The information presented in this section is based on [1].

1.1 Tumour Group 1 Innate Lymphoid Cells

In [1] the authors present three distinct phenotypes of so-called group 1 innate lymphoid cells. These are phenotypically similar immune cells, which have different roles in the immune system's response to cancer, namely control of tumour growth and metastasis. Here, we briefly outline the different group 1 innate lymphoid cell phenotypes identified, and highlight the differences between them:

1.1.1 NK cells

Natural killer (NK) cells play a critical role in the immune system's response to cancer. They act by spontaneously killing cells thought to be dangerous to the host, such as cancer cells, and thus are crucial to cancer immunosurveillance. Such is their effectiveness at controlling tumour initiation and metastasis, that there has been growing interest in exploiting NK cells in cancer immunotherapy [5]. These cells are primarily found in the blood or within secondary lymphoid tissues, such as the spleen. They develop from NK cell precursors.

1.1.2 Type 1 Innate Lymphoid Cells (ILC1s)

Type 1 Innate Lymphoid Cells ($ILC1$s) are phenotypically similar to NK cells, but originate from different progenitor cells and are not typically found in blood or lymphoid organs. These cells are well-characterised in the livers of adult mice. ILC1-like cells have also been identified among tumour-infiltrating lymphocytes and peripheral blood mononuclear cells of human cancer patients. They do not appear to be able to control tumour growth and metastasis. They can be distinguished from NK cells by the expression of the integrins CD49a and CD49b (see Figure 1).

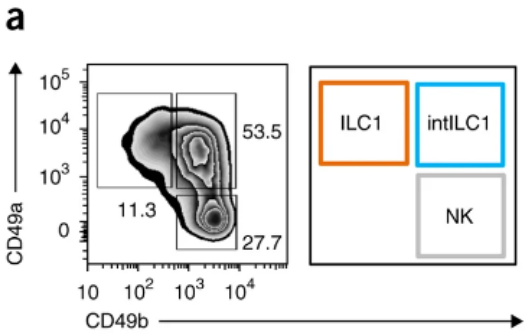

Figure 1: Figure 1a of [1]: NK cells, ILC1s, and int-ILC1s in mouse tumour environment, as distinguished by the markers CD49a and CD49b.

(At this point, the reader's attention is drawn to the difference in terminology between "group 1 innate lymphoid cells" which covers all three immune cell phenotypes discussed in this thesis - i.e. NK cells, ILC1s, and intermediate-ILC1s (below) - and "type 1 innate lymphoid cells" which refers only to the phenotype discussed here. To avoid confusion, the abbreviation "ILC1" will only be used to refer to the phenotype of *type* 1 innate lymphoid cells.)

1.1.3 Intermediate ILC1s

As their name suggests, intermediate ILC1s (*int-ILC1s*) share phenotypic similarities with both NK cells and ILC1s. They have been identified as a distinct group 1 innate lymphoid cell phenotype in mouse livers. Transcriptomic analysis of NK cells, ILC1s, and int-ILC1s from tumours showed that each of these subsets has a distinct gene-expression profile. Again, int-ILC1s can be distinguished from other group 1 innate lymphoid cells by the expression of the integrins CD49a and CD49b. Like ILC1s, they do not appear to be able to control tumour growth and metastasis.

1.2 TGF-β

TGF-β is a cytokine which is known to suppress anti-tumour properties of NK cells. Both in vitro and in vivo evidence suggests that NK cells are less prevalent in environments with higher concentrations of TGF-β and more prevalent in environments with lower concentrations of TGF-β (in terms of cell frequencies). Conversely, int-ILC1s and ILC1s are more prevalent in environments with higher concentrations of TGF-β (see Figure 2). A plausible explanation for this is that TGF-β induces additional switching from NK cells to other group 1 innate lymphoid cell phenotypes - for example [6], and [7] substantiate NK cells' ability to switch phenotype in the presence

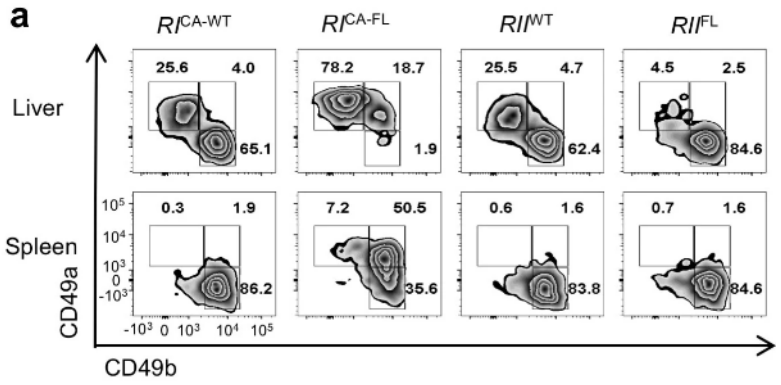

Figure 2: Supplementary figure 2a of [1]. $\text{RI}^{\text{CA-FL}}$ refers to mice with constitutively active TGF-β, RII^{FL} to mice with ablated TGF-β, whereas $\text{RI}^{\text{CA-WT}}$ and RII^{WT} are the respective control groups.

of TGF-β. However, the models considered in this thesis could also be used to explore other possible mechanisms to explain the numbers in Figure 2, such as TGF-β decreasing the growth-rate of NK cells and increasing the growth-rate of int-ILC1s and ILC1s.

As mentioned at the beginning of this section, it was not possible to obtain sufficient data to fit separate model parameters which depend on TGF-β concentration. Thus, the model fitted in Section 4 implicitly assumes a fixed concentration of TGF-β. However, as an important regulator of the dynamics considered in this thesis, I mention it here, and identify it as an area for future model improvement. This is explored further in Section 5.

1.3 Available Data

The data available is contained in Figure 2a of [1]. On day 0 of the experiment, 5×10^5 splenic NK cells were injected into mice livers. A breakdown of cell frequencies by phenotype on days 7, 14, and 21 was then obtained using flow cytometry. The results are summarised in Figure 3, where the numbers on the graph represent percentage frequencies. The quadrants correspond to the cell phenotypes as in Figure 1, i.e. top-left corresponds to ILC1s, top-right to int-ILC1s, and bottom-right to NK cells.

We neglected the bottom-left quadrant (which has not been characterised as a distinct phenotype and is therefore not modelled), and scaled the remaining frequencies so that they sum to 1 at each time point. This gives the data given by Table 1.

The available data suffer from two main deficiencies. Firstly, the model we propose estimates absolute cell numbers, whereas the available data is only in terms of cell frequencies. This was problematic for a number of

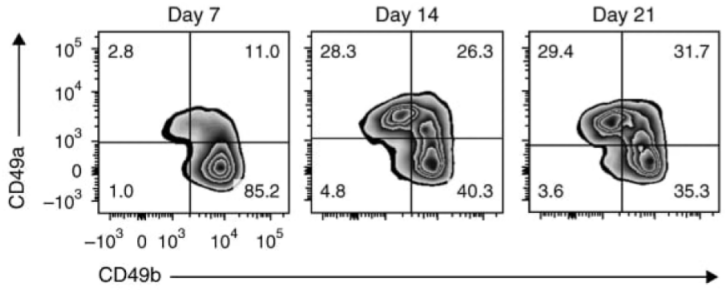

Figure 3: Figure 2a of [1] showing cell frequencies by phenotype in the mouse liver environment on days 7, 14, and 21 of the experiment respectively.

Day	NK frequency	int frequency	ILC frequency
0	100%	0%	0 %
7	86.06%	11.11%	2.83%
14	42.47%	27.71%	29.82%
21	36.62%	32.88%	30.50%

Table 1: Available data after scaling frequencies to sum to 100%.

model-fitting strategies considered, especially those which aim to give weight to earlier data-points. This is discussed further in Section 4.2. It is also not ideal from a model-validation perspective, since we can only validate against cell frequencies, meaning that a model which significantly over- or underestimates cell growth but with proportions similar to the raw data at the available time-points would still appear to perform well, despite not accurately modelling the cell dynamics in which we are interested.
Secondly, with only three data points it is difficult to credibly validate the fitted model. Indeed, even with absolute cell numbers data, the fitting methods considered in Section 4.2 would likely be ineffective due to the limited number of time-points available.

2 Stochastic Individual-based Models with Phenotypic Switching

The stochastic individual-based models considered in this section are based on those outlined in [3]. In [3] the processes are constructed as point processes. This approach has the advantage that the models are more easily generalised to infinite (even continuous) trait spaces. However, for the purpose of this thesis we are only interested in a finite trait space, and modelling the number of individuals in a population with a given trait in continuous time. Thus it is sufficient (and, to my taste, more intuitive) to consider the state space as $\mathbb{N}_0^{|\mathcal{X}|}$ for the original process, and $\frac{1}{K}\mathbb{N}_0^{|\mathcal{X}|}$ for the re-scaled process. This is also the construction followed in Chapter 11 of [8], and is mentioned as an alternative representation of the population state in Remark 1. (i) of [3].
The processes constructed in [3] also include features that were not used in this thesis, either because they were not thought to be relevant to the biological process we are trying to describe (in the case of mutations at birth) or because there was insufficient data to fit additional parameters (in the case of induced switch-rates). Therefore in this section we re-construct the process of interest using only features that will be used later, and with the state spaces mentioned above, thus avoiding any redundant notation, and focusing the reader's attention to a suitable level of generality for the modelling task at hand. The additional available features are mentioned at the end of this section, for the reader's interest and to give context for the possible model extensions discussed in Section 5.

2.1 Population process

Let $\mathcal{X}$ be a finite set, which we call the trait space for the population. We define the following rates:

- The birth rate $b_j \geq 0$, $j \in \mathcal{X}$, which is the rate at which an individual with trait j gives rise to a new individual of trait x;
- The natural death rate $d_j \geq 0$, $j \in \mathcal{X}$ which is the rate at which an individual with trait j dies naturally;
- A competition kernel $(c_{j,i})_{j,i\in\mathcal{X}}$ where $c_{j,i} \geq 0$ represents the competitive pressure felt by an individual with trait j from an individual with trait i in the population; and
- A switch rate $s_{j,i} \geq 0$ for each $j \neq i \in \mathcal{X}$ which is the rate at which an individual with trait j switches phenotype to i.

Having defined these switch-rates, we can define a Markov process with state space $\mathbb{N}_0^{|\mathcal{X}|}$ characterized by the following Q-matrix (in the sense of cf [9]):

$$\begin{aligned} q(n, n+l) &= \begin{cases} n_j b_j & l = \delta_j,\ j \in \mathcal{X} \\ n_j \left(d_j + \sum_{i\in\mathcal{X}} c_{j,i} n_i\right) & l = -\delta_j,\ j \in \mathcal{X} \\ n_j s_{j,i} & l = -\delta_j + \delta_i,\ j, i \in \mathcal{X} \\ -\sum_{j,i}(\text{above rates}) & l = 0 \end{cases} \\ &=: \beta_l(n) \end{aligned} \tag{1}$$

where $n = (n_j)_{j\in\mathcal{X}} \in \mathbb{N}_0^{|\mathcal{X}|}$ and δ_j denotes the standard unit vector in $\mathbb{N}_0^{|\mathcal{X}|}$ with value 1 in component j and 0 everywhere else.
From this Q-matrix we can construct a Markov jump process on $\mathbb{N}_0$ the standard way, namely:

1. Assume the population is in some (possibly random) state $\tilde{n}(0)$ at time 0.

2. Construct a discrete time Markov chain with transition probability, conditional on $\tilde{n}_{k-1}$ given by:

$$\mathbb{P}(\tilde{n}(k) = n + l | \tilde{n}(k-1) = n) = \frac{\beta_l(n)}{\sum_{l'\neq 0} \beta_{l'}(n)}$$

   for $l \in \{\delta_j, -\delta_j, -\delta_j + \delta_i : i, j \in \mathcal{X}\}$, and $k \in \mathbb{N}$.

3. Define the clock process

$$S(m) := \sum_{k=0}^{m-1} \frac{e_k}{\sum_{l'\neq 0} \beta_{l'}(\tilde{n}(k))}$$

   where $(e_k)_{k\in\mathbb{N}_0}$ are standard i.i.d exponential random variables.

4. Set

$$n(t) := \tilde{n}(S^{-1}(t))$$

   where $S^{-1}(t) := \inf\{m \in \mathbb{N}_0 : S(m) \geq t\}$. Then (provided $S^{-1}(t)$ is a.s well-defined for all $t \geq 0$) $(n(t))_{t\geq 0}$ is a Markov jump process with generator given by:

$$\begin{aligned} \mathcal{L}f(n) :=& \sum_{j\in\mathcal{X}} \left(f(n+\delta_j) - f(n)\right) n_j b_j \\ &+ \sum_{j\in\mathcal{X}} \left(f(n-\delta_j) - f(n)\right) n_j (d_j + \sum_{i\in\mathcal{X}} c_{j,i} n_i) \\ &+ \sum_{j\in\mathcal{X}} \sum_{i\in\mathcal{X}} \left(f(n-\delta_j+\delta_i) - f(n)\right) n_j s_{j,i} \\ =& \sum_{l} \left(f(n+l) - f(n)\right) \beta_l(n) \end{aligned} \tag{2}$$

defined on bounded, measurable functions $f : \mathbb{N}_0^{|\mathcal{X}|} \to \mathbb{R}$.

The above construction is well-defined for all times $t \geq 0$ if, with probability 1, the clock process $(S(m))_{m\in\mathbb{N}_0}$ becomes arbitrarily large with probability 1 (i.e. $S(m) \xrightarrow{m\to\infty} \infty$ a.s.). This is ensured by our choice of transition rates, so long as the population size itself does not explode in finite time. The following lemma assures us that this does not occur. We state this without proof, and refer to its analog in the construction via point processes, Theorem 2.5 (ii) of [2].

Lemma 2.1. *Let $T < \infty$ and assume $\mathbb{E}[(\sum_{x\in\mathcal{X}} \tilde{n}_x(0))^2] < \infty$. Then:*

$$\mathbb{E}[\sup_{t\leq T}(\sum_{x\in\mathcal{X}} \tilde{n}_x(t))^2] < \infty \tag{3}$$

The above lemma tells us that the value $\sum_{l'\neq 0} \beta'_l(n(t))$ remains almost surely finite for any $T < \infty$, and hence the above construction is well-defined.

2.2 Re-scaled process and Law of Large Numbers

We now re-scale the process as follows. Let $K > 0$, called the *carrying capacity*. In applications, this parameter represents the maximum size to which the population can grow. We now define $n^K(t) := \frac{1}{K}n(t)$. The re-scaled process is again a Markov jump process with rates given by the Q-matrix:

$$q^{(K)}(n^K, n^K + \frac{1}{K}l) = K\beta_l(n^K)$$

This describes a Markov process with generator given by:

$$\begin{aligned}\mathcal{L}f(n^K) :=& \sum_{j\in\mathcal{X}} \left(f(n^K + \frac{1}{K}\delta_j) - f(n^K)\right) K n_j^K b_j \\ &+ \sum_{j\in\mathcal{X}} \left(f(n^K - \frac{1}{K}\delta_j) - f(n^K)\right) K n_j^K (d_j + \sum_{i\in\mathcal{X}} c_{j,i} n_i^K) \\ &+ \sum_{j\in\mathcal{X}}\sum_{i\in\mathcal{X}} \left(f(n^K + \frac{1}{K}(-\delta_j + \delta_i)) - f(n^K)\right) K n_j^K s_{j,i} \\ =& \sum_l \left(f(n^K + \frac{1}{K}l) - f(n^K)\right) K\beta_l(n^K)\end{aligned} \tag{4}$$

for bounded, measurable $f : \frac{1}{K}\mathbb{N}_0^{|\mathcal{X}|} \to \mathbb{R}$.
The key result, on which the analysis in later sections depends, is the Law of Large Numbers for this process. As the name suggest, this theorem states that the re-scaled process tends to a deterministic limit as the carry-capacity K goes to infinity. The deterministic limiting process is the solution to a

system of ordinary differential equations characterised by the transition rates $\beta_l(n^K)$. The precise statement is as follows:

Theorem 2.2 (Law of Large Numbers)**.** *For $K > 0$ let $(n^K(t))_{t\geq 0}$ be the Markov jump process with generator given by* (4) *and suppose* $\lim_{K\to\infty} n^K(0) = \eta_0$ *a.s. where $\eta_0 \in \mathbb{R}^{|\mathcal{X}|}$ (i.e. the processes converges almost surely to a deterministic limit). Let $\eta(t)$ be a solution to the system of ordinary differential equations given by:*

$$\begin{aligned} \dot{\eta}(t) =& \sum_l l\beta_l(\eta(t)) \\ \eta(0) =& \eta_0 \end{aligned} \tag{5}$$

i.e.

$$\begin{aligned} \dot{\eta}_j(t) =& \eta_j(t)\left(b_j - d_j - \sum_{i\in\mathcal{X}} c_{j,i}\eta_i(t) - \sum_{i\in\mathcal{X}} s_{j,i}\right) + \sum_{i\in\mathcal{X}} \eta_i(t)s_{i,j} \ (j \in \mathcal{X}) \\ \eta(0) =& \eta_0 \end{aligned} \tag{6}$$

where $\eta_j(t)$ denotes the value of $\eta(t)$ in component $j \in \mathcal{X}$. Then, for all $t \geq 0$:

$$\lim_{K\to\infty} \sup_{s\leq t} |n^K(s) - \eta(s)| = 0 \quad a.s$$

Proof. This is essentially a direct application of Theorem 2.1 from Chapter 11 of [8]. The only additional hypotheses in the referenced theorem are that for each compact $B \subset \mathbb{R}^{|\mathcal{X}|}$, we have $\sum_l |l| \sup_{n\in B} \beta_l(n) < \infty$ and that there exists M_B such that $|\sum_l l(\beta_l(n) - \beta_l(m))| \leq M_B|n - m|$. These are easily seen to be satisfied by the transition rates used here. □

2.3 Additional model features - induced switching and mutations

The model we have constructed above is the model on which Sections 3 and 4 are based. However, it is worth briefly presenting additional biological mechanisms that are often included in stochastic individual-based models of cell populations, which could be useful for model extensions in the future (see Section 5). The two additional features we discuss here are:

- Induced switches; and
- Mutations at birth.

To consider these features we return to the setting of 2.1 and introduce the following additional parameters:

- A probability $\mu_K m(j) \in [0,1]$ representing the probability that a mutation occurs, at a given birth event of an individual with trait $j \in \mathcal{X}$. Here μ_K is a scaling parameter. (Of course, the probability of no mutation occurring, i.e. that the individual being added to the population at the birth event also has trait j, is then $1 - \mu_K m(j)$.)
- For each $j \in \mathcal{X}$ a probability mass $M(j, \cdot)$ on $\mathcal{X}$, where $M(j,i)$ is the probability that, given a mutation event, the individual with trait j mutates to trait i.
- For each $h \in \mathcal{X}$, an induced switching kernel $(s_h^{ind.}(j,i))_{j,i\in\mathcal{X}}$ where $s_h^{ind.}(j,i) \geq 0$ represents the rate at which an individual with trait j changes trait to i induced by an individual of trait h.

Biologically, the main difference between a mutation and a switch is that a mutation event occurs at the birth of an individual, whereas switches occur during the lifetime of an individual. Mathematically, this is expressed by adding δ_i to the population in the case where an individual with trait j gives rise to an individual of trait i (via a mutation at birth) and adding $-\delta_j + \delta_i$ to the population in the case where an individual with trait j switches to an individual with trait i.
With these additional features, we need to amend the rates defined in 2.1. The Q-matrix for a system with the above features is given as follows:

$$\begin{aligned} q(n, n+l) &= \begin{cases} n_j b_j (1 - \mu_K m(j)) + \sum_{i\in\mathcal{X}} n_i b_i \mu_K m(i) M(i,j) & l = \delta_j,\ j \in \mathcal{X} \\ n_j (d_j + \sum_{i\in\mathcal{X}} c_{j,i} n_i) & l = -\delta_j,\ j \in \mathcal{X} \\ n_j (s_{j,i} + \sum_{h\in\mathcal{X}} s_h^{ind.}(j,i) n_h) & l = -\delta_j + \delta_i,\ j,i \in \mathcal{X} \\ -\sum_{i,j} (\text{above rates}) & l = 0 \end{cases} \\ &=: \beta_l(n) \end{aligned} \tag{7}$$

With this Q-matrix we can follow the same construction as in 2.1 to get a Markov jump process with generator given by:

$$\begin{aligned} \mathcal{L}f(n) :=& \sum_{j\in\mathcal{X}} (f(n+\delta_j) - f(n)) \left(n_j b_j (1 - \mu_K m(j)) + \sum_{i\in\mathcal{X}} n_i b_i \mu_K m(i) M(i,j) \right) \\ &+ \sum_{j\in\mathcal{X}} (f(n-\delta_j) - f(n))\, n_j (d_j + \sum_{i\in\mathcal{X}} c_{j,i} n_i) \\ &+ \sum_{j\in\mathcal{X}} \sum_{i\in\mathcal{X}} (f(n-\delta_j+\delta_i) - f(n))\, n_j (s_{j,i} + \sum_{h\in\mathcal{X}} s_h^{ind.}(j,i) n_h) \\ =& \sum_l (f(n+l) - f(n))\, \beta_l(n) \end{aligned} \tag{8}$$

for bounded, measurable $f : \mathbb{N}_0^{|\mathcal{X}|} \to \mathbb{R}$.
Re-scaling the rates as before, i.e. replacing $\beta_l(n)$ with $K\beta_l(n^K)$, we again get a re-scaled process with generator given by:

$$\mathcal{L}f(n^K) = \sum_l \left(f(n^K + \frac{1}{K}l) - f(n^K) \right) K\beta_l(n^K)$$

for bounded, measurable $f : \frac{1}{K}\mathbb{N}_0^{|\mathcal{X}|} \to \mathbb{R}$. The Law of Large numbers in the form of (5) still holds, again by applying Theorem 2.1 from Chapter 11 of [8].

3 Model and Results

In this section, we fix a state space $\mathcal{X} := \{\text{NK}, \text{int}, \text{ILC}\}$ (representing NK cells, int-ILC1s, and ILC1s respectively, as described in Section 1) and implement a model based on the stochastic individual-based models with phenotypic switching constructed in Section 2.
The relatively straightforward dynamics modelled mean that the limiting process of the re-scaled population size is not influenced by rare stochastic fluctuations (e.g. a rare mutant population invading and fixating in the population, or a small initial population dying out). Therefore, it is the deterministic limiting process given by Theorem 2.2 which is modelled in the remainder of this thesis. This significantly improves computational efficiency, and, due to the above remark, does not result in a large loss of accuracy compared to an exact simulation via the Gillespie algorithm (cf. [10]) or a hybrid algorithm which combines stochastic and deterministic effects (e.g a Tau-leap algorithm, cf. [11]).
As sufficient data were not available to fit separate parameters for natural switch-rates and switch-rates induced by TGF-ß, we only look at natural switch-rates in this section and the section that follows. In doing this, we implicitly assume that the system we are modelling contains some fixed concentration of TGF-ß which remains constant over time, and that the natural switch rates we fit reflect this concentration.
We present two models for consideration: first an *equal growth rates model* where we assume that NK cell, int-ILC1, and ILC1 populations all have the same natural growth rate (defined as the difference between the natural birth and natural death-rates), and a *varied growth rates model* where we relax this assumption (within constraints determined by some other modelling assumptions imposed). We discuss some rationale around the parameter choices made and present results of the simulated deterministic system with these parameters. A more detailed description of the fitting method employed is given later.

3.1 The Model

We consider the stochastic individual-based model described by the transition diagram in Figure 3.1, with trait space $\mathcal{X} := \{\text{NK}, \text{int}, \text{ILC}\}$ as above. We denote the number of cells with trait j alive at time t by $n_j(t)$. In words,

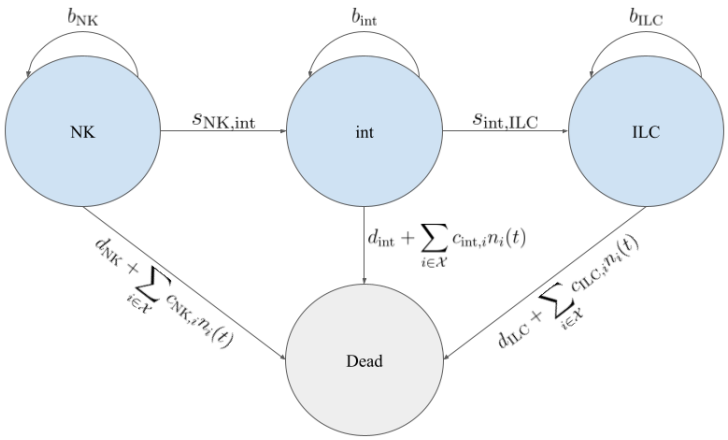

Figure 4: Transition diagram describing group 1 innate lymphoid cell plasticity as a Markov jump process.

the above diagram means that a cell with trait $j \in \mathcal{X}$ can:

- divide with rate: $n_j(t) b_j$;
- die with rate: $n_j(t) \left(d_j + \sum_{i \in \mathcal{X}} c_{j,i} n_i(t) \right)$; and
- switch phenotype to $i \in \mathcal{X}$ with rate: $n_j(t) s_{j,i}$.

As indicated by the diagram, We assume that:

- NK-cells can switch phenotype to int-ILC1;
- int-ILC1s can switch phenotype to ILC1; and
- no other phenotypic switches are possible, i.e.

$$s_{j,i} = 0 \text{ for all } (j,i) \in \mathcal{X} \times \mathcal{X} \setminus \{(\text{NK}, \text{int}), (\text{int}, \text{ILC})\}$$

This means that once a cell has phenotype ILC1, it remains with this phenotype until it dies. In her master's thesis [12], Lisa Li recently considered a similar model in which backwards switches from ILC1 to int-ILC1 and from int-ILC1 to NK are also allowed.

The Q-matrix, rate function $\beta_l(n)$, and generator $\mathcal{L}$ are exactly as in Section 2.1. We re-scale the above process as in Section 2.2 and apply the Law of Large Numbers given by Theorem 2.2
Thus the re-scaled process can be approximated by the deterministic system of ordinary differential equations:

$$\begin{aligned}\dot{n}_j^K(t) &= n_j^K(t)\left(b_j - d_j - \sum_{i\in\mathcal{X}} s_{j,i} + \sum_{i\in\mathcal{X}} c_{i,j} n_i^K(t)\right) + \sum_{i\in\mathcal{X}} s_{i,j} n_i^K(t), \quad j \in \mathcal{X} \\ \mathbf{n^K}(0) &= \mathbf{n}_0^K\end{aligned} \tag{9}$$

where $\mathbf{n}_0^K := \frac{1}{K}\mathbf{n}_0 := \frac{1}{K}(n_j(0))_{j\in\mathcal{X}}$ is the initial population state.

3.2 Additional Assumptions

To fit the model described above, we need to choose a reasonable carrying capacity K (which should reflect the maximum size to which the cell population can grow), and then fit the following 17 parameters to the data: b_{NK}, b_{int}, b_{ILC}, d_{NK}, d_{int}, d_{ILC}, $c_{\text{NK,NK}}$, $c_{\text{NK,int}}$, $c_{\text{NK,ILC}}$, $c_{\text{int,NK}}$, $c_{\text{int,int}}$, $c_{\text{int,ILC}}$, $c_{\text{ILC,NK}}$, $c_{\text{ILC,int}}$, $c_{\text{ILC,ILC}}$, $s_{\text{NK,int}}$, and $s_{\text{int,ILC}}$. Since the birth-rates b_j and natural death-rates d_j are by themselves not identifiable in the deterministic model, we instead fit a growth-rate r_j for each trait, defined as the difference between the birth- and natural death-rate associated with that trait, i.e.

$$r_j := b_j - d_j$$

This leaves 14 parameters as well as the carrying capacity to be chosen.
The carrying capacity was chosen as $K = 10^7$. This is the same carrying capacity as chosen in [4] in the context of tumour cells, and appears to fit the data well in the context of NK cells. It is worth noting that, since the available data is in terms of cell frequencies rather than absolute cell numbers, the choice of carrying capacity only impacts the equilibrium position (as a result of the below assumptions) and the initial condition.
While the number of NK-cells injected at time $t = 0$ is known in terms of absolute cell numbers (5×10^5), there is significant uncertainty around how many of the injected NK-cells actually take hold in the cell population. Thus the initial condition for the absolute number of NK-cells alive in the population at time $t = 0$ was varied between 10^4 and 5×10^5. In the end, the initial number was chosen as 10^5, i.e. it is assumed that 20% of the injected cells take hold in the population.
As the available data only contains three time-points, it is necessary to make some simplifying assumptions in order to reduce the parameter space so that we can fit a model. The simplifying assumptions made are as follows:

1. NK cells do not compete with int or ILC cells, i.e. $c_{\mathrm{NK},j} = c_{j,\mathrm{NK}} = 0$ for $j = \mathrm{int}, \mathrm{ILC}$. This assumption is based an experimental observation from Supplementary Figure 2c of [1] (Figure 5), whereby the total number[1] of int-ILC1s and ILC1s does not increase with a corresponding large decrease in the number of NK cells.

2. The rate of self-competition is the same for all cell phenotypes, i.e. $c_{j,j} =: c_{\mathrm{self}}$ for all $j \in \mathcal{X}$.

3. All cell phenotypes have equal growth rate, i.e. $r_j =: r$ for all $j \in \mathcal{X}$.

4. NK cells switch phenotype to int-ILC1s at the same rate as int-ILC1s switch phenotype to ILC1s, i.e. $s_{\mathrm{NK,int}} = s_{\mathrm{int,ILC}} =: s$.

Then, in order to have a strategy to fit the remaining parameters, we additionally assume that the cell population has stabilised at the last available timepoint, i.e.

5. The system is in equilibrium at time $t = 21$.

This assumption appears to be consistent with the available data, in the sense that the change in frequencies between days 14 and 21 is not as significant as between days 7 and 14. It was also deemed to be biologically reasonable in discussions with researchers in experimental immunology.
Moreover, since the available data is in terms of cell frequencies rather than cell numbers, we need to make an assumption about the total number of cells in the population at this time, in order to have an equilibrium value in terms of cell numbers. Thus, finally, we assume:

6. The total cell population is K at time $t = 21$, i.e. the population has reached its carrying capacity at this time.

Assumptions 1-4 reduce the dimension of the parameter space from 14 to 5. That is, the remaining parameters to be fit are r, s, c_{self}, $c_{\mathrm{int,ILC}}$, and $c_{\mathrm{ILC,int}}$. These were chosen by first fixing a common switch-rate s, then a common growth-rate r, and then solving for the 3 remaining competition parameters by setting the system of ordinary differential equations (9) to 0 at $t = 21$, which is equivalent to assumption 5 above. The set of parameters which best fit the experimental data was then chosen as our first model, which we call the *equal growth rates model*.

Remark: Assumption 1 is somewhat counter-intuitive in the sense that all group 1 innate lymphoid cells may be expected to compete for the same resources. This choice was made as an attempt to simultaneously explain the

[1] This is a rare case in which absolute cell numbers data were available, however only for one time-point meaning that the data were not suitable for model-fitting purposes.

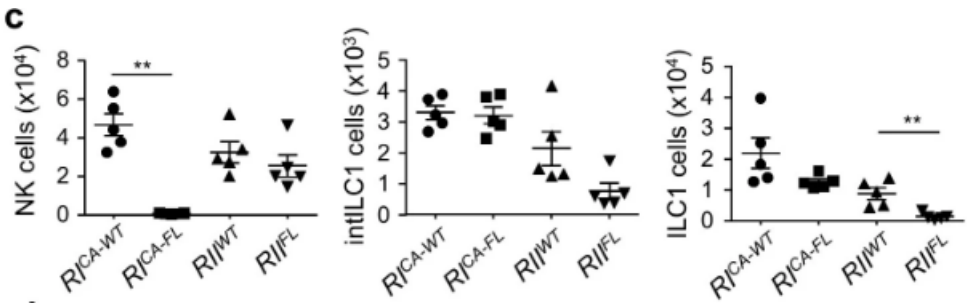

Figure 5: Supplementary figure 2c of [1], showing that a decrease in absolute number of NK-cells did not translate to a corresponding increase in absolute number of int-ILC1s or ILC1s.

somewhat surprising results from Figure 5 and reduce the parameter space. However, it was not the only choice that could have been made. For example, in [12] Lisa Li fits a similar model with a constant competition kernel which also achieves a good fit to experimental data. This constant competition assumption is arguably more consistent with the biological mechanism considered. Nevertheless, we continue with the assumption above in order to produce an alternative perspective, and recommend that both approaches are considered in future modelling ventures, especially if more data should become available to drawer stronger conclusions.

3.3 Results - Equal growth rates model

The common switch-rate s was initially varied between 10^{-6} and 0.1. At each value of s, 500 values for r ranging between 0.01 and 3 were tested. (The remaining competition parameters c_{self}, $c_{\text{int,ILC}}$, and $c_{\text{ILC,int}}$ are determined uniquely by the choice of s, r, and the above assumptions.) The system with parameters determined by the combination (r, s) was then simulated using Runge-Kutta of order 4, and the frequencies at times $t = 7$, $t = 14$, $t = 21$ were compared to the frequencies in the raw data. The value of r which minimised the squared error compared to the raw data was then chosen for each value of s. The resulting parameters and root mean-squared-error (RMSE) of the corresponding models are displayed in Table 2. Each model's fit to the data is plotted in Figure 6. Raw data are shown as scatter-plots over the simulated models, plotted as curves.

While the parameters corresponding to a switch-rate $s = 0.001$ give the least RMSE, the curve corresponding to these parameters displays surprising behaviour between between time-points $t = 7$ and $t = 14$. Therefore, in the absence of further data to support this behaviour, we might favour parameters corresponding to a switch-rate $s = 0.01$. We therefore repeated the analysis, this time with 100 switch-rates between 0.004 and 0.014, and 100

s	r	c_{self}	$c_{\text{int,ILC}}$	$c_{\text{ILC,int}}$	RMSE
10^{-6}	2.1791	5.9508	0.7287	1.1076	0.07225
10^{-5}	1.8795	5.1327	0.6285	0.9554	0.03753
10^{-4}	1.6278	4.4451	0.5447	0.8280	0.03508
0.001	1.2324	3.3627	0.4154	0.6322	0.03353
0.01	0.8848	2.3891	0.3290	0.5078	0.04365
0.1	0.4054	0.8342	0.4673	0.7873	0.13338

Table 2: Squared-error-minimising parameter values found for different values of common switch-rate s under equal growth-rates assumption.

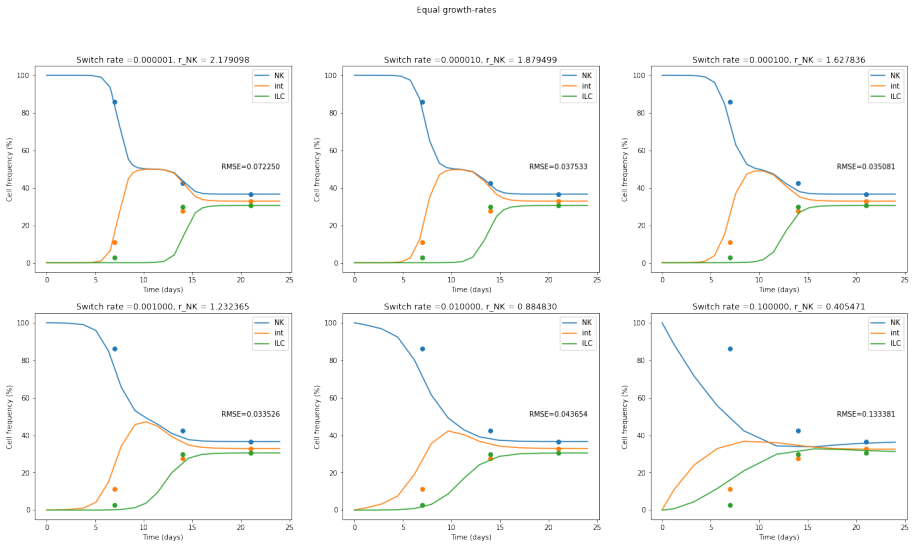

Figure 6: Best-fit simulated model and raw data for various switch rates under equal growth-rates assumption.

growth rates between 0.2 and 2. The final parameters chosen for the equal growth-rates model and the corresponding RMSE are provided in Table 3. The fit to the raw and the re-scaled cell numbers predicted by this model are displayed in Figure 7.

s	r	c_{self}	$c_{\text{int,ILC}}$	$c_{\text{ILC,int}}$	RMSE
0.00482	0.9273	2.5191	0.3261	0.4993	0.03387

Table 3: Chosen parameters and RMSE for equal growth-rates model.

3.4 Varying the growth rate

The equal growth-rates model consistently over-estimates the frequency of int-ILC1s, especially at earlier time-points. Therefore, we may wish to aban-

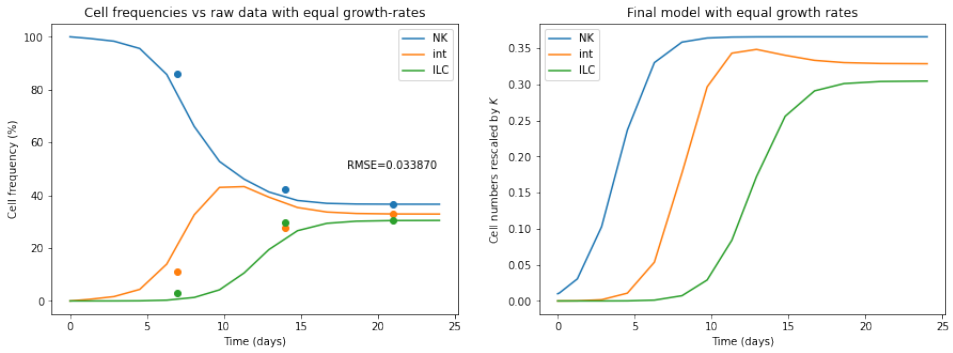

Figure 7: Chosen equal growth-rates model: predicted cell frequencies vs. raw data (left) and predicted re-scaled cell numbers (right).

don the assumption that each trait has an equal growth-rate, and try to fit a model where the growth rate of int-ILC1s, r_{int}, is reduced.
However, under the assumption that competition kernel is non-negative and assumptions 1-2 above, r_{int} may not deviate "too far" from the growth-rate of NK cells, r_{NK}.
More precisely, fixing growth-rates $(r_{\text{NK}}, r_{\text{int}}, r_{\text{ILC}})$ and switch-rates $(s_{\text{NK,int}}, s_{\text{int,ILC}})$ (no longer assumed to be equal for all phenotypes), and setting the ODE system to 0 at the assumed equilibrium state, denoted $(E_{\text{NK}}, E_{\text{int}}, E_{\text{ILC}})$ (where we assume $E_j > 0$ for each $j \in \mathcal{X}$), the resulting system of linear equations can be expressed as:

$$\sum_{i\in\mathcal{X}} c_{j,i}E_i = r_j - \sum_{i\in\mathcal{X}} s_{j,i} - \sum_{i\in\mathcal{X}} \frac{E_i s_{i,j}}{E_j} \tag{10}$$

By assumptions 1 and 2 above (no competition between NK cells and int-ILC1s or ILC1s, and equal self-competition) we immediately see that

$$c_{\text{self}} = \frac{r_{\text{NK}} - s_{\text{NK,int}}}{E_{\text{NK}}} \tag{11}$$

Back substitution into (10) yields the solution:

$$c_{\text{ILC,int}} = \frac{1}{E_{\text{int}}}\left(r_{\text{ILC}} - (r_{\text{NK}} - s_{\text{NK,int}})\frac{E_{\text{ILC}}}{E_{\text{NK}}} + s_{\text{int,ILC}}\frac{E_{\text{int}}}{E_{\text{ILC}}}\right) \tag{12}$$

and

$$c_{\text{int,ILC}} = \frac{1}{E_{\text{ILC}}}\left(r_{\text{int}} - s_{\text{int,ILC}} - (r_{\text{NK}} - s_{\text{NK,int}})\frac{E_{\text{int}}}{E_{\text{NK}}} + s_{\text{NK,int}}\frac{E_{\text{NK}}}{E_{\text{int}}}\right) \tag{13}$$

We pay particular attention to the second of these equations, as r_{int} is the parameter we wish to vary. The constraint that the quantity $c_{\text{int,ILC}} \geq 0$

requires:

$$r_{\text{int}} \geq r_{\text{NK}} \frac{E_{\text{int}}}{E_{\text{NK}}} + s_{\text{int,ILC}} - s_{\text{NK,int}} \left(\frac{E_{\text{NK}}}{E_{\text{int}}} + \frac{E_{\text{int}}}{E_{\text{NK}}} \right) \tag{14}$$

When $s_{\text{NK,int}} \geq s_{\text{int,ILC}}$ this is achieved (for any positive equilibrium value) by:

$$r_{\text{int}} = r_{\text{NK}} \frac{E_{\text{int}}}{E_{\text{NK}}} \tag{15}$$

Therefore, we replace the equal growth-rates assumption with the assumption that the relationship between r_{int} and r_{NK} described by (15) holds.

3.5 Results - Varied growth rates model

Following the same procedure as in 3.3, the square-error-minimizing parameters for different values of s between 10^{-6} and 0.1 are given by Table 4, and the corresponding models are plotted in Figure 8. Here, r corresponds to the growth-rate of NK cells (and ILC1s), while the growth rate of int-ILC1s is given by $r\frac{E_{\text{int}}}{E_{\text{NK}}}$.

s	r	c_{self}	$c_{\text{int,ILC}}$	$c_{\text{ILC,int}}$	RMSE
10^{-6}	2.3469	6.4090	0.000003	1.1929	0.03896
10^{-5}	2.1072	5.7545	0.000033	1.0711	0.02848
10^{-4}	1.6159	4.4124	0.000332	0.8219	0.03778
0.001	1.1365	3.1009	0.003317	0.5835	0.03577
0.01	0.8069	2.1763	0.033169	0.4683	0.03396
0.1	0.4234	0.8833	0.331689	0.7694	0.12111

Table 4: Squared-error-minimising parameter values found for different values of common switch-rate s under varied growth-rates assumption.

For similar reasons to before, switch-rates close to 0.01 were considered further. The parameters eventually chosen for the varied growth-rates model are given in Table 5. The fit to the raw data and the re-scaled cell numbers predicted by this model are displayed in Figure 9.

s	$r_{\text{NK}} = r_{\text{ILC}}$	r_{int}	c_{self}	$c_{\text{int,ILC}}$	$c_{\text{ILC,int}}$	RMSE
0.005061	1.000	0.8980	2.7171	0.0168	0.5377	0.027832

Table 5: Chosen parameters and RMSE for varied growth-rates model.

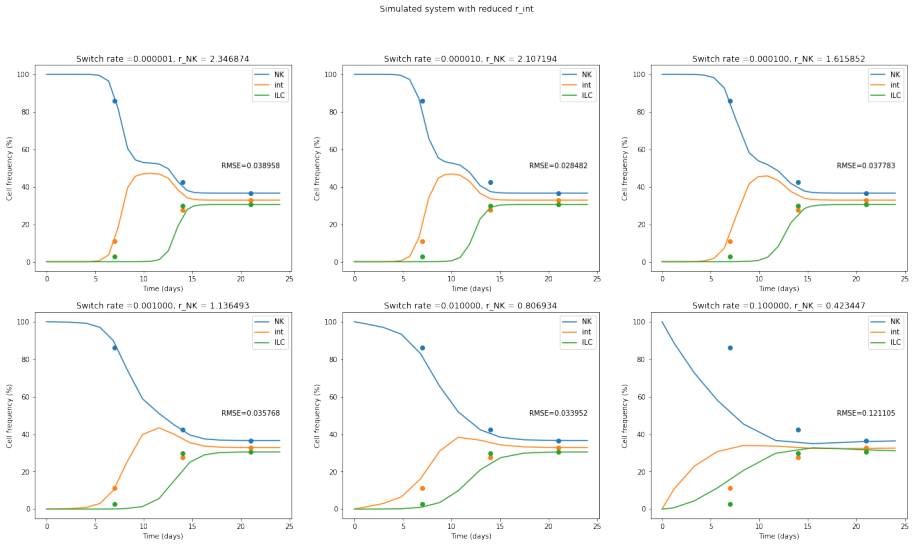

Figure 8: Best-fit simulated model and raw data for various switch rates under varied growth-rates assumption.

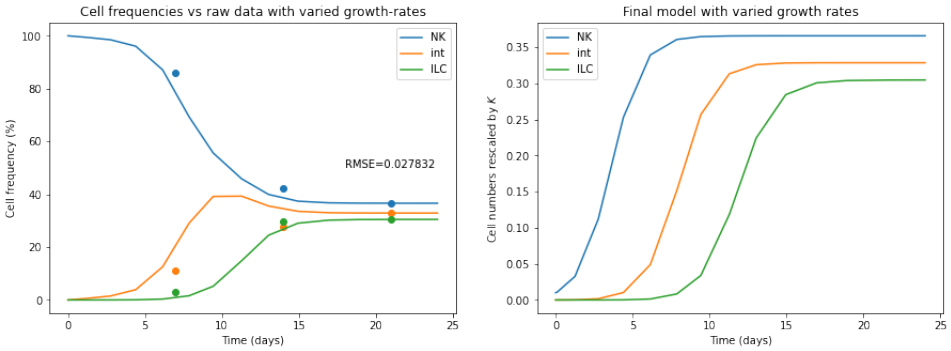

Figure 9: Chosen varied growth-rates model: Predicted cell frequencies vs. raw data (left) and predicted re-scaled cell numbers (right).

4 Fitting Methods and Parameter Choice

In this section we describe in more detail the fitting method used in Section 3 and explore an alternative approach that could be employed should more data become available via numerical optimization. These methods were tried out for simulated data generated by the varied growth-rates model fit in the previous section, and the parameter values obtained from each approach compared.

4.1 Fitting Method Used: Enforced Equilibrium at $t = 21$

Here, we describe in more detail the fitting method used throughout Section 3.
The strategy goes as follows:

1. Assume 1 and 2 as in Section 3.2, and that system is in equilibrium at some time T. (Here, $T = 21$.)
2. Fix switch-rates[2] $(s_{\text{NK,int}},\ s_{\text{int,ILC}})$.
3. Fix growth-rates $(r_{\text{NK}},\ r_{\text{int}},\ r_{\text{ILC}})$.
4. Set the ODE system (9) to 0 at time T and solve the resulting system of linear equations to determine the remaining parameters $(c_{\text{self}},\ c_{\text{int,ILC}},\ c_{\text{ILC,int}})$ for the given switch- and growth-rates.
5. Simulate the ODE system given by (9) using Runge-Kutta of Order 4. For this we used the `integrate.solve_ivp` function from the SciPy Python library (see relevant documentation [13]).
6. Calculate the root-mean-square error[3] between the predicted cell frequencies and the raw data, namely:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j\in\mathcal{X}}\left(f_j(t_i) - \tilde{f}_j(t_i)\right)^2}{m|\mathcal{X}|}}$$

   where $t_1, \ldots, t_m$ are the time-points for which data is available, $\tilde{f}_j(t_i)$ is the measured frequency of cells with phenotype j at time-point t_i,

[2]The choice of fixing switch-rates and then growth-rates rather than in the reverse order was not for any particular reason.

[3]Of course, minimizing the RMSE is the same as minimizing the sum of square errors, since the number of data points is constant, and taking square-roots is a monotone transformation. RMSE simply has the advantage that the error has the same unit as the modelled data (i.e. cell frequencies), and is therefore more straightforward to interpret.

and $f_j(t_i)$ is the frequency of cells with phenotype j at time-point t_i in the simulated model, i.e.:

$$f_j(t_i) = \frac{\bar{n}_j^K(t_i)}{\sum_{x\in\mathcal{X}} \bar{n}_x^K(t_i)}$$

where $(\bar{n}_j^K(t))_{j\in\mathcal{X}}$ denotes the Runge-Kutta simulation of the ODE system given by (9).

7. Vary the growth-rates and repeat steps 4-6, recording the parameters and calculated RMSE at each iteration.

8. Choose the parameter set corresponding to the minimal squared error as the best candidate model for the given switch-rates.

9. Vary the switch-rates and repeat steps 2-8.

10. Choose the model which best fits the raw data by:

    - visually inspecting the fit; and
    - comparing RMSE values.

This procedure was followed with the simplifying assumptions described in Sections 3.2 and then 3.4, so that the switch-rates and growth-rates depended only on the single parameters s and r respectively. Following an iterative process to determine regions which may contain the optimal values for s and r, 10,000 iterations (100 s, and 100 r values) were carried out with s varying between 0.005 and 0.015 and r varying between 0.4 and 0.6. This gave 100 candidate models (one for each value of s considered). In both the equal growth-rate and the varied growth-rates case, candidate models with the lowest RMSE were considered and the best fit was chosen.

4.2 Alternative Fitting Method: Numerical Optimization

The fitting method used in Section 3 places significantly more weight on the last data point ($t = 21$) which entirely determines the parameters c_{self}, $c_{\text{int,ILC}}$, and $c_{\text{ILC,int}}$) up to the choice of growth- and switch-rates. The drawback of this can be seen in the poorer fit at earlier time-points compared to the last time-point.
Moreover, the fitting method employed is somewhat ad-hoc in nature, and computationally inefficient. A better approach may be to use some optimization algorithm which can identify the optimal parameters more efficiently, and generalises better to new data, so that the analysis does not need to be repeated manually with each new data-set that is obtained. These methods would also not require a unique solution from setting the LLN to 0 at $t = T$, meaning that some of the simplifying assumptions in Section 3.2 could be

dropped.
In this section we discuss an alternative approach, based on numerical optimization. In the following, we drop the simplifying assumptions described in Section 3.2 unless otherwise stated.

4.2.1 Optimization Problem

From an optimization perspective, the fitting problem is as follows:

- Let $\theta := ((r_j)_{j\in\mathcal{X}},\ (c_{i,j})_{i,j\in\mathcal{X}},\ s_{\mathrm{NK,int}},\ s_{\mathrm{int,ILC}})$ be a parameter vector in $\mathbb{R}^{14}_{+}$.
- Let f be the function such that the ODE (9) from the L.L.N of a switch-model with parameters θ is given by:

$$\begin{aligned}\dot{\mathbf{n}}^K(t) &= f\left(\mathbf{n}^K(t);\theta\right)\\ \mathbf{n}^K(0) &= \mathbf{n}_0^K\end{aligned} \tag{16}$$

- Let $t_1 < \cdots < t_m$ be time-points for which (absolute cell-number) data is available, and let $(n_j^{K,\mathrm{meas}}(t_i))_{i=1,j\in\mathcal{X}}^m$ denote the re-scaled measurements at these times.
- Let $\tilde{n}_j^K(t;\theta)$ denote the simulated solution of (16) at time t in component $j \in \mathcal{X}$

Define the objective function $J(\theta)$ by:

$$J(\theta) = \sum_{i=1}^{m}\sum_{j\in\mathcal{X}}\left(\tilde{n}_j^K(t_i;\theta) - n_j^{K,\mathrm{meas}}(t_i)\right)^2 \tag{17}$$

Then the optimal parameter θ is given by:

$$\tilde{\theta} = \mathrm{argmin}_{\theta\in\mathbb{R}^{14}_{+}} J(\theta)$$

Algorithms: Since we require that the parameters be positive, this is a constrained optimization problem. For this we can employ any of the bound-constrained and constrained minimization methods of the
`scipy.optimize.minimize` Python function from the SciPy library. (See [13] and accompanying references.)

4.2.2 Simulating Absolute Cell-Numbers Data

At first, the optimization algorithms used in this section could not successfully be employed with the available data due to a combination of the following factors:

1. The available data is not in terms of absolute cell numbers.

2. There is an insufficient number of available data-points.

Since the data is in terms of cell frequencies rather than cell numbers, we attempted to minimize the adjusted objective function:

$$\tilde{J}(\theta) = \sum_{i=1}^{N} \sum_{j \in \mathcal{X}} \left(\frac{\tilde{n}_j^K(t_i;\theta)}{\sum_{l \in \mathcal{X}} \tilde{n}_l^K(t_i;\theta)} - F_j^{\text{meas}}(t_i) \right)^2 \tag{18}$$

where $F_j^{\text{meas}}(t_i)$ denotes the measured frequency of phenotype $j \in \mathcal{X}$ at the time-point t_i, and $\tilde{n}_j^K(t_i;\theta)$ is the value of the simulated system (16) for phenotype $j \in \mathcal{X}$ at time t_i, as before. Attempting to minimize this loss-function using the Python library mentioned above was not possible. Often, the algorithm would choose a search direction after which the ODE system given by (16) would explode, meaning the RK-4 method did not produce simulated values at times $t = 7, 14, 21$ which are necessary to evaluate the loss function. More precise reasons for this (for example whether this was due to the small number of data-points, or whether the adjusted objective function (18) has in some sense "worse numerics" than the usual loss-function for absolute cell numbers data (17)) were not thoroughly investigated.
In order to be able to fit parameters using a numerical optimization approach, we used the varied growth-rates model fitted in Section 3 to simulate cell-numbers data that are consistent with the actual cell-frequencies data available. This was achieved by the following procedure:

1. The ODE system (16) was simulated using Runge-Kutta of order 4 with parameter vector θ corresponding to the chosen varied growth-rates model from Section 3 at a desired number m of fixed time-points $t_0, \ldots, t_{m-1}$.

2. Random noise was added to the data in order to make the data seem more realistic, and so that the optimal parameters returned are not simply the parameters used to generate the data. We used centred normal additive noise with constant variance given by $\sigma^2 = (\text{RMSE} \times \bar{N})^2$ where $\bar{N}$ is the average re-scaled cell-population size over all times simulated, i.e. $\bar{N} = \frac{1}{m} \sum_{i=0}^{m-1} \sum_{j \in \mathcal{X}} \tilde{n}_j^K(t_i;\theta)$. (This was seen as a reasonable guess for the variance, since the RMSE can be interpreted as a measure of how far the predicted cell frequencies varied on average from the actual cell frequencies, and $\tilde{N}$ is the average re-scaled population size. Thus the product of these can be seen as a rough estimate of the standard deviation in terms of re-scaled cell numbers.) Where this resulted in negative cell-numbers (at smaller times, where the int-ILC1 and ILC1 cell populations are small), negative values were replaced by 0, so that only non-negative cell-populations were allowed. This gives initial simulated data $(\tilde{D}_j(t_i))_{i=1,j \in \mathcal{X}}^m$.

3. For times t_{i_l} at which actual cell frequency data is available, the simulated data for the phenotype $j \in \mathcal{X}$ was replaced with $F_j^{\text{meas}}(t_{i_l}) \times \sum_{j \in \mathcal{X}} \tilde{D}_j(t_{i_l})$, so that the frequencies in the simulated data agree with the actually frequencies data wherever this data is available.

The above process was followed with $m = 22$, $t_i = i$ for $i = 0, \ldots 21$, and the available cell frequencies data.

4.2.3 Parameter fitting via numerical optimization on simulated data

A bound-constrained optimization was performed on the simulated data using limited-memory BFGS (`L-BFGS-B`) [14], truncated Newton (`TNC`) [15], constrained optimization by linear approximation (`COBYLA`) [16], and trust-region for constrained optimization (`trust-constr`) [17] methods, with an initial parameter guess chosen as the parameters for the varied growth-rates model, rounded to the nearest whole number, explicitly:

$$\theta_0 = (1, 1, 1, 3, 0, 0, 0, 3, 0, 0, 1, 3, 0, 0)$$

(Other methods were excluded due to long[4] run-times.) The truncated Newton and L-BGFS-B methods employed require the gradient of the objective function to be approximated at each step. For this we used both the built-in 2-step finite differences and 3-step finite differences gradient approximation approaches. The optimization results are summarized in Table 6.

Method	Gradient Approx	Success	RMSE sim-data	RMSE actual-data[5]
TNC	2-step f.d.	False	0.089	0.126
TNC	3-step f.d.	True	0.024	0.0561
L-BFGS-B	2-step f.d.	False	0.023	0.0398
L-BFGS-B	3-step f.d.	False	0.032	0.035
COBYLA	Not applicable	True	0.034	0.072
trust-constr	Not applicable	True	0.213	0.421

Table 6: Comparison of optimization results using different optimization and gradient approximation methods.

The algorithm only exited successfully for the truncated Newton method with 3-step finite-differences, the COBYLA method, and the trust region method for constrained optimization. However, while it is of course preferable for the algorithms to terminate successfully, the purpose of the minimization algorithm is to minimize the objective function. It is therefore more

[4] that is, seemingly infinite

[5] RMSE to Simulated vs. Actual data are not directly comparable, since simulated data are cell numbers re-scaled by K, whereas actual data are cell frequencies

relevant that the L-BFGS-B method with 2-step finite differences gradient approximation achieved the lowest objective function value before terminating. (The fact that the algorithm did not converge[6] however suggests that better parameter values may - and probably do - exist.) Therefore we take the parameters obtained via the L-BFGS-B algorithm with 2-step finite differences gradient approximation as our optimal parameter values. We compare the model with these parameters to the parameters obtained in Section 3.

4.2.4 Comparison to parameters via enforced equilibrium method

Table 7 shows the parameters corresponding to the two models fitted by the enforced equilibrium method and by the L-BFGS-B constrained minimization algorithm respectively. Perhaps unsurprisingly (both the simulated data and the initial parameter guess are based on the parameters in the left-hand column) the parameter values fit by each method are relatively close to one another. However, as discussed above, the numerical optimization approach allows us to drop some of the simplifying assumptions we imposed in Section 3.2 in order to reduce the size of the parameter space. For example, we do not impose any specific relationship between the growth-rates, and do not require that NK cells do not compete with cells of other phenotypes. Thus the parameters fit via the numerical optimization are appear more "varied" than in the varied growth-rates model from Section 3. More precisely, the parameter vector chosen by the L-BFGS-B exists outside the parameter space considered in Section 3. In theory, considering a larger parameter space could allow us to achieve a lower objective function value.

Nevertheless, inspecting the RMSE values of each fitted model compared

Parameter	Enforced equilibrium method	Numerical optimization
r_{NK}	1.000	1.107
r_{int}	0.898	0.969
r_{ILC}	1.000	1.131
C	$\begin{pmatrix} 2.7171 & 0 & 0 \\ 0 & 2.7171 & 0.0168 \\ 0 & 0.5377 & 2.7171 \end{pmatrix}$	$\begin{pmatrix} 2.9715 & 0.0089 & 0.0060 \\ 0.0118 & 3.0129 & 0.0087 \\ 0 & 0.9521 & 2.9643 \end{pmatrix}$
$s_{\text{NK,int}}$	0.00506	0.00262
$s_{\text{int,ILC}}$	0.00506	0.00492

Table 7: Comparison of varied growth-rates parameters to the parameters obtained via L-BFGS-B algorithm with 2-step finite-differences gradient approximation performed on simulated data.

to the simulated cell numbers data, and the actual cell frequencies data, we

[6]The precise message received says that the linesearch terminated abnormally.

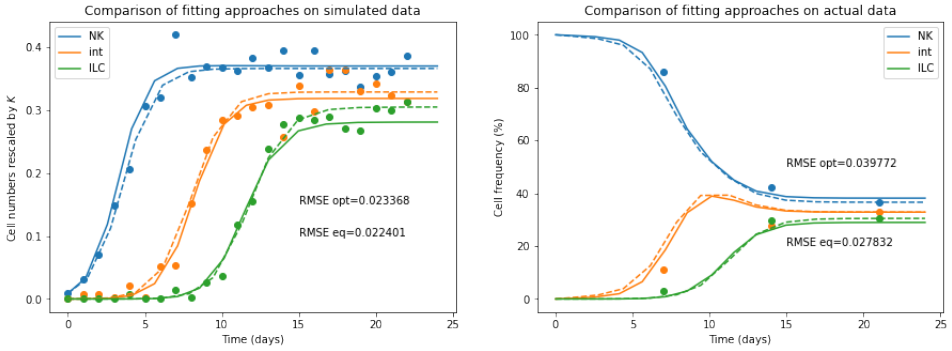

Figure 10: Comparison of model fitted by enforced equilibrium method (dashed) and L-BFGS-B optimization algorithm (solid). The left-hand graph shows the fits to simulated cell numbers data, whereas the right-hand graph compares the fits to actual cell frequencies data.

see that the numerical optimization approach did not yield better parameter values than the parameters from the varied growth-rates model. Again, this is likely due to the fact that the varied growth-rates model was used to generate the simulated cell-numbers data. Therefore, the varied growth-rates model parameters are already a very good fit to the simulated data. (Indeed, the only sources of error are the random noise added to the model, and the adjustment of the simulated data to agree with the actual cell frequencies data at times where such data are available.) Moreover, for time-points where actual cell frequency data are not available, the simulated data inherits any bias and error that exists in the varied growth-rates model, and the objective function used penalises any deviation from this simulated data. Therefore, a significantly improved fit to the actual cell frequency data was unlikely. The fits obtained by each approach are displayed in Figure 10.

5 Outlook and Discussion

5.1 Data Quality

As mentioned in Section 1.3, the available data suffers from two major shortcomings, namely:

- Only cell-frequency rather than cell-numbers data are available; and
- Data is only available at 3 time-points.

When absolute cell-numbers data with more timepoints become available, it would be of interest to repeat the analysis carried out in Section 3 and to try the fitting methods proposed in Section 4.2 on real (rather than simulated) data. This would allow stronger conclusions to be drawn about the appropriateness of the phenotypic plasticity model for modelling NK-cell dynamics.
The Python code to perform this analysis is available with this thesis.

5.2 Role of TGF-ß

As mentioned in Section 1, the cytokine TGF-ß is thought to have a significant role in driving the phenotypic plasticity among group 1 innate lymphoid cells. Experimental evidence suggests that int-ILC1 and ILC1 populations are more concentrated in the presence of constitutively active TGF-ß whereas environments with with ablated TGF-ß display higher concentrations of NK-cells (recall Figure 2). However, due to lack of data availability, no model has been fit which takes TGF-ß concentration as an explicit input. Therefore, it would be of interest to fit parameters to data-sets with different concentrations of TGF-ß, and propose how the parameters may depend on TGF-ß concentration. Some possible dynamics are considered below.

5.2.1 Induced switch-rates by TGF-ß

We take as our trait-space $\mathcal{X} := \{\mathrm{TGF}, \mathrm{NK}, \mathrm{int}, \mathrm{ILC}\}$ where of course TGF represents the population of TGF-ß molecules in the system. Although TGF-ß is itself a cytokine (a protein with a role in cell signalling) it is convenient to model it as part of the overall cell population (following the example of [4]).
We then define birth-, natural death-rates, natural switch-rates, and a competition kernel as before. Since TGF-ß is not a cell, we may assume $c_{\mathrm{TGF,i}} = 0$ for each $i \in \mathcal{X}$, and even $b_{\mathrm{TGF}} = 0$, since the mechanism of cell-division does not apply[7].

[7] In [4] cytokines are secreted along-side a birth event of CD8 T-cells. However, while NK cells and ILC1s are known to secrete other cytokines (cf. [18]), they are not associated with TGF-ß secretion. Therefore, we would need to consult biologists to determine the

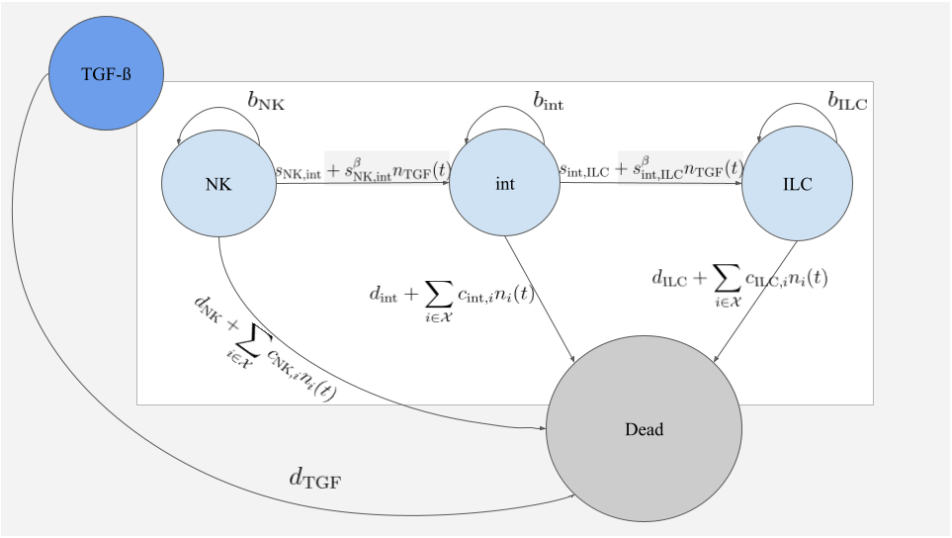

Figure 11: Transition diagram for a system now with switches induced by TGF-ß as well as natural switches. (The white background is used to show the model considered in earlier sections.)

In addition to natural switches from NK cells to int-ILC1s and from int-ILC1s, we also allow switches that are induced by TGF-ß. We do this by introducing a TGF-ß-induced switching kernel, in the sense of Section 2.3. We assume $s_h^{ind.}(j,i) = 0$ whenever $h \neq \mathrm{TGF}$ or $(j,i) \notin \{(\mathrm{NK},\mathrm{int}),(\mathrm{int},\mathrm{ILC})\}$ and denote $s_{\mathrm{TGF}}^{ind.}(j,i) =: s_{j,i}^{\beta} > 0$ for the remaining switches.
In total, the Markov jump process used to model the system is described by the transition diagram in Figure 11.

5.2.2 Growth rates depending on TGF-ß

An explanation for the increased prevalence of int-ILC1s and ILC1s in environments with higher concentrations of TGF-ß could be that TGF-ß concentration in some way decreases the growth-rate of NK cells and increases the growth-rate of int-ILC1s and ILC1s. To model this we could take the trait space $\mathcal{X}$ to be as in the induced switch-rates model above. Then we could assume, for example, strong competition felt by NK cells from TGF-ß molecule, i.e. $c_{\mathrm{NK,TGF}} > 0$, and some "induced birth" rate b_j^{β} for int-ILC1s and ILC1s, representing the additional proliferation of these cells caused by additional TGF-ß molecules. Explicitly, this would give rates for int-ILC1

most realistic way to model the development of TGF-ß in such a population, but for now we assume a fixed number of molecules at the beginning which degenerates at some "death" rate d_{TGF}.

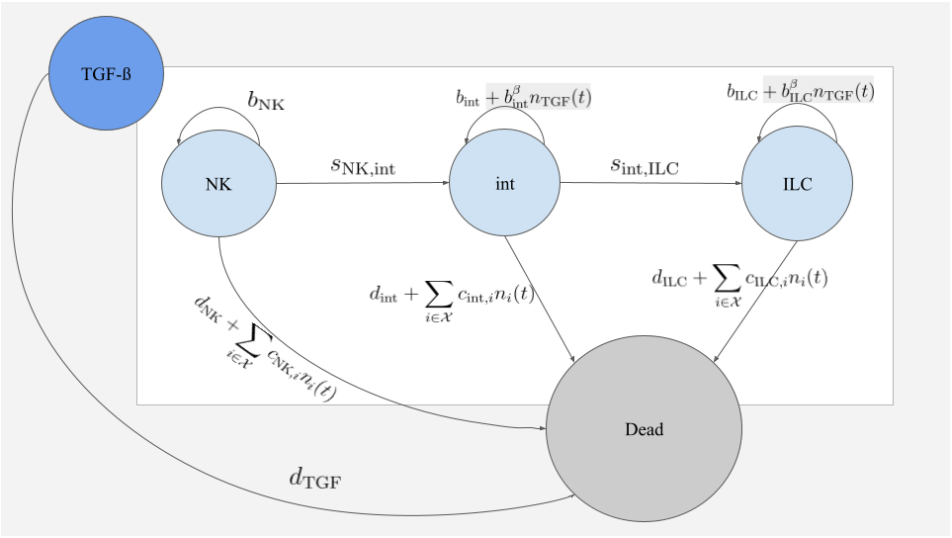

Figure 12: Transition diagram for system where growth-rates depend linearly on amount of TGF-ß in the system.

and ILC1 births by:

$$\beta_{\delta_j}(n) = n_j(b_j + b_j^{\beta} n_{\text{TGF}}), \ j \in \{\text{int}, \text{ILC}\}$$

. The other rates are as in the original model constructed in Section 2.1, with $c_{\text{TGF},i} = 0$ for each $i \in \mathcal{X}$, $b_{\text{TGF}} = 0$, and $c_{i,\text{TGF}} = 0$ for $i \neq \text{NK}$. The dynamics under this model are described by the transition diagram in Figure 12. The deterministic system given by the Law of Large Numbers for this process would be the same as for a system considering *negative* competition parameters for $c_{\text{int,TGF}}$ and $c_{\text{ILC,TGF}}$ respectively[8].

5.3 Implementation into Full Tumour Model

As mentioned at the beginning, the motivation for studying group 1 innate lymphoid cell plasticity is due to the role of NK cells in controlling tumour initiation and metastasis.
The model proposed in this paper could be implemented into a full tumour model, which takes into account the superior anti-cancer properties of NK cells compared to int-ILC1s and ILC1s. Here, we illustrate what such a model could look like.
We consider a system involving the three phenotypes of group 1 innate

[8]Of course, we could also directly implement a model with negative competition parameters, meaning that TGF-ß concentration is assumed to decrease the death-rate of int-ILC1s and ILC1s, rather than increasing the birth-rate. This would not affect the deterministic system, however could cause a problem of negative transition rates in the stochastic model for sufficiently high concentrations of TGF-ß.

lymphoid cells considered before, TGF-ß molecules (TGF), and tumour cells (tum). This gives a trait space of $\mathcal{X} := \{\text{tum}, \text{TGF}, \text{NK}, \text{int}, \text{ILC}\}$. Building on a model with TGF-ß-induced switch-rates (Figure 11), say, we can construct a Markov jump process as in Section 2 with the following rate function β given by:

$$q(n, n+l) = \begin{cases} n_j b_j & l = \delta_j,\ j \in \mathcal{X} \\ n_j \left(d_j + \sum_{i \in \mathcal{X}} c_{j,i} n_i\right) & l = -\delta_j,\ j \in \mathcal{X} \\ n_j (s_{j,i} + s^{\beta}_{j,i} n_{\text{TGF}}) & l = -\delta_j + \delta_i,\ j, i \in \mathcal{X} \\ -\sum_l (\text{above rates}) & l = 0 \end{cases} \tag{19}$$
$$=: \beta_l(n)$$

Here, again, we impose that $s_{j,i} = s^{\beta}_{j,i} = 0$ for all $(j,i) \notin \{(\text{NK}, \text{int}), (\text{int}, \text{ILC})\}$. We also suppose that TGF-ß does not compete with other cell-types, i.e. $c_{\text{TGF},i} = c_{i,\text{TGF}} = 0$ for all $i \in \mathcal{X}$, and that NK cells, int-ILC1s, and ILC1s do not feel competition from tumour cells, i.e. $c_{i,\text{tum}} = 0$ for all $i \neq \text{tum}$. Lastly, to emphasise that the dynamics experienced by tumour cells from NK cells, int-ILC1s, and ILC1s are the cells being killed rather than competition, we denote $c_{\text{tum},i} = k_i$ for $i \in \{\text{NK}, \text{int}, \text{ILC}\}$. Mathematically this does not have an impact, except that we may expect k_{NK} to be significantly larger that the other competition parameters, since these cells are actively killing tumour cells. (Based on the experimental evidence in [1] we may not expect similarly large values for k_{int} or k_{ILC}, and these parameters may even be set to 0 if that is deemed the most biologically realistic.)
The dynamics of such a model are summarised by the transition diagram in Figure 13.

5.4 Conclusions

Individual-based stochastic models with phenotypic switches are a potentially useful tool for analysing the behaviour of group 1 innate lymphoid cells and describing phenotypic plasticity among this cell group. Using the deterministic limit of these processes given by their Law of Large Numbers, we were able to fit a model with explains the limited available data relatively well. These models fit the data better when growth-rates of int-ILC1s and NK cells are in proportion to each phenotypes equilibrium population size, rather than assuming that all growth-rates are equal. We can consider the models here as a "proof of concept", and believe that collection of data which are more tailored to the dynamics and relationships we would like to describe would be worthwhile. We also described and provide Python code for fitting approaches that could be utilised should experimental data be provided in the future. Lastly, we have described how such models could be extended to take into account the influence of the cytokine TGF-ß, and even

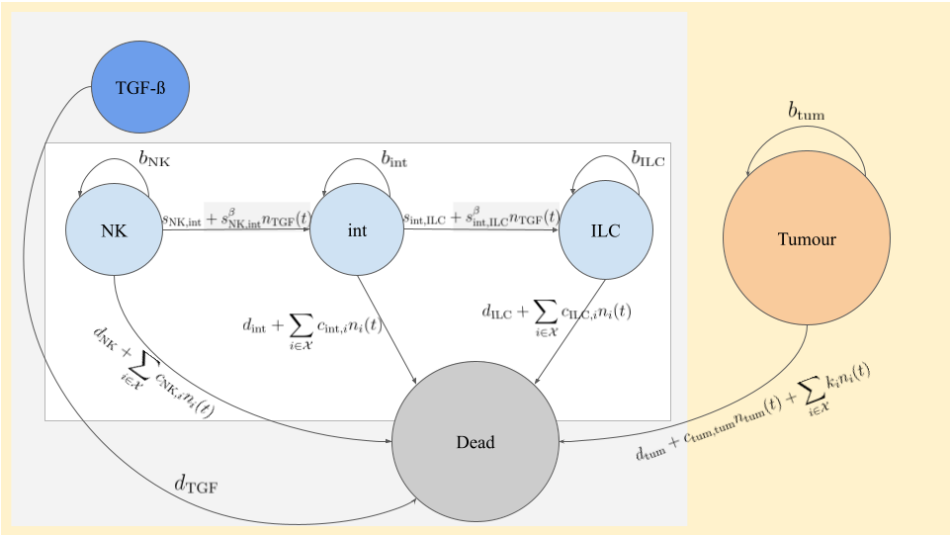

Figure 13: Illustration of transition diagram for a full tumour model. Parts with a white background correspond to the model considered for the majority of this thesis, in particular Sections 3 and 4. Grey background corresponds to the extension to a model with TGF-ß-induced switches discusses in Section 5.2.1. The orange background is the full tumour model.

to interact with tumour cells, thus bringing us closer to our ultimate goal of simulating and predicting how cancer outcomes for patients may change if some model inputs or parameters could be varied - for example, the impact of reducing the amount of TGF-ß active in the system.

References

[1] Y. Gao, F. Souza-Fonseca-Guimaraes, T. Bald, and et al, "Tumor immunoevasion by the conversion of effector NK cells into type 1 innate lymphoid cells," *Nature Immunology*, vol. 18, no. 9, pp. 1004–1015, Jul. 2017. [Online]. Available: https://doi.org/10.1038/ni.3800

[2] N. Champagnat, R. Ferrière, and S. Méléard, "From individual stochastic processes to macroscopic models in adaptive evolution," *Stoch. Models*, vol. 24, no. sup1, pp. 2–44, Nov. 2008.

[3] M. Baar and A. Bovier, "The polymorphic evolution sequence for populations with phenotypic plasticity," *ELECTRONIC JOURNAL OF PROBABILITY*, vol. 23, pp. 1–, 07 2018.

[4] N. Glodde, A. Kraut, D. van den Boorn-Konijnenberg, S. Vadder, F. Kreten, J. L. Schmid-Burgk, P. Aymans, K. Echelmeyer, M. Rumpf, J. Landsberg, T. Bald, T. Tüting, A. Bovier, and M. Hölzel, "Experimental and stochastic models of melanoma t-cell therapy define impact of subclone fitness on selection of antigen loss variants," *bioRxiv*, 2019. [Online]. Available: https://www.biorxiv.org/content/early/2019/11/30/860023

[5] C. Guillerey, N. D. Huntington, and M. J. Smyth, "Targeting natural killer cells in cancer immunotherapy," *Nature Immunology*, vol. 17, no. 9, pp. 1025–1036, Aug. 2016. [Online]. Available: https://doi.org/10.1038/ni.3518

[6] V. S. Cortez, L. Cervantes-Barragan, M. L. Robinette, J. K. Bando, Y. Wang, T. L. Geiger, S. Gilfillan, A. Fuchs, E. Vivier, J. C. Sun, M. Cella, and M. Colonna, "Transforming growth factor- signaling guides the differentiation of innate lymphoid cells in salivary glands," *Immunity*, vol. 44, no. 5, pp. 1127–1139, May 2016. [Online]. Available: https://doi.org/10.1016/j.immuni.2016.03.007

[7] D. B. Keskin, D. S. J. Allan, B. Rybalov, M. M. Andzelm, J. N. H. Stern, H. D. Kopcow, L. A. Koopman, and J. L. Strominger, "TGF promotes conversion of CD16 sup/sup peripheral blood NK cells into CD16 sup-/sup NK cells with similarities to decidual NK cells," *Proceedings of the National Academy of Sciences*, vol. 104, no. 9, pp. 3378–3383, Feb. 2007. [Online]. Available: https://doi.org/10.1073/pnas.0611098104

[8] S. N. Ethier and T. G. Kurtz, *Markov Processes*, ser. Probability & Mathematical Statistics S., S. N. Ethier and T. G. Kurtz, Eds. Nashville, TN: John Wiley & Sons, May 1986.

[9] T. Liggett, "Continuous time markov processes: An introduction," in *Graduate Studies in Mathematics*, ser. Graduate studies in mathematics. Providence, Rhode Island: American Mathematical Society, Mar. 2010.

[10] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, no. 4, pp. 403–434, Dec. 1976. [Online]. Available: https://doi.org/10.1016/0021-9991(76)90041-3

[11] D. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *Journal of Chemical Physics*, vol. 115, pp. 1716–1733, 07 2001.

[12] L. Li, "Mathematical Modelling of NK-cell plasticity," Master's thesis, Universität Bonn, Germany, 2022.

[13] scipy.org, *SciPy User Guide v1.9.1 (stable)*, 2022. [Online]. Available: https://docs.scipy.org/doc/scipy/tutorial/index.html

[14] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal of Scientific Computing*, vol. 16, pp. 1190–1208, Sep. 1995.

[15] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed., ser. Springer Series in Operations Research and Financial Enginee. Springer Science+Business Media, Jan. 2006.

[16] M. J. D. Powell, "A direct search optimization method that models the objective and constraint functions by linear interpolation," in *Advances in Optimization and Numerical Analysis*. Springer Netherlands, 1994, pp. 51–67. [Online]. Available: https://doi.org/10.1007/978-94-015-8330-5$_4$

[17] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM Journal on Optimization*, vol. 9, no. 4, pp. 877–900, Jan. 1999. [Online]. Available: https://doi.org/10.1137/s1052623497325107

[18] S. K. Panda and M. Colonna, "Innate lymphoid cells in mucosal immunity," *Frontiers in Immunology*, vol. 10, May 2019. [Online]. Available: https://doi.org/10.3389/fimmu.2019.00861