

© Copyright 2026

Obinna Ukogu

Design Principles of the T Cell Immune Response

Obinna Ukogu

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee:

Armita Nourmohammad, Chair

Ivana Bozic

Eric Shea-Brown

Program Authorized to Offer Degree:

Applied Mathematics

University of Washington

Abstract

Design Principles of the T Cell Immune Response

Obinna Ukogu

Chair of the Supervisory Committee:

Armita Nourmohammad

Department of Physics

The immune system protects its host by detecting and neutralizing evolving pathogens and malignant cells. To prevent self-damage while ensuring host survival, this response must be both proportionate and rapid, requiring the massive expansion of a small pool of specialized cells. In jawed vertebrates, T cells are the central architects of this balance: they kill infected or malignant cells, coordinate B cell activity, and maintain self-tolerance. Despite recent progress in T cell immunology, the design principles that guide T cell responses, and their limitations, remain unclear.

Here, we present three mathematical models—incorporating ecological dynamics, stochastic processes, and optimization theory—to study T cell responses across three distinct contexts: acute infection, cancer, and the complex microbial ecology of the gut. By synthesizing these frameworks, we demonstrate how the mechanistic design (e.g., gene regulation and repertoire allocation) of the T cell response influences functional objectives, such as threat clearance, memory formation and avoidance of self-damage. Crucially, we identify trade-offs between objectives and show how T cell responses can be re-engineered to navigate them. Our results provide a theoretical basis for the rational design of immunotherapies, vaccines, and microbial therapies by pinpointing high-leverage mechanistic targets to enhance or redirect T cell responses.

TABLE OF CONTENTS

	Page
List of Figures	iii
List of Tables	v
Chapter 1: Introduction	1
1.1 Scope of this dissertation	1
1.2 Previous publication of dissertation content.	5
1.3 A brief primer on the T cell response.	6
1.4 Modeling the T cell response	10
1.4.1 Ecology	11
1.4.2 Stochasticity	12
1.4.3 Optimality	14
Chapter 2: Modeling T cell memory decision-making strategies	16
2.1 Introduction	16
2.2 Results	18
2.2.1 Flexible decision model.	18
2.2.2 Early decision model	27
2.2.3 Late decision model (obligate reversible decision model)	28
2.3 Discussion	30
Chapter 3: Design principles of the cytotoxic T cell response	31
3.1 Introduction	31
3.2 Model	34
3.3 Results	42
3.4 Discussion	54

3.5	Materials and Methods	56
Chapter 4:	Rational T cell repertoire allocation to control the gut microbiota	59
4.1	Introduction	59
4.2	Model	61
4.3	Results	63
4.4	Discussion	79
Chapter 5:	Conclusion	81
5.1	Main contributions of this dissertation	81
5.1.1	Commensurate memory formation.	82
5.1.2	The control logic of the T cell response.	82
5.1.3	Immune control of the gut microbiota	83
5.2	Future research directions.	85
5.2.1	Modeling circuit failures in peripheral tolerance	85
5.2.2	Spatial models of immune-mediated tissue homeostasis	85
Appendix A:	Supplement for Chapter 3	87
A.1	Model of T cell response to infections	87
A.2	Numerical implementation of stochastic T cell response	101
A.3	Optimization of CD8 ⁺ T cell response to pathogens	109
Appendix B:	Supplement for Chapter 4	128
B.1	Data analysis	128
B.2	Models of optimal repertoire allocation.	131
B.3	Estimating cluster harmfulness	152
Bibliography		186

LIST OF FIGURES

Figure Number		Page
1.1	The many facets of the immune system	2
1.2	The origin and life of a T cell	7
2.1	The stereotypical CD8 ⁺ T cell response.	17
2.2	Flexible decision model allows memory to scale with infection severity.	19
2.3	Approximating memory magnitude during rapid pathogen replication	23
3.1	CD8 ⁺ T cell response to infections.	34
3.2	Design parameters shape T cell response and infection clearance	39
3.3	Toxicity-infection harm trade-off across T cell regulatory designs	41
3.4	Pareto-optimal designs balancing harm from acute infections and autoimmunity	44
3.5	Elbow-optimal designs recapitulating macro-dynamics of T cell expansion	48
3.6	Engineering immune response for cancer immunotherapy	50
4.1	A model of adaptive immune control of the gut microbiota.	60
4.2	T cell repertoire reactivity and coverage in a complex microbial community	64
4.3	Repertoire coverage predicts compositional shifts of strain clusters	69
4.4	Model of repertoire allocation explains strain clustering.	70
4.5	Microbial ecology can modulate the precision of immune control	76
A.1	Schematic of modeled early T cell division bursts and differentiation.	115
A.2	Design parameters shape T cell response and infection clearance	116
A.3	Near-optimal performance and designs for each immune challenge	117
A.4	Dependence of harm on baseline rates in different archetypes	118
A.5	Dependence of harm on signal sensitivities in different archetypes	119

A.6	Pareto-optimal design balancing harm from autoimmunity and other pathogen archetypes	120
A.7	Dependence of T cell macro-dynamics on baseline rates	121
A.8	Dependence of T cell macro-dynamics on signal sensitivities	122
A.9	Impact of cross-reactivity on engineered immune responses for cancer immunotherapy	123
A.10	Parameter sensitivity analysis for total harm	124
A.11	Parameter sensitivity analysis for effector fold expansion	125
A.12	Sensitivity of harm magnitude and T cell expansion to stochastic cell-fate decisions	126
B.1	T cell repertoire composition	154
B.2	Composition of microbiota clusters.	155
B.3	Robust relationship between microbiota composition and repertoire coverage	156
B.4	Properties of the optimal repertoire	157
B.5	Intra-cluster trends in strain compositional changes	157

LIST OF TABLES

Table Number		Page
2.1	Variables and initial conditions for the flexible decision model.	21
2.2	Parameters for the flexible decision model.	21
3.1	Endogenous design of CD8 ⁺ T cell response.	47
A.1	Model parameters for simulations	127

ACKNOWLEDGMENTS

Nigerian poet Niyi Osundare once said, “May your road be rough! That is my wish for you. Because only rough roads lead to heights worth climbing.” Regarding a PhD, he was spot on. The last few years have been equal parts trying and rewarding. But the journey has been worth it and I owe much of what has been good to the following people.

I would like to express my deep gratitude to my advisor, Armita Nourmohammad, for her support, guidance, and engagement with my work over the years. Every aspect of this dissertation was improved by her feedback and encouragement. Thank you to my final exam committee members—Ivana Bozic, Eric Shea-Brown, Hao Yuan Kueh, William DeWitt and Alison Feder—for overseeing my dissertation and PhD defense. My warmest thanks to my many lab-mates over the years: Zachary Montague, Michael Pun, Gian Marco Visani, Assya Trofimov, Kevin Borisiak, Ella Carlander, Maya Gatt Harari, Darin Momayezi, Manali Sawant, Rhea Grover, Quinn Bellamy and Jakub Otwinowski. Thank you for your thoughtful feedback and discussions about my research.

I’m grateful to the Department of Applied Mathematics for providing me with a lively community. Thank you Bernard Deconinck, Tony Garcia, Katherine McDermott, Ingrid Richter and Sarah Riley for keeping the department running, and allowing me to focus on research. Thank you to the friends and teachers I had in the Department: Hong Qian, Katherine Owens, Ying-Jen Yang, Nora Gilbertson, Erin Angelini, Biraj Pandey, Natalie Wellen, Damien Beecroft, Ziyu Lu, Michael Kupperman, Ruibo Zhang and Ike Griss Salas.

Thank you to my undergraduate advisors and mentors at Amherst College—Ashley Carter, Tanya Leise, Amanda Folsom, Rob and Danielle Benedetto—for giving me the confidence to pursue a PhD. Thank you Ashley for introducing me to the joys of research. Thank you Danielle for convincing me that I could be a mathematician. Thank you Tanya for showing me how to be an applied mathematician. The training I received from you all serves me to this day.

Thank you to the friends who supported me and took care of me as I worked on my PhD: Leonard Yoon, John Michael, Kevin Deptula, Cristian Ramirez, Anand Hemmady, Rui Wang, Eli Lilleskov, Colin Jacobs, Jonathan Che, Andie Jones, Motasem ElGamel, Andrew Pyo, Rohan Rele, Manish Kumar, Will Cummins, Kim Bourne, Todd Obilor, Konso Mbakire and Chong Xia. Thank you Georgia Seltzer for always being a kind, supportive, and patient partner. You helped me get through the hardest moments and enjoy the best ones.

Most importantly, thank you to my family. Even though we've been far apart these last few years, you always make me feel loved and supported. Thank you to my uncles, aunts, cousins and grandparents for their steadfast material and emotional support over the years. Thank you Mom and Dad for believing in me and always allowing me to pursue my passions. Thank you to my siblings—Chinelo, Ugo, Chiazam and Dumebi—for being constant sources of joy and respite in my life. I couldn't do this without you.

DEDICATION

To my grandparents:

Gregory and Theresa Okafor, and Pius and Cecilia Ukogu

Chapter 1

INTRODUCTION

1.1 Scope of this dissertation

In vertebrates, the immune system is a collection of organs and cells that surveil its host for threats to healthy homeostasis and act to remove them (Fig. 1.1). It is insurance in a dangerous world where a host is exposed to myriad external (e.g., viruses) and internal (e.g., cancer) threats. Physical barriers like the skin and mucosal surfaces in airways or gastrointestinal tracts are a first line of defense against external threats. Additionally, tissue cells have built in mechanisms to limit the mutations, uncontrolled divisions and longevity that lead to cancer. When these fail, the immune system steps in to clear the threat and keep the host alive. As evidenced by the fact that many of us have recovered from multiple illnesses, the immune system works. It is this remarkable efficacy that inspired this dissertation.

The fundamental agents of the vertebrate immune system are specialized cells that recognize molecular patterns associated with a threat and act to remove the detected threat [2]. Traditionally, these cells are delineated into two arms: an innate and an adaptive system [3]. First, innate immune cells—such as neutrophils, dendritic cells and natural killer cells—recognize a largely fixed repertoire of evolutionarily stable/conserved threat-associated patterns that are encoded in the host’s genome. Its response to a potential threat is typically fast-acting (i.e., hours) and is the first wave of immune defense. We can think

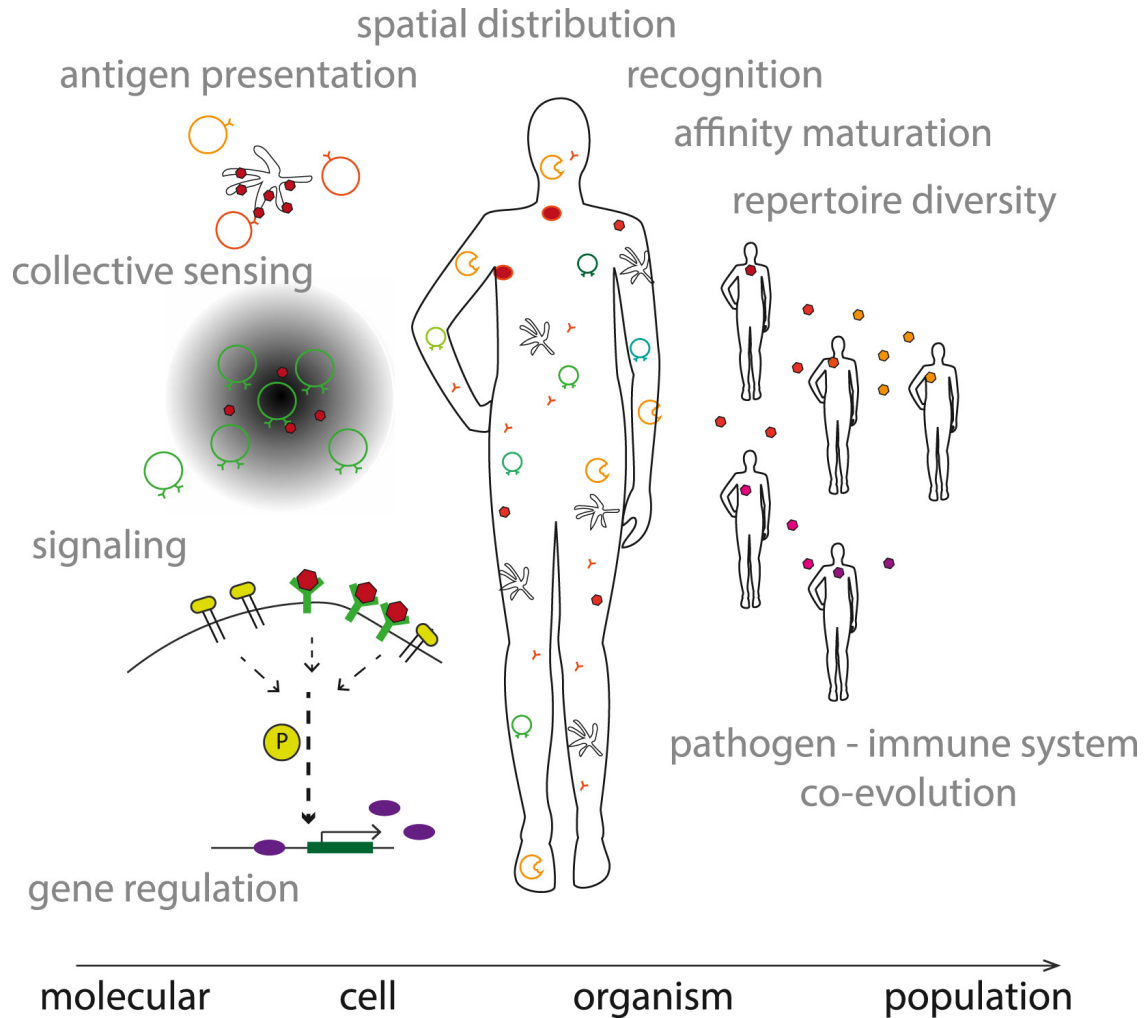


Figure 1.1: The many facets of the immune system. The immune system protects the body against threats by integrating molecular signals to influence gene regulation and marshal a multi-cellular response tailored to the threat. The immune response also adapts to its threat environment on multiple scales. Over generations, the innate immune response adapts in a population to better match its pathogen environment. In the adaptive immune system, repertoire diversity, memory formation and affinity maturation rapidly co-evolve within an individual's lifetime to keep up with threats. Reproduced from [1] (CC BY).

of it as traditional insurance for predictable risks: high probability, routine threats. For instance, it reflects the likelihood of encountering a historically prevalent strain of bacteria in the natural environment. Like insurance, innate immunity has costs. It is inherited and so occupies space in the genome. Maintaining a specialized set of cells requires energy, and these cells sometimes malfunction by attacking healthy host tissues, causing auto-inflammatory

diseases. More quantitatively, we can think of the innate immune system as a Bayesian actor with a strong prior over the distribution of possible patterns that require a response. “Trust, but verify.”

Second, the adaptive immune system consists of B and T cells, which through processes of stochastic genetic recombination and mutation, are able to recognize a much broader set of molecular patterns than the innate immune system [4]. However, because of this randomization, there is no guarantee that a pattern recognized by a B or T cell corresponds to a threat the host is likely to encounter [5]. Fortunately, cells of the adaptive immune system have an incredible capacity to learn, forming memories of which patterns are likely harmful (mostly foreign) and which are benign (mostly self, food or commensal microbes) over a lifetime. Consequently, compared to the innate response, the adaptive response is slow (i.e., days) and more specific. As insurance, adaptive immunity is protection from unpredictable risks, such as previously unseen pathogens and evolved variants of previously encountered threats. In terms of cost, the adaptive immune repertoire (i.e., the collection of recognized patterns) is generally not inherited [6], and the genes responsible for generating the random repertoire take up less genomic space than the innate immune system [3]. However, maintaining these many cells requires energy, and autoimmune diseases caused by a malfunctioning adaptive response are orders of magnitude more prevalent than auto-inflammatory (innate immune) diseases: 1 in 10 of the global population versus 1 in 100,000 [7–9]. In the Bayesian framework, the adaptive immune system is an actor with an initially weak prior on the distribution of possible patterns which it iteratively updates: “Live and learn.”

B and T cells play distinct roles in the adaptive immune system. B cells (humoral immunity) generate and secrete Y-shaped receptors/antibodies which can recognize complex chemical/structural features of threats. Antibodies bind to viruses, bacteria and other threatening agents and prevent them from replicating or mark them for degradation by other parts of the immune system. Additionally, antibodies can also undergo a process of Dar-

winian selection called affinity maturation, wherein they are iteratively mutated and tested to improve their binding to a desired target. On the other hand, T cells (cellular immunity) display receptors on their surface that can only bind to short peptide sequences presented on the surface of other cells. These short peptide sequences reflect the internal state of the displaying cell since they are generated by chopping up proteins produced within that cell. Consequently, an unpredictable risk can be managed, as the adaptive immune system tries to cover the space of possible simple patterns (T cells) and generate antibodies that can recognize arbitrary complex patterns (B cells). Furthermore, combining the capabilities of the innate and adaptive immune systems, the immune system accomplishes a few major goals: (1) stop threats from growing; (2) if possible, remove threats; and (3) remember threats to improve future. These facts make it clear that the immune system learns molecular signatures of pathogens “associatively” over time [10, 11].

The focus of this dissertation is understanding how the T cell response works and how it fails in the varied contexts of acute infection, memory and the complex microbial ecology of the gut. These are our so-called “design principles.” The rest of this chapter dives into the biology of the T cell response and briefly surveys the mathematical ideas used in this work. In the following chapters, we ask: (1) How can T cells make memory commensurate with the severity of an infection (Chapter 2)? (2) How should T cells rationally use the many signals generated during an infection to best protect the host (Chapter 3)? (3) How does the T cell repertoire control the composition of microbes that colonize the gut to maintain health (Chapter 4)? These questions are multifaceted and have been studied experimentally. Here, we developed simple and intuitive models grounded in the idea that the T cell response is trying to “do something”: to achieve some function(s) [12]. In particular, we relate the dynamic and statistical properties of the T cell response to the optimization of objectives: memory formation, pathogen clearance/management and avoiding self-damage/immunopathology. Normative approaches like this have a long tradition in

mathematical biology, particularly in computational neuroscience and biophysics. They are appealing because mathematical optimization can reduce the degrees of freedom inherent in many biological systems by identifying a few optimal configurations in parameter space. However, this approach is not perfect: one is always at risk of telling quaint, “just-so” stories. Responsible application requires consistent dialogue with other quantitative approaches, such as mechanistic modeling, and experimental validation. We have threaded this needle by working closely with experimentalists—Dr. Hao Yuan Kueh, Dr. Grégoire Altan-Bonnet and Dr. Angelina Bilate—to “liv[e] dangerously by turning our thinking into formal mathematical predictions and confronting that math with experiments that have not yet been done” [13] (Chapter 5).

1.2 *Previous publication of dissertation content*

This dissertation incorporates ideas, text, and materials from the following previously published manuscripts, with contributions from all authors:

- O. A. Ukogu, A. Bilate, A. M. Walczak, T. Mora, A. Nourmohammad, “Rational T cell repertoire allocation to control the gut microbiota,” *in preparation*.
- O. Ukogu, Z. Montague, G. Altan-Bonnet, A. Nourmohammad, “Design principles of the cytotoxic CD8+ T-cell response,” PNAS (2026), pnas.2528202123.
- K. Abadie, . . . , O. Ukogu et al., “Reversible, tunable epigenetic silencing of TCF1 generates flexibility in the T cell memory decision,” Immunity (2024), j.immuni.2023.12.006.
- R. Zhang, O. Ukogu, I. Bozic, “Waiting times in a branching process model of colorectal cancer initiation,” Theoretical Population Biology (2023), j.tpb.2023.04.001.

1.3 *A brief primer on the T cell response*

In this work, we focus on T cells, which have the ability to rapidly proliferate and form memory in response to a perceived threat [14, 15]. In adult humans, the collection of T cells (i.e., the repertoire) in the body numbers roughly 10^{12} cells. The repertoire consists of many millions ($10^8 - 10^9$) of distinct (i.e., genetically distinct), “clonal” sub-populations, each consisting of $10 - 10^4$ cells [16, 17]. T cell precursors are generated in the bone marrow through the progressive differentiation of stem-like progenitor cells during a process called hematopoiesis [18] (Fig. 1.2A). Hematopoiesis also generates blood, innate immune and B cell precursors. Afterwards, T cell precursors migrate to an organ called the thymus where they begin to express their surface receptors (TCRs) and gain the ability to recognize short amino acid sequences bound to major histocompatibility complexes (MHCs) presented on the surface of other cells (Fig. 1.2B).

The T cell receptor (TCR) protein consists of two variable regions (α and β chains). Each T cell has multiple versions of three gene segments (V, D and J genes) that are essentially concatenated—then transcribed and translated—to build the TCR: V-J for α and V-D-J for β chain. In the thymus, T cell precursors undergo V(D)J recombination; during which, versions/alleles of V, D and J segments are randomly combined together, and random insertion and deletion mutations are made at the connecting junctions between the gene segments. V(D)J recombination thus fixes the TCR sequence of a given cell while generating many different TCR sequences. Sub-populations of T cells with identical TCRs (called “clones”) emerge from this process. Estimates of the theoretical diversity of generated sequences suggest over 10^{15} possible TCR receptors, compared to the $10^8 - 10^9$ clones that are typically present in humans [16, 19]. Notably, two-thirds of the TCR sequences generated are not functional [3], and V(D)J recombination is not entirely random so that certain sequences have relatively high probabilities of generation [20, 21]. Thus, even though differ-

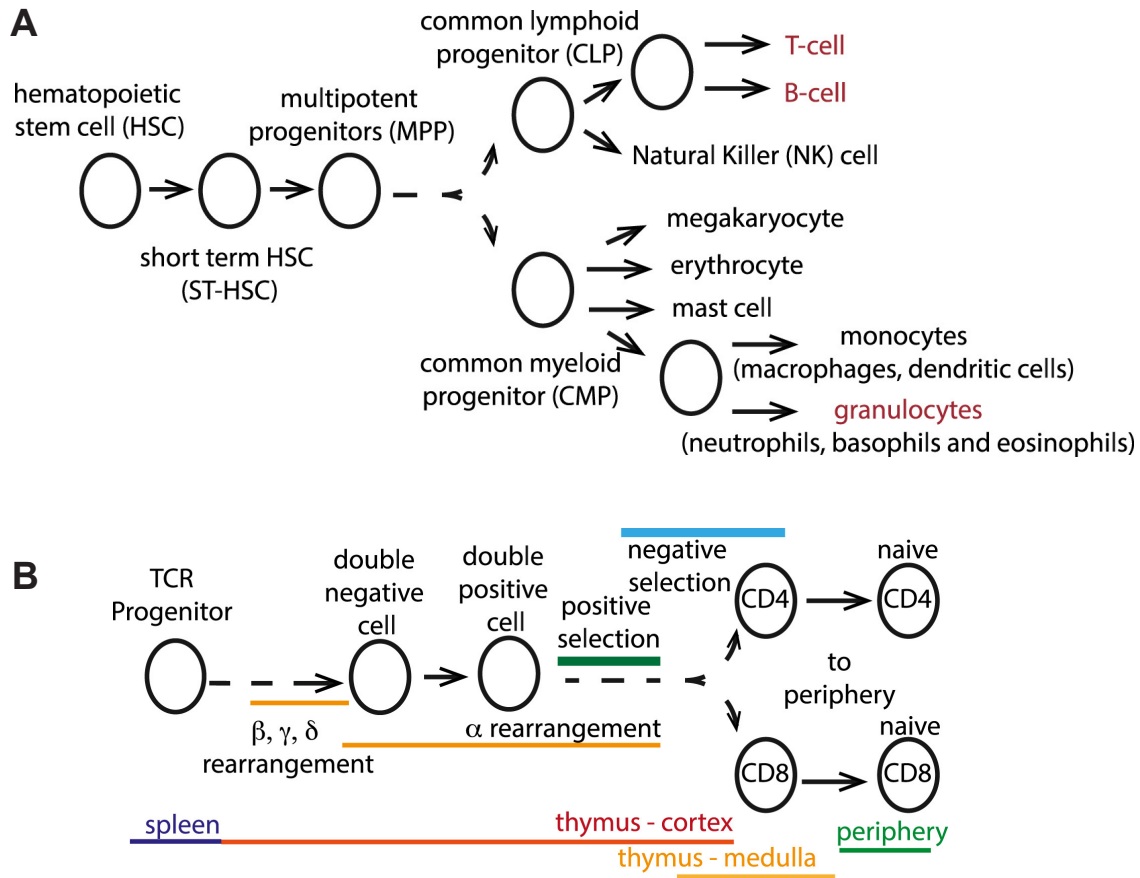


Figure 1.2: The origin and life of a T cell. (A) In the bone marrow, hematopoiesis generates T cells along with many other cell types. (B) In the thymus, V(D)J recombination and central tolerance generate a repertoire of differentiated T cells with diversified functions and pattern recognition capabilities. The repertoire is screened to minimize the risk of self-reactivity, however, many self-reactive cells still enter peripheral tissues and must be calibrated there (peripheral tolerance). Reproduced from [1] (CC BY).

ent individuals have largely different repertoires, some “public” TCR sequences are shared. Finally, a major consequence of these stochastic processes is that the host does not yet know which particular T cells are likely to be efficacious against threats or likely to recognize and attack self-tissues.

After V(D)J recombination, a T cell receptor’s efficacy and safety is screened in the thymus over a few days. First, during positive selection, T cells are systematically exposed to self-peptides bound to major histocompatibility complexes (MHCs). T cells that do not bind sufficiently strongly to self-peptide-MHC undergo apoptosis, removing them from the

repertoire. Approximately 5% of functional T cells make it through positive selection [22]. Next, during negative selection, T cells that bind strongly to self-peptides are selectively killed, removing them from the repertoire. However, negative selection is leaky, and many potentially self-reactive T cells still make it into the repertoire. Finally, based on their affinity for one of two MHC variants (class I and II) T cells differentiate into either cytotoxic $CD8^+$ T cells or helper $CD4^+$, so called because of their expression of the surface protein markers $CD4^+/CD8^+$ [23] (Fig. 1.2B). $CD8^+$ T cells (the focus of Chapters 2 and 3) have the ability to kill cells expressing a recognized pattern/epitope. On the other hand, $CD4^+$ T cells (the focus of Chapter 4) act as signal relays and orchestrators, helping B cells, innate immune cells and other T cells respond appropriately to stimuli.

From the thymus, these “naïve” or inexperienced T cell clones enter the periphery: blood, tissues and lymphatic vessels. Naïve T cells spend most of their time in specialized vessels called secondary lymphoid organs (SLOs)—such as the spleen, tonsils, and lymph nodes—where they sample the antigens present in the associated tissues [24, 25]. This process of circulation and exposure establishes peripheral self-tolerance, wherein T cell clones are further filtered and trained to be nonreactive to self-antigens encountered in the absence of danger signals. During an infection at a tissue site, antigen presenting cells (APCs) and other immune cells secrete cytokines which recruit T cells to the site(s) of infection, where they are systematically exposed to the self and foreign peptides [26–28]. Typically, of the many T cells in the repertoire, only a small subset will be responsive to a particular antigen: about 1 in a million [17]. Thus, the immune response must quickly solve a “needle in a haystack problem” and identify adequate responders [17, 24, 29].

During a first encounter with a pathogen (“primary infection”), a naïve T cell is activated to respond only if it binds strongly to a cognate peptide on an APC [30]. T cells integrate the intensity and duration of TCR signaling through biochemical pathways that produce unstable intermediates in proportion to stimulation [31–34]. Activation proceeds after a

threshold level of these intermediates has been accumulated. Only then is a complex program of cellular proliferation and ligand secretion initiated [1]. An activated T cell responds to a threat by rapidly dividing to greatly expand its numbers. For instance, a T cell clone that initially consisted of a few hundred cells can expand a thousand-fold in a few days [15, 34, 35]. The realized extent of proliferation is determined by the nature of the threat and regulated by the innate immune response [36].

After activation, responding T cells differentiate into phenotypically distinct subtypes with specialized functions [37–44]. $CD8^+$ T cells differentiate into an effector or a memory fate. Effector $CD8^+$ T cells travel to the sites of infection and display enhanced proliferation capacity, anabolic metabolism for building macromolecules, and secretion cytolytic molecules (cytokines) that can degrade targeted cells [45]. Comparatively, memory $CD8^+$ cells are not similarly enhanced but pair the longevity of naïve cells (months to years, versus days to weeks) with lower thresholds of activation and/or residence in the affected tissues. These qualities—longevity, sensitivity and proximity—constitute the potency of T cell memory responses during reinfection.

Similarly, activated $CD4^+$ T cells differentiate into effector/conventional, memory or an additional regulatory fate [46, 47]. Effector $CD4^+$ cells further specialize in two non-exclusive ways:

1. Lineage specification: T helper 1 (Th1), Th2 or Th17 subsets characterized by distinct cytokine profiles,
2. Functional localization: T follicular helper (Tfh) cells, which migrate to B cell follicles to direct antibody production, or non-follicular effectors that direct $CD8^+$ T and innate cells.

$CD4^+$ effectors are complemented by regulatory $CD4^+$ T cells (Treg) that suppress the activities of other T cell types to maintain self-tolerance and homeostasis [48]. Notably, while

most other $CD4^+$ effector fates differentiate exclusively in the periphery, Treg fates can be determined either in the thymus or induced from effector T cells in peripheral tissues [47]. Finally, after pathogen clearance, a $CD4^+$ T cell of any fate may become a long-lived memory cell.

However, T cell fates are not set in stone. Recent studies have shown that cell-fate decisions are reversible, so that effector cells may differentiate into a memory-like state (“effector memory”) [49, 50]. Eventually, the immune response ends (typically within a few days to weeks) with most activated cells dying or entering an exhausted state, while a small fraction become memory [14]. This allows the immune system to avoid clonal deletion of its (best) responders after an infection clears. The memory cells thus formed are long-lived and can respond (i.e., initiate proliferation) more rapidly to subsequent (“secondary”) infections [51]. Unfortunately, the formation of persistent immune memory has also been linked to the development of allergies and autoimmune diseases [52].

We conclude this section by noting that although the processes we have described are regulated and stereotyped to a degree, they are shaped by the chemical signaling and pathogenic context in which they take place. For instance, recent studies have posited that T cell differentiation is regulated by integrated signals despite having intrinsic rates [35, 36, 53–55]. Thus, the T cell response is malleable and its functions can be effectively perturbed.

1.4 Modeling the T cell response

Quantitative immunology is a decades-old and rapidly-growing field, with sub-disciplines tracking the development of new experimental techniques [1, 56–58]. Data-driven modeling approaches in immunology draw on ideas from dynamical systems, information theory, statistical physics, probability theory, population genetics and optimization theory. Here, we describe our main quantitative approaches for modeling the T cell response.

1.4.1 Ecology

Dynamical systems models readily describe the population dynamics resulting from interactions between agents in a closed system [59]. These models successfully described numerous aspects of the T cell response, and every project in this dissertation adopts this approach in some way [60–66]. For instance, a simple predator-prey model of the T cell response (T) to a growing pathogen (v) can be written as:

$$\begin{aligned}\frac{dv}{dt} &= (\gamma_v - \delta_v T)v, \\ \frac{dT}{dt} &= (\gamma_T v - \delta_T)T,\end{aligned}\tag{1.1}$$

where γ_v is the growth rate of the pathogen, δ_v is the rate at which T cells remove/kill the pathogen, γ_T is the rate at which T cell interactions with the pathogen lead to growth, and δ_T is the death rate of the T cells. Following the standard solution of a predator-prey model, the populations of the virus and T cells will alternatively rise and fall, never going extinct [59]. This model departs from the biological reality of the T cell response in key ways. First, most pathogens do not grow indefinitely in the host as they quickly exhaust the population of susceptible tissues or otherwise kill the host, precluding a T cell response. In other words, growth saturates and there is a carrying capacity. For instance, we could instead set $\gamma_v(v) \propto (1 - v/v_{\max})$. Second, although transient oscillations are sometimes encountered in the T cell response, a successful immune response removes the pathogen or arrests its growth [67]. For instance, we could instead set $\gamma_v(v) \propto \Theta(v - v_{\min})$, where $\Theta(x)$ is the Heaviside function. We could also raise similar objections about T cell growth and killing dynamics. The point is that starting from a simple archetypal model, we can make modifications/extensions and eventually build up to something more realistic.

One such extension is to expand the model to more viruses or antigens, $v_a, a = 1, \dots, n_a$

and more T cell clones, $T_i, i = 1, \dots, n_T$:

$$\begin{aligned}\frac{dv_a}{dt} &= \left(\gamma_v - \delta_v \sum_i c_{a,i} T_i \right) v_a, \\ \frac{dT_i}{dt} &= \left(\gamma_T \sum_a c_{a,i} v_i - \delta_T \right) T_i,\end{aligned}\tag{1.2}$$

where the new parameters $c_{a,i} \geq 0$ quantify the interaction strength between virus/antigen a and T cell clone i . More generally, we can include additional agents, such as susceptible and infected cell populations, different types of T cells and components of the innate immune system—enriching model relevance but complicating analysis (see Chapter 2). To study the resulting high-dimensional and non-linear models, one can often analyze the steady-state solutions by setting $\dot{v}_a = 0$, $\dot{T}_i = 0$. However, the biological phenomena we are often most interested in are the transient dynamics of the system. To capture these, we can readily turn to computational/simulation methods to investigate transient behaviors. Simulation methods allow us to fine-tune the interactions between agents to more closely match true biology and efficiently search model parameter space [68].

1.4.2 Stochasticity

As we have seen, the T cell response is stochastic. At the outset, the naïve repertoire is constructed by stochastic developmental processes (VDJ recombination, positive and negative selection, and fate specification) that bias the distribution of naïve cells. This bias in turn determines the likelihood that an individual can mount an effective response to a particular pathogen [20, 21]. Starting from such a repertoire, naïve cells stimulated under similar conditions vary in their extent of proliferation and patterns of differentiation [15, 51, 69]. Sources of this variation include intrinsic molecular noise (gene regulation, protein expression, signal transduction, etc.), variation in local environmental conditions like APC availability and

timing, and TCR heterogeneity in the population of responding cells. Like many biological systems, T cells use chemical and physical signaling to reliably transmit information and control properties of the distribution of responding cells [70–74]. For instance, consider an irreversible transition regulated by a time-dependent signal $\sigma(t)$:

$$\begin{aligned} X &\xrightarrow{b(t)} Y, \\ b(t) &= b_{\max} \times \sigma(t), \end{aligned} \tag{1.3}$$

where the probability density of the transition time is given by

$$p_{X \rightarrow Y}(t) = b(t) \exp \left[- \int_0^t b(s) ds \right]. \tag{1.4}$$

Although the transition is stochastic, its speed and statistics (mean and variance) can still be controlled. This is the case for many biological processes [75–77].

Next, suppose there are different T cells $i = 1, \dots, n_T$ in the repertoire each with a possibly different affinity $c_i > 0$ for the signal $\sigma(t)$. Then the probability density of the transition time for T cell i is

$$p_{X \rightarrow Y}^{(i)}(t) = c_i b(t) \exp \left[- \int_0^t c_i b(s) ds \right]. \tag{1.5}$$

Thus, by controlling the signal magnitude, certain cells can be preferentially induced to transition. If there are many T cells, a reliable response can then emerge despite stochasticity following the law of large numbers [78, 79]. Similarly, multiple transitions can be regulated by signals to control the T cell response. Suppose, we have

$$X \xrightarrow{b(t)} Y \quad \text{and} \quad Y \xrightarrow{d(t)} Z.$$

The rates $b(t)$ and $d(t)$ can be regulated to dynamically control the probability of being in

the intermediate state Y . In this dissertation, we will combine stochastic chemical reaction networks like this with systems of ordinary differential equations to model aspects of the T cells response (see Chapter 3).

1.4.3 Optimality

The T cell response is thought to have undergone a process of evolutionary selection akin to stochastic optimization [2, 4, 80]. Understanding its limitations helps clarify the logic of cellular decision-making and guide therapeutic perturbations for improving efficacy [63, 81, 82]. Although, the T cell response is likely not optimal in many of its functions, it is sometimes useful to pretend it is, or to ask what it might look like (mechanistically) if it were. Concretely, given degrees of freedom $\vec{p}$, one first identifies a plausible objective $F(\vec{p})$ and the admissible configurations of the system, $\vec{p} \in \Omega$, specified by biologically relevant constraints. Solving a problem of the form,

$$\max_{\vec{p} \in \Omega} F(\vec{p}), \quad (1.6)$$

analytically or numerically may then yield one or more optimal solutions/configurations of the system [83]. For instance, the degrees of freedom could be the parameters/rates of a system of ordinary differential equations (ODEs) or chemical reaction network, $\dot{\vec{x}}(\vec{p})$. We can then ask what dynamics/statistics the optimal configurations exhibit, and relate them to the behavior of true system (see Chapter 4). Deviations from optimality may point to missing constraints or objectives and can suggest possibilities for improving system performance [84]. We may account for additional objectives by performing multi-objective optimization (MOO) of functions, $F_k, k = 1, \dots, n_F$:

$$\max_{\vec{p} \in \Omega} \sum_k s_k F_k(\vec{p}), \quad (1.7)$$

where the weights s_k quantify the relative importance of the different objectives [83]. Different weights may then admit different solutions. For instance, while a system may not be optimal in any one function, it may achieve a holistic balance between different objectives.

This approach has been applied successfully in developmental biology, where it explains the high fidelity of embryonic development in fruit flies [76, 85], and in immunology where it has provided insights into the long-term dynamics of immune memory [62, 63, 80, 81]. Here, as we seek design principles of the T cell response, we are faced with another question: “For which objectives is the T cell response optimized?” In general, we do not know and inverse-inferences based on the observed functions of the evolved T cell response are difficult to justify [86]. Thus, we have focused on the easier, though still illuminating, task of selecting a few pertinent functions and evaluating the performance of the T cell response against these heuristic metrics [87–89] (Chapter 3 and 4).

Chapter 2

MODELING T CELL MEMORY DECISION-MAKING STRATEGIES

The content of this chapter was previously published in [49] in collaboration with Dr. Hao Yuan Kueh's lab. For this project, we performed original mathematical analysis of ordinary differential equations (ODE) models of the T cell response to acute infection. In this chapter, we adapt the sections most relevant to our contributions, incorporating simulations performed by Dr. Kueh.

2.1 Introduction

T cells are the watchdogs of the adaptive immune system and humans possess billions of T cells capable of recognizing unique molecular targets called antigens. $CD8^+$ T cells are a subset of T cells that help resolve immune threats by killing infected and malignant cells that present targeted antigens on their cell surface. These T cells ordinarily circulate through the body or reside in tissues in a inactive state. However, priming by the innate immune system activates T cells, prompting them to differentiate into rapidly dividing, short-lived, cytotoxic cells (effectors) or a slower-dividing, long-lived, non-cytotoxic cells (memory) [14]. After expanding to clear the infection, effectors die leaving behind a relatively small number of memory cells.

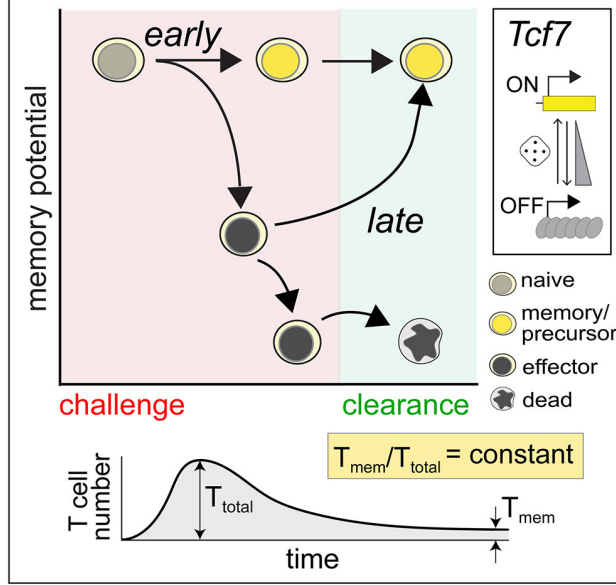


Figure 2.1: The stereotypical $CD8^+$ T cell response. Activation triggers an early but reversible cell-fate decision into memory or effector fates. Abadie et al. (2024) demonstrated that this flexibility is regulated by the epigenetic silencing of the *Tcf7* gene. Effectors (T_{eff}) rapidly divide to clear the threat before dying. Memory cells (T_{mem}) divide more slowly but their numbers are bolstered by a small fraction of effector cells which de-differentiate into memory cells. Experiments indicate that, in the end, a small but consistent fraction of the total responding population ($T_{total} = T_{eff} + T_{mem}$) remains as memory: $T_{mem}/T_{total} = \text{constant}$. Reproduced from [49] (CC BY).

It has long been known that the magnitude of memory formed during a response scaled with the severity of the infection (Fig. 2.1), but the mechanisms behind it were initially unknown [90]. In particular, the timing of memory formation was disputed, with two leading theories positing that memory was formed exclusively either early [91–93] or late [50, 94, 95] in the immune response. Our collaborators in the Kueh lab performed in vitro and in vivo experiments in mice to interrogate these mechanisms. They found that memory differentiation was reversibly regulated by the epigenetic silencing of the *Tcf7* gene in response to stimulatory signals. Silencing *Tcf7* prompts T cells to differentiate into the effector fate early in the immune response and later reversing *Tcf7* silencing prompts effectors to de-differentiate into a memory state. They thus unified the two previous theories into a single model of flexible T cell differentiation.

To rationalize these observations, we considered three mathematical models of CD8^+ T cell memory decision making strategies in response to an acute infection: (1) memory decisions are made in a flexible manner, occurring both early and late (“Flexible decision model”); (2) memory decisions are made early and irreversibly (“Early decision model”) [91–93]; and (3) memory decisions occur late after naive cells, following antigen encounter, differentiate into effector cells (“Late decision model”) [50, 94, 95]. The models account for replication of pathogen, its removal by effector T cells and innate immune cells, as well as the differentiation of naive cells into effector and memory cells following different decision making strategies. For simplicity, the model neglects CD4^+ T and B cells which also contribute to pathogen clearance. In this chapter, we used simulations and mathematical analysis to compare the ability of the different strategies to robustly produce memory that scales with infection severity. We found that the observed flexible decision-making strategy is the only one that satisfies this requirement, suggesting why it may have been selected for during evolution.

2.2 Results

2.2.1 Flexible decision model

We begin by describing the reversible model of T cell memory decision making (Fig. 2.2A), wherein naive cells first transition into memory cell precursors in the presence of pathogens. These memory precursors then transition into effector cells with a probability that increases with pathogen load. Both memory precursors and effector cells proliferate at similar rates that increase with pathogen load, as previously observed [96]; however, only effector cells undergo apoptosis as a result of activation-induced cell death, a reflection of their short-lived nature. In addition to T cells, we also model the innate immune system in both its inactive and active states. Effector T cells and innate cells mediate pathogen killing, both

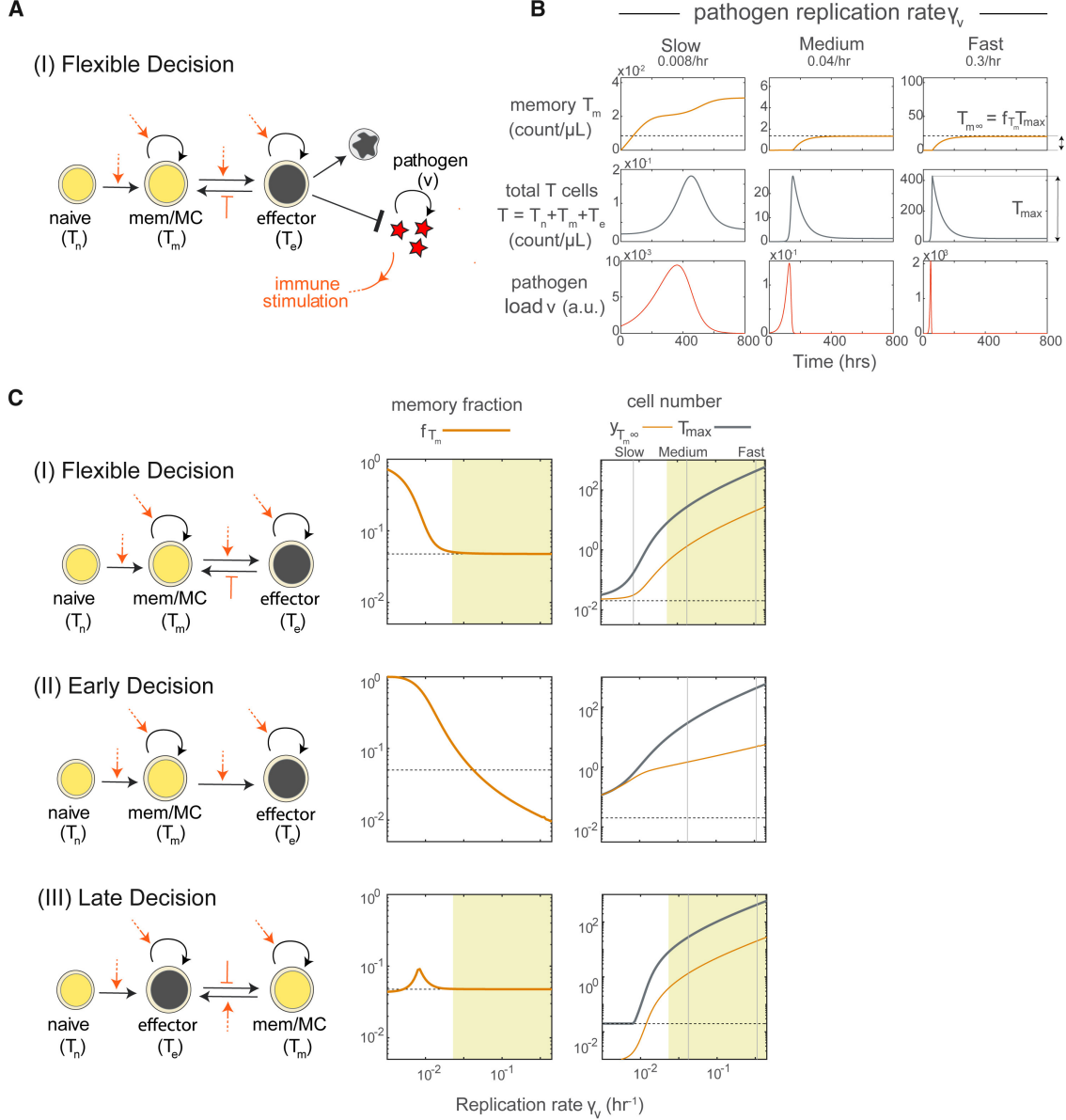


Figure 2.2: Flexible decision model allows memory to scale with infection severity. (A) Schematic of the flexible decision model. Arrows indicate fate transitions/divisions/pathogen clearance (black) and pathogen-load-mediated regulation. (B) Simulations of memory response (top), total T cells (middle) and pathogen load (bottom) over time (hrs) for varying pathogen replication rates (γ_v). Dotted lines (top) indicate memory concentration at $f_{T_m} = 0.05$ of the peak response (middle). (C) Comparison of flexible (top), early (middle) and late (bottom) decision strategies. Left: model schematics as in (A). Middle: memory fraction (f_{T_m}) vs. γ_v . Dotted as in (B). Right: T cell peak (T_{max} ; black) and final memory (y_{T_m} ; orange) vs. γ_v . Dotted lines mark initial concentration of naive cells, and shading marks the scalable memory regime. Solid, vertical lines correspond to the pathogen replication rates in (B). Reproduced from [49] (CC BY).

independently from each other and cooperatively. This model is described by the following equations:

$$\begin{aligned}
\text{naive T cell:} & \quad \frac{dT_n}{dt} = -\alpha_n v T_n \\
\text{memory T cell:} & \quad \frac{dT_m}{dt} = \alpha_n v T_n + \left(\frac{v}{v + K_m}\right) \gamma_m T_m - \left(\frac{v}{v + K_{m,e}}\right) \alpha_{m,e} T_m + \left(\frac{K_{e,m}}{v + K_{e,m}}\right) \beta_{e,m} T_e \\
\text{effector T cell:} & \quad \frac{dT_e}{dt} = \left(\frac{v}{v + K_e}\right) \gamma_e T_e + \left(\frac{v}{v + K_{m,e}}\right) \alpha_{m,e} T_m - \left(\frac{K_{e,m}}{v + K_{e,m}}\right) \beta_{e,m} T_e - \delta_e T_e \\
\text{pathogen:} & \quad \frac{dv}{dt} = \left(\frac{v}{\epsilon + v}\right)^N \gamma_v v - (\delta_{v_1} T_e + \delta_{v_2} z^* + \delta_{v_3} z^* T_e) v \\
\text{inactive innate cell:} & \quad \frac{dz}{dt} = -\alpha_z v z \\
\text{active innate cell:} & \quad \frac{dz^*}{dt} = \alpha_z v z + \left(\frac{v}{v + K_{z^*}}\right) \gamma_{z^*} z^* - \delta_{z^*} z^*
\end{aligned} \tag{2.1}$$

Here, $T_{(\cdot)}$ with subscripts n , m , and e denote naive, memory, and effector T cell types, respectively, v denotes the pathogen population, and z and z^* denote inactive and active innate immune cells, respectively. The rates $\alpha_{x,y}$ denote differentiation rates from cell type x to y cell (for $x, y \in \{n, m, e\}$), $\beta_{e,m}$ denotes the de-differentiation rate from effector to memory (i.e., the only reversed differentiation in the model), γ_x denotes the proliferation rate of cell type x , and δ_x denotes the death rate of cell type x . The parameters $K_{(\cdot)}$ denote the pathogen load required for a half maximal rate of the process indicated in the subscript.

A detailed description of model variables, parameters and initial conditions are given in the two tables below (Tables 2.1, 2.2). Parameters were chosen based on the immune compartment sizes as measured in mice, as well as T cell biological parameters that the Kueh lab and others have measured. The initial conditions were chosen to reflect the initial onset of an infection by a pathogen for which no prior immunological memory has been developed; specifically, antigen-specific naive cells are present at low amounts, effector and memory cells are absent and the initiating pathogen is introduced at a low initial level.

This system is fundamentally a Lotka-Volterra model where immune cells are predators and pathogens are prey (see Chapter 1). However, we have modified this framework to

Table 2.1: Variables and initial conditions for the flexible decision model.

	Variable description	Initial value
T_n	naive T cell	0.02 cells/ μL
T_m	memory T cell	0 cells/ μL
T_e	effector T cell	0 cells/ μL
v	pathogen load	10^{-3} units/ μL
z	inactive innate cell	1 cell/ μL
z^*	activated innate cell	0 cells/ μL

Table 2.2: Parameters for the flexible decision model.

	Parameter description	Value
α_n	naive cell activation rate constant	$5/(\text{units}/\mu\text{L})/\text{hr}$
γ_m	maximal memory cell proliferation rate constant	0.25/hr
K_m	pathogen load for half maximal memory cell proliferation	0.1 units/ μL
$\alpha_{m,e}$	maximal effector differentiation rate constant	0.25/hr
$K_{m,e}$	pathogen load for half maximal effector differentiation	0.1 units/ μL
γ_e	maximal effector cell proliferation rate constant	0.25/hr
K_e	pathogen load for half maximal effector proliferation	0.1 units/ μL
$\beta_{e,m}$	maximal rate of effector de-differentiation	$8 \times 10^{-4}/\text{hr}$
$K_{e,m}$	pathogen load for half maximal effector de-differentiation	2.5×10^{-3} units/ μL
δ_e	rate constant for effector cell death	0.016/hr
γ_v	rate of pathogen replication	0.003-0.45/hr
ϵ	pathogen load for extinction	10^{-4} / μL
N	sharpness of extinction effect for pathogen	100
δ_{v_1}	rate constant for T cell pathogen killing	$4.5 \times 10^{-3}/(\text{cells}/\mu\text{L})/\text{hr}$
δ_{v_2}	rate constant for innate cell pathogen killing	$1.5 \times 10^{-3}/(\text{cells}/\mu\text{L})/\text{hr}$
δ_{v_3}	rate constant for T-cell assisted innate cell pathogen killing	$1.5 \times 10^{-3}/(\text{cells}/\mu\text{L})^{-2}/\text{hr}$
α_z	innate cell activation rate constant	$2.5/(\text{units}/\mu\text{L})/\text{hr}$
γ_{z^*}	maximal activated innate cell proliferation rate	0.02/hr
K_{z^*}	pathogen load for half maximal innate cell proliferation	0.01 units/ μL
δ_{z^*}	turnover rate for activated innate cell	$2 \times 10^{-4}/\text{hr}$

describe the immune response in the following ways: first, we have incorporated saturation terms in the rates of pathogen-induced T cell and innate cell division (K_m , K_e), as well as T cell effector to memory differentiation ($K_{m,e}$) and de-differentiation ($K_{e,m}$). The values for these saturation terms are chosen to reflect the biological upper-bounds for these cellular processes. Second, we incorporate a threshold pathogen load, ϵ , below which the pathogen

replication rate drops to zero, reflecting extinction of the pathogen. This pathogen extinction threshold ensures that this deterministic system of equations, when simulated, has a well-defined behavior and is not subject to numerical integration errors at very low pathogen loads [97]. However, we note that the dynamics of simulated responses do not generally depend on the exact value of the extinction coefficient chosen.

From numerical simulations, we see that the flexible decision model reproduces the canonical dynamics of the adaptive immune response (Fig. 2.2B). Initially, pathogens grow exponentially, giving rise to a subsequent expansion of the T cell numbers from their initial low levels in the naive cell population. This expansion occurs concomitantly with pathogen clearance, and is followed by a decline in T cell numbers to a stable elevated baseline, reflecting the generation of long-lived memory cells that can survive following pathogen clearance. Next, we wish to ascertain whether this system can produce memory cells in numbers that scale linearly with the peak T cell numbers during an infection, as observed experimentally [90]. To this end, we performed simulations of eq. 2.1 with different values of pathogen replication rate, γ_v , as a means to vary pathogen virulence. We found that pathogens with different replication rates gave rise to different degrees of T cell expansion and contraction, with faster-replicating pathogens generating a stronger T cell response, as expected (Fig. 2.2B, middle row).

However, the fraction f_{T_m} of memory cells to the total number of T cells present at the expansion peak approaches a fixed number as the viral replication rate increases (Fig. 2.2B, C: shaded area). Specifically, in the regime where there is substantial T cell expansion ($> 10^2$ fold relative to naive cell numbers), the memory fraction f_{T_m} remains constant for a broad range of viral replication rates γ_v , spanning an order of magnitude (Fig. 2.2C, top row, shaded area). On the other hand, for slowly growing pathogens (small γ_v), the memory fraction f_{T_m} is elevated, with a non-linear inverse dependence on the viral replication rate (Fig. 2.2C); in this regime, effector differentiation is slow so the number of

memory cells depends strongly on the initial number of naive cells present.

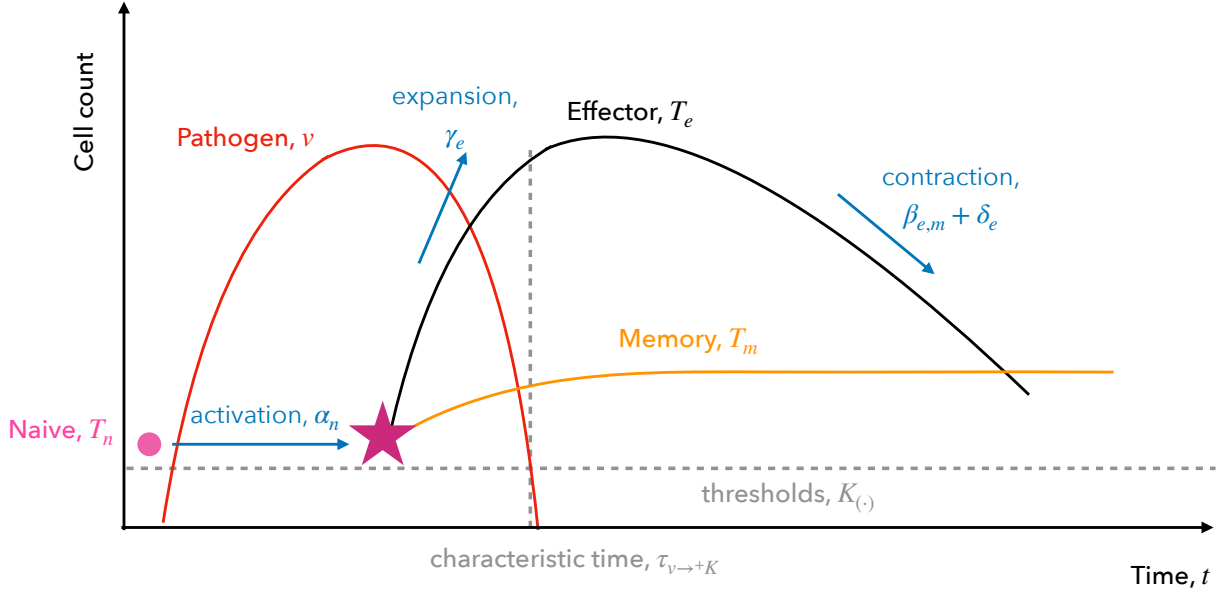


Figure 2.3: Schematic model of T cell dynamics and memory magnitude under rapid pathogen replication. When a pathogen (v , red curve) replicates rapidly relative to the immune response thresholds ($K_{(\cdot)}$, grey horizontal dashed line), the pathogen load significantly exceeds these thresholds by the time naive T cells (T_n , magenta circle) undergo activation (α_n , star). Following activation, effector T cells (T_e , black curve) undergo rapid expansion (γ_e) and some memory cells are formed. The effector response clears the infection, driving the pathogen population down towards the response threshold at a characteristic time $\tau_{v \rightarrow +K}$ (grey vertical dashed line), which aligns closely with the peak of the effector response. After the pathogen is cleared, the effector population undergoes contraction from its peak ($\beta_{e,m} + \delta_e$). This phase contributes to the long-term T cell memory pool (T_m).

To generalize these insights, we considered the dynamics of memory T_m and effector T_e populations, given by eq. 2.1. We begin by identifying the dominant processes in different regimes of accumulated viral load v . Specifically, we compare the viral load with the half maximal loads ($K_{(\cdot)}$'s in eq. 2.1), necessary for different processes. Below the terms that are relevant for the dynamics of memory and effector populations in the high viral load

($v \gg K_m, K_{m,e}$) and the low viral load ($v \ll K_{e,m}$) regimes are indicated:

$$\text{memory T cell: } \frac{dT_m}{dt} = \underbrace{\alpha_n v T_n + \left(\frac{v}{v + K_m}\right) \cdot \gamma_m T_m - \left(\frac{v}{v + K_{m,e}}\right) \cdot \alpha_{m,e} T_m}_{v \gg K_m, K_{m,e}} + \underbrace{\left(\frac{K_{e,m}}{v + K_{e,m}}\right) \cdot \beta_{e,m} T_e}_{v \ll K_{e,m}} \quad (2.2)$$

$$\text{effector T cell: } \frac{dT_e}{dt} = \underbrace{\left(\frac{v}{v + K_e}\right) \cdot \gamma_e T_e + \left(\frac{v}{v + K_{m,e}}\right) \cdot \alpha_{m,e} T_m}_{v \gg K_e, K_{m,e}} - \underbrace{\left(\frac{K_{e,m}}{v + K_{e,m}}\right) \cdot \beta_{e,m} T_e}_{v \ll K_{e,m}} - \delta_e T_e. \quad (2.3)$$

We can formally integrate over the dynamical equation in eq. 2.2 to find a formal solution for the number of memory cells $T_m(t)$ at time t post infection,

$$T_m(t) = T_m(0) \exp \left(\int_0^t \frac{\gamma_m v(s)}{v(s) + K_m} - \frac{\alpha_{m,e} v(s)}{v(s) + K_{m,e}} ds \right) + \left[\int_0^t \exp \left(\int_s^t \frac{\gamma_m v(s)}{v(s) + K_m} - \frac{\alpha_{m,e} v(s)}{v(s) + K_{m,e}} ds \right) \times \left(\alpha_n v(s) T_n(s) + \frac{\beta_{e,m} T_e(s) K_{e,m}}{v(s) + K_{e,m}} \right) ds \right]. \quad (2.4)$$

Because we are interested in a naïve response to an infection, we set the initial condition $T_m(0) = 0$. If the pathogen grows rapidly before being cleared by the immune response such that the typical viral load over the course of the infection is much higher than the rate thresholds $K_m, K_{m,e}, K_e$, and $K_{e,m}$ (Fig. 2.3), the time-dependent rates can be approximated by their maximal rates:

$$T_m(t) \approx \int_0^{\min(t, \tau_{v \rightarrow +K})} \exp [(\gamma_m - \alpha_{m,e})(\min(t, \tau_{v \rightarrow +K}) - s)] \alpha_n v(s) T_n(s) ds + H(t - \tau_{v \rightarrow +K}) \int_{\tau_{v \rightarrow +K}}^t \beta_{e,m} T_e(s) ds, \quad (2.5)$$

where $\tau_{v \rightarrow +K}$ is the time when the pathogen is driven down to $K = \max\{K_m, K_{m,e}, K_e, K_{e,m}\}$ from above, and $H(t - \tau_{v \rightarrow +K})$ is a heaviside step function that takes value 1 for $t > \tau_{v \rightarrow +K}$,

and 0, otherwise. Because pathogen expansion and clearance are exponentially fast, the time spent in the transition windows $v \sim K_{(\cdot)}$ is negligible compared to the total duration of the expansion and contraction phases (Fig. 2.3). Moreover, given parameters used for the model (Table 2.2), we can neglect the effective growth of the memory pool, i.e., $\gamma_m - \alpha_{m,e} \approx 0$. With these assumptions, the size of the memory pool from eq. 2.5 follows,

$$T_m(t) \approx \int_0^{\min(t, \tau_{v \rightarrow +K})} \alpha_n v(s) T_n(s) ds + H(t - \tau_{v \rightarrow +K}) \int_{\tau_{v \rightarrow +K}}^t \beta_{e,m} T_e(s) ds,$$

and the following terms determine the size of the memory pool:

1. $\tau_{v \rightarrow +K}$: time to effectively clear the infection;
2. $\beta_{e,m}$: the transition rate from effector to memory.

Our goal is to estimate the asymptotic (long-term) fraction of memory ($T_m(\infty)$) to the total number of T cells (primarily effector cells) at the expansion peak ($T_e^{\max}$), $f_{T_m} = T_m(\infty)/T_e^{\max}$. The asymptotic amount of memory follows,

$$\lim_{t \rightarrow \infty} T_m(t) \approx \int_0^{\tau_{v \rightarrow +K}} \alpha_n v(s) T_n(s) ds + \int_{\tau_{v \rightarrow +K}}^{\infty} \beta_{e,m} T_e(s) ds.$$

Then assuming the number of effectors peaks as the number of pathogens falls below detectable levels, $\tau_{v \rightarrow +K} \approx \tau_{e^{\max}}$ [60, 98] (Fig. 2.2B, 2.3):

$$\begin{aligned} T_m(\infty) &\approx \int_0^{\tau_{e^{\max}}} \alpha_n v(s) T_n(s) ds + \int_{\tau_{e^{\max}}}^{\infty} \beta_{e,m} T_e(s) ds \\ &\approx \int_0^{\tau_{e^{\max}}} \alpha_n v(s) T_n(s) ds + \int_{\tau_{e^{\max}}}^{\infty} \beta_{e,m} T_e^{\max} e^{-(\beta_{e,m} + \delta_e)s} ds \\ &\approx \int_0^{\tau_{e^{\max}}} \alpha_n v(s) T_n(s) ds + \beta_{e,m} \frac{T_e^{\max}}{\beta_{e,m} + \delta_e} \end{aligned} \tag{2.6}$$

where we used the relation $T_e(t > \tau_{e,\max}) = T_e e^{-(\beta_{e,m} + \delta_e)s}$, indicating an exponential decay of effector cells after the peak of the response ($v \rightarrow^+ K_{e,m}$), from eq. 2.3. From eq. 2.1, we can also formally express the size of the naive pool as,

$$T_n(t) = T_n(0) \exp \left[-\alpha_n \int_0^t v(s) ds \right],$$

Therefore, the first term in the solution of eq. 2.6 follows,

$$\alpha_n \int_0^{\tau_{e,\max}} v(s) T_n(s) ds = T_n(0) \left[1 - e^{-\alpha_n \tilde{V}} \right], \quad \text{with } \tilde{V} = \int_0^{\tau_{e,\max}} v(r) dr. \quad (2.7)$$

Here, $\tilde{V}$ reflects the total amount of pathogens accumulated during the infection. Putting it all together, we find

$$T_m(\infty) \approx T_n(0) \left[1 - e^{-\alpha_n \tilde{V}} \right] + \beta_{e,m} \frac{T_e^{\max}}{\beta_{e,m} + \delta_e}. \quad (2.8)$$

The memory fraction is then,

$$f_{T_m} = \frac{T_m(\infty)}{T_e^{\max}} \approx \frac{T_n(0)}{T_e^{\max}} \left[1 - e^{-\alpha_n \tilde{V}} \right] + \frac{\beta_{e,m}}{\beta_{e,m} + \delta_e}. \quad (2.9)$$

When $T_e^{\max} \gg T_n(0)$, we recover the constant memory fraction $f_{T_m} \approx \frac{\beta_{e,m}}{\beta_{e,m} + \delta_e}$.

So far we have assumed that $\gamma_m - \alpha_{m,e} \approx 0$. When this assumption does not hold, following eq. 2.5, the first term in our expression for f_{T_m} becomes,

$$C_0 = \frac{\alpha_n T_n(0)}{T_e^{\max}} \int_0^{\tau_{e,\max}} \exp \left[(\gamma_m - \alpha_{m,e})(\tau_{e,\max} - s) - \alpha_n \int_0^s v(r) dr \right] v(s) ds \quad (2.10)$$

with a strong dependence on $\gamma_m - \alpha_{m,e}$. Note that the innate immune dynamics does not explicitly determine the memory fraction f_{T_m} , however, it influences the magnitude of $\tilde{V}$ and so it is expected to be important in the low viral replication γ_v regime.

These results demonstrate how a flexible switching strategy for T cell memory generation allows the amount of T cell memory generated to scale with the magnitude of the T cell effector response (and thus the severity of the infection).

2.2.2 *Early decision model*

Previous studies have proposed that memory cells originate from cells that have undergone little or no effector differentiation, and that memory precursors, upon silencing the memory regulator gene *Tcf7* are committed to becoming short-lived effectors [91, 99]. Using the mathematical modeling framework developed above, we evaluated whether this irreversible effector decision strategy enables long-term memory T cell numbers ($T_m(\infty)$) to scale linearly with the peak T cell number ($T_e^{\max}$).

We performed simulations of the above model, rendering effector differentiation irreversible by setting the rate of effector de-differentiation $\beta_{e,m}$ to zero. To perform a suitable comparison of the flexible and early decision models [100], we chose a rate of effector differentiation for the latter model to be $\alpha_{m,e} = 0.12/\text{hr}$, such that the fraction of memory cells generated under conditions of moderate pathogen virulence ($\gamma_m = 0.04/\text{hr}$) were equivalent for the two models with $f_{T_m} = 0.05$. All other parameters were kept constant. These simulations show that the early switching model is unable to generate a constant fraction of memory cells amid changes in pathogen replication rates (Fig. 2.2C, middle row). The memory fraction was upheld at a similar value $f_{T_m} = 0.05$ at moderate pathogen replication rates $\gamma_v = 0.04/\text{hr}$; however, this fraction decreased steadily with increasing γ_v , eventually approaching less than 0.01 at high pathogen replication rates (Fig. 2.2C, middle).

Indeed, following the analysis from the previous section (eq. 2.7) and setting $\beta_{e,m} = 0$, we found that the memory cell fraction has an inherently inverse dependence on the peak T

cell population size, as follows:

$$f_{T_m} \approx \frac{T_n(0)}{T_e^{\max}} \left[1 - e^{-\alpha_n \tilde{V}} \right], \quad (2.11)$$

where α_n is the rate at which naive cells transition to the memory fate and $\tilde{V}$ reflects the total amount of pathogens accumulated during the infection (eq. 2.7). This dependence cannot be flexibly scaled by infection severity when the fraction, $T_n(0)/T_e^{\max}$, is much smaller than unity, which is the case in the regime of high pathogen replication wherein large number of effectors is needed to clear the pathogen. Thus, from these results, we conclude that the early T cell decision making strategy cannot produce memory T cell numbers that scale proportionally with the peak T cell population sizes.

2.2.3 Late decision model (obligate reversible decision model)

An alternate strategy for memory differentiation is that, upon activation, all naive cells must first pass through the effector stage before differentiating into memory (Fig. 2.2C, bottom row). This view was supported by the evidence that cells with a history of effector gene expression can become memory cells [94], and that cells on the path to forming memory retain chromatin signatures of the effector state, while harboring the ability to re-activate memory genes that are silenced during effector differentiation [50]; this model can be thought as an obligate reversible decision model.

To evaluate such a decision making strategy, we alter the model above (eq. 2.1), such that naive cells, upon activation, directly transition to an effector state instead of a memory precursor state (Fig. 2.2, bottom). Following the model in eq. 2.1, the ordinary differential equations describing the T cell populations in the late decision model are modified as follows:

$$\begin{aligned}
\text{naive T cell: } \frac{dT_n}{dt} &= -\alpha_n v T_n \\
\text{memory T cell: } \frac{dT_m}{dt} &= \left(\frac{v}{v + K_m}\right) \cdot \gamma_m T_m - \left(\frac{v}{v + K_{m,e}}\right) \cdot \alpha_{m,e} T_m + \left(\frac{K_e}{v + K_e}\right) \cdot \beta_{e,m} T_e \\
\text{effector T cell: } \frac{dT_e}{dt} &= \alpha_n v T_n + \left(\frac{v}{v + K_e}\right) \cdot \gamma_e T_e + \left(\frac{v}{v + K_{m,e}}\right) \cdot \alpha_{m,e} T_m \\
&\quad - \left(\frac{K_{e,m}}{v + K_{e,m}}\right) \cdot \beta_{e,m} T_e - \delta_e T_e
\end{aligned}$$

For direct comparison of this decision making strategy to the model with flexible switching, we keep all the parameters unchanged. From simulations, we find this obligate reversible switching strategy can generate constant fractions of memory cells over a range of pathogen proliferation rates, but fails to generate any substantial memory when pathogens replicate slowly (small γ_v) and the ensuing immune responses are mild (Fig. 2.2C, bottom). When pathogens replicate rapidly and give rise to a substantial T cell expansion, memory cells form robustly at a defined fraction, similar to the flexible switching model (Fig. 2.2C, bottom); however, when pathogens proliferate very slowly, such that there is minimal amount of T cell expansion, the number of formed memory cells constitutes only a small fraction of the starting naive cells. Consequently, in this regime, the ability of T cells to respond to a secondary challenge is greatly diminished.

Following the analysis of the flexible switching model, we can again identify the dominant terms in different regimes of viral load. Following the derivations of eqs. 2.4-2.6 with $\alpha_n = 0$ and $T_m(0) = 0$, we have

$$f_{T_m} = \frac{T_m(\infty)}{T_e^{\max}} \approx \frac{\beta_{e,m}}{\beta_{e,m} + \delta_e}. \quad (2.12)$$

which holds in the regime of rapid viral dynamics, consistent with results from simulations (Fig. 2.2).

2.3 Discussion

By analyzing ordinary differential equation (ODE) models, we investigated how effector and memory CD8⁺ T cell populations scale during infection. While experimental studies have cataloged the molecular and cellular actors involved, the logic governing commensurate scaling of the memory response had remained unclear [50, 90–95]. We used mathematical models to rationalize the experimental observation that memory lineage decisions are reversibly regulated by the *Tcf7* gene locus [49]. Our modeling demonstrates that rigid early or late lineage commitment models fail to produce the near-constant proportional scaling observed in vivo. Instead, we show that flexible timing of differentiation allows the memory pool to robustly scale with infection severity, ensuring a reservoir of cells that is commensurate with the initial threat.

Chapter 3

DESIGN PRINCIPLES OF THE CYTOTOXIC T CELL RESPONSE

The contents of this chapter were previously published in [101].

3.1 Introduction

Cellular immunity hinges upon the regulated activation, expansion, and contraction of lymphocytes that clear an infection while limiting collateral tissue damage. We focus on CD8⁺ T cells—also known as killer or cytotoxic T cells—which are crucial for immune surveillance and defense against infections and cancer. Pathogen recognition is mediated by T cell receptors (TCRs) that bind short pathogen-derived peptides (or antigens) presented by major histocompatibility complexes (MHCs) on antigen-presenting cells (APCs) and other target cells. To recognize a multitude of pathogens, the immune system generates a diverse repertoire of TCRs through a stochastic process known as V(D)J recombination, producing hundreds of billions of distinct T cell clonotypes (i.e., T cells with unique receptors) [16, 102, 103].

Pathogen recognition triggers a response program in naïve T cells, tuned by antigen signal, pro- and anti-inflammatory signaling molecules (cytokines), and other environmental signals generated by the infection and the host’s response (Fig. 3.1A) [104–108]. These signals determine the extent of burst-like proliferation of short-lived effector T cells versus differen-

tiation into long-lived, slowly dividing memory subsets. Effector T cells have the capacity to rapidly divide—up to 19 cell divisions [107, 109–111]—and secrete cytolytic chemical molecules that can degrade infected or unwanted cells [28, 112]. While most activated naïve cells differentiate into terminal effectors that undergo apoptosis once infection is cleared, a small fraction survive as long-lived effector or central memory cells [49, 50, 112, 113].

The logic of signal integration for immune decision-making throughout a response remains the central open problem. Simple additive models of antigen and cytokine inputs can reproduce experimental patterns of clonal expansion in some settings [106–108]. However, genetic and single-cell analyses reveal far richer control circuits governing activation, differentiation, and contraction of T cells that involve intricate genetic programs and probabilistic regulatory rules [34, 49, 50, 96, 111, 114–116].

Pathogens vary in their immunogenicity—the ease with which they are recognized as “non-self”—and in the degree of harm they impose via secreted toxins, inflammation, and disruption of cellular function. The CD8^+ T cell response must have the potential to curb “harm” across diverse infection scenarios [117, 118], yet the evolutionary objectives and biophysical constraints shaping their decision-making program remain unclear. Here, we formalize one plausible objective for T cell decision-making: to minimize cumulative host healthy-cell loss due to infection and collateral death of healthy cells due to the immune response (immunopathology or toxicity).

We model the T cell decision-making program as a flexible, coarse-grained signal-integration protocol. Specifically, we cast each CD8^+ T cell as a feedback controller whose intra-cellular regulatory program modulates activation, differentiation, proliferation, and death in response to intra-host cues, such as antigen levels and cytokine signals, which themselves are dynamic in the course of an infection [64, 107].

We then ask how alternative immunological contexts require distinct effector programs. First, we parameterize a broad class of feedback-controller designs, with varied signal sensi-

tivities and integration rules. Next, by systematically sampling designs, we optimize for rapid pathogen clearance while minimizing immunopathology. We find that optimal designs lie on a Pareto front: further reductions in infection-induced harm entail increased immunopathology. Accounting for the imperative to separate infection—microbial or malignant—from transient inflammatory events (e.g., tissue damage) [119–123], we show how a trade-off between pathogen-clearance and immunopathology (toxicity) may be encoded into the logic of T cell responses. Optimizing these competing objectives yields effector programs consistent with experiments with vertebrate organisms. Finally, applying our framework to a minimal tumor-growth model, we demonstrate how modifying the endogenously encoded effector program can improve the effectiveness of T cell cancer therapies, pointing to signal-processing adjustments through genetic engineering that could enhance T cell-based cancer immunotherapy.

The immune system’s true performance objective, and how evolution has fine-tuned it, may ultimately be unknowable. Formulating and testing principled heuristics provides a baseline to gauge the performance of real immune systems in responding to infections. This perspective has yielded deep insights in other biological contexts, revealing the architectural logic of genetic and regulatory networks [61, 68, 124–127]. In immunology, high-level mathematical models have demonstrated how a finite lymphocyte repertoire can be distributed to best confront a diverse pathogenic landscape [128–130] and how immune memory should be allocated to counter future threats [10, 81, 131–133]. By extending this perspective to the dynamics of CD8⁺ T cell signal integration and response, our work provides a quantitative framework that links observed decision-making to operating principles and trade-offs.

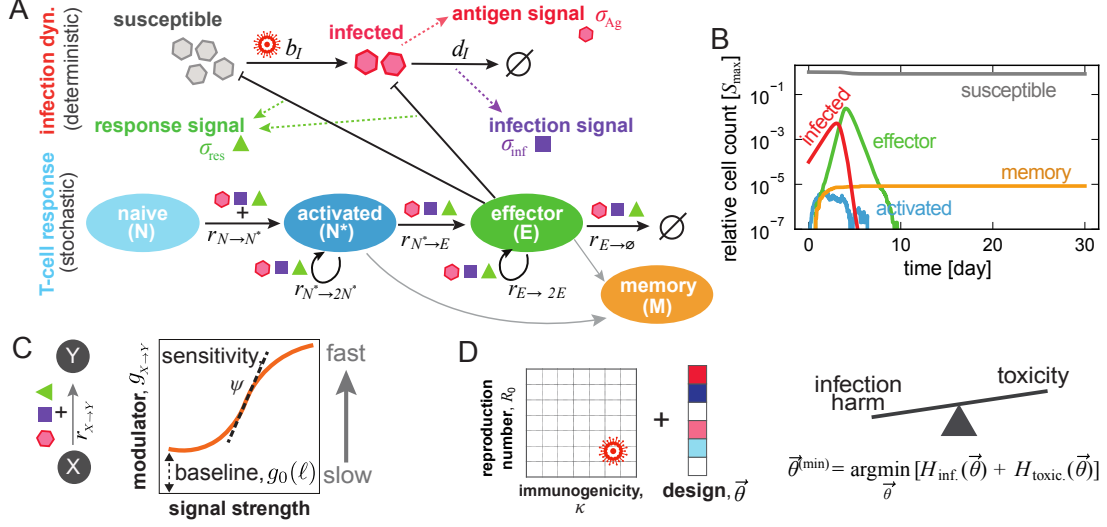


Figure 3.1: CD8⁺ T cell response to infections. (A) The schematic shows how upon infection, susceptible cells S become infected I at rate b_I . Infected cells provide an antigen signal $\sigma_{Ag} \propto I$, and die at rate d_I , which triggers an infection harm signal σ_{inf} . In response to infection, naïve CD8⁺ T cells (with initial clone size of $N_{lin.}$) are activated $N \rightarrow N^*$, and can then differentiate into effector or memory fates (Fig. A.1). Effector cells can proliferate and kill infected and bystander cells, generating a response-harm signal σ_{res} , and finally become terminal or differentiate into memory. Consequently, T cell transitions are modulated by the three signals (σ_{Ag} , σ_{inf} , σ_{res}). Infection dynamics (S , I) are modeled by ODEs coupled to stochastic dynamics of the T cell response. Memory transitions (gray arrows) are neglected in our model. (B) Representative trajectories of susceptible, infected, effector, and memory CD8⁺ T cells are shown over time. Cell counts are reported in units of the maximum number of susceptible cells S_{max} . (C) Each T cell transition has a rate $r_i = r^{max} g_i(\sigma_{Ag}, \sigma_{inf}, \sigma_{res}; \vec{\psi}, \ell)$, modulated by a sigmoidal function with baseline ℓ and signal sensitivities $\vec{\psi}$ (eqs. 3.4, 3.5). The regulatory parameters define the design of the effector program $\vec{\theta}$ (eq. 3.6). (D) We assess design performance over diverse infections sampled from a broad range of reproduction numbers R_0 (eq. 3.1) and immunogenicities κ (eq. 3.2). For each infection, we can find an optimal design $\vec{\theta}^{(min)}$ that minimizes total cell loss (harm). Harm decomposes into infection harm H_{inf} . (cells killed by the pathogen plus infected cells eliminated by effectors), and toxicity H_{toxic} . (healthy bystanders killed by effectors), highlighting an intrinsic trade-off (eq. 3.7). Simulation parameters in (B): $S_{max} = 10^7$, $N_{lin.} = 100$, $K_I = 10^{-3} S_{max}$, $K_S = S_{max}$, $K_H = 10^{-3} S_{max}$, $b_I = 2.0 \times 10^{-7} \text{day}^{-1}$ per cell, $d_I = 0.5 \text{day}^{-1}$, $I_0 = 10^{-4} S_{max}$, $T = 30 \text{days}$, $\vec{\theta} = (\psi_{Ag} = 2, \psi_I = 2, \psi_E = -0.25, \ell_{N \rightarrow N^*} = -0.75, \ell_{N^* \rightarrow E} = -0.75, \ell_{E \rightarrow \emptyset} = 0.75)$.

3.2 Model

To characterize the immune response to an infection, we model the large-population dynamics of susceptible and infected cells in a host as deterministic processes, coupled to a stochastic, time-inhomogeneous birth-death process for the CD8⁺ T cell response (Fig. 3.1A, B and

Appendix A).

Intra-host infection dynamics. We model a population of healthy, susceptible cells $S(t)$. Pathogen exposure seeds an infected subpopulation $I(t)$, which grows over time t via encounters with an otherwise homeostatically stable susceptible cell population, and declines through infected-cell death, triggering an immune response (Figs. 3.1A, B, A.1). Activated effector T cells $E(t)$ can kill both infected and susceptible cells, with clearance governed by *recognition thresholds* K_I and K_S , respectively, which quantify how readily effector cells recognize each cellular subpopulation; higher thresholds indicate that more effector cells are needed for recognition and killing. We model this dynamics as a deterministic birth-death process with competitive interaction of T cells with infected and healthy cells; see Materials and Methods and Appendix A.

We quantify within-host reproductive success of a pathogen by the *intra-host basic reproduction number*,

$$R_0 = \frac{b_I S_{\max}}{d_I}, \quad (3.1)$$

where b_I/d_I is the ratio of birth to death rates for infected cells, and $S_{\max}$ is the maximum number of susceptible cells in a tissue (Materials and Methods). R_0 measures the expected number of secondary infected cells generated by a single infected cell in an otherwise fully susceptible cell population. This definition mirrors the standard epidemiological reproduction number [134, 135].

We model the T cell repertoire as initially consisting of a number of identical naïve T cells, and assume that the T cell recognition threshold for infected cells is lower than that for susceptible cells ($K_I \leq K_S$), such that T cells more readily respond to infected cells. Accordingly, we define the *immunogenicity* of an infection as,

$$\text{immunogenicity: } \kappa = K_I^{-1} K_S. \quad (3.2)$$

This dimensionless quantity reflects how readily cognate T cells recognize antigens presented by infected cells as “non-self” relative to those presented by healthy cells (self). We consider infection scenarios with broad ranges of immunogenicity (κ) and basic reproduction numbers (R_0) (Appendix A).

Signal cues to modulate T cell decision-making. We express damage during an infection in terms of cell deaths,

$$h(t) = h_{\text{inf}}(t) + \underbrace{h_{E,I}(t) + h_{E,S}(t)}_{=h_{\text{res}}(t)}, \quad (3.3)$$

where $h_{\text{inf}}(t)$ represents the number of cells killed directly by the pathogen (infection), and $h_{E,I}(t)$ and $h_{E,S}(t)$ denote the number of infected and healthy cells, respectively, killed by the effector T cell response, amounting to the total damage from the response $h_{\text{res}}(t)$.

Cell deaths can produce different types of signals. Unprogrammed cell death (e.g., pyroptosis triggered by viruses and bacteria) can lead to the release of pro-inflammatory danger-associated molecular patterns (DAMPs) [119]. On the other hand, cell death initiated by CD8⁺ T cells generates a heterogeneous signal that generates pro- and anti-inflammatory cytokines (e.g., IFN γ , TNF α , IL-1 α , IL-4, IL-10, TGF- β) [119, 120, 123].

We simplify this picture by partitioning non-antigen cues into (i) infection-induced and (ii) response-induced harms, whose accumulated effects trigger infection and response harm-induced signals, $\sigma_{\text{inf}}(t)$ and $\sigma_{\text{res}}(t)$, respectively. Presented antigens on APCs serve as an additional cue. We take the infected-cell burden $I(t)$ as a proxy for the instantaneous antigen signal $\sigma_{\text{Ag}}(t)$, which together with the other two signals, modulate the cellular decision-making processes in Fig. 3.1A; see Materials and Methods and Appendix A for details. By disentangling these sources of (potentially redundant) signals, we probe how CD8⁺ T cells use these cues to make fate decisions. We do not pre-assign any signals to a pro- or anti-

inflammatory role but allow optimization goals to determine their influence across diverse infection scenarios.

We should emphasize that CD8^+ responses do not occur in isolation: the innate immune system, regulatory T cells (Tregs), helper T cells, and B-cells can all directly or indirectly shape CD8^+ dynamics. While these interactions are not modeled explicitly, we interpret the harm and antigen signals as coarse-grained proxies for inputs from other immune compartments, capturing their net pro- and anti-inflammatory effects. In the same vein, our immunogenicity parameter κ (eq. 3.2) may also subsume regulatory constraints, e.g. through Treg-mediated suppression that limits effector responses to self-like antigens, which in our framework correspond to low-immunogenicity targets. Incorporating these additional components explicitly would be an important direction for future work, enabling the study of richer feedbacks and more granular immune regulation and dynamics.

Design of T cell programs for infection response. Naïve T cells cycle through lymphoid tissues, becoming activated when they durably engage cognate peptide-MHC complexes on activated APCs [42, 136]. Activated T cells can differentiate into an effector phenotype, which participates in infection clearance, or into a long-lived memory state (Figs. 3.1A, A.1). Here, we examine the course of CD8^+ T cells response during an infection, dividing the process into four canonical response modules (Fig. 3.1A):

1. Activation: priming of naïve T cells, $N \rightarrow N^*$,
2. Differentiation: commitment of the activated T cell to an effector phenotype, $N^* \rightarrow E$,
3. Expansion: division of effector cells, $E \rightarrow 2E$,
4. Contraction: programmed loss of effector cells, $E \rightarrow \emptyset$.

These response processes map onto partially distinct genetic programs: the MYC pathway drives activation and proliferation [34]; the Wnt-Tcf7 pathway regulates effector versus

memory fate decisions [49, 137]; contraction can be attributed to effector cell apoptosis controlled by the Bcl-2-family of proteins, or the differentiation of effectors into effector-memory cells [49]. The latter is regulated by the balance of T-bet/Blimp-1 (effector) versus Bcl-6/Eomes (memory) programs, as well as metabolic pathways such as mTORC1 [138]. To reduce model degeneracy, we do not distinguish among these distinct biological sources of contraction in our analysis.

For each process i , we represent its genetic regulation by a time-dependent transition rate $r_i(t)$, modulated by the instantaneous signals $\vec{\sigma}(t)$ (eq. 3.10) received by the T cell (Fig. 3.1C),

$$r_i(t) = r^{\max} g_i \left(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}}; \vec{\psi}, \ell_i \right), \quad (3.4)$$

where the maximal rate $r^{\max}$ is capped by the cell-cycle speed [28, 42, 116], and g_i is a monotonic *response function* of signal input, chosen from the modified Monod-Wyman-Changeux functions (Hill-type family) [75, 139]:

$$g_i(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}}; \vec{\psi}_i, \ell_i) = \frac{1}{1 + \exp \left[-L_i(\vec{\sigma}; \vec{\psi}_i, \ell_i) \right]}, \quad (3.5)$$

with,

$$L_i(\vec{\sigma}; \vec{\psi}_i, \ell_i) = \ell_i + \sum_{\alpha \in \{\text{Ag}, \text{inf}, \text{res}\}} \psi_i^\alpha \log [1 + \sigma_\alpha].$$

The parameter ℓ_i sets the baseline propensity of transition i : in the absence of external signals, the regulatory factor evaluates to $g_{0,i} = (1 + \exp(-\ell_i))^{-1}$ (Fig. 3.1C). The weights $\{\psi_i^{\text{Ag}}, \psi_i^{\text{inf}}, \psi_i^{\text{res}}\}$ encode the sensitivity and polarity (positive / negative = up / down-regulation) of response to the antigen (σ_{Ag}) and harm signals (σ_{inf} and σ_{res}). This formulation abstracts away biochemical details while implementing the fold-change response to stimuli observed in many biological circuits [61, 74, 140–142].

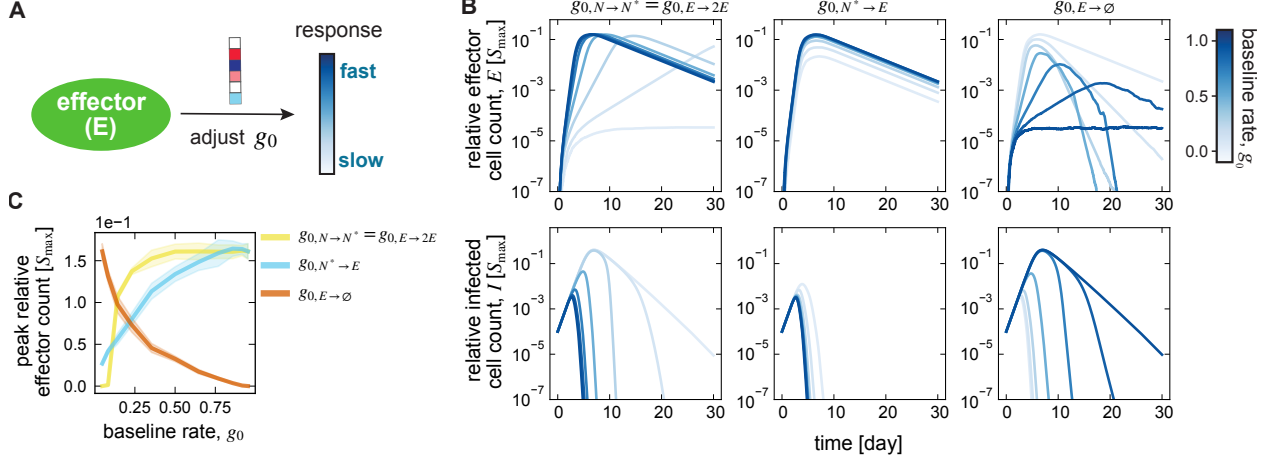


Figure 3.2: Design parameters shape T cell response and infection clearance. (A) Varying the design parameters influences effector trajectories. (B) Trajectories are shown for the population size of effectors (top) and infected cells (bottom) for designs that vary a single baseline transition rate $g_{0,i}$ (color) indicated atop each panel. Both quantities are measured in units of the maximum number of susceptible cells, $S_{\max}$, which is a convenient reference scale for different cell sub-populations. All other design parameters are fixed at: fast activation, $g_{0,N \rightarrow N^*} = 0.95$; fast differentiation, $g_{0,N^* \rightarrow E} = 0.95$; slow contraction, $g_{0,E \rightarrow \emptyset} = 0.05$; and zero signal feedback, $\psi_{\text{Ag}} = \psi_{\text{inf}} = \psi_{\text{res}} = 0$. Trajectories reflecting the impact of varying signal sensitivities ψ 's are shown in Appendix A, Fig. A.2. (C) Peak effector count, measured in units of maximum number of susceptible cells $S_{\max}$, is shown by varying each of the three baseline transition rates (colors), while keeping the rest of the parameters fixed (similar to (B)). Solid lines show the mean of effector counts across $n = 5$ simulation replicates and shaded envelopes indicate one standard deviation (Materials and Methods). Parameters are the same as Fig. 3.1, except for $K_I = 10^{-2} S_{\max}$.

To reflect shared regulatory programs while controlling complexity, we impose two simplifications. First, activation and early expansion are both MYC-driven [36]; we therefore require the modulatory factors, $g_{N \rightarrow N^*}$ and $g_{E \rightarrow 2E}$, to share parameter values, even though their realized rates will differ because signals vary in time, as cells transition between naive and effector states. Second, logically, since extracellular cues should converge on coherent effector programs, we enforce a common set of signal sensitivities across modules. With the sign convention that positive sensitivities promote effector production, for each signal σ , we set $\psi_{\sigma} \equiv \psi_{N \rightarrow N^*}^{\sigma} = \psi_{E \rightarrow 2E}^{\sigma} = \psi_{N^* \rightarrow E}^{\sigma} = -\psi_{E \rightarrow \emptyset}^{\sigma}$.

The parameters of these constrained response functions define a *response design* $\vec{\theta}$ as,

$$\vec{\theta} = \underbrace{(\psi_{\text{Ag}}, \psi_{\text{inf}}, \psi_{\text{res}})}_{\text{signal feedback}}, \underbrace{(\ell_{N \rightarrow N^*} = \ell_{E \rightarrow 2E}, \ell_{N^* \rightarrow E}, \ell_{E \rightarrow \emptyset})}_{\text{module-specific baselines}}. \quad (3.6)$$

We systematically sample $\vec{\theta}$'s to assess performance of designs against different infections (Fig. 3.1D; Appendix A). While $\vec{\theta}$ is broadly sampled, we hold the remaining biological rates and design parameters underlying the decision-making program in Fig. 3.1A within empirically supported ranges reported in Appendix A, Table S1.

Cell economics of T cell response. Tracking all clinical sequelae of an infection is rarely feasible, but tissue integrity degrades as cells are lost. Cumulative cell death correlates with macroscopic readouts such as weight change, pathogen burden, and temperature dynamics [143, 144]. We therefore adopt a cell-economics objective in which harm is measured by the total number of cells lost over the course of infection. We evaluate the performance of a given design $\vec{\theta}$ by simulating the response dynamics (Fig. 3.1D) and computing the *total cell death* (or inflicted harm) as

$$H(\vec{\theta}) = \underbrace{\int_0^T [h_{\text{inf}}(t, \vec{\theta}) + h_{E,I}(t, \vec{\theta})] dt}_{H_{\text{inf.}}(\vec{\theta})} + \underbrace{\int_0^T h_{E,S}(t, \vec{\theta}) dt}_{H_{\text{toxic.}}(\vec{\theta})}, \quad (3.7)$$

where $H_{\text{inf.}}(\vec{\theta})$ aggregates infection-related harm (cells killed by the pathogen, and infected cells eliminated by effectors), $H_{\text{toxic.}}(\vec{\theta})$ captures immunopathology (healthy bystander cells killed by effectors), over the fixed time window T . In our main simulations, we set $T = 30$ days to reflect the typical timescale of acute infections and the associated immune response (Appendix A, Table S1), but we also assess the sensitivity of our results to variation in T (Materials and Methods). This single, interpretable metric of harm (eq. 3.7) balances clearance benefits against collateral damage, enabling principled comparison across designs

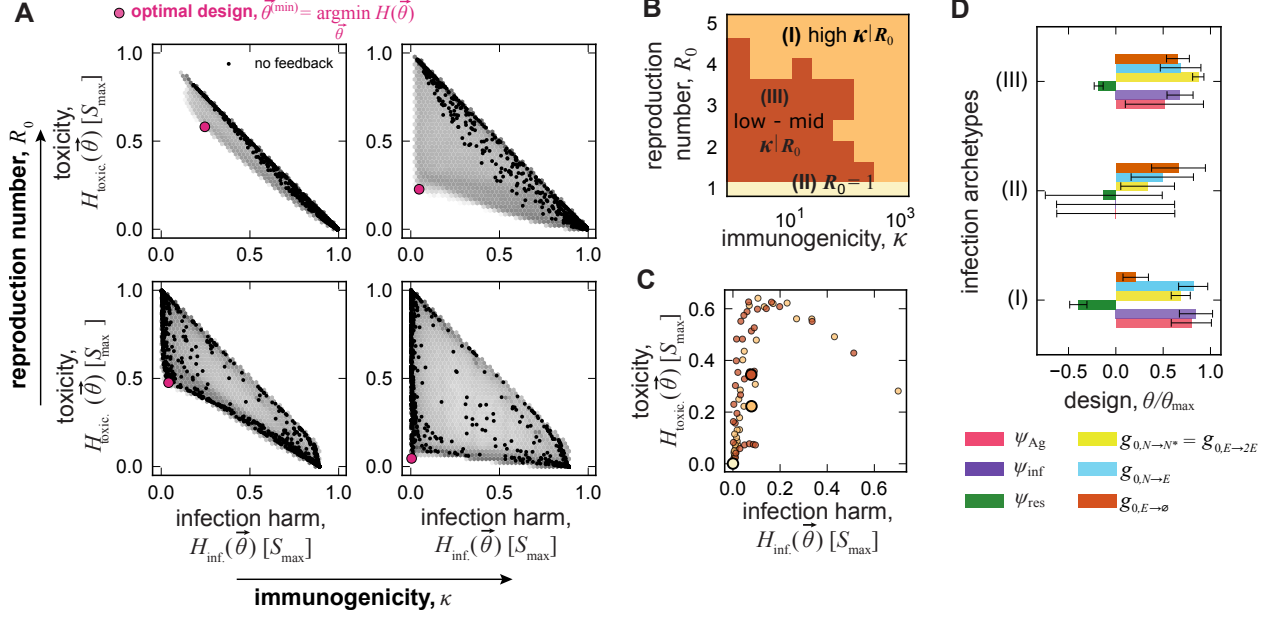


Figure 3.3: Toxicity-infection harm trade-off across T cell regulatory designs. (A) Density plots of toxicity H_{toxic} versus infection harm H_{inf} (eq. 3.7)—measured in units of the maximum number of susceptible cells S_{max} —for 4.8×10^6 T cell regulatory designs, $\vec{\theta}$, uniformly grid-sampled across four infection scenarios (panels; columns: $\kappa = 5.6$ (left), $\kappa = 178$ (right); rows: $R_0 = 2.5$ (bottom), $R_0 = 5.0$ (top)). Designs are sampled uniformly across signal sensitivities ψ 's $\in [-3, 3]$ and baseline rates ℓ 's $\in [-3, 3]$ (eq. 3.6). Infection scenarios are sampled by changing K_I and b_I , while holding S_{max} , d_I , and K_S fixed. Black points denote no-feedback designs ($\vec{\psi} = 0$), and the large magenta circle in each panel marks the design that minimizes total harm, $\vec{\theta}^{(\min)}$. (B) For many different infection scenarios (κ, R_0), we identify an ensemble of near-optimal designs $\Theta(\kappa, R_0)$, whose total harm is within $10^{-2} \times S_{\text{max}}$ of the minimum sampled total harm (as in the magenta point in A). By computing the ensemble-averaged design parameters $\langle \vec{\theta} \rangle_{\Theta(\kappa, R_0)}$ for each infection scenario, and using these averages to cluster infections, we identify three pathogen “archetypes” (colors): I–III, as indicated (Appendix A). (C) Toxicity vs. infection harm associated with the optimal harm-minimizing designs for each pathogen archetype in (B) is shown. Large points indicate archetype means; small semi-transparent points show individual immune challenges sampled in (B). Reducing infection harm (horizontal axis) comes at the expense of increased toxicity (vertical axis). (D) Bars indicate the ensemble-mean of design parameters, $\langle \vec{\theta} \rangle_{\Theta(i)}$, computed by identifying the ensemble of designs $\Theta^{(i)}$ which minimize cluster-averaged total harm for each archetype within $10^{-2} \times S_{\text{max}}$ of the minimum sampled cluster-averaged total harm (similar to (B)). Error bars denote the standard deviation within this ensemble. All design variables are normalized by their respective bounds ($\psi^{\text{max}} = 3$, $\ell^{\text{max}} = 3$). Unless otherwise noted, results represent the average of $n = 5$ simulation replicates with parameters: $S_{\text{max}} = 10^7$, $N_{\text{lin}} = 10^{-5} S_{\text{max}}$, $K_S = S_{\text{max}}$, $K_H = 10^{-3} S_{\text{max}}$, $d_I = 0.5 \text{ day}^{-1}$, $I_0 = 10^{-4} S_{\text{max}}$.

and infections (Fig. 3.1D).

3.3 Results

Cell-fate transitions determine effector response and pathogen clearance

How do rates of cell-fate transitions, determined by the regulatory design $\vec{\theta}$ (eq. 3.6), impact the immune response and pathogen clearance? To isolate the effect of baseline propensities ($\vec{\ell}$) independently of signal feedback ($\vec{\psi}$), we first simulated "hard-wired" designs in which all signal sensitivities were set to zero ($\vec{\psi}_i = 0$). Each transition then happens at a fixed fraction of its maximal rate $g_{0,i} = [1 + \exp(-\ell_i)]^{-1}$, independently of antigen or harm signals; here, $i \in \{N \rightarrow N^*/E \rightarrow 2E, N^* \rightarrow E, E \rightarrow \emptyset\}$ indicates the transition module.

Varying these baseline rates produced strikingly different effector trajectories (Fig. 3.2). Promoting activation/proliferation ($g_{0,N \rightarrow N^*}, g_{0,E \rightarrow 2E} \uparrow$), or lowering the contraction baseline rate ($g_{0,E \rightarrow \emptyset} \downarrow$) generated a rapid, high-amplitude effector burst that cleared the pathogen quickly, after which effector numbers decay (Fig. 3.2B, C). In contrast, slowing activation and division ($g_{0,N \rightarrow N^*}, g_{0,E \rightarrow 2E} \downarrow$), or accelerating contraction ($g_{0,E \rightarrow \emptyset} \uparrow$), yielded sluggish effector accumulation, delayed clearance, and in extreme cases pathogen persistence, accompanied by a prolonged plateau of few effectors. Interestingly, varying the baseline rate of differentiation $g_{0,N^* \rightarrow E}$ had a weaker impact on the overall effector profile and pathogen clearance (Fig. 3.2B, C), suggesting its relevance may lie elsewhere in the response program.

Because effector and pathogen abundances determine the levels of antigen and cumulative harm-associated (cytokine) signals, these baseline choices not only shape the immediate effector response but also alter the available antigen and cytokine signals, which in turn impact the T cell response through feedback regulation (Appendix A, Fig. A.2).

Effector designs trade infection harm for toxicity

Previous research has identified a trade-off between maximizing infection clearance and minimizing response toxicity (immunopathology) in the immune response [122, 123, 145]. Con-

sistently, by systematically sampling the space of designs (Appendix A), we found the same trade-off: plotting infection harm $H_{\text{inf.}}(\theta)$ (i.e., infected cells killed by pathogen or by effector response) versus the toxicity $H_{\text{toxic.}}(\theta)$ (i.e., bystander cells killed by effector response) revealed that decreasing $H_{\text{inf.}}$ comes at the cost of increasing $H_{\text{toxic.}}$ (Fig. 3.3A). The extent of this trade-off depends on pathogen characteristics. This trade-off is stronger for infections with a higher intra-host reproductive number R_0 or lower immunogenicity κ . Interestingly, for scenarios with low R_0 , many of the hard-wired response programs without signal feedback ($\vec{\psi}_i = 0$; black dots) perform near-optimally, laying close to the *Pareto front*. Here, the Pareto fronts at the boundaries of these graphs indicate the set of strategies for which neither harm nor toxicity can be reduced without increasing the other, representing the best achievable balance between these competing objectives (see Appendix A). For aggressive high- R_0 threats—be it acute infections (high κ), or highly metastatic cancers (mid to low κ)—these same hard-wired designs fall far from the Pareto front and well inside the sub-optimal region. Proportionate control in these cases requires designs with appropriate sensitivities to antigen and cytokine/harm signals ($\vec{\psi} \neq 0$) to minimize infection harm while limiting immunopathology.

Next, we explored regulatory design space by evaluating responses of different designs against diverse infection scenarios, sampled on a grid of immunogenicities (κ) and reproduction numbers (R_0) (Fig. 3.3B). For each infection scenario (κ, R_0) , we identified the ensemble of near-optimal designs, $\Theta(\kappa, R_0)$. A design $\vec{\theta}$ belongs to this ensemble if its incurred total harm (eq. 3.7) is within $10^{-2} \times S_{\text{max}}$ of the minimum realized total harm from all sampled designs for that infection; note that the maximum number of susceptible cells S_{max} provides a convenient reference scale against which to quantify harm (cell death). Finally, we clustered infections based on their ensemble-averaged design parameters $\langle \vec{\theta} \rangle_{\Theta(\kappa, R_0)}$ (Appendix A, Fig. A.3). Three coherent design classes emerged, loosely corresponding to intuitive pathogen “archetypes” (Fig. 3.3B): **I**—highly immunogenic bacteria and viruses (high κ or high R_0),

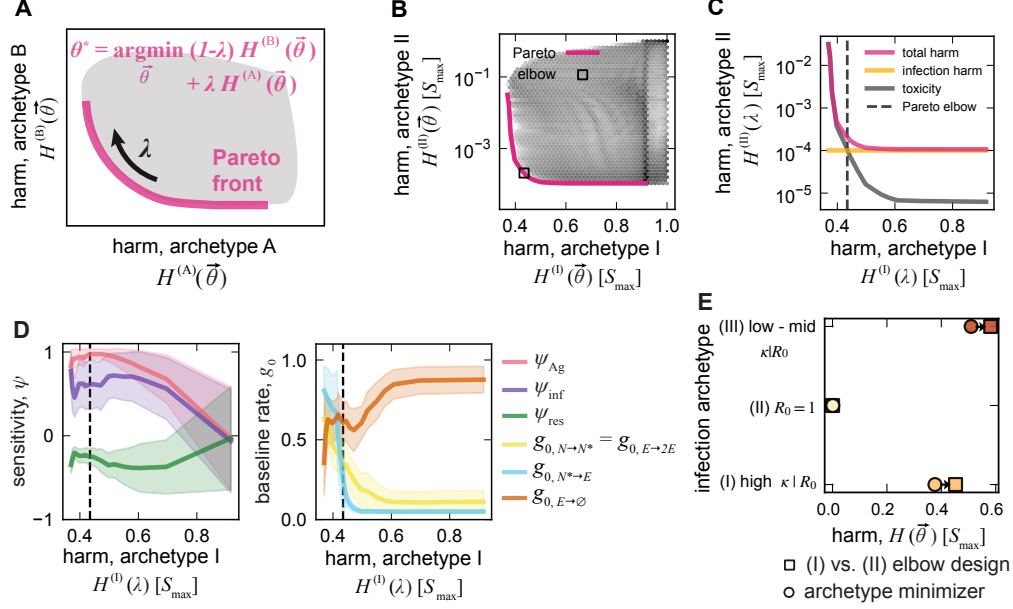


Figure 3.4: Pareto-optimal designs balancing harm from acute infections and autoimmunity. (A) Bi-objective optimization to minimize the weighted sum of harms associated with T cell regulatory designs $\vec{\theta}$ across two pathogen archetypes (A, B) yields a Pareto front (magenta curve); eq. 3.8. Points on the curve are Pareto-optimal designs, parameterized by the weight $\lambda \in [0, 1]$, quantifying the emphasis on archetype A; moving left to right corresponds to decreasing λ (less emphasis on archetype A). (B) The density plot shows harms (in units of the maximum number of susceptible cells $S_{\max}$) incurred by 4.8×10^6 uniformly sampled designs in their response to archetype I (immunogenic infection; high κ) versus archetype II (non-replicating, self-like/ autoimmune inducing or allergenic antigens; $R_0 = 1$), as defined in Fig. 3.3B. Designs are sampled on a uniform Cartesian grid of sensitivities $\psi_{\sigma} \in [-3, 3]$ and baseline rates $\ell_i \in [-3, 3]$ (as in Fig. 3.3A). For many values of λ , we minimize the objective in (A), and trace out the Pareto front (magenta). For each point λ on the front, we identify the ensemble of designs closest to the Pareto front ($\Theta(\lambda)$), where a design belongs to the near-Pareto ensemble at λ if its incurred harm against archetypes I and II simultaneously lie within 5% of the λ -Pareto-optimal harm of the respective archetypes. The black square indicates the elbow of the Pareto curve λ^* : the point where further reductions in archetype-II harm incur larger increases in archetype-I harm (Appendix A). (C) Decomposition of total harm along the Pareto front in (B) into infection-related harm and toxicity (eq. 3.7). The black dashed line indicates the position of the elbow from (B). (D) The magnitudes of the signal sensitivities ψ (left) and the baseline rates $g_0(\ell)$ (right), each scaled to variable bounds ($\psi^{\max} = 3$, $\ell^{\max} = 3$), are shown as a function of harm incurred under archetype I along the Pareto curve (parameterized by λ). Solid and shaded envelopes show the mean and standard deviation of the design parameters, respectively, across the near-Pareto ensemble of designs $\Theta(\lambda)$, $\lambda \in [0, 1]$ traced out by the Pareto front. (E) For each archetype in Fig. 3.3B, total harm is shown for the archetype-specific minimizer $\vec{\theta}^{(\min)}$ (circle) and for the average over the near-elbow ensemble in (B). See Appendix A, Fig. A.6 for Pareto fronts formed by other pairs of archetypes. All simulation parameters other than the varied K_I , b_I , and $\vec{\theta}$ match Fig. 3.1.

associated with type-I immunity [146]; **II**— non-replicating, self-like/auto-immune inducing or allergenic antigens ($R_0 = 1$), in part associated with type-II immunity [146]; **III**— mid-range threats (low-medium κ and R_0) including difficult to detect pathogens and slowly growing tumors. The balance between minimizing toxicity and clearing infection (i.e., reducing harm) differs across these archetypes, resulting in different design strategies: signal-sensitive designs optimized for archetypes I and III prioritize pathogen clearance at the cost of increased immunopathology, with the severity of the trade-off varying by scenario, whereas the largely pre-programmed designs optimized for archetype II suppress toxicity to avoid autoimmunity (Fig. 3.3C).

Design parameters of effector programs minimizing harm vary across archetypes (Fig. 3.3D). For archetype II (non-replicating antigens) the optimal design is pre-programmed suppression of the immune response: signal sensitivities $\vec{\psi}$ are close to 0, and the baseline rates are tuned to slow activation, proliferation, and differentiation, and enforce rapid effector deletion. This damped response completely avoids toxicity when the threat does not grow so infection harm is determined by the initial exposure and cannot be decreased. Across the other two archetypes a common pattern emerges: the response-derived cytokine signal σ_{res} acts as an anti-inflammatory brake ($\psi_{\text{res}} < 0$), with its magnitude tuned to context. Moreover, antigen sensitivity ψ_{Ag} grows with immunogenicity, with the designs optimized for archetype I (strong immunogenicity) showing the largest sensitivity to antigens, while those for archetype II (non-immunogenic) show no antigen sensitivity, and are largely hard-wired. Overall, across all archetypes, near-optimal designs distinctively tune both the baseline rates g_0 's and the infection-signal sensitivity ψ_{inf} to balance clearance against toxicity and minimize harm (Appendix A, Figs. A.4, A.5).

Endogenous T cell programs are shaped by minimizing harm from autoimmunity and acute infections

Although the relative evolutionary pressures exerted by different threats are unknown, we can still pose a principled multi-objective problem for the effector program, i.e., find a design $\vec{\theta}_{\min}$ (eq. 3.6) that minimizes the weighted sum of harms across a desired set of pathogen archetypes:

$$\vec{\theta}_{\min} = \arg \min_{\vec{\theta}} \sum_{i:\text{archetypes}} \lambda_i H^{(i)}(\vec{\theta}), \quad (3.8)$$

where $\lambda_i \in [0, 1]$ with the constraint $\sum_i \lambda_i = 1$. The weight λ_i specifies how strongly we prioritize minimizing harm associated with the i^{th} pathogen archetype (i.e., an objective), when identifying the optimal design $\vec{\theta}_{\min}$. For m archetypes, the solution to this family of problems maps an $m - 1$ -dimensional simplex of $\vec{\lambda} = (\lambda_1, \dots, \lambda_m)$ values to a set of designs that achieve the best possible trade-off between the different archetypes (Fig. 3.4A; Appendix A). When optimizing performance against one archetype necessarily worsens performance against another (i.e., a tradeoff), these optimal designs form a *Pareto front*: a set of solutions for which no objective can be improved without degrading at least one other (Fig. 3.4A). In contrast, if there are no such conflicts among archetypes, the Pareto front collapses to a single design that is simultaneously optimal for all objectives. This framework naturally accommodates uncertainty in the prevalence of pathogenic challenges and in the effective “preferences” imposed by evolution or environment, in finding the optimal solution. Similar approaches have been successfully applied in other biological systems [87, 147].

Even though optimization can be done simultaneously over all three archetypes, we proceed by examining pairwise trade-offs that generate a one dimensional Pareto front, parameterized by a single scalar λ (Fig. 3.4A). A biologically salient case contrasts archetype I (immunogenic pathogens) with archetype II (non-replicating antigens). Non-replicating antigens plausibly arise from un-programmed cell deaths or external trauma inducing inflam-

mation and may be implicated in autoimmune flares. This bi-objective harm minimization aims to find designs balancing effective clearance of immunogenic viral or bacterial infections and suppression of autoimmune responses; see Appendix A, Fig. A.6 for Pareto fronts formed by optimizing other pairs of clusters.

Fig. 3.4B shows harm from archetype II plotted against that of archetype I for a broad range of sampled designs and highlights the Pareto front (magenta). Designs close to the Pareto front achieve near-best total protection for a given compromise weight λ ; moving to the right along the Pareto front shifts the emphasis from clearance of immunogenic pathogens ($\lambda \uparrow$) to tolerance ($\lambda \downarrow$). Decomposing total harm into infection-related harm and toxicity (eq. 3.7) shows that along this Pareto front it is relatively easy to reduce toxicity to non-replicating antigens without substantially worsening outcomes against immunogenic infections; this is demonstrated in Fig. 3.4C as substantial decreases in toxicity are achieved with only minor changes in infection harm.

Next, we evaluated immune designs near the Pareto front (Fig. 3.4D; Appendix A). For each point λ , we identified an ensemble of designs near the Pareto front, $\Theta(\lambda)$. For a design to belong to the near-Pareto ensemble, its incurred harm against archetypes I and II must simultaneously lie within 5% of the Pareto optimal harm of those respective

Design Parameter	Effect	Source
baseline rates:		
1. activation & prolifer., $g_{N \rightarrow N^*}, g_{E \rightarrow 2E}$	low	[34, 104]
2. differentiation, $g_{N^* \rightarrow E}$	low	[49]
3. contraction, $g_{E \rightarrow \emptyset}$	moderate	[34, 60]
sensitivity to signal:		
4. antigen, ψ_{Ag}	promote	[104, 106, 148]
5. infection, ψ_{inf}	promote	[105, 123, 149]
6. response, ψ_{res}	suppress	[122, 123]

Table 3.1: Endogenous design of CD8⁺ T cell response.

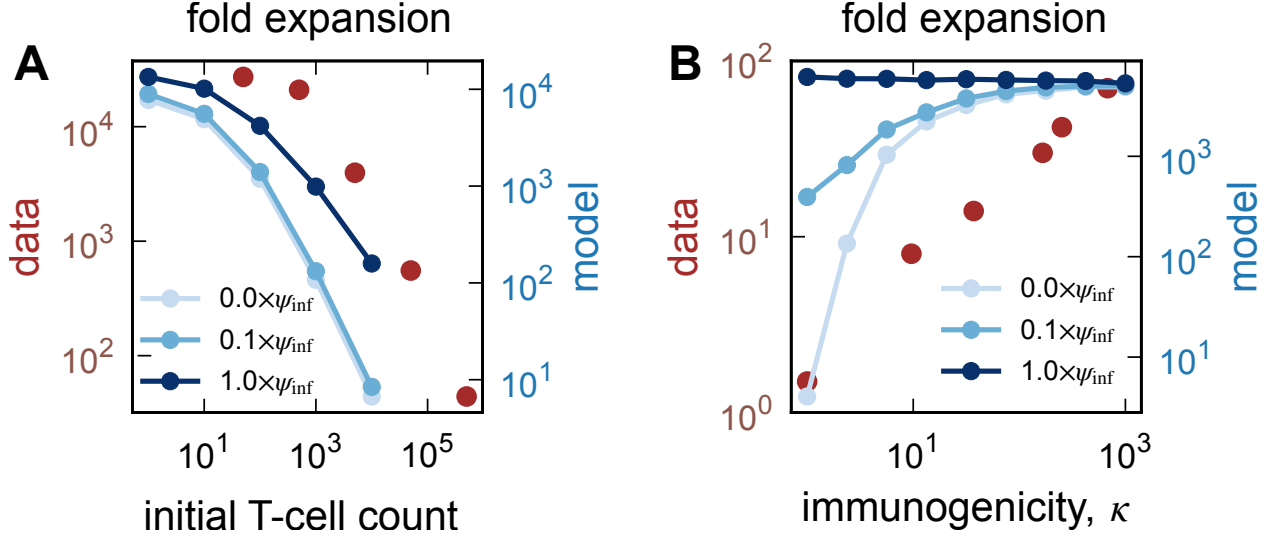


Figure 3.5: Elbow-optimal designs recapitulating macro-dynamics of T cell expansion. (A,B) In maroon (left vertical axis), fold expansion (i.e, peak clone size divided by the initial precursor number) of antigen-specific CD8⁺ T cells in mice plotted against (A) the initial precursor number (data from [109]), and (B) the relative pMHC concentration required for half-maximal activation EC₅₀ (data from [148]); in our model, EC₅₀ is proxied by K_I . In blue (right vertical axis), simulated mean fold expansion for designs belonging to the near-elbow ensemble from Fig. 3.4B ($\Theta(\lambda^*)$) shown as a function of the initial precursor T cell number (A), and the immunogenicity of the antigen κ (B). Their sensitivity to the infection-harm signal (ψ_{inf}) is varied (shades of blue). For each design in the near-elbow ensemble identified in Fig. 3.4, $\theta \in \Theta(\lambda^*)$, we change $\psi_{\text{inf}} \rightarrow c \times \psi_{\text{inf}}$ for constants $c \in \{0, 0.1, 1.0\}$, and simulate responses against the broad sample of infections (Fig. 3.3B). Trend lines are obtained by averaging over infections and designs in the near-elbow ensemble (Appendix A, Figs. A.7, A.8). Elbow designs with reduced sensitivity to infection-derived harm qualitatively match the experimental macro-dynamics of T cell expansion. Unless otherwise indicated, simulation parameters: $S_{\text{max}} = 10^7$, $K_S = S_{\text{max}}$, $K_H = 10^{-3} S_{\text{max}}$, $d_I = 0.5 \text{ day}^{-1}$, $I_0 = 10^{-4} S_{\text{max}}$.

archetypes at the same compromise value λ . The *elbow* of the Pareto front represents the point on the curve, λ^* , at which further decreases in harm from archetype II will incur disproportionately large increases in harm from archetype I (Fig. 3.4B; Appendix A). The ensemble of designs near the elbow $\Theta(\lambda^*)$ encodes a compelling logic: sensitivity from antigen and infection cues promotes effector production ($\psi_{\text{Ag}}, \psi_I > 0$), whereas response-derived cues act like a brake with a predominantly anti-inflammatory effect on the immune response ($\psi_{\text{res}} < 0$); Fig. 3.4D. Sensitivity to these signals diminishes as λ decreases (i.e., as avoidance of autoimmunity is prioritized); see Fig. 3.4D. Baseline rates (g_0 's) of the designs at the

elbow favor moderate to slow activation and proliferation of effectors, slow differentiation to effector state, and moderate contraction (Fig. 3.4D). As λ increases (leftward; clearance weighs more), all baseline rates shift toward promoting more effector cells. As λ decreases below the elbow (rightward; toxicity weighs more), the already slow differentiation baseline rate falls toward zero, while activation/proliferation and contraction are tuned to suppress effector production, reflecting designs that minimize collateral damage even at the expense of incomplete clearance. Notably, despite being optimized only for archetypes I and II, designs in the near-elbow ensemble $\Theta(\lambda^*)$ perform competitively for archetype III, incurring harm close to the archetype-specific optimal designs (Fig. 3.4E; Appendix A, Fig. A.6).

From these analyses alone, it is not possible to ascertain the compromises encoded in the endogenous T cell response program. However, experiments on T cell fate decisions in endogenous immunological programs (Table 3.1) are qualitatively consistent with designs that jointly minimize harm from immunogenic pathogens and auto-immune responses to transient injuries (elbow in Fig. 3.4B).

The choice of design impacts the dynamics of the immune response, which we can compare to experimental data. Prior experiments in mice [109, 148] indicate that peak fold-expansion of T cell populations in response to a pathogen scales inversely with the initial number of cognate T cells, and with the pMHC concentration for half-maximal response (proxied by K_I in our model); see Fig. 3.5A, B. This pattern was recapitulated by explicitly modeling competition for antigens in polyclonal T cell responses [63]. Without any parameter fitting, near-elbow designs $\vec{\Theta}(\lambda^*)$ from Fig. 3.4 reproduce this inverse dependence on initial clone size but underestimate the steepness, and overestimate the magnitude of the trend with respect to immunogenicity (κ) (Fig. 3.5A, B).

Reducing the sensitivity to infection harm ψ_{inf} of these designs by an order of magnitude restores the observed dynamics range (Fig. 3.5B) while having only a minor effect on total harm, and hence, our optimization objective (Fig. A.8). Interestingly, adjusting the other

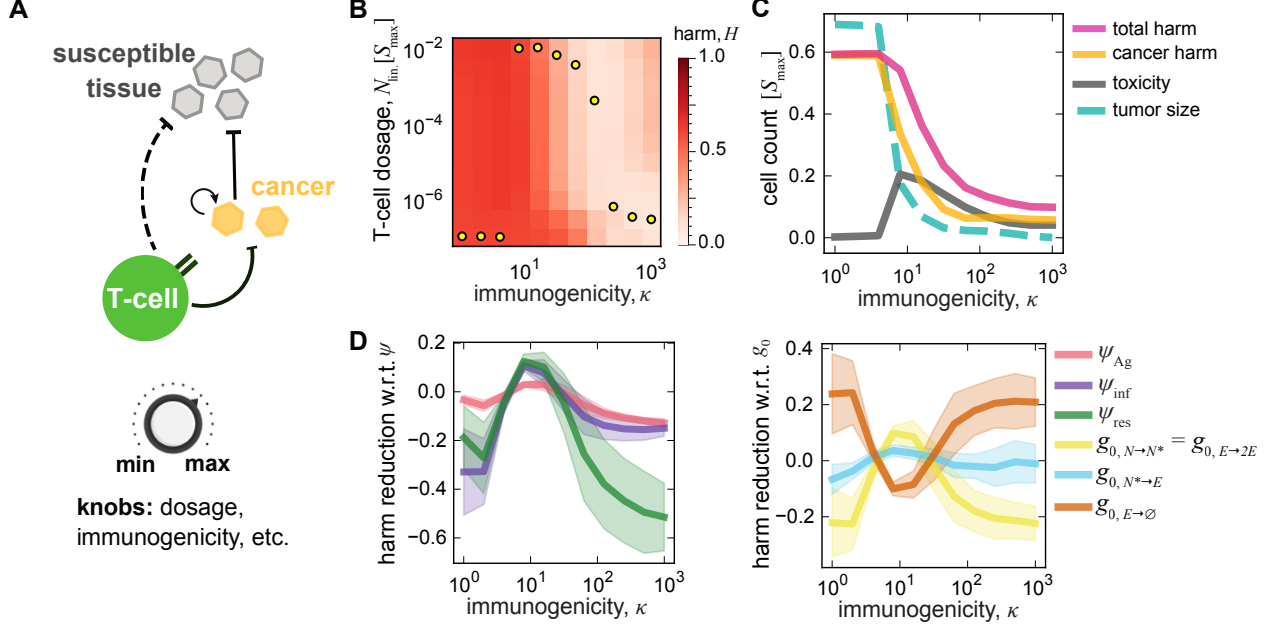


Figure 3.6: Engineering immune response for cancer immunotherapy. (A) The schematic shows the potential clearance of replicating cancer tumor cells by T cells (e.g., infused during immunotherapy), and the potential collateral damage of therapy to susceptible tissue cells. The accessible levers in an immunotherapy include the dosage and antigenicity of the infused cells. Antigenicity is a property of the interaction between a specific TCR-antigen pair, and it can be increased by either changing the antigen or the TCR. Given a specific cognate T cell, immunogenicity is the ratio of the antigenicity of a cancer cell to that of a healthy cell. (B) Averaged over the tuned near-elbow ensemble $\tilde{\Theta}(\lambda^*)$, the heatmap displays harm (in units of the maximum number of susceptible cells S_{max}) as a function of cancer antigenicity and dosage of administered T cells. Yellow points correspond to the optimal dosage of T cells for a given immunogenicity/antigenicity. (C) Total harm, harm from cancer, immune toxicity, and the final tumor size following the immune response are shown as functions of cancer immunogenicity (κ) at the optimal T cell dosage (yellow circles in B). (D) Clearance response—defined as the derivative of negative harm (i.e., protection) with respect to each design parameter θ_i , $-\partial H(\tilde{\theta})/\partial \theta_i$ for $\tilde{\theta} \in \tilde{\Theta}(\lambda^*)$ —averaged over the tuned near-elbow designs $\tilde{\Theta}(\lambda^*)$, is shown for different levels of tumor immunogenicity κ ; Left: $\vec{\psi}$ parameters; right: $\vec{g}_0$ parameters (colors). Shaded envelopes denote the standard deviation of outcomes over $\theta \in \tilde{\Theta}(\lambda^*)$. For this analysis, we assume that T cells are administered at high dosage ($N_{\text{lin.}} = 10^{-2} S_{\text{max}}$). Simulation parameters: $S_{\text{max}} = 10^7$, $K_S = S_{\text{max}}$, $K_H = 10^{-3} S_{\text{max}}$, $d_I = 0.0 \text{ day}^{-1}$, $I_0 = 10^{-1} S_{\text{max}}$, $b_I = 0.1 \text{ day}^{-1}$.

sensitivities ($\psi_{\text{Ag}}, \psi_{\text{res}}$), or the baseline rates g_0 's does not reproduce the dynamic range observed in the experiments (Figs. A.7, A.8). Improved quantitative agreement may be achieved by adjusting a subset of the model's non-design parameters (Appendix A, Table S1), an analysis that would benefit from a larger dataset to enable robust fitting and validation;

see Appendix A and Figs. A.10–A.12 for the sensitivity of model performance to variation in non-design parameters and to simulation noise.

To align responses with the observed macro-dynamic trends, subsequent analyses consider near-elbow designs $\vec{\theta} \in \Theta(\lambda^*)$ with their infection-signal sensitivity reduced to one-tenth of their original value: for $\psi_{\text{inf}} \mapsto 0.1 \psi_{\text{inf}}$. This ensemble of *tuned* designs is indicated by $\tilde{\Theta}(\lambda^*)$.

Designing an effector response to a growing tumor

CD8⁺ T cells are central to cancer surveillance and elimination of mutated pre-cancerous cells. We therefore asked how the tuned elbow-optimal designs $\vec{\theta} \in \tilde{\Theta}(\lambda^*)$ identified for infectious threats (Fig. 3.4B, Fig. 3.5) respond to cancer, which we modeled as a growing, low-immunogenic tumor that kills healthy cells as it expands (Fig. 3.6A; Appendix A). We quantified harm as the total number of healthy cells lost, combining cell death caused directly by tumor expansion (replacing healthy cells) and collateral killing mediated by the immune response (Fig. 3.6B,C).

Similar to the case of infection control, a clearance-toxicity trade-off arises, but it is sharpened in cancer because many tumor antigens are only marginally distinct from self. Even high-affinity recognition risks cross-reactivity and bystander damage. Tuned near-elbow designs incur high total harm for low-immunogenic tumors (Fig. 3.6B, C). As tumor immunogenicity increases, harm decreases due to reduced cross-reactive toxicity and enhanced tumor clearance (Fig. 3.6C).

Cancer immunotherapy leverages and improves upon the body’s own immune system to target tumors [150]. Patient-specific tumor-infiltrating lymphocytes (TILs) and TCR-engineered T cells (TCR-T’s) purified and expanded ex-vivo, and infused in patients to recognize tumor peptide-MHC’s have shown clinical benefits. Chimeric antigen receptor (CAR) T cells, on the other hand, target surface antigens without MHC restriction and can

be deployed more broadly when targets are shared across patients. Contemporary CARs incorporate co-stimulatory domains (e.g., CD28 or 4-1BB) that enhance activation, proliferation, survival, and effector function [151]. In practice, three levers are pulled to engineer a T cell therapy: (i) the antigen-recognition module (native TCR, TCR-T, or CAR) to specifically target the tumor; (ii) the intracellular signaling architecture (co-stimulation and engineered circuits) to sustain a robust and durable response; and (iii) the infused T cell dose/schedule, tuned to maximize tumor control while limiting toxicity.

Our model predicts that the optimal dose should be tuned to tumor immunogenicity (Fig. 3.6B), consistent with prior findings on antigen-dependent T cell dosing [152]. Examining the tuned near-elbow designs, we find: at low immunogenicity, clearance is inefficient and toxicity dominates, so smaller infusion doses minimize harm; at high immunogenicity, recognition and in-situ expansion are efficient, so even modest doses suffice; at intermediate immunogenicity, larger initial doses minimize total harm by accelerating tumor elimination while keeping toxicity tolerable (Fig. 3.6C).

TCR engineering raises an additional nuance: increasing TCR-tumor affinity often increases affinity to self [153], so that increasing tumor immunogenicity is difficult. To capture this, we evaluated treatment scenarios where the affinities/reactivities of tumor and self to the engineered TCR/CAR are proportional, leaving tumor immunogenicity unchanged ($\kappa = K_I^{-1}K_S = \text{constant}$; Appendix A, Fig. A.9). We quantify affinity or reactivity of a cell by the ratio of the maximum healthy cell counts $S_{\max}$ to the number of target cells required for half-maximal T cell response (EC_{50}), i.e., $K_I^{-1}S_{\max}$ for tumor cells, and $K_S^{-1}S_{\max}$ for self targets. Provided that tumor immunogenicity κ is not too small and self-reactivity remains moderate, the dose-response as a function of tumor antigenicity (Fig. A.9 A) mirrors its dependence on immunogenicity in Fig. 3.6B. Nevertheless, these cross-reactive regimes (Fig. A.9B) yield higher total harm than scenarios where self-interaction is fixed so that increasing tumor antigenicity increases immunogenicity (Fig. 3.6C). This highlights the cost

of affinity-driven cross-reactivity in TCR engineering for cancer therapy [154–156].

Finally, we asked which aspects of the endogenous CD8⁺ T cell program could be modulated to improve tumor control. Using high-dose infusions for a broad range of immunogenicities, we computed the sensitivity of negative harm (i.e., protection) to small changes in design parameter values, i.e., $-\nabla H(\vec{\theta})$ for $\vec{\theta} \in \tilde{\Theta}(\lambda^*)$ (Fig. 3.6D). For very high-affinity T cells (those able to respond at $\leq 1/100$ of the number of healthy cells), gains from further engineering are limited. For intermediate affinities—arguably more relevant clinically—several levers are impactful: increasing activation/proliferation baseline rates ($g_{0,N \rightarrow N^*}, g_{0,E \rightarrow 2E} \uparrow$) [157] and decreasing effector death ($g_{0,E \rightarrow \emptyset} \downarrow$) provide strong improvements but may entail perturbing oncogene-linked pathways (e.g., MYC, BIM) and warrant caution [158]. Boosting antigen and infection sensitivities ($\psi_{\text{Ag}}, \psi_{\text{inf}} \uparrow$), and tempering the anti-inflammatory brake ($|\psi_{\text{res}}| \downarrow$) also enhance clearance, albeit with a risk of runaway inflammation [159]. Increasing the baseline rate of differentiation ($g_{0,N^* \rightarrow E} \uparrow$) yields smaller gains. This suggests that reducing effector differentiation—e.g., via Tcf7 modulation [49, 157]—may provide a tractable knob to promote memory production and reduce relapse risk. Reduced effector differentiation may then be compensated for by tuning effector expansion and longevity.

Interestingly, recent CRISPR screens show that among other candidates, knocking out FAS and PTPN2 genes in CAR T cells reduces apoptosis ($g_{0,E \rightarrow \emptyset} \downarrow$) and redirects differentiation toward effector programs ($g_{0,E \rightarrow 2E} \uparrow$), thereby enhancing CAR T cell efficacy [160–163]. Our modeling framework can help prioritize targets for such screens to rationally identify candidate genes that improve T cell immunotherapies. It should be noted that the achievable benefit of any perturbation for T cell engineering will depend on patient- and tumor-specific constraints. These include the extent to which affinity-driven TCR cross-reactivity can be mitigated (Fig. A.9) and may require patient-specific measurements of T cell interactions and signaling to calibrate designs for robust, durable therapy.

3.4 Discussion

In this work, we modeled the CD8⁺ T cell response as a feedback-controlled program that integrates three cues (i.e., instantaneous antigen, time-integrated infection-derived harm, and time-integrated response-derived harm) to minimize cumulative cell death during an immune challenge. Modulation of response proportional to the antigen signal and the time-integrated harm signals mirrors proportional-integral-derivative (PID) control of engineered systems, a strategy known for its stable and robust steering of dynamical processes toward desired outcomes [164].

Exploring an ensemble of biologically motivated controllers (designs) and optimizing for pathogen clearance while minimizing immunopathology exposed a Pareto trade-off: a reduction in infection harm generally leads to an increased collateral damage to healthy tissues. This pattern persisted across diverse immune challenges defined by their immunogenicity and within-host basic reproduction number.

Three design principles emerged from our analyses: (i) Pre-programmed (baseline-tuned) designs help restrain autoimmunity and slowly proliferating threats but are inefficient against fast-proliferating infections or metastatic tumors; effective control there requires signal integration and feedback. (ii) Among signal sources, antigen and response-harm dominate; sensitivity to infection-harm contributes comparatively little once CD8⁺ effector T cell expansion is underway. (iii) Feedback from response-harm acts as an anti-inflammatory brake, enabling strong antigen-driven responses while limiting collateral damage when the response overshoots—a characteristic consistent with endogenous effector programs [122, 123]. Table 3.1 further summarizes the characteristics of our inferred optimal designs and relates them to supporting evidence from endogenous effector programs.

We posited that effector programs should attempt to solve a multi-objective problem to minimize harm (cell death) inflicted by diverse pathogenic challenges. Interestingly, we

found that a Pareto-optimal design that suppresses responses to frequent, self-like injuries while clearing recognizable, replicating (e.g., acute) infections mirrors the endogenous T cell programs in vertebrates: strongly antigen-driven with an anti-inflammatory brake, and conservative cell-fate transition speeds that prevent runaway effector responses. We note that, beyond minimizing harm due to cell death, additional factors may also shape the objectives optimized by effector T cell programs, including the metabolic costs of effector activity and cytokine production, the cumulative burden of inflammation, and the long-term value of generating memory cells. Although the relative importance of these contributions remains unclear, they could be incorporated and analyzed within the same Pareto-optimization framework we introduced in this work.

Memory formation is a key feature of CD8⁺ T cell immunity, but here we focus on acute infection control and therefore do not model memory expansion or recall explicitly. In endogenous responses, immunological memory arises in part from the differentiation of a small fraction of effector cells ($\sim 5 - 10\%$ of the peak response [45, 49]) into effector-memory cells, effectively removing them from the active response. In our model, this loss of effector cells is captured by the contraction step, which can represent either effector apoptosis or differentiation into memory. By aggregating these processes into a single contraction term, we reduce degeneracy in the design parameters while retaining interpretability. Nonetheless, memory “quality” depends on differentiation timing and clone avidity [50, 116, 165, 166], which would be relevant for the strength and durability of protection upon secondary immune challenges. Extending our framework to an explicit allocation trade-off, where diverting cells to memory reduces primary effector output but seeds future protection, will be an important next step toward linking control designs to memory formation and the evolutionary pressures shaping effector programs.

Using a minimal tumor model, we showed how the design objective informs rational immunotherapy protocols. First, it suggests that infused T cell dosage should depend on tumor

immunogenicity—balancing clearance against collateral damage, as previously emphasized in the literature [150, 152, 155]. Second, treating the design parameters as engineering targets yields testable strategies to improve efficacy and safety, which is highly desirable [167–169]. For moderately immunogenic tumors, promoting effector output by raising baseline rates of activation/proliferation or increasing antigen and infection sensitivities, while tempering the anti-inflammatory brake, may produce robust responses.

To operationalize these ideas, abstract signal sensitivities and baseline rates must be mapped to gene programs as targets of cell engineering (Table 3.1). Combining principled control-theoretic design with patient-specific measurements of antigenicity and damage signaling may enable tailored effector programs that respect the fundamental trade-offs of response from this study, while bending them toward better outcomes [170].

3.5 *Materials and Methods*

Data and code for all the analyses can be found in the Github repository: <https://github.com/StatPhysBio/designimmune>. Previously published data from refs. [109, 148] were used for this work.

Intra-host infection dynamics. We model a population of healthy, susceptible cells $S(t)$, initially consisting of $S_{\max}$ identical cells. Pathogen exposure seeds an infected subpopulation $I(t)$, which grows via encounters with susceptible cells at rate b_I , and declines by infected-cell death at rate d_I , triggering an immune response (Figs. 3.1A,B, A.1). The dynamics of the susceptible and infected cells follow,

$$\begin{aligned}\dot{S} &= -b_I IS - d_E \frac{E}{K_S + I \frac{K_{S+E}}{K_I+E} + E + S} S \\ \dot{I} &= b_I IS - d_I I - d_E \frac{E}{K_I + S \frac{K_{I+E}}{K_S+E} + E + I} I,\end{aligned}\tag{3.9}$$

where d_E is the maximal clearance rate of infected/susceptible cells by effector T cells E , and K_I and K_S are *recognition thresholds*, quantifying how readily effector cells can recognize infected and susceptible cells, respectively, and consequently kill them; a larger K_I or K_S implies more effector cells are needed for recognition [171–173]. The functional form of the last terms in these equations is motivated by considering competitive binding of T cells to infected and healthy cells (Appendix A). Note that these equations describe dynamics during the immune challenge and exclude the slow birth-death processes that homeostatically maintain the susceptible cells in the absence of infection.

Stochastic T cell response model. We model the CD8⁺ T cell response to an immune challenge as a time-inhomogeneous birth-death process coupled to deterministic dynamics of susceptible and infected cells. Beginning with an initial population of monoclonal naïve T cells, we track individual T cell lineages through recruitment of naïve cells to interact with antigen-presenting cells (APCs; $N \rightarrow \mathbf{APC} \cdot N$), activation ($\mathbf{APC} \cdot N \rightarrow N^*$), differentiation to effector ($N^* \rightarrow E$), expansion ($E \rightarrow 2E$), and contraction ($E \rightarrow \emptyset$). Naïve T cells are recruited to APCs at a rate modulated by infection harm signals, and upon sustained engagement, become activated and undergo 2-3 burst-like divisions largely independent of continued antigenic input. During this early proliferative window, each activated cell commits to an effector or memory fate according to the first arrival of a time-inhomogeneous Poisson process with a signal-dependent rate; the number differentiating into effectors is drawn as a binomial random variable, with remaining cells adopting a memory phenotype. Effector cells subsequently divide and die at rates governed by time-varying antigen and harm signals. Proliferation is capped by a maximum division number [107, 109–111], which we set to 15 rounds, beyond which cells undergo terminal arrest (see Appendix A for details).

In our model, inter-lineage variability arises solely from stochastic differences in transition timing, capturing with prior observations [34, 49, 50, 96, 111, 114–116]. To quantify

this stochasticity and demonstrate robustness, results are averaged over five independent full-simulation replicates, with variability reported. The coefficient of variation averaged over immune challenges is small (C.V. = 0.05; Fig. A.12), indicating that macroscopic quantities such as total harm and effector peak are robust to stochasticity in T cell response dynamics. We also characterize sensitivity to simulation parameters (Figs. A.10, A.11), which vary across hosts and infection scenarios.

Infection and response signals. T cell activation relies on the integration of antigen and harm signals. Cell deaths can induce harm in the host and produce pro- or anti-inflammatory signals, depending on the underlying cause. We simplify this by distinguishing infection-induced and response-induced harms, which in a cumulative form, generate signals $\sigma_{\text{inf}}(t)$ and $\sigma_{\text{res}}(t)$, respectively. The instantaneous level of presented antigens, proxied by the infected-cell burden $I(t)$, defines a separate antigen signal $\sigma_{\text{Ag}}(t)$. Together, scaled by their recognition thresholds, K_I for antigens and K_H for non-antigens (harm), the three signals available to T cells follow,

$$\begin{aligned} \text{antigen signal: } \sigma_{\text{Ag}}(t) &= K_I^{-1} I(t) \\ \text{harm signals: } \sigma_{\text{inf/res}}(t) &= K_H^{-1} \int_0^t e^{-(t-s)d_H} h_{\text{inf/res}}(s) \, ds. \end{aligned} \tag{3.10}$$

where $h_{\text{inf}}(s)$ is the number of cells killed by the pathogen (infection), and $h_{\text{res}}(s)$ denotes the number of cells killed by the effector T cell response at time s (eq. 3.3). The time-integrated harm-induced signals are exponentially discounted at rate d_H , reflecting the decay of cytokines in the absence of ongoing stimulation (Appendix A).

Chapter 4

RATIONAL T CELL REPERTOIRE ALLOCATION TO CONTROL THE GUT MICROBIOTA

The contents of this chapter are being prepared for submission to a journal.

4.1 Introduction

The gut microbiome is a complex ecological community that influences its host’s metabolism, development, immunity and other critical physiological processes [174–179]. Given its easy accessibility, it is a promising delivery route for clinical therapies, such as antibiotics, probiotics, and fecal matter transplants (FMT). However, microbiota composition varies among healthy individuals, making it difficult to identify stable drivers of health and disease [179–181]. Recently, it was reported that despite compositional variation, the production of many metabolites is stable across healthy individuals on similar diets [176]. Another study found that host-secreted antibody-coating of strains can reliably identify colitogenic, harmful strains in the gut [177]. These examples suggest certain properties of the microbiota—metabolism, immunity, and perhaps others—are stably regulated by healthy hosts. Still, given the complex ecological interactions in the gut, it is unclear how host health can be reliably influenced by perturbing microbiota composition. Understanding the logic behind host-control of microbiota composition could thus enhance the effectiveness of clinical ther-

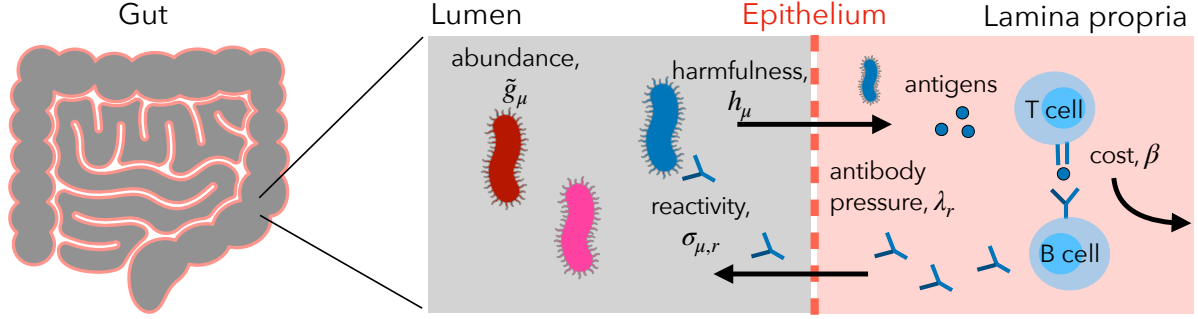


Figure 4.1: A model of adaptive immune control of the gut microbiota. Microbial abundance in the gut lumen ($\tilde{g}_\mu$) is influenced, among other factors, by microbiota ecology and host immunity. As strains cross the epithelium at rate ($h_\mu \tilde{g}_\mu$) and potentially trigger infection, they stimulate cognate T cells that can bind to their antigens. T cells direct B cells to secrete antibodies (λ_r) into the lumen, which bind specifically ($\sigma_{\mu,r}$) to strains and decrease their abundance. The T-cell-dependent immune response may also trigger off-target harms to the host at a per antibody cost β .

apies [12, 182–184].

Few quantitative models have linked the immune response to microbiota composition in the gut [182]. In the digestive tract, environmental and dietary factors determine which strains are likely to be present, limiting the host’s control of microbiota composition (Fig. 4.1). Colonizing strains become entrenched by attaching to the gut epithelium and may harm their host by breaching this barrier. Hosts, in turn, influence the abundance of invaders by modulating their internal environment through metabolism and immunity, not unlike the action of clinical pro- and anti-biotic treatments. The immune response recognizes microbial signatures/epitopes, and can inhibit microbe growth through inhibitory mechanisms such as secretion of bactericidal molecules and antibodies, or tolerate strains to maintain their abundance [174, 177, 185]. The balance between inhibition and tolerance mechanisms directed towards a microbe is thought to influence its abundance in the gut, with certain strains reliably inducing tolerizing or inhibitory immune responses [177, 185–198]. Conversely, given the complex microbial ecology in the gut, and the likelihood of runaway immune responses triggering gut dysbiosis, allergies or other pathologies, it is unclear how immune mechanisms rationally control microbiota composition [174, 192, 199–205].

Here, we develop a quantitative model linking the T cell repertoire to microbial composition in the gut. Our approach is motivated by, and leverages data from, recent experiments that measured immune interactions with strains in germ-free mice inoculated with a defined microbial community [179, 206, 207]. We find that the immune response partly determines microbial abundances, but that other factors, such as ecological interactions between strains, may also matter. Our model further rationalizes the experimentally observed pattern of sparse yet broad T-cell microbe interactions, showing it to be optimal for microbial control when maintaining an immune repertoire is costly. We demonstrate that this optimal repertoire allocation is biologically plausible and can emerge from competition between T cells for microbe-derived epitopes. Finally, we propose experiments to further refine the principles governing immune control of gut microbiota composition.

4.2 Model

How can the T cell repertoire optimally control the composition of the gut microbiota to maintain homeostasis? To answer this question, we developed a model wherein a T-cell-driven immune response is allocated to jointly minimize the risk of microbe-driven systemic infection and the cost of maintaining the repertoire (Fig. 4.1). We quantify each microbe’s engagement with the immune system through its immune coverage $C_\mu(\vec{\lambda})$ (for $\mu = 1, \dots, M$) which modulates microbial abundances:

$$\text{immune coverage:} \quad C_\mu(\vec{\lambda}) = \sum_{r=1}^R \sigma_{\mu,r} \lambda_r \quad (4.1)$$

Here, λ_r (for $r = 1, \dots, R$) is the inhibitory pressure exerted by the immune cell (here, a T cell) carrying receptor r , and $\sigma_{\mu,r}$ is the reactivity between microbe μ and the immune receptor r , defining a non-negative $M \times R$ interaction/reactivity matrix $\Sigma = [\sigma_{\mu,r}]$ (Fig. 4.1). Mechanistically, this pressure (λ_r) corresponds to the magnitude of T-cell-dependent,

high-affinity immunoglobulin A (IgA) secreted by B cells at the gut mucosal surface, which can bind to a cognate epitope expressed by targeted strains and inhibit their growth [177, 208, 209]. The coverage C_μ thus reflects the cumulative inhibitory pressure from all such antibodies generated by the repertoire targeting detectable epitopes of microbe μ .

Given this immune coverage, we model the abundance g_μ of each microbe $\mu = 1, \dots, M$ with logistic dynamics. Specifically, by exploiting niches in the gut, microbe μ can grow to a carrying capacity $g_\mu^{\max}$ that is modulated both by its repertoire/immune coverage $C_\mu(\vec{\lambda})$ and ecological interactions $s_{\mu,\alpha}$ with other strains $\alpha \neq \mu$,

$$\dot{g}_\mu = g_\mu \left[\frac{1}{1 + K_C^{-1} \left(C_\mu(\vec{\lambda}) + \sum_{\alpha \neq \mu} s_{\mu,\alpha} g_\alpha \right)} - \frac{g_\mu}{g_\mu^{\max}} \right] \quad (4.2)$$

where K_C is the half-maximal constant of inhibition of microbe abundance. We note that this formulation assumes a well-mixed population of strains and immune cells. Given the spatial heterogeneity and the separation of microbial and immune niches in the gut, this is a strong assumption that should be addressed in future work.

Because immune repertoire dynamics are slow (weeks [177, 207]) relative to microbial turnover (days), we treat the repertoire coverage C_μ as a slow variable that sets the steady-state abundances $\tilde{g}_\mu$ (for $\mu = 1, \dots, M$), holding it fixed over the fast reorganization of microbial composition. Neglecting ecological interactions between strains yields the simplified expression:

$$\tilde{g}_\mu(C_\mu) = \frac{g_\mu^{\max}}{1 + K_C^{-1} C_\mu}. \quad (4.3)$$

Each microbe μ has the potential to inflict systemic harm on the host by breaching the gut epithelium at rate $h_\mu \tilde{g}_\mu$, with the constant $h_\mu > 0$ reflecting the microbe's potential harmfulness (Fig. 4.1). For instance, harmfulness could be related to the rate at which a

microbe breaches the gut epithelium [177, 195, 210]. We assume that the repertoire response is mobilized to minimize harm by controlling the steady-state population of strains. Given the reactivities $\Sigma = [\sigma_{\mu,r}]$ between strains (μ) and receptors (r), we seek a repertoire that minimizes the objective

$$\min_{\vec{\lambda} \geq 0} F(\vec{\lambda}; \Sigma) = \min_{\vec{\lambda} \geq 0} \left\{ \sum_{\mu=1}^M h_{\mu} \tilde{g}_{\mu}(\vec{\lambda}) + \beta \sum_{r=1}^R \lambda_r \right\}, \quad (4.4)$$

where the first term on the right represents the cumulative steady-state rate of harm inflicted by the microbiota, and the second term captures the cost of maintaining the immune repertoire, arising from increased risk of autoimmunity associated with having more receptors, the difficulty of generating particular receptors, competition among receptor-bearing cells for limited survival signals and/or metabolic costs [174, 203, 211] (Fig. 4.1). The constant β specifies the marginal cost of adding to the repertoire, and the linear form of the second term is chosen for simplicity. Readers may also recognize that the linear form of the second term, chosen for simplicity, corresponds to a sparsity promoting L1 regularizer [83]. Here, we solve this convex optimization problem in several illustrative cases to explain features of the experimentally observed CD4+ T cell repertoire in the gut (Appendix B).

4.3 Results

Sparse, generalist T cell coverage of the gut microbiota.

In recent experiments by Fischbach and colleagues, germ-free mice were inoculated with a 97-member microbial community ("hCom1") [179, 206, 207]. Community composition from fecal samples was profiled at weekly intervals, and the responding T cell repertoire was profiled two weeks after inoculation by harvesting gut-associated lymph nodes. Responding T cell clones/receptors found in lymph nodes were then systematically sequenced and profiled

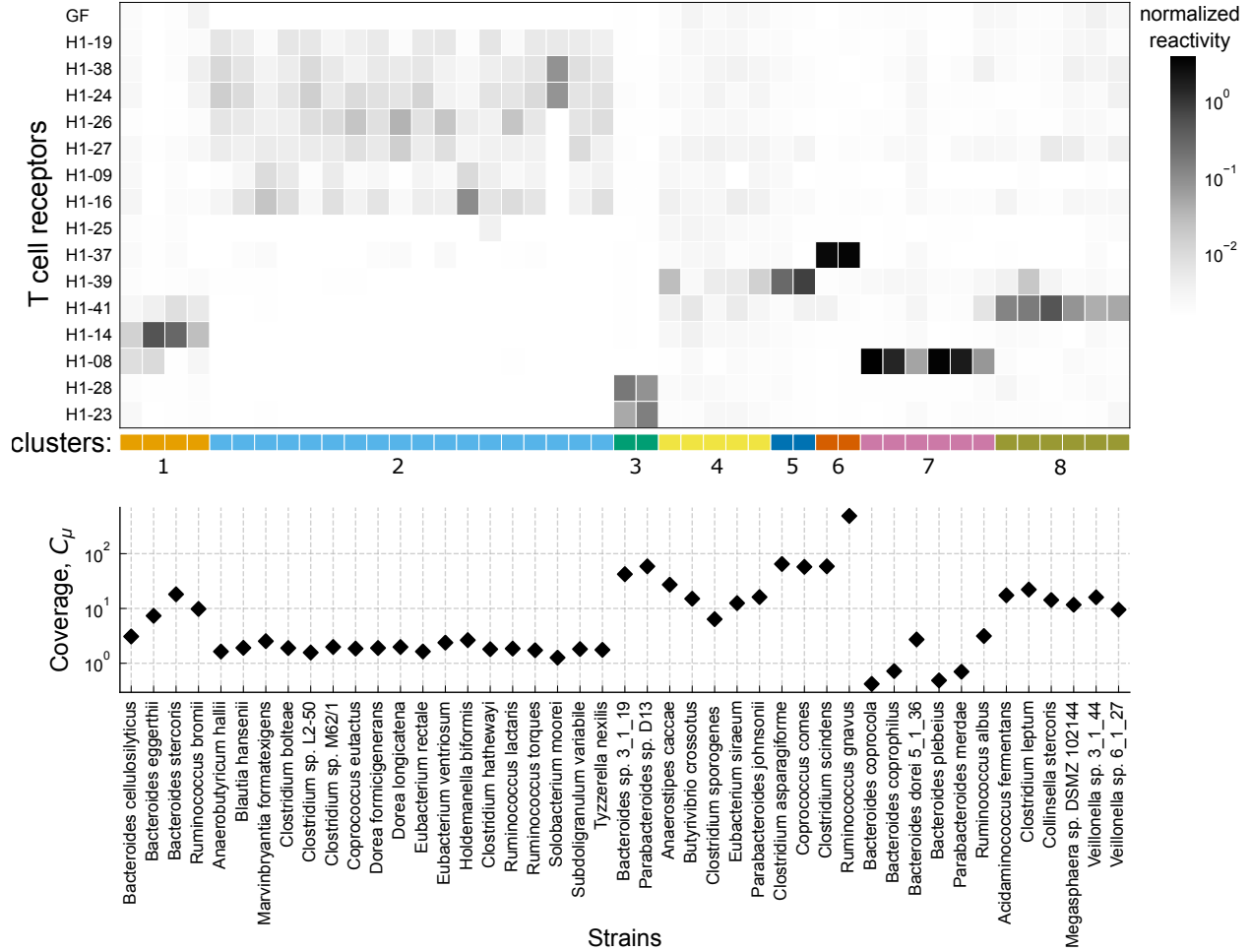


Figure 4.2: T cell repertoire reactivity and coverage in a complex microbial community. (A) Top: Heatmap showing strain reactivity (secreted IL-2 concentrations; columns L-1 normalized) of 45 microbial strains from a 97-member community (hCom1) to 15 distinct T cell receptors. The “GF” row provides a baseline for non-specific responses. K-means clustering based on these reactivities identifies 8 reactive groups (clusters 1-8), while 52 non-reactive strains (cluster 0) are omitted. Bottom: Estimated repertoire coverage (C_μ) per strain calculated as a reactivity-weighted ratio of effector-to-regulatory cells (eq. 4.6). Coverage varies significantly across strains, with potential pathobionts receiving higher coverage than commensals.

for their reactivity to individual strains in the community, generating an empirical reactivity matrix (Fig. 4.2, top). Individual T cell clones recognized food-derived epitopes or limited groups of taxonomically diverse strains expressing a common epitope, some of which could be directly linked to strain metabolism. Furthermore, taxonomically related strains, which are expected to express similar epitopes, were often not targeted by the same T cell clone.

Thus, the T cell repertoire consisted of generalists, which sparsely recognized taxonomically diverse sets of strains.

T cells responding to gut strains are phenotypically heterogeneous, even among cells sharing the same T cell receptor $r \in \{1, \dots, R\}$ (i.e., belonging to the same clone). Within a single clone, cells differentiate into regulatory subsets (Treg: tTreg and pTreg), effector subsets (Teff: Th1, Th2, Th17, and other effectors), memory cells, and additional phenotypes (Appendix B, Fig. B.1). Assuming each clone recognizes a specific microbe-expressed epitope through its unique TCR r , the balance of effector to regulatory cells within that clone reflects the net immune pressure it exerts on its target [186–192, 212, 213]. To relate our theoretical model to data, we define the immune pressure from a clone with receptor r as

$$\lambda_r := [\text{Teff}]/[\text{Treg}]. \quad (4.5)$$

We exclude memory and other cell types from the analysis since their roles in the immune response are more subtle. Using eq. 4.5 for immune pressure induced by each T-cell and an empirically measured reactivity matrix $\Sigma = [\sigma_{\mu,r}]$ between strains μ and receptors r (Fig. 4.2, *top*), we can then evaluate the repertoire coverage of microbe μ as the reactivity-weighted sum of clonal pressures $C_\mu = \sum_r \sigma_{\mu,r} \lambda_r$.

To evaluate microbe-specific coverage in practice, we turned to data from Nagashima and colleagues, who measured reactivities for $R = 15$ T-cell clones against $M = 97$ microbial strains [206]; 45 of these strains elicited a response and form the basis of our analysis (Fig. 4.2, *top*). The remaining 52 strains were nonreactive to all sampled clones, and additional profiled clones reactive to food-derived epitopes or taken from germ-free mice were likewise excluded. Two features of the data complicate the direct use of eq. 4.2 to compute immune coverage. First, cells were pooled from three replicate mice, so we cannot determine whether the multiplicity of T cell clones reactive to similar groups of strains (Fig. 4.2) reflects a

genuine within-host signature or an artifact of combining distinct repertoires. Second, per-clone Teff and Treg counts ranged from zero to 15, making immune pressure λ_r (eq. 4.5) measured as Teff-to-Treg ratio noisy, and undefined when $[\text{Treg}] = 0$ (Appendix B, Fig. B.1).

To mitigate both issues, we aggregated counts across clones before taking the ratio, approximating the relative repertoire coverage as

$$C_\mu \approx \sum_{r=1}^R \sigma_{\mu,r}[\text{Teff}] \bigg/ \sum_{r=1}^R \sigma_{\mu,r}[\text{Treg}]. \quad (4.6)$$

This expression retains the interpretation of coverage as a reactivity-weighted Teff-to-Treg ratio while reducing sensitivity to sparsely sampled clones (Appendix B, Fig. B.1). Using this measure, coverage varied substantially across strains (Fig. 4.2, *bottom*). For instance, *Bacteroides coprocola*, a commensal, exhibited the lowest coverage, while *Ruminococcus gnavus*, a potential pathobiont, exhibited the highest [214, 215].

Targeted repertoire coverage controls microbiota composition.

Recent research has shown that strain clustering approaches based on genetic similarity can predict the metabolic outputs of complex communities [181, 216]. Here, we asked if clustering strains based on their reactivities to the shared repertoire could identify the relationship between composition dynamics of the hCom1 community and immune coverage. Clustering, we identified nine microbial groups (0-8) in hCom1 (Fig. 4.2, *middle*; Appendix B), whereby strains in cluster 0 (52 total) were nonreactive to experimentally extracted T cells, while those in clusters 1-8 (45 total) were reactive to at least one T cell clone [206]. Clusters varied in their relative abundance and exhibited distinct and reproducible compositional dynamics (Fig. 4.3A and Appendix B, Fig. B.2), even as they contained multiple taxa (Fig. 4.2 and Appendix Fig. B.2). Strains of the same taxon were often dispersed across different

clusters (Fig. 4.2), and clusters were dissimilar in taxon membership when measured by their overlap in strain genera (Fig. 4.3B). Such immune-based clustering is distinct from taxonomic relationship between strains and defines an aggregation of taxonomically diverse strains in the microbiota based on their potential immune responses.

In vitro, the composition of hCom1 stabilizes within 48 hours in the absence of perturbations [207], while in mice inoculated with a different community, the T cell response takes weeks to develop [177]. We therefore interpret changes from week 0 to week 1 as reflecting ecology and innate immune responses alone, and changes from week 1 to subsequent weeks as additionally shaped by the adaptive immune response. First, we define the repertoire coverage of the microbial cluster k , as the strain-abundance-weighted average of repertoire coverage of its constituent strains using relative abundances $g_\mu / \sum_\alpha g_\alpha$ at week 2, when the T cell repertoire was profiled (Appendix B, Fig. B.2)

$$\langle C_\mu \rangle_k = \frac{1}{\sum_{\mu \in \text{cluster } k} g_\mu / \sum_\alpha g_\alpha} \sum_{\mu \in \text{cluster } k} \frac{g_\mu}{\sum_\alpha g_\alpha} C_\mu. \quad (4.7)$$

Next, we quantify cluster-level compositional shifts in the microbiota by computing the centered log-ratio (CLR) transform of cluster relative abundances at each time $t > 0$. For each cluster $k = 0, \dots, 8$, we have

$$\text{CLR}_k(t) = \log_{10} \left(G_k(t) / \sqrt[9]{\prod_{j=0}^8 G_j(t)} \right), \quad (4.8)$$

where $G_k(t) = \sum_{\mu \in \text{cluster } k} g_\mu(t) / \sum_{\mu=1}^M g_\mu(t)$ is the relative abundance of cluster k . This transformation minimizes spurious correlations introduced by the sum-to-one constraint of relative abundance data of the M microbial strains. CLR can be computed analogously for individual strains (Appendix B). Changes in abundance were then computed between sampled time points (t_1 to t_2) following inoculation, $\Delta \text{CLR}_k(t_1, t_2)$.

Plotting the strain cluster compositional shifts $\Delta\text{CLR}_k(1, 2)$ from week 1 to week 2 shows a strong inverse relationship with log-cluster coverage $\log\langle C_\mu \rangle_k$ ($R^2_{\text{cluster}} = 0.66$; Fig. 4.3C), indicating that higher immune coverage more strongly constrains the abundance of a strain cluster. This relationship persisted when comparing week 1 to weeks 3 and 4 and was robust to dropping individual clusters or biological replicates (Appendix B, Fig. B.3). Notably, clusters 1 and 2, which are the most abundant clusters at week 1, have the lowest coverages and the largest CLR increases (Fig. 4.3A), consistent with previous reports that strains in these clusters are likely commensals engaged in nutrient metabolism [206]. Furthermore, the predominant ($> 60\%$) strains in these clusters are *Bacteroides cellulosilyticus* (commensal) and *Clostridium hathewayi* (commensal/opportunistic) [217, 218]. On the other hand, cluster 6 exhibits the highest repertoire coverage, the largest CLR decrease and persisted in the microbiota (Fig. 4.3A). This cluster contains *R. gnavus*, a well-known pathobiont associated with inflammatory bowel disease (IBD) [215]. Thus, high-coverage clusters were suppressed but not eliminated, suggesting costs or diminishing returns to coverage. Together, these results indicate that T cell repertoire coverage distinguishes commensal from harmful strains and shapes microbiota composition accordingly. However, the inverse relationship observed at the cluster level is absent at the level of individual strains (Fig. 4.3D), a disparity we investigate later in this work.

Model of repertoire allocation explains repertoire sparsity and establishment of generalist immune receptors

Next, we asked whether our model could explain two features of observed repertoires in the gut: the prevalence of sparsely binding generalists and the variability in coverage across microbial strains. For these analyses, we make the convenient change of variables, $K_C^{-1}C_\mu \mapsto C_\mu$, expressing repertoire coverage in units of the unknown half-maximal constant of inhibition (K_C). First, we consider the simplest case wherein a single receptor is available to cover

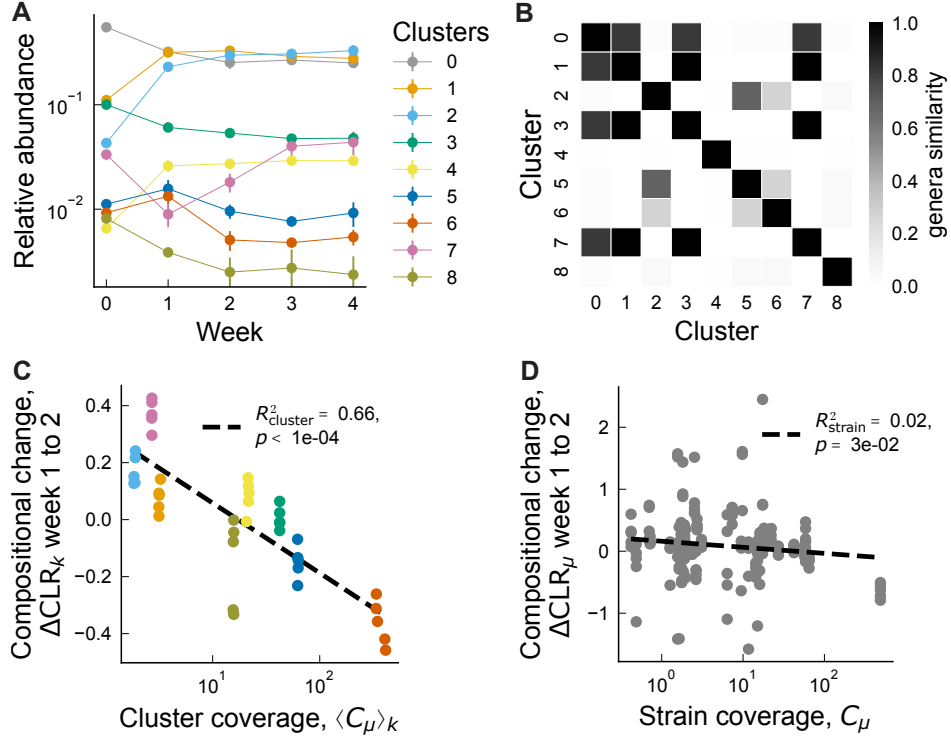


Figure 4.3: Repertoire coverage predicts compositional shifts of strain clusters. (A) Longitudinal tracking of average cluster relative abundances over 4 weeks following inoculation. Error bars indicate 1 standard deviation across 5 biological replicates. (B) Bray-Curtis similarity heatmap showing low taxonomic (genera) overlap between all 9 clusters. (C) Scatter plot showing relationship between cluster repertoire coverage ($\langle C_\mu \rangle_k$) and the change in centered log-ratio (ΔCLR ; eq. 4.8) from week 1 to week 2 ($R^2_{\text{cluster}} = 0.66$). Colors are as in (A) and points of the same color are biological replicates. Dotted line represents fit to $\Delta \text{CLR}_k = m \log_{10}(\langle C_\mu \rangle_k) + b$. Statistical significance was quantified through a hierarchical permutation test by shuffling cluster and biological replicate labels $n = 10000$ times while preserving the grouping of replicates. (D) No relationship between individual strain composition changes (ΔCLR_μ) from week 1 to 2 and repertoire coverage (C_μ ; $R^2_{\text{strain}} = 0.02$). Points indicate strains and their biological replicates. Dotted line represents fit to $\Delta \text{CLR}_\mu = m \log_{10}(C_\mu) + b$. Statistical significance was quantified as in (C) at the strain level.

each microbe and each receptor is reactive to a single microbe. For a given microbe-receptor pair with reactivity $\sigma \geq 0$, repertoire coverage and microbe abundance are given by $C = \sigma \lambda$ and $\tilde{g}(\lambda) = g^{\max}(1 + \sigma \lambda)^{-1}$ (eq. 4.3), respectively; $g^{\max} > 0$ is the maximum achievable abundance of the microbe. The optimal solution which minimizes the costs of microbial

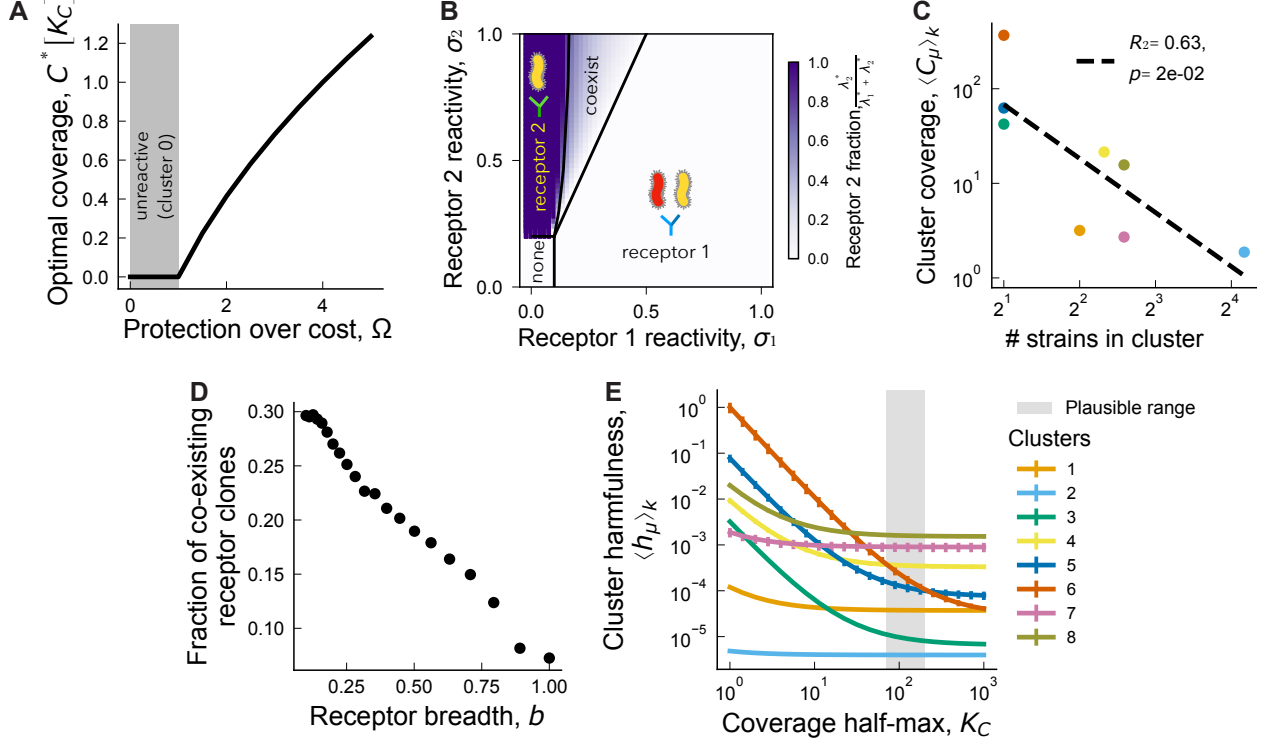


Figure 4.4: Model of repertoire allocation explains strain clustering and infers harmfulness of strain clusters. (A) Optimal repertoire coverage (C^*) as a function of the marginal protection-to-cost ratio ($\Omega = \sigma h g^{\max} \beta^{-1}$) in a repertoire where each receptor is reactive to a single strain. When this ratio falls below one, no receptor is deployed. (B) Phase diagram of receptor co-existence for two receptors as a function of their reactivities for two strains. Receptor 1 is a generalist equally reactive to both strains (σ_1) while receptor 2 is a specialist reactive to a single strain (σ_2). The heatmap shows the fraction of receptor 2 in the optimal repertoire ($\lambda_2^*/(\lambda_1^* + \lambda_2^*)$). The region of viability for receptor 2 is highly restricted because receptor 1 tends to out-compete it in the optimal repertoire. Parameters: $h_1 = h_2 = 5$. (C) Relationship between cluster repertoire coverage and cluster diversity in hCom1. Each colored [as in (E)] dot represents a cluster, and the dotted line shows a log-log fit between coverage ($\langle C_\mu \rangle_k$) and number of strains ($|k|$), $\log_{10}(\langle C_\mu \rangle_k) = m \log_{10}(|k|) + b$, in each cluster ($k = 1, \dots, 8$). (D) Co-existing fraction of the potential repertoire as a function of average receptor breadth (b). A random of $M \times R$ reactivity matrix, $\{\Sigma\}_b$, are generated by drawing R columns corresponding to a receptor and its strain reactivities, $\vec{\sigma}_r \sim \text{Unif}[0, 1] \times \text{Binomial}(M, b)$. Solving eq. 4.4, optimal repertoire coverage is evaluated for each matrix and averaged over an ensemble of random matrices. As receptor breadth (b) increases, the fraction of the available repertoire deployed decreases. (E) Relative harmfulness of strain clusters (eq. 4.11) as a function of the half-maximal constant of immune inhibition of microbial abundance (K_C). The gray region denotes the ± 1.96 standard deviation parameter range obtained by fitting a curve to data in Fig. 4.3C (Appendix, Fig. B.4). Unless otherwise specified, parameters are $M = 10$, $R = 20$, $h_\mu = g_\mu^{\max} = \beta = 1$ for $\mu = 1, \dots, M$.

pathogenicity and immune response in the host (eq. 4.4) satisfies,

$$C^* = \max \left[\Omega^{\frac{1}{2}} - 1, 0 \right] \quad \text{with,} \quad \Omega = \frac{\sigma h g^{\max}}{\beta} \quad (4.9)$$

where $h > 0$ is the harmfulness of the microbe, and β is the marginal cost of keeping a receptor in repertoire, and thus, Ω is the protection-to-cost ratio. A similar square-root dependence of the immune coverage has been obtained in models of optimal immune memory allocation and optimal security screening, where it was interpreted as flattening the distribution of allocated memory/surveillance [129, 219]. Here, it reflects the diminishing returns of marginal increases of repertoire coverage (eq. 4.3).

Following eq. 4.9, a receptor is allocated, i.e., $C^* > 0$, if $\Omega > 1$ (Fig 4.4A); that is, if its protective potential is greater than its cost. Recall that previous experiments reported only 45 out of 97 strains were strongly reactive to detected T cells and that this reactive subset did not include many high abundance strains in the microbiota [206]. Similar observations have been made for immunoglobulin A (IgA) reactive strains in the gut [177]. In the non-reactive cluster (0), the three most abundant strains in cluster 0 (*Bacteroides* sp., *Bacteroides thetaiotaomicron* and *Akkermansia muciniphila*) together account for nearly 50% of this cluster's abundance, and are well-established human gut commensals [220, 221], against which mounting an immune response may not justify the cost. With the caveat that not all responding T cells may have been sampled, this suggests that responses to the nonreactive strains may be insufficiently cost-effective to warrant allocation.

Next, we asked if competitive exclusion between T cell clones could explain the sparsity of the observed repertoires, where a few generalist clones cover taxonomically diverse strains. Neglecting ecological interactions, we considered two strains $\mu \in \{1, 2\}$, and two T cell clones with receptors $r \in \{1, 2\}$, coupled by a fixed 2×2 reactivity matrix Σ ; we allow a receptor to be reactive to both strains. For each microbe-receptor pair (μ, r) , we can define

a protection-to-cost ratio, $\Omega_{\mu,r} = \sigma_{\mu,r} h_r g_{\mu}^{\max} / \beta$. As in previous case, we found that whenever the reduction in harm from deploying either receptor exceeds the marginal cost of immunity $\sum_{\mu \in \{1,2\}} \Omega_{\mu,r} > 1$ for $r \in \{1,2\}$, it is optimal to allocate at least one receptor to cover the strains: i.e., $\lambda_1 + \lambda_2 > 0$ (see Appendix B for derivation). Furthermore, for both receptors to co-exist in the repertoire, however, neither can be more reactive to both strains than the other. Because satisfying this constraint requires fine-tuning of receptor reactivities, under generic conditions, the optimal allocation favors generalist receptors that recognize common epitopes shared across strains (Fig. 4.4B and Appendix B, Fig. B.4). This parallels prior findings that broadly neutralizing antibodies, rather than antigen-specific ones, are favored when responding to diverse viral strains that share common epitopes [222, 223].

In the case of an optimal repertoire with a single generalist receptor, if the reactivity of the generalist receptor r^* is approximately equal for both strains $\sigma_{\mu,r^*} \approx \sigma > 0$ (for $\mu \in \{1,2\}$), the coverage of the two strains by the generalist receptor in the optimal repertoire follows

$$C^* \approx \max \left[\left(\sum_{\mu} \Omega_{\mu,r^*} \right)^{1/2} - 1, 0 \right] \quad (4.10)$$

where Ω_{μ,r^*} is the protection-to-cost ratio for each microbe-receptor pair, and depends on the potential harmfulness of the targeted strains h_1, h_2 . Assuming comparable receptor reactivities σ and microbial niche sizes (maximum abundance) $g^{\max}$, we can ask how a microbe's harmfulness shapes immune coverage. A highly harmful microbe ($\Omega > 1$) can on its own justify a dedicated receptor against it in the repertoire, as implied by eq. 4.9. Less harmful strains, by contrast, can collectively justify a generalist receptor r^* targeting their shared epitope whenever $\sum_{\mu} \Omega_{\mu,r^*} > 1$, as implied by eq. 4.10.

These results suggest that immune-reactive clusters containing many diverse strains tend to contain less harmful strains that justify immune coverage collectively, whereas clus-

ters with few strains tend to be associated with more harmful strains that justify coverage individually. This is indeed what we observe in hCom1, where the number of strains in a cluster is strongly inversely related to the repertoire coverage averaged over the cluster’s constituent strains (eq. 4.7; Fig. 4.3C). Although these conclusions rest on a simplified model of repertoire reactivities, computational simulations and analysis of more complex microbiota-repertoire compositions with arbitrary reactivity generalize these findings (Appendix B). In particular, Fig. 4.3D shows that as the fraction of strains in the microbiota that the average receptor reacts to (receptor breadth, $b \in [0, 1]$) increases, the fraction of the repertoire deployed decreases and the efficiency of immune control increases (Appendix B Fig. B.4).

Taken together, our results suggest that through competitive exclusion, the availability of broadly expressed epitopes biases the repertoire towards a subset of minimally overlapping (i.e., sparse), broadly-reactive receptors. Indeed, accounting for TCR cross-reactivity, the experimentally observed repertoire does not contain any receptors that bind to just one microbe and there is little overlap in receptor reactivity patterns between clusters (Fig. 4.2). Competitive exclusion has similarly been proposed as a principle shaping optimal memory and regulatory T cell repertoires [66, 129]. This model of repertoire allocation is biologically plausible as its optimal solutions can be mapped to the steady-states of a model for consumer-resource competition between T cell clones for limited microbe epitopes (Appendix B) [66, 129]. In doing so, the model recapitulates the biological intuition that the visibility of microbial epitopes to the immune system is determined by factors beyond microbial abundance; here described by the harmfulness constant, h_μ .

T cell repertoire coverage reflects cluster-level microbial harm

We next sought to infer a measure of the relative harm of microbial clusters from their T cell repertoire coverage. While the physiological roles of strains in complex communities are difficult to determine directly, the immune response provides a promising readout for

distinguishing harmful from commensal strains [177, 182, 185, 195, 198]. Generalizing eqs. 4.9 and 4.10 to an arbitrary cluster of strains co-reactive to the same T cell, we relate the expected harmfulness of strains in cluster k to the observed cluster coverage $\langle C_\mu \rangle_k$ as,

$$\langle h \rangle_k(K_C) \approx \beta \frac{(K_C^{-1} \langle C_\mu \rangle_k)^2 + 1}{\sigma_k \sum_{\mu \in \text{cluster } k} g_\mu^{\max}}, \quad (4.11)$$

where K_C is the (unmeasured) half-maximal constant of abundance inhibition by immune coverage (eq. 4.2), β is the (unmeasured) marginal cost of immunity, σ_k is the constant reactivity of strains within the cluster to a common single receptor, and $g_\mu^{\max}$ is the maximum niche size of microbe μ (see Appendix B for a detailed derivation).

The cost of immunity β and total microbial abundance at week 1 are unknown. For simplicity, we assume both are uniform across clusters, in which case they cancel when harm is scaled by its maximum to compute relative harm. The reactivity of strains within the cluster (σ_k) is approximated using data from Fig. 4.2 as the cluster-averaged maximum reactivity of strains within the cluster to all receptors r in the repertoire, $\sigma_k \approx \langle \max_r \{ \sigma_{\mu,r} \} \rangle_k$. The pre-coverage niche size of each cluster in the denominator of eq. 4.11 is approximated as proportional to its relative abundance at week 1 with data from Fig. 4.3A. We then systematically varied the remaining unknown parameter in eq. 4.11—the constant K_C —which determines how strongly coverage determines the harmfulness of a cluster in eq. 4.11 (Fig. 4.4E): at high K_C , harmfulness tracks inverse niche size alone (abundance at week 1); at low K_C , harmfulness tracks both microbial abundance and repertoire coverage.

We note that although the precise value of K_C is unknown, regulatory T cells are known to constitute 5 – 10% of peripheral CD4+ T cells in healthy individuals, suggesting a homeostatic baseline ratio of conventional to regulatory T cells (i.e., immune coverage) of $C^{\text{low}} \approx 10 - 20$ [188, 224]. The half-maximal constant K_C is expected to exceed this baseline level. Using eqs. 4.3 and 4.8, we derived an analytical expression for the compositional

change (ΔCLR_k) from pre- to post-immune coverage which was fit the data in Fig. 4.3C. We obtained a plausible range (± 1.96 standard deviations) for the value of $K_C \approx 70.7 - 204$ (Fig. 4.4E and Appendix B, Fig. B.4), which compares favorably to the known homeostatic baseline. Within this plausible range for K_C , clusters 1-3 (commensals) are robustly inferred as low-harm, and clusters 5 and 6 (pathobionts) as high-harm across a broad range of K_C (Fig. 4.4E). This is consistent with prior work showing that *Ruminococcus*, *Coprococcus*, and *Clostridium*—taxa in clusters 5 and 6—are highly IgA-coated in fecal samples from patients with IBD [177]. Experiments that simultaneously titrate coverage and measure abundance changes could pin down K_C and tighten these inferences [185]. Finally, we note that the harmfulness inferred here cannot be unambiguously linked to the tendency of a microbe to breach the gut epithelium, and is likely only one of many factors determining the abundance of strains in the host gut.

Microbial ecology can limit immune control of microbiota composition.

A puzzling feature of our data is that the strong inverse relationship observed between cluster compositional changes and repertoire coverage ($R^2_{\text{cluster}} = 0.66$; Fig. 4.3C) all but disappears when clusters are disaggregated into their constituent strains ($R^2_{\text{strain}} = 0.02$; Fig. 4.3D). Repertoire coverage thus predicts the abundance dynamics of strain clusters (and their targeted epitopes), but not that individual strains within them. One explanation is spatial compartmentalization: because mucosal immune responses are localized to sites of epithelial breaches [225], strains far from the breach that express the targeted epitope may escape immune pressure. Evaluating this hypothesis requires experiments that collect spatial data on microbial composition together with direct measurements of microbe–T cell interactions, for instance, using methods such as LIPSTIC labeling [226].

Nonetheless, it is striking that the discrepancy is concentrated among low-abundance strains. Progressively removing the lowest-abundance strains from each cluster restores

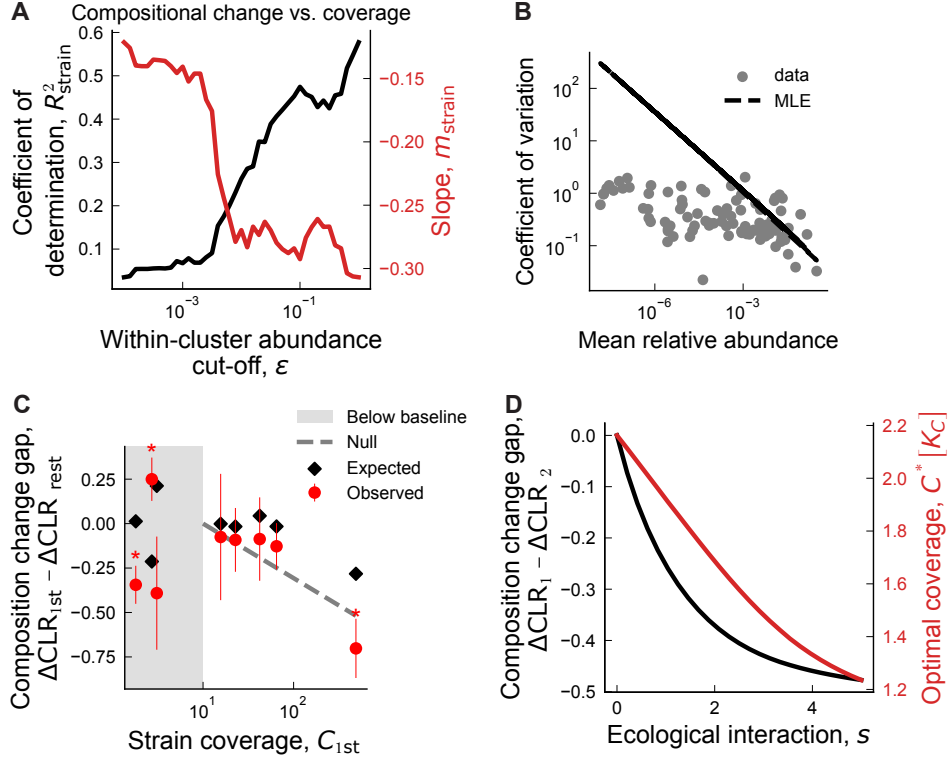


Figure 4.5: Microbial ecology can modulate the precision of immune control. (A) Recovery of strain-level coverage-abundance relationship (R^2_{strain} ; left) and regression slope (m_{strain} ; right) as rarer strains whose relative abundance falls below ϵ times that of the most abundant strains in that cluster are removed. Regression analysis as in Fig. 4.3D, $\Delta\text{CLR}_{\mu} = m \log_{10}(C_{\mu}) + b$. (B) Relationship between coefficient of variation (CV) and mean suggests non-neutral sampling noise. The relationship between replicate mean and CV (gray data points) deviates from the linear dependence expected under a Dirichlet noisy sampling model (black line, MLE fit). (C) The composition change gap ($\Delta\text{CLR}_{1\text{st}} - \Delta\text{CLR}_{\text{rest}}$) between the most abundant strain and all other strains in a cluster plotted against the immune coverage of the most abundant strain in each cluster ($C_{1\text{st}}$). Grey region indicates clusters whose observed coverage was below the biological baseline ($C^{\text{low}} = 10$). Gray dotted line and black diamonds indicate the expected gap for each cluster estimated using the fit in A (at $\epsilon = 1$) and coverage data from Fig. 4.2. For the gray line, only the most abundant cluster is assumed to be covered by the immune response and the coverage of all other strains set to the biological baseline, $C^{\text{low}} = 10$. For the black diamonds, it is assumed that the observed coverage reflects the true coverage of each strain. Red points indicate the observed gap and error bars indicate 1 standard deviation. Statistical significance of gap ($p < 0.05$; *) was determined through a paired t-test of 5 biological replicates and p -values were corrected for multiple testing (Benjamini-Hochberg). (D) Numerical solution (eq. 4.4) of a minimal two-strain system under shared immune pressure where strain 1 suppress the abundance of strain 2. Left axis (black): optimal composition change gap as a function of ecological competition s between two strains (eq. 4.13); right axis (red): optimal coverage (C^*) of both strains. Parameters: $h_1 = h_2 = 5$, $g_{\mu}^{\text{max}} = \beta = 1$.

the coverage–abundance relationship at the strain level (Fig. 4.5A). To test whether this low-abundance discrepancy simply reflects stochastic measurement noise, we fit a Dirichlet distribution to the biological replicate data, a standard model for noise in compositional measurements [227]. The model fit was poor, systematically over-estimated the observed replicate variance (Fig. 4.5B). This suggests that variation across biological replicates might not be explainable by measurement noise alone.

We next quantified the systematic divergence within clusters by calculating the composition change gap ($\Delta\text{CLR}_{\text{1st}} - \Delta\text{CLR}_{\text{rest}}$) between the most abundant strain (“1st”) and the remainder of the cluster (Fig. 4.5C). Because individual strains within a cluster experience different amounts of coverage (Fig. 4.2) their independent compositional changes are expected to diverge. Using the empirical fit from Fig. 4.5A (with $\epsilon = 1$), we constructed a null model where only the most abundant strain in a cluster interacts with repertoire coverage while the other strains experience tolerant baseline coverage ($C^{\text{low}} = 10$; Fig. 4.5C) [188, 224]. Additionally, with the same empirical fit, we constructed a model of the expected compositional change gap for each cluster based on observed difference between the immune coverage of the most abundant strain and the abundance-weighted mean immune coverage of the remainder strains (Fig. 4.5C). Intriguingly, the observed gap was frequently negative, deviating from the null model and exceeding the expected gap for highly covered clusters (Fig. 4.5C). This effect is most pronounced in cluster 6, which contains two strains involved in arginine metabolism [179]. Here, the dominant pathobiont *R. gnavus* declines relative to the entire microbiota while its less abundant cluster-mate *Clostridium scindens* expands (Appendix B, Fig. B.5). This trend is absent in the less immunogenic, more taxonomically diverse commensal clusters (1, 2 and 7). These results suggest that the strain-level breakdown reflects a systematically biased gap between the compositional change of the most abundant strain and the remainder strains in each cluster.

Ecological interactions among strains might provide another, non-exclusive explana-

tion for these puzzling patterns. We therefore asked whether less abundant strains might be shaped by ecological cascades triggered when repertoire coverage drives changes in the abundance of more dominant strains within their cluster, akin to “competitive release” [183, 228–231]. To formalize this, we modeled the steady-state abundances of a minimal two-strain system (eq. 4.2) under shared immune coverage (C), where a dominant strain 1 suppresses a subdominant strain 2 with interaction strength $s > 0$:

$$\tilde{g}_1(C) = \frac{g_1^{\max}}{1 + C}, \quad \tilde{g}_2(C) = \frac{g_2^{\max}}{1 + C + s\tilde{g}_1(C)}, \quad (4.12)$$

where $g_{1/2}^{\max}$ are their respective carrying capacities. For any combination of immune coverage (C) and ecological interaction (s), the composition change gap between the two strains is given by (Appendix B):

$$\Delta\text{CLR}_1 - \Delta\text{CLR}_2 = -\log_{10} \left(\frac{1 + sg_1^{\max}}{1 + sg_1^{\max}/(1 + C)^2} \right). \quad (4.13)$$

For competitive interactions ($s > 0$), as immune coverage increases, eq. 4.13 predicts an increasingly negative composition change gap between strain 1 and 2, consistent with Fig. 4.5C. Numerically solving eq. 4.4 for the optimal immune coverage as a function of interaction strength, recapitulates this negative gap (Fig. 4.5D). As ecological competition intensifies, immune coverage becomes less effective at controlling the composition of the two strains and the optimal immune coverage decreases.

Definitively attributing the strain-level breakdown to ecology requires targeted experiments. The strain pair in cluster 6 (*R. gnavus* and *C. scindens*) represents an ideal candidate. Co-culture, selective depletion of the dominant partner, or pairwise reconstitution in germ-free hosts would test whether subdominant strain dynamics are strongly driven by their dominant partner, as the model predicts. Such experiments can be used to decouple ecological and immune forces shaping the microbiota.

4.4 Discussion

We modeled the gut-associated T cell repertoire as a rational allocation system that minimizes infection risks alongside the metabolic and regulatory costs of immunity (Fig. 4.1, eq. 4.4). This framework rationalizes why the repertoire targets shared epitopes across taxonomically diverse microbial clusters rather than individual strains (Fig. 4.2, 4.3, eqs. 4.9, 4.10). This model offers a comparative basis for immune repertoire allocation across host-associated mucosal barriers. For instance, because microbial harmfulness (h) scales inversely with epithelial layer thickness, our framework predicts that baseline immune coverage should be higher in the thin-walled lower lung than in the skin [232–235].

Furthermore, our model offers design principles for minimally disruptive microbiota therapies. Altering a strain’s ecological niche shifts its protection-to-cost ratio ($\Omega_{\mu,r}$), which can reorganize the repertoire (eq. 4.10, Fig 4.4A), and vice versa [192]. This implies that compatibility for interventions like fecal matter transplants may not require precise strain-level matches; instead, aligning the broad epitope compositions of donor and host may suffice. Additionally, mirroring the experimental approach of [177], we demonstrated how the inherent harmfulness of strain clusters could be inferred from their steady-state repertoire coverage (Fig. 4.4E). Finally, integrating ecological interactions into this framework, we clarified how top-down adaptive immune control can be limited by bottom-up inter-strain competition (Fig. 4.5D).

The observed sparsity of the T cell repertoire could simply reflect a limited availability of cognate receptors or microbial epitopes; however, two lines of evidence suggest otherwise. First, the peripheral repertoire cycles through the gut epithelium, making a diverse, organism-wide receptor pool available for microbe recognition [236]. In mice, this peripheral repertoire contains approximately 10^6 unique clones, orders of magnitude larger than the handful of responding clones identified in the data we analyzed [206, 237]. Second, de-

spite the immense structural complexity and multi-epitope nature of taxonomically diverse strains [238], the T cell response to a systemic infection is capable of targeting multiple epitopes derived from a single pathogen [17, 239, 240]. The targeted focus on specific clusters observed here thus points to a rational strategy of allocation rather than happenstance.

Despite these insights, our study relies on a modest sample size ($n = 5$ mice for compositional data; $n = 3$ mice for pooled T cell repertoires), which precludes independent verification of whether these specific microbial clusters are universal across all hosts. Furthermore, our assumption that the host-microbiota interaction actively optimizes repertoire allocation requires mechanistic validation. Nevertheless, our results demonstrate that a data-driven mathematical approach can successfully rationalize complex immune-microbiota interactions [65, 66, 81, 129, 183, 241–244]. Future studies pairing longitudinal, single-animal microbiota profiling with single-cell TCR sequencing, metabolic outputs and spatial tracking tools will be essential [179, 185, 226, 245, 246]. These efforts would benefit from control studies in immuno-compromised (e.g., $Rag^{-/-}$) mice to isolate purely ecological dynamics [192, 247, 248], alongside targeted strain invasions or depletion experiments. These studies will be essential for determining how the immune repertoire reorganizes in response to a changing environment.

Chapter 5

CONCLUSION

5.1 Main contributions of this dissertation

Scientific understanding of the immune response has historically lagged behind our ability to shape it. The earliest vaccines were discovered in the late 18th century, well before the cellular basis of immunity was understood [249]. However, modern immunology has accelerated the development of safer, more targeted interventions. It has also refined early efforts, moving from broad-spectrum prophylaxis to specialized regimens such as the childhood polio series, seasonal influenza updates, and age-targeted shingles vaccines. Likewise, contemporary immunotherapies leverage our growing understanding of immune memory [150, 250], augmenting immune cells to combat malignant cells or reeducating them to prevent autoimmunity. These advancements hinge on the discovery of “design principles” encoded within the immune response. This dissertation has posited that a quantitative understanding of these principles can optimize vaccination schedules to develop long-lasting immune memory, engineer T cells capable of infiltrating immunosuppressive tumor microenvironments, and enable personalized therapies that synergize with natural immune homeostatic programs in the gut.

5.1.1 *Commensurate memory formation*

By analyzing ordinary differential equation (ODE) models, this work investigated how effector and memory CD8^+ T cell populations scale during infection. We showed that flexible timing of differentiation allows the memory pool to robustly scale with infection severity, ensuring a reservoir of cells that is commensurate with the initial threat (Chapter 2). This project contributes to a body of research showing that the immune response can be actively shaped [53, 114, 251, 252]. This plasticity means that the immune response is influenced by its history of previous exposures, making it a significant source of inter-individual variation [188]. Recent work has shown that the timing of memory formation during an immune response determines its longevity and tissue residency. Our work thus provides a blueprint for modulating the quantity and quality of memory formed by perturbing the *Tcf7* gene locus and its associated regulatory pathways [157].

5.1.2 *The control logic of the T cell response*

In Chapter 3, we recast CD8^+ T cell effector response as a feedback-controlled program, wherein the rates of activation, proliferation, differentiation and death are regulated through antigenic, pro- and anti-inflammatory cues. By exploring a broad class of feedback-controller designs as potential immune programs, we demonstrated how the speed and magnitude of T cell responses emerge from optimizing signal-feedback to protect against diverse infection settings. We recovered an inherent trade-off: infection clearance at the cost of immunopathology. We showed how this trade-off is encoded into the logic of T cell responses by hierarchical sensitivity to different immune signals. Notably, we found that T cell designs that balance the harm from acute infections and autoimmunity produce immune responses consistent with the experimentally observed patterns of T cell effector expansion in mice.

Extending this framework to immune-based T cell therapies for cancer tumors, we quan-

tified the trade-off between the affinity for tumor antigens (“quality”) and the abundance (“quantity”) of infused T cells necessary for effective treatment. Finally, we showed how therapeutic efficacy can be improved by targeted genetic perturbations to T cells. Our findings offer a unified control-logic for cytotoxic T cell responses and point to specific regulatory programs that can be engineered (e.g., using CRISPR-Cas9 to enhance or suppress specific gene programs) for more robust T cell therapies.

Key questions follow from this work. Our coarse-grained model elided many details of tissue biology and micro-environment: how do tissue contexts impact the state of resident cells and shift clearance-toxicity trade-offs [253, 254]? Endogenous immune responses are polyclonal [17]: how do the proposed designs shape information flow and selection of dominant clones in polyclonal responses? Do the optimization principles posited here predispose the immune response to certain vulnerabilities, and what mechanisms might have evolved to mask them? Would explicitly modeling interactions between CD8⁺ T-cells and other immune compartments, such as regulatory T-cells and innate immune system, enhance the robustness of cytotoxic responses across different pathogenic challenges? Addressing these issues through the lens of cell economics and immunological trade-offs may clarify the roles of chance, contingency, and necessity in the evolution of effector programs.

5.1.3 Immune control of the gut microbiota

In Chapter 4, we transitioned from the response to acute threats to the management of the gut microbiota. This microbial ecosystem provides essential metabolic benefits to its host while posing a systemic threat if the intestinal barrier is breached [174–176]. A central challenge in microbiome research is predicting properties of the microbiota, such as metabolic output and harmfulness, from its taxonomic composition. Large-scale composition profiling has revealed inter-individual taxonomic variation in microbiota, even as certain metabolic outputs remain stable [179–181]. This suggests that despite ecological complexity, simple

mappings between microbial composition and host-relevant properties may exist.

We showed that a host's CD4^+ T cell repertoire implements a phenotypic mapping of the microbiota, toggling between inhibition and tolerance to control its composition: rather than tracking individual species, it targets molecular epitopes. We developed a mathematical model of optimal immune control designed to minimize host harm. We demonstrated that by regulating epitope abundance using generalist CD4^+ T cell receptors reactive to taxonomically diverse subsets of the microbiota, the immune system collapses microbial complexity into distinct microbe clusters based on potential harmfulness. This strategy minimizes the cost of immune surveillance while effectively maintaining homeostasis. Furthermore, our model explained observed microbiota dynamics and CD4^+ T cell repertoire structure, while revealing how inter-microbial ecological interactions limit the precision of immune control. These findings redefine host-microbe interactions as the strategic management of harm-associated epitopes rather than species-specific recognition. For instance, the efficacy of fecal matter transplants and probiotic interventions may be more robustly evaluated by their influence on epitope rather than strain composition. Similarly, this model provides a framework for understanding how the rapid depletion of specific taxonomic groups by antibiotics can disrupt the allocation of immune surveillance in the gut microbiota, providing opportunities for harmful microbes.

Future models/experiments should integrate high-throughput fecal sampling with comprehensive mapping of antibodies or T cells to specific strains, as demonstrated in a recent studies [177, 185, 206]. Because microbiota composition and antibody coverage can be measured from fecal samples, such approaches open the door to more human-focused studies.

5.2 *Future research directions*

As noted in Chapter 1, the immune system operates on many spatial and temporal scales. This dissertation has focused primarily on acute and homeostatic temporal scales (spanning days to months) while coarse-graining the underlying spatial architecture of lymphoid and peripheral tissues. These limitations present two promising opportunities for future work.

5.2.1 *Modeling circuit failures in peripheral tolerance*

First, how T cells are educated in the periphery following imperfect central tolerance in the thymus remains an open question [5, 255]. Recent studies indicate that the dynamic balance between conventional and regulatory T cells, alongside the persistence of short-lived innate immune memory, can predict when peripheral tolerance fails [187, 188, 191, 256, 257]. Formalizing this multicellular circuit into a mathematical framework would clarify how immune homeostasis is established or disrupted. This could help identify new targets for managing autoinflammatory and autoimmune diseases [252].

5.2.2 *Spatial models of immune-mediated tissue homeostasis*

Second, a critical gap in immunology is understanding the sources of inter-individual variation of responses to infection and treatment. This variation partly reflects the fragility of immune-controlled tissue homeostasis [258]. Tissues are not passive sites of immunity; they are complex spatial structures with collective physiological functions. Beyond pathogen surveillance, tissue-resident immune cells (e.g., macrophages, T cells, etc.) monitor paracrine signaling gradients in their local tissue environments and drive cell state changes in nearby cells to regulate tissue functions [259]. Dysregulation of immune-mediated homeostasis is implicated in the development of autoimmune disorders [260]. On short timescales, the spatial structure of a tissue can be encoded as a non-uniform graph of cells (nodes) with

gene expression profiles (node features), and spatial interactions (edges). Given the new widespread availability of spatial transcriptomic data, there is now an opportunity to develop graph-based dynamical models of the immune response in tissues (e.g., networked dynamical systems [261] and graph neural networks [262, 263]). In this setting, models have the potential to identify cell types and cell-communication channels that are essential for maintaining/perturbing specific tissue microenvironments.

Appendix A

SUPPLEMENT FOR CHAPTER 3

A.1 Model of T cell response to infections

Competitive binding of T cells to healthy and infected cells

T cells are cross-reactive and can bind to a broad range of antigens. A cognate T cell for a pathogen-derived antigen can be cross-reactive to antigens derived from self-proteins. Consequently, when analyzing T cell recognition, it is essential to account for the potential competition between binding to cells presenting cognate non-self antigens and to self-antigens.

To model this effect in a simple way, we consider the following coupled reactions:



where $E^{(u)}$ denotes the number of unbound effector T cells, $I^{(u)}$ and $S^{(u)}$ represent the numbers of unbound infected and susceptible cells, respectively, and C_I and C_S correspond to the bound complexes of effector T cells with those cells. Note that we take the number of infected and susceptible cells as a proxy for the amount non-self and self antigens presented, respectively. The parameters $k_{(\cdot)}^a$ and $k_{(\cdot)}^d$ are the rate constants for the association and

dissociation of complexes $C_{(\cdot)}$, where the placeholder $(\cdot)$ can be substituted with either I (infected) or S (susceptible healthy). The dynamics of these complexes over time follows,

$$\begin{aligned}\frac{dC_I}{dt} &= k_I^a E^{(u)} I^{(u)} - k_I^d C_I \\ \frac{dC_S}{dt} &= k_S^a E^{(u)} S^{(u)} - k_S^d C_S.\end{aligned}$$

We assume a separation of timescales: binding and unbinding events occur much more rapidly than T cell proliferation and effector killing of bound cells. The bound complexes can be treated as being in quasi-equilibrium at all times, i.e., $dC_{(\cdot)}/dt = 0$, which yields,

$$E^{(u)} I^{(u)} = K_I C_I, \quad E^{(u)} S^{(u)} = K_S C_S, \quad (\text{A.2})$$

where $K_I = k_I^d/k_I^a$ and $K_S = k_S^d/k_S^a$ are the dissociation constants of the effector cells from infected and healthy cells, respectively.

Expressing eq. A.2 in terms of the total concentrations of effectors $E = E^{(u)} + C_I + C_S$, infected cells $I = I^{(u)} + C_I$, and healthy cells $S = S^{(u)} + C_S$, we arrive at,

$$\begin{aligned}0 &= (E - C_I - C_S)(I - C_I) - K_I C_I \\ 0 &= (E - C_I - C_S)(S - C_S) - K_S C_S\end{aligned}$$

Assuming that the number of bound complexes is small relative to that of unbound cells, we can neglect the second order terms $\mathcal{O}(C_I^2)$, $\mathcal{O}(C_S^2)$, $\mathcal{O}(C_I C_S)$, and solve for these abundances in terms of the total number of effectors, and the infected and healthy cells,

$$C_I = \frac{EI}{E + I + K_I + \frac{E+K_I}{E+K_S} S}, \quad C_S = \frac{ES}{E + S + K_S + \frac{E+K_S}{E+K_I} I}. \quad (\text{A.3})$$

The correction terms in the denominators ($\frac{E+K_I}{E+K_S}S$, and $\frac{E+K_S}{E+K_I}I$) capture the competition between binding to self- and non-self-derived antigens. In the limit $K_S \gg K_I$ (i.e., when T cells bind more strongly to infected cells than healthy ones), the antagonistic effect vanishes and C_I approaches the form for response in the presence of a single-antigen only, consistent with ref. [63, 171]. However, as $K_S \rightarrow K_I$ (self and non-self antigens become comparable in affinity), the fraction of effector cells bound to infection-derived antigens decreases, reflecting the impact of self/non-self cross-reactivity and antagonism on recognition. Prior work has explored the impact of antagonistic effects on T cell response in much more detail [72, 136]. Here, we only adopt a high-level perspective to capture how such effects shape responses to low-immunogenicity infections and cancer.

The relative magnitude of infection- to self-derived antigenicities (i.e, the ratio of the respective recognition thresholds) determine the *immunogenicity* of an infection, which we define as,

$$\text{immunogenicity: } \kappa = K_I^{-1} K_S. \quad (\text{A.4})$$

This dimensionless measure reflects how readily cognate T cells recognize antigens presented by infected cells as “non-self” relative to those presented by uninfected cells (self).

Dynamics of growing (acute) infections

In the absence of infection, the population of susceptible cells (S) are homeostatically maintained by slow birth-death processes at a steady state, $S(t = 0) = S_{\max}$. Upon infection, there are initially a small number of infected cells— $I(t = 0) = I_0$ —but susceptible cells gradually become infected as they come into contact with other infected cells at rate b_I . The infected state of cells and their death, at rate d_I , produce chemical signals, namely the antigen signal $\sigma_{\text{Ag}}(t)$ and the infection signal $\sigma_{\text{inf}}(t)$, which are detected and transmitted by the innate immune system. The innate response then triggers naïve T cells in the adaptive

immune response (N) to differentiate into effector cells (E). Differentiated effector cells kill infected and susceptible cells, producing chemical response signals ($\sigma_{\text{res}}(t)$) that feedback on the response. The dynamics of susceptible and infected cells follow,

$$\begin{aligned}\dot{S} &= -b_I I S - d_E C_S, \\ \dot{I} &= b_I S I - d_I I - d_E C_I,\end{aligned}\tag{A.5}$$

where d_E is the maximal clearance rate of infected/susceptible cells by effector T cells, and $C_S = \frac{ES}{E+S+K_S+\frac{E+K_S}{E+K_I}I}$ and $C_I = \frac{EI}{E+I+K_I+\frac{E+K_I}{E+K_S}S}$ represent the number of susceptible or infected cells bound to an effector cell at time t , respectively; see section A.1 for a derivation of these quantities. This framework closely reflects viral or intracellular bacterial infections, but also captures features of extracellular pathogens that damage or invade cells indirectly (e.g., via toxins); we do not distinguish these mechanisms here.

The extent of pathogen reproduction during the lifetime of an infected cell is determined by the product of the infection rate and the characteristic survival time of the cell. Specifically, over the infected-cell lifetime $\tau_{\text{death}} = d_I^{-1}$, an infected cell produces on average $b_I S \cdot \tau_{\text{death}} = b_I S / d_I$ new infections. In epidemiology, this term is often referred to as the *effective reproduction number*, whose maximum value—attained when $S = S_{\text{max}}$ —is the *basic reproduction number*, $b_I S_{\text{max}} / d_I$ [134, 135]. This motivates defining the *intra-host basic reproduction number*,

$$R_0 = \frac{b_I S_{\text{max}}}{d_I},\tag{A.6}$$

which measures the expected number of secondary infected cells generated by a single infected cell in an otherwise fully susceptible cell population. Note that R_0 scales with S_{max} , the total number of cells in the modeled tissue. We simulate these dynamics over a time horizon $T > 0$ days, chosen to match the typical timescale of acute immune responses (approximately 2–3 weeks). A window that is too short underestimates cumulative harm, whereas a longer

window increases computational cost. Figs. A.10, A.11 demonstrate how total harm and effector cell fold expansion depend on the choice of T , over a wide range from 1 to 100 days. In our main simulations, we set $T = 30$ days to ensure that most responses generated by our model are fully captured [109, 148, 264]. For related approaches see refs. [171, 265, 266].

Dynamics of slow-growing tumors

In Fig. 3.6, we show the efficacy of different T cell response designs for tumor clearance. In this context, we model a tumor as a slowly growing cell population—maximum rate b_I —which grows by out-competing healthy tissue cells for limited space/resources, resulting in a logistic growth of tumor cells [267]. We assume that at the time of treatment, the tumor is already at detectable size, $I_0 = 10\% S_{\max}$, and that tumor cells die at negligible rates during our simulation window, $d_I = 0$. The resulting dynamics are

$$\begin{aligned}\dot{S} &= -b_I \left(1 - \frac{I}{S_{\max}}\right) I - d_E C_S, \\ \dot{I} &= b_I \left(1 - \frac{I}{S_{\max}}\right) I - d_E C_I.\end{aligned}\tag{A.7}$$

Sampling pathogen space

To model diverse immune challenges, we sample a broad range of immunogenicities (κ), and basic reproduction numbers (R_0). We take $K_S = S_{\max} \geq K_I$ so that effector cells typically bind more strongly to infected cells than healthy cells. We note that K_I and K_S are effective parameters reflecting the dynamics of antigen detection in the complex tissue environment. The killer T cell response is delayed with respect to pathogen exposure—typically around a few days [60, 264]—so there are limits on its effectiveness against fast-replicating pathogens. We sample a range of reproduction numbers (R_0) to include non-

replicating immune challenges $R_0 = 1$ and fast replicating immune challenges which deplete the reservoir of susceptible cells. Quantitatively, we can consider when the infected cell population peaks in the absence of an immune response $T_{I_{\max}}$ by setting $dI/dt = 0$ and $E = 0$ in eq. A.5 to find:

$$S(T_{I_{\max}}) = \frac{S_{\max}}{R_0}.$$

In our simulations, we choose the maximum R_0 such that $S(T_{I_{\max}}) \ll S_{\max}$, representing a pathogen that quickly infects most susceptible cells. Here, we chose $(K_I, R_0) \in [10^{-3}S_{\max}, S_{\max}] \times [1, 5]$, discretized into a 9×9 grid, to reflect the aforementioned considerations; see Table A.1 for details on simulation parameter ranges.

Harm during T cell response to pathogens

We express harm (damage) to a host in terms of cell deaths. This depends on the details of the programmed immune response, which we parametrize by $\vec{\theta}$ (see below for details). The rate of harm at a given time point can be expressed as

$$h(t) = h_{\text{inf}}(t, \vec{\theta}) + \underbrace{h_{E,I}(t, \vec{\theta}) + h_{E,S}(t, \vec{\theta})}_{=h_{\text{res}}(t)}, \quad (\text{A.8})$$

where $h_{E,I}(t) = d_E C_I(t)$ and $h_{E,S}(t) = d_E C_S(t)$ denote the loss of infected and healthy cells, respectively, caused by the effector T cell response, amounting to the total damage by the response $h_{\text{res}}(t)$. Here, $C_I(t)$ and $C_S(t)$ are the expected number of infected and susceptible cells bound to effectors at time t (eq. A.3). We consider two scenarios:

- **Harm associated with growing (acute) infections.** The infection harm $h_{\text{inf}}(t)$ captures pathogen-induced harm, i.e., cells directly killed by the infection itself. Its functional form depends on the biological context. For an acute, proliferating infection

(eq. A.5), it follows:

$$h_{\text{inf}}(t, \vec{\theta}) = d_I I(t, \vec{\theta}) \quad (\text{A.9})$$

reflecting cell loss due to the death of infected cells. The *total cell death* (or inflicted harm) then follows

$$H(\vec{\theta}) = \underbrace{\int_0^T \left[h_{\text{inf}}(t, \vec{\theta}) + h_{E,I}(t, \vec{\theta}) \right] dt}_{H_{\text{inf.}}(\vec{\theta})} + \underbrace{\int_0^T h_{E,S}(t, \vec{\theta}) dt}_{H_{\text{toxic.}}(\vec{\theta})}, \quad (\text{A.10})$$

where $H_{\text{inf.}}(\vec{\theta})$ aggregates infection-related harm (cells killed by the pathogen and infected cells eliminated by effectors), $H_{\text{toxic.}}(\vec{\theta})$ captures immunopathology (healthy bystander cells killed by effectors), and $T > 0$ is the simulation window. This single, interpretable metric balances clearance benefits against collateral damage, enabling principled comparison across designs and infections.

- **Harm associated with slow-growing tumors.** For logistically growing cancer tumors (eq. A.7), the harm from cancer takes the form

$$h_{\text{cancer}}(t) = b_I \left(1 - \frac{I(t)}{S_{\text{max}}} \right) I(t), \quad (\text{A.11})$$

reflecting net healthy cell loss due to tumor growth (eq. A.7). In this case, we define the total harm as

$$H(\vec{\theta}) = \underbrace{\int_0^T h_{\text{cancer}}(t, \vec{\theta}) dt}_{H_{\text{cancer}}(\vec{\theta})} + \underbrace{\int_0^T h_{E,S}(t, \vec{\theta}) dt}_{H_{\text{toxic.}}(\vec{\theta})}. \quad (\text{A.12})$$

Here, $H_{\text{cancer}}(\vec{\theta})$ quantifies the loss of healthy cells as the tumor expands and replaces normal tissue, while $H_{\text{toxic.}}(\vec{\theta})$ captures collateral healthy-cell death caused by the en-

suing immune response. Note that in the case of acute infections (eq. A.10), loss of healthy cells is represented by $H_{\text{inf.}}$, the sum of cells killed directly by the pathogen and infected cells eliminated by effectors. In cancer, by contrast, a slowly growing tumor replaces healthy cells with malignant cells. Accordingly, harm is tied to the “birth” of cancer cells captured by the time-integral of the birth term $h_{\text{cancer}}(t)$ in eq. A.11.

Signal cues for T cell response

We assume that T cells can sense three types of signals during infection:

- **Antigen signal** $\sigma_{\text{Ag}}(t)$, which is proportional to the number of infected cells, and expressed in units of the antigen recognition threshold K_I ,

$$\text{antigen signal: } \sigma_{\text{Ag}}(t) = K_I^{-1} I(t) \leq K_I^{-1} S_{\text{max}} \quad (\text{A.13})$$

- **Infection-induced harm signal** $\sigma_{\text{inf}}(t)$, reflecting cumulative damage caused directly by the pathogen, and expressed in units of the harm recognition threshold K_H ,

$$\text{infection harm signal: } \sigma_{\text{inf}}(t) = K_H^{-1} \int_0^t e^{-(t-s)d_H} h_{\text{inf}}(s) \, ds \leq K_H^{-1} S_{\text{max}}. \quad (\text{A.14})$$

- **Response-induced harm signal** $\sigma_{\text{res}}(t)$, reflecting cumulative damage caused by the effector response, and expressed in units of the harm recognition threshold K_H ,

$$\text{response harm signal: } \sigma_{\text{res}}(t) = K_H^{-1} \int_0^t e^{-(t-s)d_H} h_{\text{res}}(s) \, ds \leq K_H^{-1} S_{\text{max}}. \quad (\text{A.15})$$

The harm-derived signals are time-integrated, reflecting the accumulation of cytokines during an immune response, and are exponentially discounted at rate d_H , reflecting the natural

decay of cytokines in the absence of ongoing stimulation. Note that K_I plays a dual role: it represents both the recognition threshold of effectors binding to and killing infected cells (see section A.1) and the recognition threshold for binding to and stimulating T cells.

Finally, although we do not explicitly model innate or regulatory T cell dynamics, these populations act primarily as sources and sinks of pro- and anti-inflammatory mediators. Their effects are therefore captured in part by our representation of antigen- and response-induced harm signal production and decay. With this interpretation, the design parameters and the latent biological factors that set their values implicitly encode the contributions of these cell populations, albeit at a coarse-grained level. Moreover, in our framework, highly immunogenic antigens (with large κ) are those that elicit strong $CD8^+$ T cell responses, whereas antigens with low immunogenicity have more subtle effects. This, in part, reflects the suppressive role of Tregs in limiting effector responses against self-antigens, which in our definition correspond to low-immunogenicity targets. Future extensions could more explicitly incorporate the time-varying impact of Tregs and the innate immune system on the dynamics of the immune response, similar to the approach in ref. [141].

Lineage-based model of $CD8^+$ T cell response

To characterize the immune response to a pathogenic challenge, we model the large-population dynamics of susceptible and infected cells in a host as deterministic processes, coupled to a stochastic, time-inhomogeneous birth-death process for the $CD8^+$ T cell response.

The T cell repertoire typically contains many precursors, and responses to a given pathogen are often polyclonal. An interesting problem would be to assess how interactions between different T cell clones impact the response in an individual, which we defer to future work. In our simulations, we model a response mediated by a single clone of naïve T cells (i.e., cells sharing the same TCR). Upon activation, each naïve cell can seed a lineage whose progeny differentiate into effector cells that act against the source of the immune challenge

(e.g., infection or tumor) or memory cells which do not participate in resolving the ongoing immune challenge.

In what follows, we consider an initial population of N_{lin} monoclonal naïve T cells, and track the lineages of their progeny in response to an immune challenge. Each cell’s decisions—to engage an antigen-presenting cell (APC), become activated, divide, differentiate, or die—are governed by time-dependent antigen and infection harm signals (eqs. A.19). We assume these signals are lineage-independent: cells in the same state but belonging to different lineages share identical transition rates. Inter-lineage variability therefore arises solely from stochastic differences in transition timing, consistent with experimental observations [35, 51, 266].

Recruitment of naïve T cells to interact with APCs ($N \rightarrow [\text{APC} \cdot N]$). Naïve T cells recirculate through blood and lymph and may also reside in tissues [27]. Upon infection, APCs become activated and acquire antigen as infected cells die [268]. This process is coupled to the production of an infection harm signal σ_{inf} , which serves as a proxy for the “danger” associated with tissue damage caused by the immune challenge. As a result, naïve T cells can be recruited to the sites of infection and interact with APCs to initiate an immune response. We model the recruitment rate of naïve T cells to bind to the APCs as

$$r_{N \rightarrow \text{APC} \cdot N}(t) \equiv r_{\text{bind}}(t) = \frac{r_{\text{bind}}^{\text{max}}}{1 + \exp[-\log(1 + \sigma_{\text{inf}}(t))]}, \quad (\text{A.16})$$

where $r_{\text{bind}}^{\text{max}} \approx 1 \text{ [day]}^{-1}$ is the upper bound on the recruitment rate of T cells to lymph nodes for sustained APC binding [29, 30, 269] (see Table A.1 for the biological range). The denominator modulates this rate according to the strength of the infection harm signal activating the APCs. We assume that recruitment of all the naïve cells to APCs is mediated by the same infection harm signal, and that these cells do not compete for APCs or infection harm signal.

Activation of naïve T cells ($[\text{APC} \cdot N] \rightarrow N^*$). When a naïve T cell engages an APC to form a complex $[\text{APC} \cdot N]$, the two cells may remain bound for several hours, leading to activation of the bound T cell (“priming phase”). During this time, the immunological synapse formed between the T cell and APC initiates a cascade of intracellular signaling events in the naïve T cell, enabling it to surpass a sharp activation threshold [31]. The strength of antigenic stimulation, together with co-stimulatory signals, modulates the activation rate of APC-bound naïve T cells $r_{N \rightarrow N^*}(t)$.

Experiments have shown that naïve T cells can remain stably bound for 12-20 hours before the first division [51] to become activated $[\text{APC} \cdot N] \rightarrow N^*$. Following activation, cells commit to a rapid proliferative program [104, 270], resulting in $k = 2 - 3$ burst-like divisions at a rate $r_{N^* \rightarrow 2N^*}$ [51] (Fig. A.1). Once initiated, these early doublings proceed largely independently of continued antigen or cytokine input [42, 51], and set an upper bound $N_{\max}^* = 2^k$ for the number of activated cells generated prior to differentiation in a lineage; we use $N_{\max}^* = 2^2$ in our simulations (Table A.1).

Differentiation of activated T cells to effector phenotype ($N^* \rightarrow E$). During an immune challenge, CD8^+ T cells follow a pathway from long-lived naïve/memory cells to short-lived effector cells. This fate is partially “reversible”: effectors can reacquire memory-like traits (e.g., longevity), and become so-called effector memory [14, 49, 50]. We consider differentiation from naïve to effector state as the primary transition for activated cells, with un-transitioned cells retaining the relative longevity and slow proliferation of naïve cells. Consistent with prior work [14, 15, 49, 51], we take the burst-like early proliferation phase as the window during which effector commitment occurs (Fig. A.1).

Formally, for a given activated cell N^* , we model effector commitment as the first arrival of a time-inhomogeneous Poisson process with instantaneous rate, $r_{N^* \rightarrow E}(t)$. Given

the stochastic time interval τ_{burst} to complete the initial burst-like divisions, the probability that an activated cell differentiates into the effector state is computed as the probability that the first arrival time is less than τ_{burst} ,

$$p_{N^* \rightarrow E} = 1 - \exp \left[- \int_0^{\tau_{\text{burst}}} r_{N^* \rightarrow E}(s) ds \right]. \quad (\text{A.17})$$

Given N_{max}^* activated cells available at the end of the burst, the number that differentiate into effectors is drawn independently as $e_{\text{differentiate}} \sim \text{Binom}(N_{\text{max}}^*, p_{N^* \rightarrow E})$, with the remaining $N_{\text{max}}^* - e_{\text{differentiate}}$ cells adopting a memory fate. This fate decision is executed independently across lineages (Algorithms A.1-A.3; Fig. A.1).

Division and death of effector T cells ($E \rightarrow 2E$, and $E \rightarrow \emptyset$). Experiments have shown that T cell lineages do not divide indefinitely [34]. By tracking proliferation markers in cell lineages, it has been estimated that a naïve precursor undergoes at most $K_{\text{max}} \approx 10 - 19$ division rounds before terminal arrest [15, 35, 109, 110], with differentiation into effector possible along the way [15, 34, 35]. This sets an upper bound on the number of effector cells in a lineage (if all cells turn into effector), which is $E_{\text{max}} = 2^{K_{\text{max}}}$, which we set to 2^{15} in our simulations (Table A.1). A comparable bound for memory cells is less certain, as establishing it would require years of longitudinal tracking.

Because effector commitment is stochastic, the maximum additional effector output beyond the initial committed pool varies by lineage. At the end of the early burst, the lineage has completed $\log_2 N_{\text{max}}^*$ semi-synchronous divisions. Each newly committed effector cell can therefore undergo at most $K_{\text{max}} - \log_2 N_{\text{max}}^* = \log_2 E_{\text{max}} - \log_2 N_{\text{max}}^*$ additional rounds. Thus, starting from $e_{\text{differentiate}}$ effectors at the end of the burst phase, proliferation halts once the effector count reaches, $e_{\text{stop prolif.}} = e_{\text{differentiate}} \times \left(\frac{E_{\text{max}}}{N_{\text{max}}^*} \right)$; see Algorithms A.1-A.3. Each differentiated effector cell undergoes division and death events at rates $r_{E \rightarrow E}(t)$ and $r_{E \rightarrow \emptyset}(t)$, respectively, which are modulated by the time-varying antigen and harm signals

(see below). In lieu of tracking the division history/tree of cells in the lineage, as effector cells divide, we keep a running count of the total number of effector division events in that lineage, $B_{\text{proliferation}}$, and set the division rate to zero once $B_{\text{proliferation}} \geq e_{\text{stop prolif}}$. (Algorithms A.1-A.3).

In summary, the dynamics for a single T cell responding to an immune challenge, through activation by engaging APC's, differentiation, expansion, and contraction are summarized as follows,

$$\begin{aligned}
&\text{exposure} \begin{cases} S + I & \xrightarrow{b_I} I + I \rightarrow \sigma_{\text{Ag}} \\ I & \xrightarrow{d_I} \emptyset \rightarrow \sigma_{\text{inf}} \end{cases} \\
&\text{clearance} \begin{cases} I/S + E & \xrightarrow{d_E} E + \emptyset \rightarrow E + \sigma_{\text{res}} \end{cases} \\
&\text{activation} \begin{cases} \text{APC} + N & \xrightarrow{r_{\text{bind}}(\vec{\sigma})} \text{APC} \cdot N \\ \text{APC} \cdot N & \xrightarrow{r_{N \rightarrow N^*}(\vec{\sigma})} N^* + \text{APC} \\ N^* & \xrightarrow{r_{N^* \rightarrow 2N^*}} 2N^* \xrightarrow{r_{N^* \rightarrow 2N^*}} 4N^* \end{cases} \quad (\text{A.18}) \\
&\text{differentiation} \begin{cases} N^* & \xrightarrow{r_{N^* \rightarrow E}(\vec{\sigma})} E \end{cases} \\
&\text{expansion} \begin{cases} E & \xrightarrow{r_{E \rightarrow 2E}(\vec{\sigma})} E + E \end{cases} \\
&\text{contraction} \begin{cases} E & \xrightarrow{r_{E \rightarrow \emptyset}(\vec{\sigma})} \emptyset \end{cases}
\end{aligned}$$

Note that the transition rates are dependent on the three time-dependent signals $\vec{\sigma} = (\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}})$; see below for details. The programmatic details of these reactions, when keeping track of different T cell lineages, are provided in Algorithms A.1-A.3.

Regulation of T cell transition rates. Activated T cells undergo differentiation, proliferation, and eventual contraction in response to antigen and other signaling molecules like cytokines [271–273]. These processes are shaped by both extrinsic noise (e.g., variability in cytokine levels and spatial heterogeneity of APCs) and intrinsic noise (e.g., stochastic gene expression and chromatin remodeling) [274]. We model the effector CD8⁺ T cell response as a sequence of four canonical modules:

1. **Activation:** priming of naïve T cells, $N \rightarrow N^*$,
2. **Differentiation:** commitment of the activated T cells to effector phenotype, $N^* \rightarrow E$,
3. **Expansion:** division of effector cells, $E \rightarrow 2E$,
4. **Contraction:** programmed loss of effector cells, $E \rightarrow \emptyset$.

We model the complex genetic regulation of each transition i as an inhomogeneous Poisson process with a transition rate $r_i(t)$, modulated by the signals that T cells receive at a given time,

$$r_i(t) = r^{\max} g_i \left(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}}; \vec{\psi}_i, \ell_i \right), \quad (\text{A.19})$$

where the maximal rate $r^{\max}$ is bounded by the cell-cycle time ($\sim 5 - 6$ hrs [28, 42, 116]), and g is a monotonic *response function* chosen from the modified Monod-Wyman-Changeux (Hill-type) family [75, 139]:

$$g_i(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}}; \vec{\psi}_i, \ell_i) = \frac{1}{1 + \exp \left[-L_i(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}}; \vec{\psi}_i, \ell_i) \right]}, \quad (\text{A.20})$$

with

$$L_i(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}}; \vec{\psi}_i, \ell_i) = \ell_i + \sum_{\alpha \in \{\text{Ag}, \text{inf}, \text{res}\}} \psi_i^\alpha \log [1 + \sigma_\alpha].$$

The parameter ℓ_i sets the baseline propensity of transition i : in the absence of external signals ($\vec{\sigma} = 0$), the regulatory factor evaluates to $g_{0,i} = (1 + \exp(-\ell_i))^{-1}$. As ℓ_i increases, this baseline transition rate approaches its ceiling $r^{\max}$. The weights $\{\psi_i^{\text{Ag}}, \psi_i^{\text{inf}}, \psi_i^{\text{res}}\}$ encode the sensitivity and polarity (positive / negative = up / down-regulation) of response to the antigen (σ_{Ag}) and harm-induced signals (σ_{inf} , and σ_{res}). Each logarithmic term $\log[1 + \sigma_\alpha]$ becomes significant only once its corresponding cue $\sigma_\alpha \sim 1$. This formulation abstracts away many layers of biochemical detail, while implementing the fold-change response to stimuli observed in many biological circuits [61, 74, 140–142].

The transition rates governing each of the four modules are parameterized by a response function (eq. A.20), whose parameters we collect into a *design vector* $\vec{\theta}$ (up to 4×4 degrees of freedom). These functions are expressive enough to include many three-signal logic circuits [139]. Introducing explicit non-linear interactions between signals would broaden the expressivity of the model but at the cost of greater complexity and limited interpretability, an extension that we leave for future work.

A.2 Numerical implementation of stochastic T cell response

Parameters governing our model and their biological ranges are given in Table A.1, and the pseudo-code describing our numerical approach is provided in Algorithms A.1-A.3. Simulating our lineage-based model of T cell response with a fully agent-based Gillespie algorithm with time-dependent rates is computationally challenging given the large number of agents ($\sim 10^7$) that must be tracked and updated [275]. We employ two approximations. First, we model the dynamics of the susceptible and infected cells, and harm signals using a piecewise, deterministic-stochastic system of ODEs, implemented numerically with first-order difference equations, and specified in Algorithms A.1-A.3. Second, we model the cellular immune response as a time-inhomogeneous Markov process implemented with a tau-leaping

scheme [276].

As long as the simulation time-step Δt is chosen carefully, this approach provides a good compromise between computational speed and accuracy. In our case, it is sufficient to choose Δt such that transition rates do not change much during this small time interval, i.e., $r(t + \Delta t) \approx r(t)$. Thus, we seek to bound

$$\frac{dr}{dt} = r^{\max} \sum_{\sigma \in \{\text{Ag}, \text{inf}, \text{res}\}} \frac{dg}{d\sigma} \frac{d\sigma}{dt}, \quad (\text{A.21})$$

where

$$\frac{dg}{d\sigma} = \frac{\psi_\sigma}{1 + \sigma} \frac{1}{1 + e^{-L(\sigma)}} \frac{e^{-L(\sigma)}}{1 + e^{-L(\sigma)}} \implies \left| \frac{dg}{d\sigma} \right| \leq \frac{|\psi_\sigma|}{1 + \sigma} \frac{1}{2^2}. \quad (\text{A.22})$$

The upper bound is observed by noting that the product of the last two terms is equivalent to $x(1 - x)$ for $x \in (0, 1)$. Next, using equations from section A.1, we compute

$$\begin{aligned} \frac{d\sigma_{\text{Ag}}}{dt} &= K_I^{-1} \frac{dI}{dt} = \sigma_{\text{Ag}} \times (b_I S - d_I - d_E C_I / I), \\ \frac{d\sigma_{\text{inf}}}{dt} &= K_H^{-1} h_{\text{inf}} - d_H \sigma_{\text{inf}} = K_H^{-1} d_I I - d_H \sigma_{\text{inf}}, \\ \frac{d\sigma_{\text{res}}}{dt} &= K_H^{-1} h_{\text{res}} - d_H \sigma_{\text{res}} = K_H^{-1} d_E C_S - d_H \sigma_{\text{res}}, \end{aligned} \quad (\text{A.23})$$

from which we conclude

$$\begin{aligned} \left| \frac{d\sigma_{\text{Ag}}}{dt} \right| &\leq \sigma_{\text{Ag}} \times (b_I S + d_I + d_E C_I / I), \\ \left| \frac{d\sigma_{\text{inf}}}{dt} \right| &\leq K_H^{-1} d_I I(t) + d_H \sigma_{\text{inf}} \sim 2d_H \sigma_{\text{inf}}, \\ \left| \frac{d\sigma_{\text{res}}}{dt} \right| &\leq K_H^{-1} d_E C_S(t) + d_H \sigma_{\text{res}} \sim 2d_H \sigma_{\text{res}} \end{aligned} \quad (\text{A.24})$$

where we rely on the homeostatic forms of the dynamical equations for σ_{inf} and σ_{res} to argue that typically, $K_H^{-1} d_I I(t) \sim d_H \sigma_{\text{inf}}(t)$ and $K_H^{-1} d_E C_S(t) \sim d_H \sigma_{\text{res}}(t)$. Thus, we see that

typically,

$$\begin{aligned}
\left| \frac{dr}{dt} \right| &\leq r^{\max} \frac{1}{4} \left(\frac{\sigma_{\text{Ag}}}{1 + \sigma_{\text{Ag}}} |\psi_{\text{Ag}}| (b_I S + d_I + d_E C_I / I) + \frac{\sigma_{\text{inf}}}{1 + \sigma_{\text{inf}}} 2d_H |\psi_{\text{inf}}| + \frac{\sigma_{\text{res}}}{1 + \sigma_{\text{res}}} 2d_H |\psi_{\text{res}}| \right) \\
&\leq r^{\max} \frac{1}{4} (|\psi_{\text{Ag}}| (b_I S + d_I + d_E C_I / I) + 2d_H |\psi_{\text{inf}}| + 2d_H |\psi_{\text{res}}|) .
\end{aligned} \tag{A.25}$$

With parameters chosen according to Table A.1, each term in eq. A.25 is at most on the order of $\mathcal{O}(10)$. Thus, taking $\Delta t = \mathcal{O}(10^{-2})$ guarantees that transition rates typically change only minimally over each time interval,

$$|r(t + \Delta t) - r(t)| \approx \left| \frac{dr}{dt} \right| \Delta t < \mathcal{O}(10^{-1}). \tag{A.26}$$

Algorithm A.1 Piecewise, deterministic-stochastic simulation of CD8⁺ T cell response to an immune challenge. Simulating the trajectories of susceptible and infected cells, and T cells.

```

1: Input:  $\vec{\psi}, \vec{\ell}, b_I, K_I, K_S, K_H, r^{\max}, r_{\text{bind}}^{\max}, r_{N^* \rightarrow 2N^*}, d_I, d_E, N_{\text{lin.}}, N_{\text{max}}^*, E_{\text{max}}, \Delta t, T$ 
2: Initialize:  $S \leftarrow S_{\text{max}} - I_0, I \leftarrow I_0, \vec{N} \leftarrow \vec{1}, \vec{\sigma} \leftarrow \vec{0}, \vec{E} \leftarrow \vec{0}, \vec{e}_{\text{stop prolifer.}} \leftarrow 0; \vec{B}_{\text{differentiate}} \leftarrow 0, \vec{B}_{\text{proliferate}} \leftarrow 0$ 
3: for  $t = 0, \dots, T - 1$  step  $\Delta t$  do
     $\triangleright$  (i) deterministic update of immune challenge dynamics, using INFECTION (Algorithm A.2)
4:  $(\vec{r}, h_{\text{inf}}, h_{E,I}, h_{E,S}) \leftarrow \text{INFECTION}(S, I, \vec{E}, \vec{\sigma}; I_0, N_{\text{lin.}}, \vec{\psi}, \vec{\ell}, b_I, K_I, K_S, K_H, r^{\max}, r_{\text{bind}}^{\max}, r_{N^* \rightarrow 2N^*}, d_I, d_E, \Delta t)$ 
     $\triangleright$  — shorthand notation:  $\vec{r} = \{r_{\text{bind}}, r_{N^* \rightarrow 2N^*}, r_{N \rightarrow N^*}, r_{N^* \rightarrow E}, r_{E \rightarrow 2E}, r_{E \rightarrow \emptyset}\}$  —
5:
6:     $\triangleright$  (ii) stochastic T cell response for all lineages  $l = 1, \dots, N_{\text{lin.}}$ , using RESPONSE (Algorithm A.3)
7:    for  $l = 1, \dots, N_{\text{lin.}}$  do
8:
9:         $(N^{(l)}, [\text{APC} \cdot N]^{(l)}, N^{*,(l)}, E^{(l)}, M^{(l)}, e_{\text{stop prolifer.}}^{(l)}, B_{\text{differentiate}}^{(l)}, B_{\text{proliferate}}^{(l)})$ 
10:        $\leftarrow \text{RESPONSE}(N, N^*, [\text{APC} \cdot N], E, M, \vec{r}, N_{\text{max}}^*, E_{\text{max}}, e_{\text{stop prolifer.}}, B_{\text{differentiate}}, B_{\text{proliferate}}; \Delta t)$ 
11:    end for
12:
13:     $\triangleright$  (iii) record trajectories of observables at time  $t$ 
14:    Record  $S, I, \vec{N}, \vec{N}^*, \vec{E}, \vec{M}, h_{\text{inf}}, h_{E,I}, h_{E,S}$  into  $[\cdot]_t$ 
15: end for
16:
17: return  $([S]_t, [I]_t, [\vec{N}]_t, [\vec{N}^*]_t, [\vec{E}]_t, [\vec{M}]_t, [h_{\text{inf}}]_t, [h_{E,I}]_t, [h_{E,S}]_t)$ 

```

Algorithm A.2 Deterministic update of immune challenge dynamics. Infection variables (S, I) , harms $(h_{\text{inf}}, h_{E,I}, h_{E,S})$, signals $(\sigma_{\text{Ag}}, \sigma_{\text{inf}}, \sigma_{\text{res}})$, and transition rates $(r_{\text{bind}}, r_{N^* \rightarrow 2N^*}, r_{N \rightarrow N^*}, r_{N^* \rightarrow E}, r_{E \rightarrow 2E}, r_{E \rightarrow \emptyset})$ are computed.

1: **function** INFECTION($S, I, \vec{E}, \vec{\sigma}; I_0, N_{\text{lin}}, \vec{\psi}, \vec{\ell}, b_I, K_I, K_S, K_H, r^{\text{max}}, r_{\text{bind}}^{\text{max}}, r_{N^* \rightarrow 2N^*}^{\text{max}}, d_I, d_E, d_H, \Delta t$)

▷ (i) compute the number of T cell bound infected and susceptible cells C_I and C_S

2:

$$C_I \leftarrow \frac{I \sum_{l=1}^{N_{\text{lin}}} E^{(l)}}{\sum_{l=1}^{N_{\text{lin}}} E^{(l)} + I + K_I + \frac{\sum_{l=1}^{N_{\text{lin}}} E^{(l)} + K_I}{\sum_{l=1}^{N_{\text{lin}}} E^{(l)} + K_S} S}, \quad C_S \leftarrow \frac{S \sum_{l=1}^{N_{\text{lin}}} E^{(l)}}{\sum_{l=1}^{N_{\text{lin}}} E^{(l)} + S + K_S + \frac{\sum_{l=1}^{N_{\text{lin}}} E^{(l)} + K_S}{\sum_{l=1}^{N_{\text{lin}}} E^{(l)} + K_I} I}.$$

▷ (ii) compute the infection harm h_{inf} , and response harms $h_{E,I}, h_{E,S}$

3:

$$h_{\text{inf}} \leftarrow d_I I \Delta t, \quad h_{E,I} \leftarrow d_E C_I \Delta t, \quad h_{E,S} \leftarrow d_E C_S \Delta t$$

▷ (iii) update the number of susceptible and infected cells

4:

$$S \leftarrow S + (-\mathbf{1}_{I \geq I(0)} b_I I S - d_E C_S) \Delta t$$

$$I \leftarrow I + (\mathbf{1}_{I \geq I(0)} b_I I S - d_I I - d_E C_I) \Delta t$$

▷ (iv) update the three signals $\vec{\sigma}$

5:

$$\sigma_{\text{Ag}} \leftarrow K_I^{-1} I$$

$$\sigma_{\text{inf}} \leftarrow \sigma_{\text{inf}} + (K_H^{-1} h_{\text{inf}} - d_H \sigma_{\text{inf}}) \Delta t$$

$$\sigma_{\text{res}} \leftarrow \sigma_{\text{res}} + (K_H^{-1} h_{E,I} + K_H^{-1} h_{E,S} - d_H \sigma_{\text{res}}) \Delta t$$

▷ (v) update the recruitment rate of naïve cells to bind to APC's r_{bind} and the transition rates r_x for all the processes $x \in \{N \rightarrow N^*, N^* \rightarrow E, E \rightarrow 2E, E \rightarrow \emptyset\}$

6:

$$r_{\text{bind}} \leftarrow \frac{r_{\text{bind}}^{\text{max}}}{1 + \exp[-\log(1 + \sigma_{\text{inf}})]}$$

7: **for** $x \in \{N \rightarrow N^*, N^* \rightarrow E, E \rightarrow 2E, E \rightarrow \emptyset\}$ **do**

$$L_x \leftarrow \ell_x + \sum_{\alpha: \{\text{Ag}, \text{inf}, \text{res}\}} \psi_\alpha \log(1 + \sigma_\alpha), \quad g_x \leftarrow (1 + \exp[-L_x])^{-1}, \quad r_x \leftarrow r^{\text{max}} g_x$$

8: **end for**

9: **return** $\vec{r} = \{r_{\text{bind}}, r_{N^* \rightarrow 2N^*}, r_{N \rightarrow N^*}, r_{N^* \rightarrow E}, r_{E \rightarrow 2E}, r_{E \rightarrow \emptyset}\}, h_{\text{inf}}, h_{E,I}, h_{E,S}, S, I$

10: **end function**

Algorithm A.3: Stochastic response of a T cell lineage. Stochastic recruitment, activation, proliferation, differentiation, and death of CD8⁺ T cells are simulated for each lineage (lineage superscripts dropped). The number of cells from different states ($N, [\text{APC} \cdot N], N^*, E, M$), and the variables determining the lineage's fate ($e_{\text{stop proliferate}}, B_{\text{differentiate}}, B_{\text{proliferate}}$) are evaluated.

```

1: function RESPONSE( $N, N^*, [\text{APC} \cdot N], E, M, \vec{r}, N_{\text{max}}^*, E_{\text{max}}, e_{\text{stop proliferate}}, B_{\text{differentiate}}, B_{\text{proliferate}}; \Delta t$ )
    ▷ — shorthand notation:  $\vec{r} = \{r_{\text{bind}}, r_{N \rightarrow N^*}, r_{N^* \rightarrow 2N^*}, r_{N^* \rightarrow E}, r_{E \rightarrow 2E}, r_{E \rightarrow \emptyset}\}$  —

2:   Initialize: ( $n_{\text{bound}}, n_{\text{activated}}, n_{\text{proliferate}}^*, n_{\text{differentiate}}^*, e_{\text{differentiate}}, m_{\text{differentiate}}, e_{\text{contract}}$ )  $\leftarrow 0$ 

    ▷ (i) recruitment of naïve cells to interact with APC for activation

3:   if  $N > 0$  then
4:      $n_{\text{bound}} \leftarrow \text{Binom}(N, r_{\text{bind}} \Delta t)$   ▷ — binomial sampling of naïve cells to bind to APC, rate  $r_{\text{bind}}$ 
    —

5:      $n_{\text{activated}} \leftarrow \text{Binom}([\text{APC} \cdot N], r_{N \rightarrow N^*} \Delta t)$   ▷ — binomial sampling to activate APC-bound
    naïves, rate  $r_{N \rightarrow N^*}$  —

6:   end if

    ▷ (ii) burst-like proliferation of activated naïves  $N^*$ , differentiation to effectors  $E$ , or memory  $M$ 

7:   if  $0 < N^* < N_{\text{max}}^*$  then
8:      $n_{\text{proliferate}}^* \leftarrow \text{Binom}(N^*, r_{N^* \rightarrow 2N^*} \Delta t)$   ▷ — binomial sampling to replicate  $N^*$ , rate  $r_{N^* \rightarrow 2N^*}$ 
    (initial burst) —

9:      $B_{\text{differentiate}} \leftarrow B_{\text{differentiate}} + r_{N^* \rightarrow E} \Delta t$   ▷ — updating likelihood of  $N^*$  differentiating to  $E$  by
    time  $t$  —

10:  else if  $N^* \geq N_{\text{max}}^*$  then
11:     $n_{\text{differentiate}}^* \leftarrow N^*$   ▷ — the number of activated naïves that can differentiate —

```

12: $e_{\text{differentiate}} \leftarrow \text{Binom}(N^*, 1 - \exp(-B_{\text{differentiate}}))$ $\triangleright$ — binomial sampling to differentiate N^*
to E during burst —

13: $m_{\text{differentiate}} \leftarrow N^* - e_{\text{differentiate}}$ $\triangleright$ — number of activated cells differentiating to memory —

14: $e_{\text{stop prolifer.}} \leftarrow \min(e_{\text{differentiate}}, N_{\text{max}}^*) \times \frac{E_{\text{max}}}{N_{\text{max}}^*}$ $\triangleright$ — maximum number of allowed effectors in the
lineage —

15: **end if**

$\triangleright$ (iii) contraction and proliferation of effectors with rates $r_{E \rightarrow \emptyset}$, and $r_{E \rightarrow 2E}$, respectively

16: $e_{\text{contract}} \leftarrow \text{Binom}(E, r_{E \rightarrow \emptyset} \Delta t)$ $\triangleright$ — binomial sampling to remove E —

17: **if** $B_{\text{proliferate}} < e_{\text{stop prolifer.}}$ **then**

18: $e_{\text{proliferate}} \leftarrow \text{Binom}(E, r_{E \rightarrow 2E} \Delta t)$ $\triangleright$ — binomial sampling to replicate E —

19: $B_{\text{proliferate}} \leftarrow B_{\text{proliferate}} + e_{\text{proliferate}}$ $\triangleright$ — running count of the number of effector cell divisions
—

20: **end if**

$\triangleright$ (iv) update the populations of naïve N , naïve bound to APC $[APC \cdot N]$, activated N^* , effector E , and
memory M cells:

21: $N \leftarrow N - n_{\text{bound}}$

22: $[APC \cdot N] \leftarrow [APC \cdot N] + n_{\text{bound}} - n_{\text{activated}}$

23: $N^* \leftarrow N^* + n_{\text{activated}} + n_{\text{proliferate}}^* - n_{\text{differentiate}}^*$

24: $E \leftarrow E + e_{\text{proliferate}} - e_{\text{contract}} + e_{\text{differentiate}}$

25: $M \leftarrow M + m_{\text{differentiate}}$

26: **return** $N, [APC \cdot N], N^*, E, M, e_{\text{stop prolifer.}}, B_{\text{differentiate}}, B_{\text{proliferate}}$

27: end function

A.3 Optimization of $CD8^+$ T cell response to pathogens

Design architecture of T cell programs for immune response

Infection clearance results from coordination of the effector trajectory $E(t)$ in response to an infection $I(t)$. Physically, this corresponds to controlling/regulating the rates, $\vec{r}(t)$ of cell state transitions, divisions and deaths, potentially varying $4 \times 4 = 16$ module-specific sensitivity and baseline parameters,

$$\begin{aligned}
\text{activation: } & \psi_{N \rightarrow N^*}^{\sigma_{\text{Ag}}}, \psi_{N \rightarrow N^*}^{\sigma_{\text{inf}}}, \psi_{N \rightarrow N^*}^{\sigma_{\text{res}}}, \ell_{N \rightarrow N^*} \in \mathbb{R}, \\
\text{differentiation: } & \psi_{N^* \rightarrow E}^{\sigma_{\text{Ag}}}, \psi_{N^* \rightarrow E}^{\sigma_{\text{inf}}}, \psi_{N^* \rightarrow E}^{\sigma_{\text{res}}}, \ell_{N^* \rightarrow E} \in \mathbb{R}, \\
\text{proliferation: } & \psi_{E \rightarrow 2E}^{\sigma_{\text{Ag}}}, \psi_{E \rightarrow 2E}^{\sigma_{\text{inf}}}, \psi_{E \rightarrow 2E}^{\sigma_{\text{res}}}, \ell_{E \rightarrow 2E} \in \mathbb{R}, \\
\text{contraction: } & \psi_{E \rightarrow \emptyset}^{\sigma_{\text{Ag}}}, \psi_{E \rightarrow \emptyset}^{\sigma_{\text{inf}}}, \psi_{E \rightarrow \emptyset}^{\sigma_{\text{res}}}, \ell_{E \rightarrow \emptyset} \in \mathbb{R},
\end{aligned} \tag{A.27}$$

to generate effector response trajectories.

Accounting for shared regulatory programs and controlling the combinatorial complexity of our model parameter space, we impose two simplifications. First, activation and proliferation are both MYC-driven [36]; we therefore constrain the modulatory factors, $g_{N \rightarrow N^*}$ and $g_{E \rightarrow 2E}$, to share parameter values. Note, however, that their realized rates will differ because signals vary in time, as cells transition between naive and effector states. Second, we make the assumption that extracellular cues converge on coherent effector programs, and thus enforce a common set of signal sensitivities across modules. With the sign convention that positive sensitivities promote effector production, for each signal σ , we set

$$\psi_{\sigma} \equiv \psi_{N \rightarrow N^*}^{\sigma} = \psi_{E \rightarrow 2E}^{\sigma} = \psi_{N^* \rightarrow E}^{\sigma} = -\psi_{E \rightarrow \emptyset}^{\sigma}.$$

Each module i nonetheless retains its own baseline parameter ℓ_i , or equivalently the baseline rate $g_{0;i} = (1 + \exp(-\ell))^{-1}$. With this simplification we reduce the number of degrees of

freedom in our model from 16 to six (three sensitivity and three baseline rates),

$$\begin{aligned}
\text{activation: } & \psi_{N \rightarrow N^*}^{\sigma_{\text{Ag}}} = \psi_{\sigma_{\text{Ag}}}, \psi_{N \rightarrow N^*}^{\sigma_{\text{inf}}} = \psi_{\sigma_{\text{inf}}}, \psi_{N \rightarrow N^*}^{\sigma_{\text{res}}} = \psi_{\sigma_{\text{res}}}, \ell_{N \rightarrow N^*} \in \mathbb{R}, \\
\text{differentiation: } & \psi_{N^* \rightarrow E}^{\sigma_{\text{Ag}}} = \psi_{\sigma_{\text{Ag}}}, \psi_{N^* \rightarrow E}^{\sigma_{\text{inf}}} = \psi_{\sigma_{\text{inf}}}, \psi_{N^* \rightarrow E}^{\sigma_{\text{res}}} = \psi_{\sigma_{\text{res}}}, \ell_{N^* \rightarrow E} \in \mathbb{R}, \\
\text{proliferation: } & \psi_{E \rightarrow 2E}^{\sigma_{\text{Ag}}} = \psi_{\sigma_{\text{Ag}}}, \psi_{E \rightarrow 2E}^{\sigma_{\text{inf}}} = \psi_{\sigma_{\text{inf}}}, \psi_{E \rightarrow 2E}^{\sigma_{\text{res}}} = \psi_{\sigma_{\text{res}}}, \ell_{E \rightarrow 2E} = \ell_{N \rightarrow N^*} \in \mathbb{R}, \\
\text{contraction: } & \psi_{E \rightarrow \emptyset}^{\sigma_{\text{Ag}}} = -\psi_{\sigma_{\text{Ag}}}, \psi_{E \rightarrow \emptyset}^{\sigma_{\text{inf}}} = -\psi_{\sigma_{\text{inf}}}, \psi_{E \rightarrow \emptyset}^{\sigma_{\text{res}}} = -\psi_{\sigma_{\text{res}}}, \ell_{E \rightarrow \emptyset} \in \mathbb{R}.
\end{aligned} \tag{A.28}$$

Note the relationship, $\psi_{E \rightarrow \emptyset}^{\sigma} = -\psi_{\sigma}$ reflects coordinated signal-regulation of effector-promoting and effector-suppressing transitions.

The parameters of these constrained response functions define a *response design* $\vec{\theta}$ as

$$\vec{\theta} = (\underbrace{\psi_{\text{Ag}}, \psi_{\text{inf}}, \psi_{\text{res}}}_{\text{signal feedback}}, \underbrace{\ell_{N \rightarrow N^*}, \ell_{E \rightarrow 2E}, \ell_{N^* \rightarrow E}, \ell_{E \rightarrow \emptyset}}_{\text{module-specific baselines}}). \tag{A.29}$$

For each transition i , we choose the baseline response parameter, ℓ_i , such that $|\ell_i| \leq \ell^{\max}$ with

$$r_i(\ell^{\max}) \sim r^{\max}, \quad r_i(-\ell^{\max}) \sim T^{-1},$$

where T is the duration of the simulation. This range is set to be wide enough that at the lower baseline rate ($-\ell^{\max}$) transitions are seldom observed during our simulation time window, whereas at the upper baseline rate ($\ell^{\max}$) they occur near the maximum rate. Here, we take $\ell^{\max} = 3.0$, so $r_i(\ell^{\max} = 3) \approx 0.95 \times r^{\max}$. For the signal-sensitivity parameter ψ_{σ} , we require that under weak stimulation ($\sigma = 1$) the signal term in the response function (eq. A.20) is below the threshold of response, set by the maximal baseline term, i.e., $\psi_{\sigma}^{\max} \cdot \log(1 + 1) < \ell^{\max}$, and choose $\psi_{\sigma}^{\max} = 3.0$.

Optimal design for an immune challenge. We define high performing designs $\vec{\theta}$ as those that tend to minimize the harm from an immune challenge. For a challenge characterized by immunogenicity κ and basic reproduction number R_0 , we define the optimal design as,

$$\vec{\theta}^{\min}(\kappa, R_0) = \arg \min_{\vec{\theta}} H(\vec{\theta}). \quad (\text{A.30})$$

where the harm $H(\vec{\theta})$ is computed by eq. A.10 for growing (acute) infections, and by eq. A.12 for slow-growing cancer tumors.

To do this optimization for a fixed pathogenic threat (κ, R_0) , we perform a uniform grid search over $\vec{\theta} \in [-\psi^{\max}, \psi^{\max}]^3 \times [-\ell^{\max}, \ell^{\max}]^3$. We use a Cartesian grid with $m = 13$ equally spaced points per coordinate, yielding a total of $13^6 \approx 4.8 \times 10^6$ candidate designs. For each $\vec{\theta}$, we simulate the immune response dynamics to compute the associated harm and identify the design that minimizes harm $\vec{\theta}^{\min}(\kappa, R_0)$.

Multi-objective design optimization across pathogen archetypes. To find designs that can perform well across pathogen archetypes, we pose a principled multi-objective problem for the immune program, i.e., choose a design $\vec{\theta}$ that minimizes the weighted sum of harms across pathogen archetypes:

$$\vec{\theta}'(\{\lambda_i\}) = \arg \min_{\vec{\theta}} \sum_{i:\text{archetypes}} \lambda_i H^{(i)}(\vec{\theta}), \quad (\text{A.31})$$

where $\lambda_i \in [0, 1]$ with the constraint $\sum_i \lambda_i = 1$. The weights $\{\lambda_i\}$ determine the relative importance of each archetype in determining the optimal design $\vec{\theta}'$. The solution to this family of problems forms a simplex over which λ_i 's vary, called the *Pareto front*; see ref. [83] for an extensive discussion on multi-objective optimization, and refs. [87, 147] for successful

application of these methods in different biological systems.

Similar to the case of optimization for a single immune challenge, we estimate the Pareto front by sampling designs on a grid and identifying the design that minimizes the weighted harm $\sum_{i:\text{archetypes}} \lambda_i H^{(i)}(\vec{\theta})$ for a given instance of $\{\lambda_i\}$.

Finding the elbow of the two-dimensional Pareto curve. Even though optimization can be done simultaneously over all three archetypes, we examine trade-offs between pairs of archetypes A and B , generating one-dimensional Pareto fronts, with single independent weights $\lambda = \lambda_A = 1 - \lambda_B$. The curvature of the Pareto front quantifies the lowest achievable increase in harm from one archetype that we must trade to decrease harm from the other archetype by tuning design parameters. Given a convex Pareto front for archetypes A and B , in some regions of the curve (values of λ) the benefit of decreasing harm from one archetype may be larger than the cost. A common heuristic for multi-objective optimization in engineering problems is to identify the point along the curve, $(H[\vec{\theta}(\lambda^*)]^{(A)}, H[\vec{\theta}(\lambda^*)]^{(B)})$, called the *elbow* or *knee*, at which no additional benefit can be obtained by tuning parameters [277]. To estimate the elbow of the Pareto front, we first interpolate the discretely sampled Pareto points to obtain a continuous curve. This procedure is sensitive to the density of sampled points near the Pareto front and requires fine-tuning. For plots involving archetype I (Fig. 3.4, A.6), we found that subsetting to points where $H(\vec{\theta})^{(A)} < 10^{-2} S_{\max}$, and linearly interpolating this subset of performed best. Finally, we estimated the elbow point (λ^*) along the interpolated curve using the Kneedle algorithm [277, 278].

Identifying the ensemble of near-optimal designs. Since we are sampling over a finite grid and our simulations are stochastic, we must contend with the likelihood that discretely sampled points will not coincide with the true minimizer in the continuous space of designs

and that stochastic effects will add noise to our measurements of harm. To capture this uncertainty, we identify a *near-optimal ensemble* of designs defined as the set of designs θ such that $H(\vec{\theta}) \leq H(\vec{\theta}') + \epsilon$, where $\vec{\theta}'$ is the sampled minimizer and $0 < \epsilon$ is a threshold. We typically choose $\epsilon = 10^{-2}S_{\max}$, and compute the mean and variance of designs in this ensemble, as shown in Fig. 3.3 D.

In the case of multi-objective optimization, first we estimate the Pareto front for a pair of archetypes $\{A, B\}$. For each point on the Pareto front, $\vec{\theta}'(\lambda)$ (equation A.31), we identify an ensemble of near-Pareto designs

$$\Theta(\lambda) = \{\vec{\theta} | H(\vec{\theta})^{(A)} \leq (1 + \alpha)H[\vec{\theta}'(\lambda)]^{(A)}, H(\vec{\theta})^{(B)} \leq (1 + \alpha)H[\vec{\theta}'(\lambda)]^{(B)}\}, \quad (\text{A.32})$$

with $\alpha = 5 \times 10^{-2}$.

Clustering of pathogen archetypes

For each infection scenario (κ, R_0) , we identify the ensemble of sampled designs $\vec{\theta}$ such that $H(\vec{\theta}) \leq H(\vec{\theta}^{\min}) + \epsilon$, where $\vec{\theta}^{\min}$ is the harm-minimizing design for that scenario and $\epsilon = 10^{-2}S_{\max}$. We call this ensemble $\Theta(\kappa, R_0)$, and compute its mean, $\langle \vec{\theta} \rangle_{\Theta(\kappa, R_0)}$ (Fig. A.3). Next, we pool the mean vectors of all infection scenarios and normalize their distribution so each design parameter has zero mean and unit variance. Finally, we use a spectral clustering algorithm implemented in Python [279] to cluster scaled mean design parameters to identify distinct pathogen archetypes. We selected the cluster number $n_{\text{cluster}} = 3$ that maximized the mean silhouette coefficient (implemented in Python [279]),

$$\sum_{(\kappa, R_0)} [b(\kappa, R_0) - a(\kappa, R_0)] / \max[a(\kappa, R_0), b(\kappa, R_0)],$$

where $a(\kappa, R_0)$ is the mean distance between $\langle \vec{\theta} \rangle_{\Theta(\kappa, R_0)}$ and every other point in its assigned cluster, and $b(\kappa, R_0)$ is the mean distance between $\langle \vec{\theta} \rangle_{\Theta(\kappa, R_0)}$ and every other point in its nearest, neighboring cluster. The mean silhouette coefficient measures cohesiveness of clusters and how distinct they are from each other.

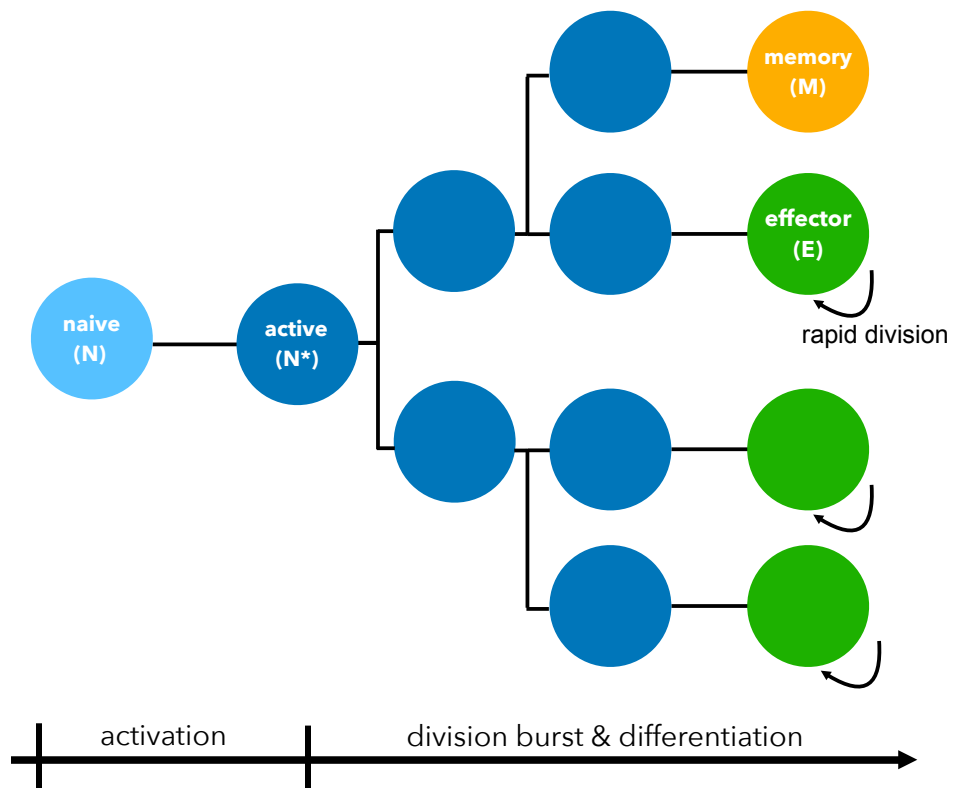


Figure A.1: Schematic of modeled early T cell division bursts and differentiation.

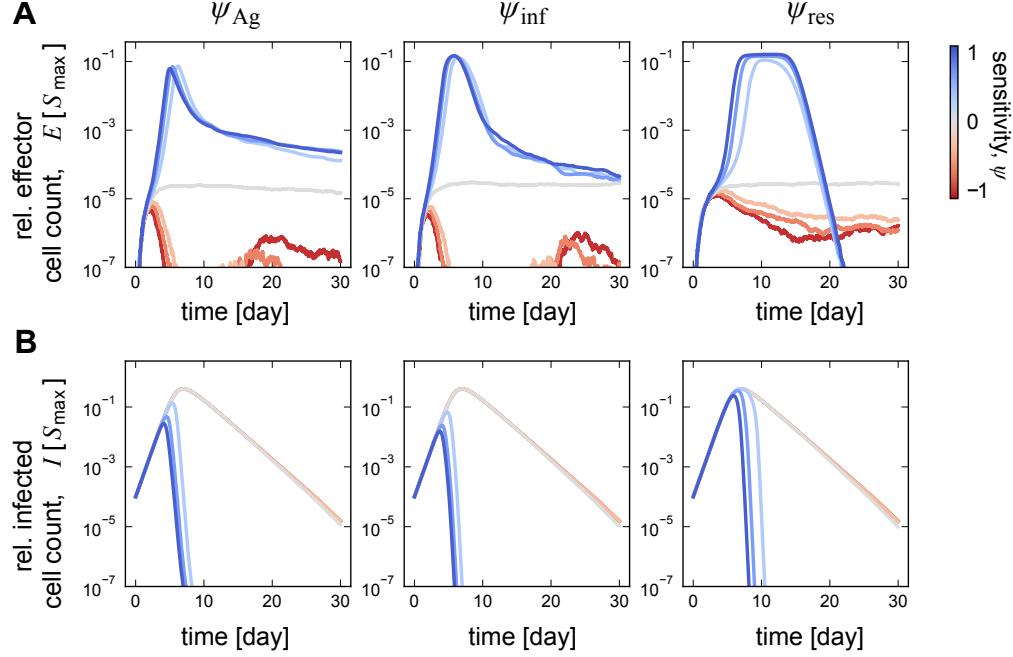


Figure A.2: Design parameters shape T cell response and infection clearance. (A, B) Time course of the relative population size of (A) effector cells E and (B) infected cells is shown for designs when varying (color) a single sensitivity parameter ψ_σ (indicated at the top of each column), while keeping the rest fixed at: moderate activation / proliferation $g_{0,N \rightarrow N^*} = g_{0,E \rightarrow 2E} = 0.5$; moderate differentiation $g_{0,N^* \rightarrow E} = 0.5$; moderate contraction $g_{0,E \rightarrow \emptyset} = 0.5$; and zero sensitivity to the other signals $\psi_{Ag} = \psi_{inf} = \psi_{res} = 0$. All population sizes are measured in units of the maximum number of susceptible cells $S_{\max}$ and sensitivities in units of $\psi_{\max}$. Other simulation parameters as in Fig. 3.2.

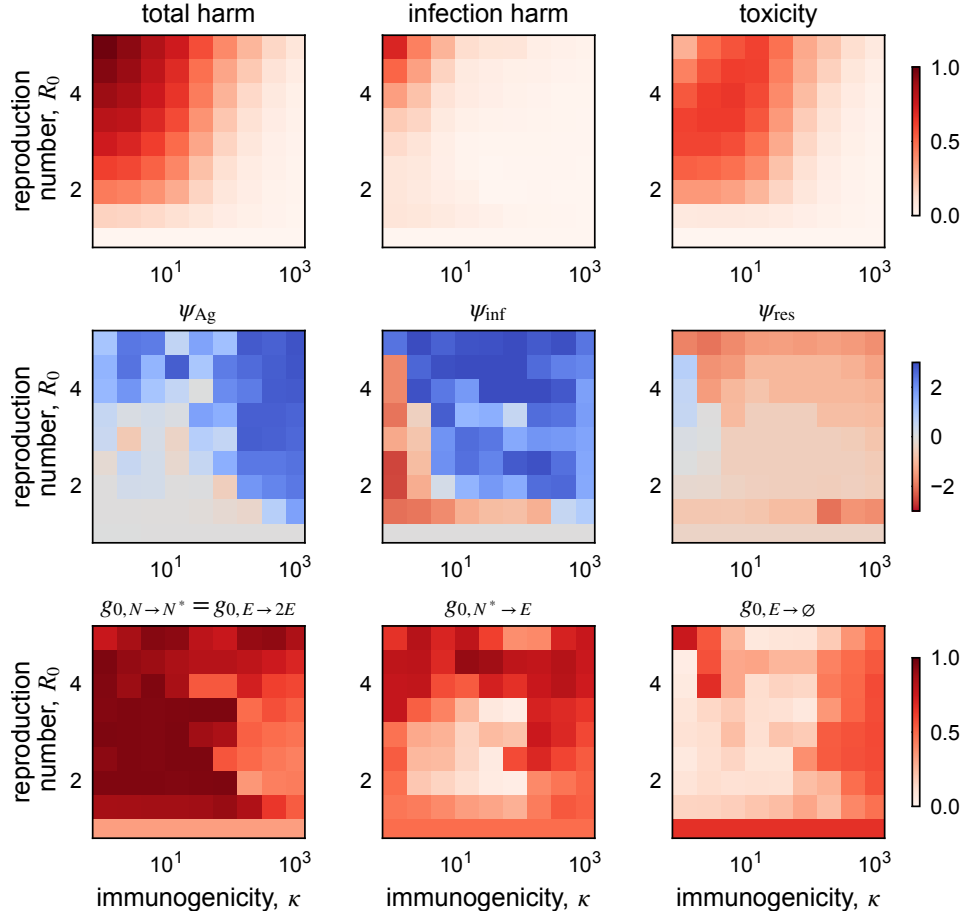


Figure A.3: Near-optimal performance and designs for each immune challenge. Heat maps showing the near-optimal outcomes and design values that achieve them for each infection scenario (κ, R_0) . The first row displays the average total harm, infection harm and toxicity incurred by the near-optimal designs for each infection scenario. The last two rows display the average near-optimal design values for each scenario, $\langle \vec{\theta} \rangle_{\Theta(\kappa, R_0)}$. The data in the last two rows is used for clustering infection scenarios in Fig. 3.3 B.

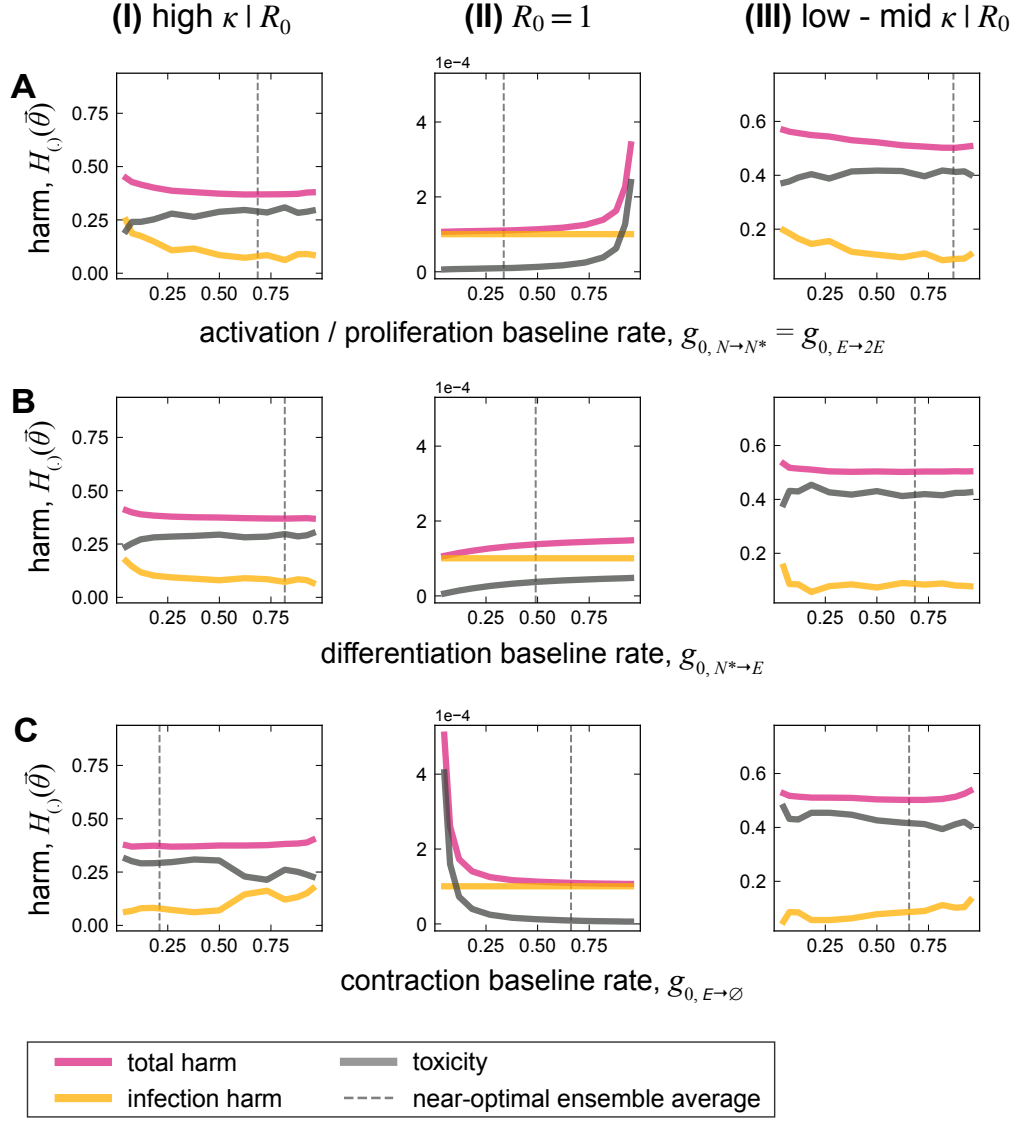


Figure A.4: Dependence of harm on baseline rates in different archetypes. Total harm, infection harm, and toxicity (colors) are shown as a function of baseline rates for **(A)** activation / proliferation $g_{0, N \rightarrow N^*} = g_{0, E \rightarrow 2E}$, **(B)** differentiation $g_{0, N^* \rightarrow E}$ and **(C)** contraction $g_{0, E \rightarrow \emptyset}$, in different archetypes as indicated above each column. In each panel, the value of the design parameter on the horizontal axis is fixed and total harm is minimized by varying the other design parameters. The dashed line in each panel indicates the near-optimal ensemble average values shown in bar charts in Fig. 3.3 D. Simulation parameters as in Fig. 3.3.

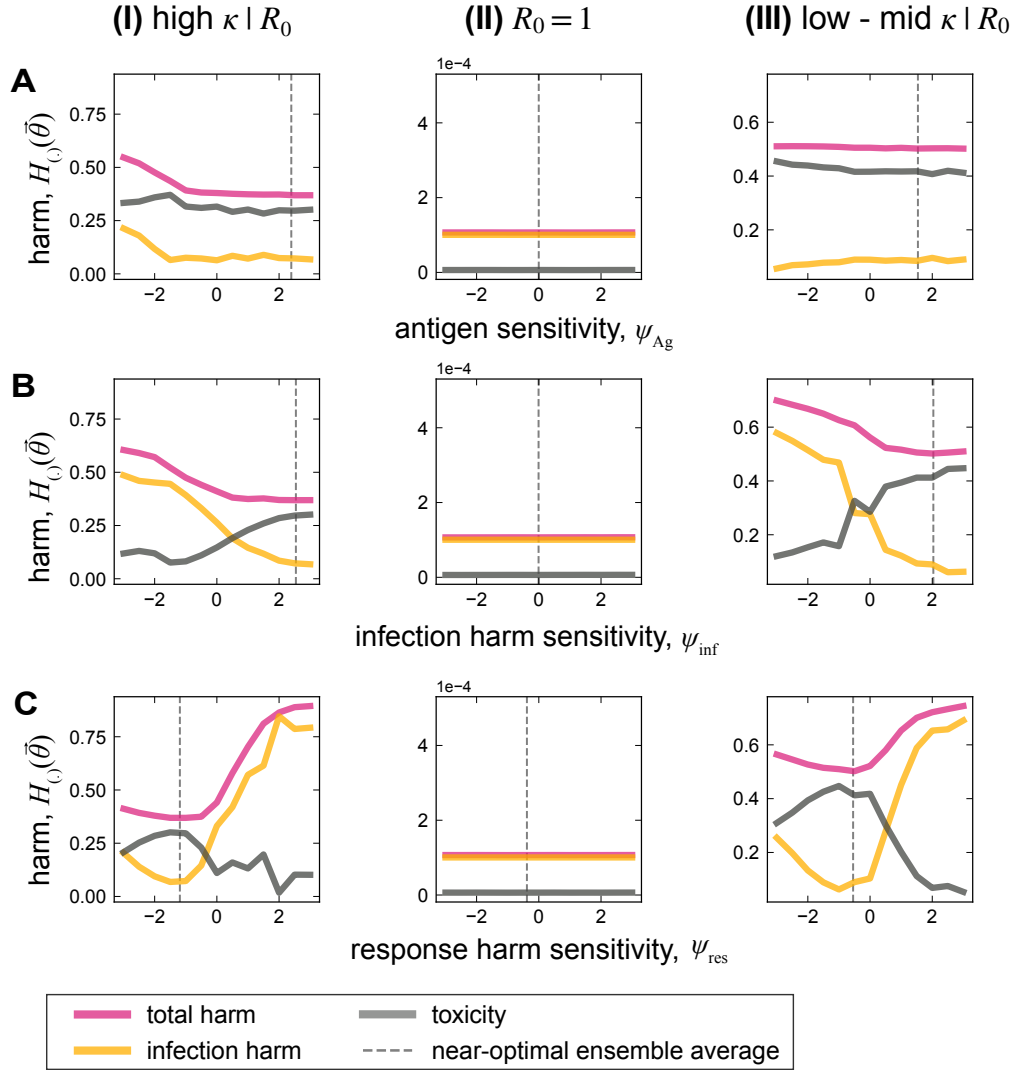


Figure A.5: Dependence of harm on signal sensitivities in different archetypes. Similar to Fig. A.4, but when varying the sensitivities to **(A)** the antigen signal ψ_{Ag} , **(B)** the infection harm signal ψ_{inf} , and **(C)** the response harm signal ψ_{res} .

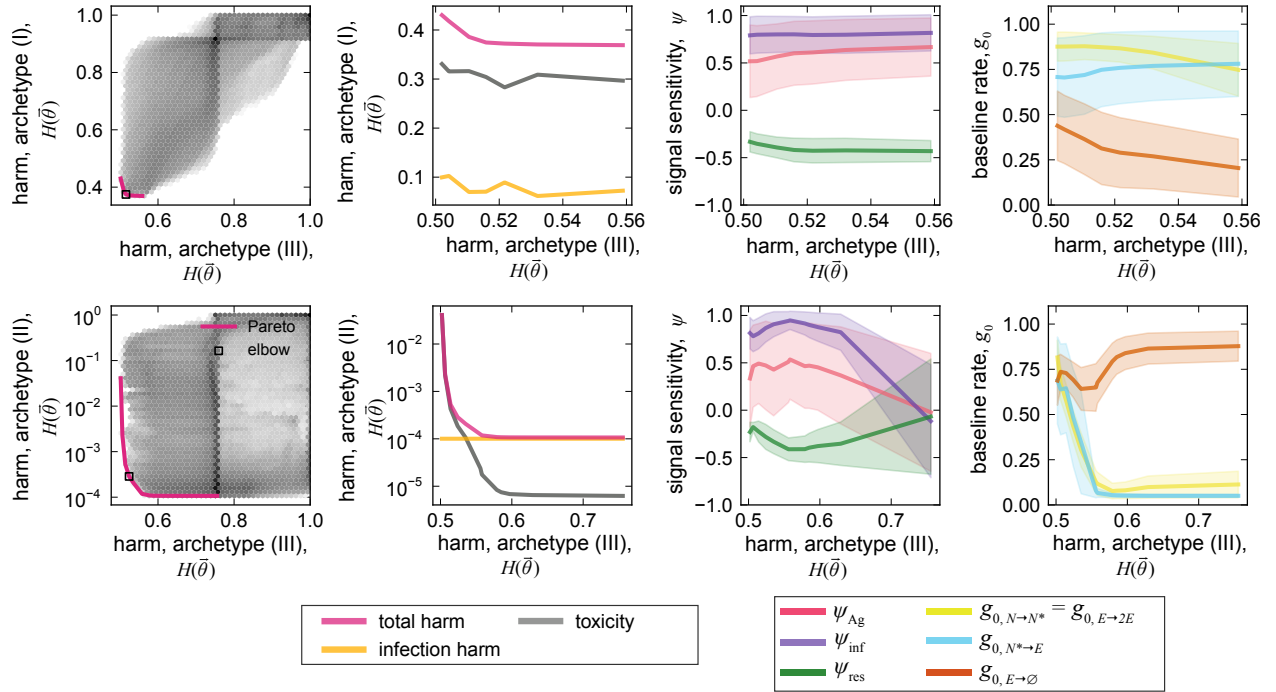


Figure A.6: Pareto-optimal design balancing harm from autoimmunity and other pathogen archetypes. Similar to Fig. 3.4, but showing the Pareto optimal designs resulting from joint optimization of pathogen archetypes I and III, and pathogen archetypes II and III (rows); joint optimization for archetypes I and II is shown in Fig. 3.4. Simulation parameters as in Fig. 3.4.

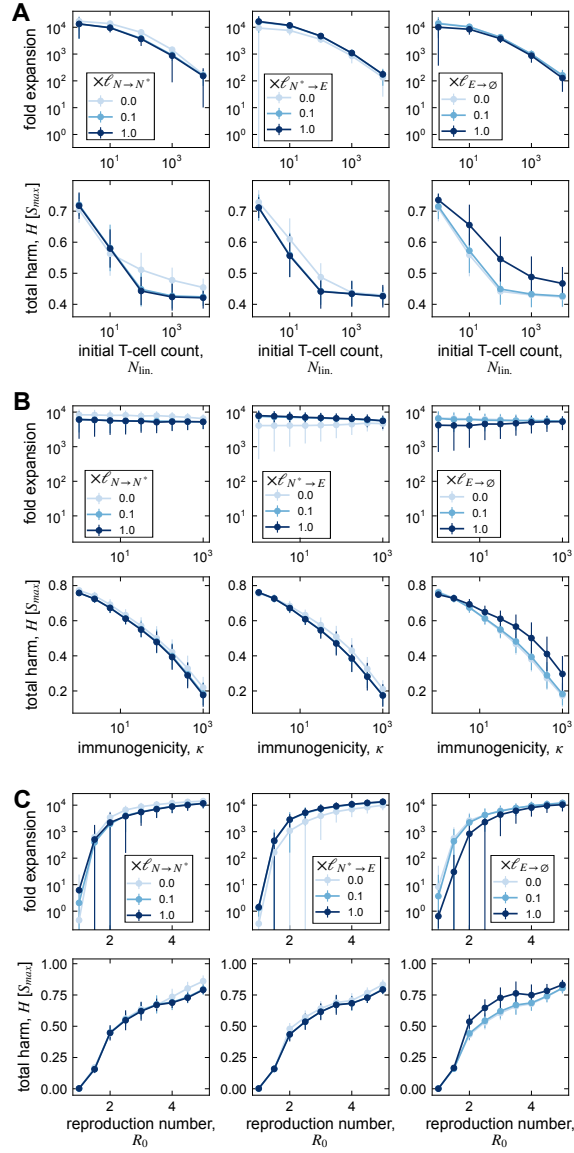


Figure A.7: Dependence of T cell macro-dynamics on baseline rates. T cell fold expansion (top) and total harm (bottom) are shown as functions of **(A)** initial T cell number $N_{\text{lin.}}$, **(B)** immunogenicity κ , **(C)** and the intra-host basic reproduction of the infection R_0 . In each panel, outcomes are shown for the “elbow” Pareto-optimal design from Fig. 3.4 B while varying a single baseline rate (colors)—activation/proliferation baseline rate $g_{0,N \rightarrow N^*} = g_{0,E \rightarrow 2E} = (1 + \exp(-\ell_{N \rightarrow N^*}))^{-1}$ (left), differentiation baseline rate $g_{0,N^* \rightarrow E} = (1 + \exp(-\ell_{N^* \rightarrow E}))^{-1}$ (center), or contraction baseline rate $g_{0,E \rightarrow \emptyset} = (1 + \exp(-\ell_{E \rightarrow \emptyset}))^{-1}$ (right)—with all other parameters held fixed. Trend lines and standard deviation error bars are obtained by averaging over sampled immune challenges (in Fig. 3.3 B) and designs in the near-elbow ensemble. Simulation parameters as in Fig. 3.5.

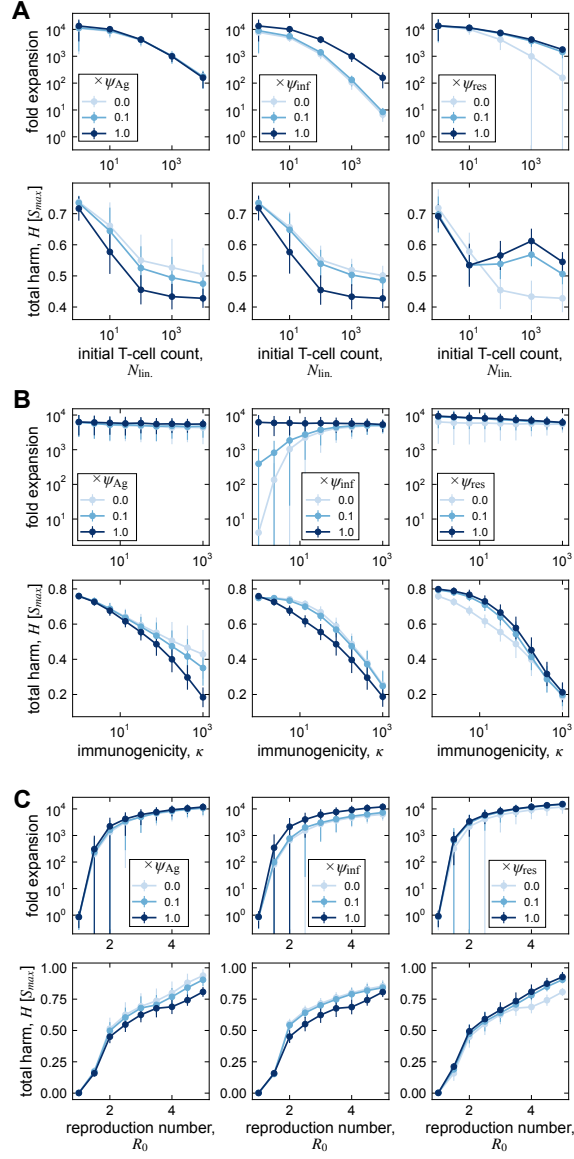


Figure A.8: Dependence of T cell macro-dynamics on signal sensitivities. T cell fold expansion (top) and total harm (bottom) are shown as functions of **(A)** initial T cell number $N_{\text{lin.}}$, **(B)** immunogenicity κ , **(C)** and the intra-host basic reproduction of the infection R_0 . In each panel, outcomes are shown for the “elbow” Pareto-optimal program from Fig. 4B while varying a single sensitivity parameter (colors)—antigen sensitivity ψ_{Ag} (left), infection-harm sensitivity ψ_{inf} (center), or response-harm sensitivity ψ_{res} (right)—with all other parameters held fixed. Trend lines and standard deviation error bars are obtained by averaging over sampled immune challenges (in Fig. 3.3 B) and designs in the near-elbow ensemble. Simulation parameters as in Fig. 3.5.

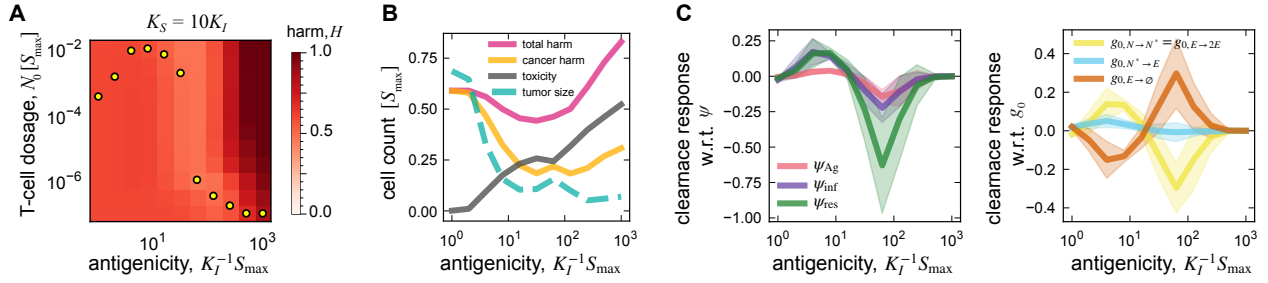


Figure A.9: Impact of cross-reactivity on engineered immune responses for cancer immunotherapy. As in Fig. 3.6, but incorporating cross-reactivity of engineered TCRs with self and tumor antigens. We model cross-reactivity by assuming tumor and self reactivities are proportional, fixing the tumor immunogenicity $\kappa = K_I^{-1} K_S = 10$ while varying the antigenicity of the tumor defined as $K_I^{-1} S_{\max}$ (proxy for inverse EC_{50} , where larger values indicate higher antigenicity). **(A)** With responses averaged over the tuned near-elbow ensemble $\tilde{\Theta}(\lambda^*)$, the heat map shows harm as a function of T cell dosage and tumor antigenicity. **(B)** Total harm, cancer harm, immune toxicity, and the final tumor size following the immune response are shown as functions of tumor antigenicity at the optimal T cell dosage (yellow circles in A). **(C)** Clearance response, defined as sensitivity of protection (negative harm) to each design parameter—evaluated and averaged over $\theta \in \tilde{\Theta}(\lambda^*)$ —is shown for different levels of tumor antigenicity for signal sensitivity parameters ψ 's (left), and for baseline parameters g_0 's (right). For this analysis, we assume that T cells are administered at high dosage ($N_{\text{lin.}} = 10^{-2} S_{\max}$). Simulation parameters as in Fig. 3.6.

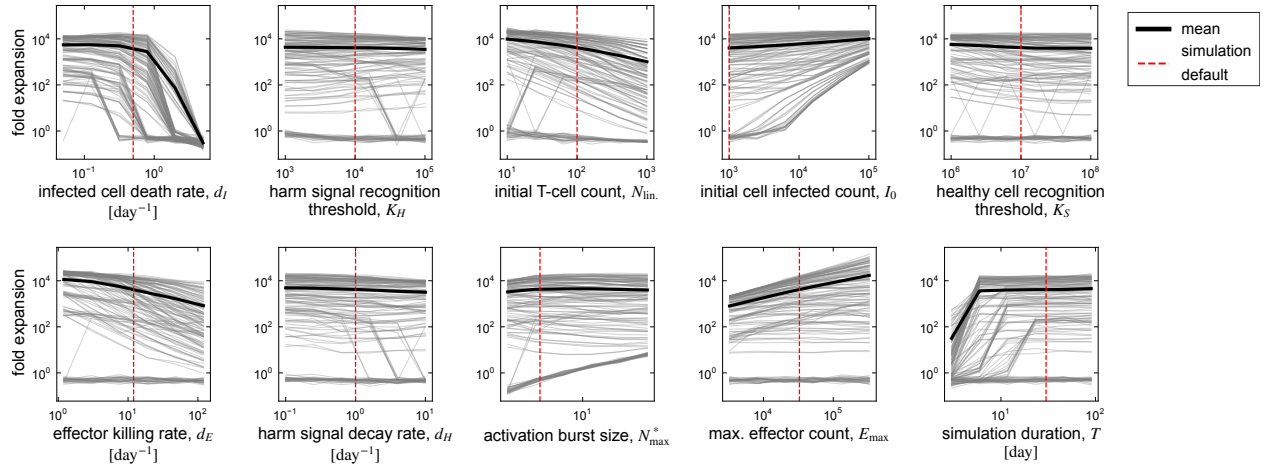


Figure A.10: Parameter sensitivity analysis for total harm. Total harm, evaluated and averaged over the near-elbow ensemble $\theta \in \tilde{\Theta}(\lambda^*)$, is shown for different levels of the indicated parameter value. Gray lines represent trend lines for each infection condition (κ, R_0) in Fig. 3.3 B, and the thick black line is the average trend over all infections. The red dotted line indicates the default parameter value used in simulations (Table A.1). Except when explicitly varied, simulation parameters are as in Fig. 3.3.

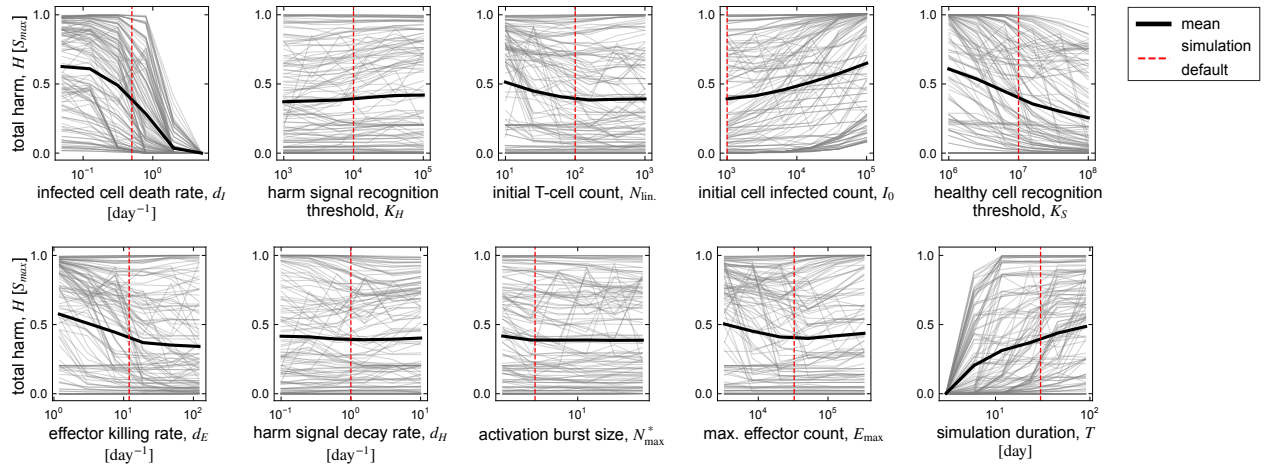


Figure A.11: Parameter sensitivity analysis for effector fold expansion. Similar to Fig. A.10 but showing the trend for effector fold expansion.

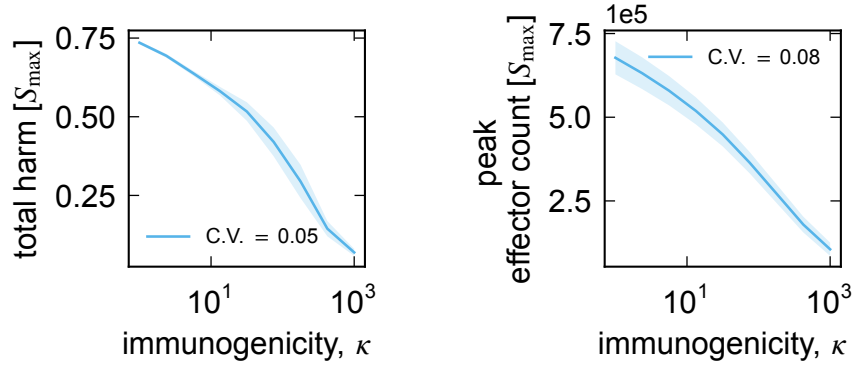


Figure A.12: Sensitivity of harm magnitude and T cell expansion to stochastic cell-fate decisions. (A) Total harm, and (B) effector-cell expansion, averaged over the near-elbow design ensemble $\theta \in \Theta(\lambda^*)$, and over immune challenges spanning the range of reproduction numbers R_0 in Fig. 3.3, are shown as functions of immunogenicity κ . Insets report the coefficients of variation (C.V.), averaged across immune challenges spanning different κ 's. Except when explicitly varied, simulation parameters are as in Fig. 3.3.

Table A.1: Model parameters for simulations. The * symbol indicates parameters that were varied as part of the analysis.

Parameter	Value	Reference(s)
initial susceptible cell count (S_0)	10^7	
initial infected cell count (I_0)	10^3	
infected cell death rate (d_I)	0.5 day^{-1}	[280, 281]
susceptible cell infection rate (b_I)*	$1 - 5 \times 10^{-7} \text{ infected cell}^{-1} \text{ day}^{-1}$	[280]
maximum cancer growth rate (b_C)	0.1 day^{-1}	
effector cell killing rate (d_E)	12 day^{-1}	[282]
infected cell recognition threshold (K_I)*	$10^4 - 10^7 \text{ cells}$	[171]
healthy cell recognition threshold (K_S)*	10^7 cells	
harm signal recognition threshold (K_H)	10^4 cells	
max. binding rate of APCs to Naive cells ($r_{\text{bind}}^{\text{max}}$)	1 day^{-1}	[60, 63, 269]
decay rate of harm signals (d_H)	1.0 day^{-1}	[283]
initial naive cognate T cell count ($N_{\text{lin.}}$)*	10^2	[17]
activation burst size (N_{max}^*)	2^2	[51]
max. effector count per lineage (E_{max})	2^{15}	[15, 35, 109, 110]
max. rate of T cell state transition (r^{max})	4 day^{-1}	[28, 42, 51, 116]
rate of activated naive cell division ($r_{N^* \rightarrow 2N^*}$)	$\frac{24}{8.8} \text{ day}^{-1}$	[51]
signal sensitivity (ψ_σ)*	$\in [-3, 3]$	
baseline parameter (ℓ_i)*	$\in [-3, 3]$	
simulation duration (T)	30 days	[109, 148, 264]
time step (Δt)	10^{-2} days	

Appendix B

SUPPLEMENT FOR CHAPTER 4

B.1 Data analysis

Data and code availability

Data on microbial composition was previously published in [179, 207], and data on T cell reactivity and counts was previously published in [206]. All code for data processing and statistical analysis will be made available on Github upon publication of the manuscript.

Clustering of strains

Reactive strains were grouped using K-means clustering based on their L -1 normalized reactivities for the responsive T cells. We selected the cluster number $K = 8$ that nearly maximized the mean silhouette coefficient (implemented in Python [279]),

$$\sum_{\mu=1}^{45} [b(\mu) - a(\mu)] / \max[a(\mu), b(\mu)],$$

where $a(\mu)$ is the mean distance between strain μ 's reactivity vector and every other point in its assigned cluster, and $b(\mu)$ is the mean distance between strain μ 's reactivity vector and every other point in its nearest, neighboring cluster. The mean silhouette coefficient

measures the cohesiveness of clusters and how distinct they are from each other.

Processing and transformation of microbial composition data

Before analysis, strains that were undetected across all replicates at week 1 were dropped from the analysis. Otherwise, missing relative abundance data was set to 10^{-8} , consistent with the reported limit of detection in the study [207]. Relative abundance data for all strains were then normalized to sum to 1, per time-point and replicate.

To quantify compositional shifts at the strain level, we computed the centered log-ratio (CLR) transform of strain relative abundances at each time $t > 0$, $g_\mu(t) / \sum_{m=1}^M g_m(t)$, $\mu = 1, \dots, M$. For each strain, the CLR is

$$\text{CLR}_\mu(t) = \log_{10} \left(\frac{g_\mu(t)}{\sum_{m=1}^M g_m(t)} \middle/ \sqrt[M]{\prod_{\alpha=1}^M \frac{g_\alpha(t)}{\sum_{m=1}^M g_m(t)}} \right), \quad (\text{B.1})$$

where $M = 97$ and the total abundance ($\sum_{m=1}^M g_m(t)$) cancels out of the numerator and denominator of the expression inside the logarithm. This transformation minimizes spurious correlations introduced by the sum-to-one constraint of relative abundance data of the M microbial strains. We can then compute the compositional change of each strain μ from time point t_1 to t_2 :

$$\begin{aligned} \Delta \text{CLR}_\mu(t_1, t_2) &= \text{CLR}_\mu(t_2) - \text{CLR}_\mu(t_1) \\ &= \log_{10} \left(\frac{g_\mu(t_2)}{g_\mu(t_1)} \right) - \frac{1}{M} \sum_m^M \log_{10} \left(\frac{g_m(t_2)}{g_m(t_1)} \right), \end{aligned} \quad (\text{B.2})$$

from which we can observed that the first term is related to the abundance fold-change of the strain $g_\mu(t_2)/g_\mu(t_1)$, and the second term depends on the fold-change of every strain in

the microbiota. If we consider the gap in the change in CLR between two strains (μ, α) ,

$$\Delta\text{CLR}_\mu(t_1, t_2) - \Delta\text{CLR}_\alpha(t_1, t_2) = \log_{10} \left(\frac{g_\mu(t_2)}{g_\mu(t_1)} \right) - \log_{10} \left(\frac{g_\alpha(t_2)}{g_\alpha(t_1)} \right), \quad (\text{B.3})$$

this quantity depends only on the abundance fold-change of the two strains. Analogous expressions hold for the CLR of microbial clusters.

Fitting microbial composition data to a Dirichlet distribution

The Dirichlet distribution, $D(\vec{\alpha})$ is a non-negative, continuous, m -dimensional probability distribution parameterized by a vector of positive numbers $\vec{\alpha} \in \mathbb{R}^m$. For $m \geq 2$, if $\vec{X} \sim D(\vec{\alpha})$ then $\sum_{i=1}^m X_i = 1$ and its probability density function is:

$$f(\vec{x}; \vec{\alpha}) = \frac{1}{B(\vec{\alpha})} \prod_{i=1}^m x_i^{\alpha_i - 1} \quad (\text{B.4})$$

with

$$B(\vec{\alpha}) = \frac{\prod_{i=1}^m \Gamma(\alpha_i)}{\Gamma(\alpha_0)},$$

where $\Gamma(\cdot)$ is the standard gamma function, and $\alpha_0 = \sum_{i=1}^m \alpha_i$. Importantly, the mean and variance of X are given by:

$$\text{E}[X_i] = \frac{\alpha_i}{\alpha_0}, \quad \text{Var}[X_i] = \frac{\text{E}[X_i](1 - \text{E}[X_i])}{\alpha_0 + 1}, \quad (\text{B.5})$$

for $i = 1, \dots, m$.

Using the *dirichlet* package in Python ([284, 285]), we fit 5 biological replicates of relative abundance data to a Dirichlet distribution to obtain a maximum likelihood estimate of the parameters $\vec{\alpha}$ consistent with data.

B.2 Models of optimal repertoire allocation.

Model definition and analysis

We consider a model for a microbial ecology with strains g_μ , $\mu = 1, \dots, M$, that can grow in the gut and achieve a maximum population size $g_\mu^{\max}$. The immune response directed at mircoobe μ , C_μ , can inhibit the strain, decreasing its population:

$$\dot{g}_\mu = f(\mathbf{g}, C_\mu; g_\mu^{\max}), \quad (\text{B.6})$$

where,

$$C_\mu(\vec{\lambda}) = \sum_{r \in \{1, \dots, R\}} \sigma_{\mu, r} \lambda_r, \quad (\text{B.7})$$

is implemented by the balance of inflammatory and tolerant cells/receptors (λ_r) specific to detectable epitopes, $r = 1, \dots, R$, derived from the microbiota. The relationship between receptors/antigens and strains is given by the entries of the matrix $\Sigma = [\sigma_{\mu, r}]$, indicating how strongly a strain μ can be targeted by receptor/epitope r .

Assuming that the immune repertoire is updated slowly compared to microbial dynamics, then for a given immune coverage vector $\vec{C}$, there may be a steady-state solution to the growth dynamics for each strain population, $\tilde{g}_\mu(\vec{C}; g_\mu^{\max})$, for $\mu = 1, \dots, M$. For concreteness, we consider

$$\tilde{g}_\mu(C_\mu) = \frac{g_\mu^{\max}}{1 + K_C^{-1} C_\mu}, \quad (\text{B.8})$$

where K_C is the half-maximal constant of inhibition of strain abundance. In the analyses that follow, we make the convenient change of variables, $K_C^{-1} C_\mu \mapsto C_\mu$, expressing repertoire coverage in units of the unknown half-maximal constant of inhibition.

We assume that once in the gut a strain causes harm to the host organism at rate proportional to h_μ , and further that the objective of the immune response is to control the steady-state population of strains in the gut to minimize harm. Given a set of available receptors Σ , we seek the solution to the optimization problem

$$\min_{\vec{\lambda} \geq 0} \left\{ F(\vec{\lambda}; \Sigma) \right\} = \min_{\vec{\lambda} \geq 0} \left\{ \sum_{\mu} h_{\mu} \tilde{g}_{\mu}(\vec{\lambda}) + I(\vec{\lambda}) \right\}, \quad (\text{B.9})$$

where the first terms is the steady-state rate of harm induced by the strain and the second term corresponds to additional costs from maintaining an immune repertoire. Here, we specify the linear cost function,

$$I(\lambda) = \sum_r \beta \lambda_r. \quad (\text{B.10})$$

For example, the costs may describe an increased risk of autoimmunity associated with having more receptors, and/or the difficulty of generating particular receptors, and/or competition among receptor-bearing cells for limited survival signals.

To find the optimal solution, we write down the Lagrangian:

$$L(\vec{\lambda}; \delta) = \sum_{\mu} h_{\mu} g_{\mu}(C_{\mu}) + \sum_r \beta \lambda_r + \sum_r \delta_r \cdot (l_r^2 - \lambda_r), \quad (\text{B.11})$$

where we have introduced slack variables, $l_r \in \mathbb{R}$, and Lagrange multipliers, $\delta_r \in \mathbb{R}$, $r = 1, \dots, R$, to account for the non-negativity constraints on $\vec{\lambda}$.

Consider a system with M strains and R receptors, and immune coverage

$$C_{\mu} = \sum_{r \in \{1, \dots, R\}} \sigma_{\mu, r} \lambda_r, \quad (\text{B.12})$$

for $\mu = 1, \dots, M$, where λ_r corresponds to an inflammatory immune receptor/antibody r .

With $\tilde{g}_\mu(\cdot)$ defined as in eq. B.8, we will solve

$$\min_{\lambda_r \geq 0} \left\{ \sum_{\mu=1}^M h_\mu \tilde{g}_\mu(C_\mu(\vec{\lambda})) + \sum_{r=1}^R \beta \lambda_r \right\} \quad (\text{B.13})$$

by considering the Lagrangian

$$L_\beta(\vec{\lambda}, \boldsymbol{\delta}) = \sum_{\mu} h_\mu \frac{g_\mu^{\max}}{1 + C_\mu} + \sum_r [\beta \lambda_r + (l_r^2 - \lambda_r) \delta_r], \quad (\text{B.14})$$

with

$$\begin{aligned} \frac{\partial L_\beta}{\partial \lambda_r} &= \sum_{\mu} h_\mu \frac{\partial C_\mu}{\partial \lambda_r} \frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) + \beta - \delta_r, \\ \frac{\partial L_\beta}{\partial \delta_r} &= l_r^2 - \lambda_r, \end{aligned} \quad (\text{B.15})$$

for $r = 1, \dots, R$, where

$$\frac{\partial \tilde{g}_\mu}{\partial C_\mu} = -\frac{g_\mu^{\max}}{(1 + C_\mu)^2} = -\frac{\tilde{g}_\mu}{(1 + C_\mu)}, \quad (\text{B.16})$$

and

$$\frac{\partial C_\mu}{\partial \lambda_r} = \sigma_{\mu,r}. \quad (\text{B.17})$$

Note that

$$0 \geq \frac{\partial \tilde{g}_\mu}{\partial C_\mu} \rightarrow 0 \text{ as } \tilde{g}_\mu \rightarrow 0.$$

Optimality conditions. From the KKT conditions, solutions satisfy

$$\begin{aligned}\frac{\partial L_\beta}{\partial \lambda_r} &= \sum_{\mu} \sigma_{\mu,r} h_{\mu} \frac{\partial \tilde{g}_{\mu}}{\partial C_{\mu}}(\vec{\lambda}) + \beta - \delta_r = 0, \\ \frac{\partial L_\beta}{\partial \delta_r} &= l_r^2 - \lambda_r = 0,\end{aligned}\tag{B.18}$$

for $r = 1, \dots, R$. If $\vec{\lambda}^*$ is a local minimizer, then $\delta_r^* \geq 0$, $r = 1, \dots, R$, and if $\delta_j^* > 0$ for some $j \in \{1, \dots, R\}$, then we must be on a boundary, $\lambda_j^* = 0$. Define the rate of change in harm from killing strain μ as

$$y_{\mu}(\vec{\lambda}) = h_{\mu} \frac{\partial \tilde{g}_{\mu}}{\partial C_{\mu}}(\vec{\lambda}) = -\frac{g_{\mu}^{\max} h_{\mu}}{(1 + C_{\mu})^2}\tag{B.19}$$

where the sign of y_{μ} is determined by h_{μ} , the harm/benefits of strain μ , and

$$\|y_{\mu}\| < \|g_{\mu}^{\max} h_{\mu}\|.\tag{B.20}$$

Note that $\|y_{\mu}\| \propto \frac{1}{C_{\mu}}$. We will focus on the first system of equations which describes the extremal conditions for λ_r . The second set of equations imposes the non-negativity constraint on λ_r , and we will check it at the end.

Recall the cross-responsiveness matrix,

$$\Sigma = [\sigma_{\mu,r}] \in \mathbb{R}^{M \times R}.\tag{B.21}$$

We can now rewrite the first system of equations (Eq B.18),

$$\delta_r - \beta = \sum_{\mu} h_{\mu} \frac{\partial C_{\mu}}{\partial \lambda_r} \frac{\partial \tilde{g}_{\mu}}{\partial C_{\mu}} = \sum_{\mu} \sigma_{\mu,r} y_{\mu} \implies \Sigma^{\top} \mathbf{y} = \boldsymbol{\delta} - \beta \mathbf{1}.\tag{B.22}$$

In eq. B.18, we need to solve a non-linear system of R equations in R variables, $\vec{\lambda}$, which

are constrained to be non-negative. Any solution to that system is a solution to the “linear” system in eq. B.22, with $\mathbf{y} \in \mathbb{R}^M$ treated as a free variable and with the same choice of $\boldsymbol{\delta} \in \mathbb{R}^R$. The converse statement is generally not true since even if Σ^T is invertible, $\mathbf{y}^* = (\Sigma^{-1})^\top(\boldsymbol{\delta} - \beta\mathbf{1})$ may not be in the range of eq. B.19.

Convexity of the cost function. Computing the second derivative for a general cost function $L_\beta(\vec{C}(\vec{\lambda}))$ satisfying $\frac{\partial^2 L_\beta}{\partial C_\mu^2}(\vec{C}(\vec{\lambda})) > 0$ for $\mu \in \{1, \dots, M\}$:

$$\begin{aligned} H_{j,r} &= \frac{\partial^2 L_\beta}{\partial \lambda_r \partial \lambda_j} = \sum_\mu \sigma_{\mu,j} \sigma_{\mu,r} \frac{\partial^2 L_\beta}{\partial C_\mu^2} \\ &= \sum_\mu \sigma_{\mu,j} \left[\frac{\partial^2 L_\beta}{\partial C_\mu^2} \right]^{\frac{1}{2}} \times \sigma_{\mu,r} \left[\frac{\partial^2 L_\beta}{\partial C_\mu^2} \right]^{\frac{1}{2}} \\ &= (A_j)^\top A_r \end{aligned} \tag{B.23}$$

where the r -th column of A is

$$A_r^\top = \left[\left(\frac{\partial^2 L_\beta}{\partial C_\mu^2} \right)^{\frac{1}{2}} \sigma_{\mu,r} \right]_{\mu=1}^M$$

and

$$A = \text{diag} \left(\left[\left(\frac{\partial^2 L_\beta}{\partial C_\mu^2} \right)^{\frac{1}{2}} \right]_{\mu=1}^M \right) \times \Sigma \tag{B.24}$$

$$H = A^\top A.$$

Then H is positive semi-definite, since

$$v^\top H v = v^\top A^\top A v = (vA)^\top A v \geq 0, \tag{B.25}$$

which is a quadratic form. This shows that the function is convex, although it is not strictly convex unless Σ is full rank. For example, when there are more receptors than strains ($R > M$), Σ is not full rank.

Solving the general problem. We want the most protective immune coverage of the microbiota with the fewest number of immune cells (and receptors). The KKT conditions require

$$0 = \beta + \Sigma^\top \mathbf{y}(\vec{\lambda}^*) - \delta^* = \nabla L_\beta(\vec{\lambda}^*),$$

which is just an expression about the derivative of the Lagrangian. If $\vec{\lambda}^*$ is a local minimizer, then $\delta_r^* \geq 0$, $r = 1, \dots, R$, and if $\delta_j^* > 0$ for some $j \in \{1, \dots, R\}$, then we must be on a boundary, $\lambda_j^* = 0$.¹ This tells us immediately that if $\delta_r^* < 0$ at the boundary then $\vec{\lambda}^*$ is not a local minima, since it means it is possible to minimize the objective further by increasing $\lambda_r^* = 0 \rightarrow \lambda_r^* > 0$. Computationally, finding the optimal solution boils down, in part, to selecting which $\delta_r^* = 0$, $r \in \kappa_+ \subseteq \{1, \dots, R\}$ (i.e. which receptors to use) and which not to use $\delta_j^* > 0$, $\lambda_j^* = 0$, $j \in \{1, \dots, R\} \setminus \kappa_+$, and solving the system of equations

$$\begin{aligned} \tilde{\Sigma}^\top \mathbf{y}(\vec{\lambda}^*) &= -\tilde{\beta} \quad \text{for } r \in \kappa_+, \\ \delta_j^* &> 0 \quad \text{for } j \in \{1, \dots, R\} \setminus \kappa_+, \quad (\text{Local minimizer}) \end{aligned} \tag{B.26}$$

for a $\mathbf{y}^*$ that is in the functional range of Equation B.19, and where $\tilde{\Sigma}^\top$ and $\tilde{\beta}$ only include rows (corresponding to receptors) in κ_+ .² Naively, there are 2^R possible receptor configurations. Generally there should be $|\kappa_+| \leq M$ receptors (the number of strains) in the optimal repertoire since we only have M variables to satisfy R constraints. However, additional receptors may be able to co-exist if their reactivities satisfy a strict condition: groups of

¹From eq. B.18, we can infer that $\delta_r^* \leq \beta_r$.

²If the rows of $\tilde{\Sigma}^\top$ are orthogonal, then the solution takes the form $\vec{y}^* = \sum_{r \in \kappa_+} \beta_r \frac{\tilde{\Sigma}_r}{\|\tilde{\Sigma}_r\|_2^2}$, where Σ_r is the r -th column of $\tilde{\Sigma}$. Since the entries of $\tilde{\Sigma}$ are non-negative, this is only possible if each receptor covers a distinct group of strains.

vectors lying on the same hyperplane in $\mathbb{R}^M$, $\sigma_j = \sum_r c_r \sigma_r$, with $\sum_r c_r = 1$ and $\sigma_j \geq 0$. Otherwise satisfying the KKT conditions would require solving an over-determined system, which is impossible.

If a vector $\mathbf{y}^* \leq 0$ can be found that solves the the reduced system of linear equations (eq. B.26), we check that it is a feasible solution for the original non-linear problem. This requires solving the non-linear equation for $\vec{C}^* \geq 0$ (eq. B.19),

$$y_\mu^*(C_\mu) = -\frac{g_\mu^{\max} h_\mu}{(1 + C_\mu(\vec{\lambda}^*))^2},$$

from which the sign of the RHS is fixed. Thus

$$C_\mu^* = \left[\frac{g_\mu^{\max} h_\mu}{|y_\mu^*|} \right]^{\frac{1}{2}} - 1. \quad (\text{Optimal immune coverage}) \quad (\text{B.27})$$

If $\vec{C} \geq 0$, to find $\vec{\lambda}^*$ we need to solve the linear system of equations,

$$C_\mu^* = \sum_{r \in \{1, \dots, R\}} \sigma_{\mu,r} \lambda_r^* \quad (\text{B.28})$$

for $\mu \in \{1, \dots, M\}$, with $\lambda_j = 0$ if $j \in \{1, \dots, R\} \setminus \kappa_+$ (i.e. $\delta_j > 0$). Finally, we must verify that the solution we found is non-negative $\vec{\lambda}^* \geq 0$.

Ecological competition for epitopes can optimize the repertoire

Following [62], we ask if competition for antigens can generate the optimal repertoire. Specifically, we claim the ecological dynamics

$$\frac{d\lambda_r}{dt} = \lambda_r \left(-\sum_\mu \sigma_{\mu,r} h_\mu \frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) - \beta \right), \quad (\text{B.29})$$

achieves this modified gradient descent of the cost function via competitive dynamics, where multiplying by λ_r ensures that its value is never negative and β acts like an attrition rate. The term

$$h_\mu \frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) \rightarrow 0, \text{ as } g_\mu \rightarrow 0 \text{ or } h_\mu \rightarrow 0,$$

corresponds to the availability of strain-derived antigens. With strain-immune response functions as in eq. B.8, we have

$$-\frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) = \frac{g_\mu^{\max}}{(1 + C_\mu)^2}$$

which models a monotonically decreasing epitope supply as immune coverage increases.

Since any point that satisfies the KKT conditions (eq. B.18) is a fixed point for the ecological dynamics, $d\lambda_r/dt = 0$, $r = 1, \dots, R$, it is immediate that the optimal repertoire solution is a fixed point of the dynamics. Next, we show that stable fixed points of the dynamics are exactly the solutions of the KKT equations. For a fixed point to be stable the Jacobian matrix computed at the point must have only negative eigenvalues. The Jacobian is given by

$$J_{r,r'} = \delta_{r,r'} \cdot \left(-\sum_{\mu} \sigma_{\mu,r} h_\mu \frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) - \beta \right) - \lambda_r \sum_{\mu} \sigma_{\mu,r} \sigma_{\mu,r'} h_\mu \frac{2g_\mu^{\max}}{(1 + C_\mu)^3}. \quad (\text{B.30})$$

When $\lambda_r = 0$, row r will have non-zero entries everywhere except the diagonal. The requirement that the eigenvalue of the elementary vector $\vec{e}_r$ be negative means that

$$\sum_{\mu} \sigma_{\mu,r} h_\mu \frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) + \beta > 0,$$

which is equivalent to the KKT conditions at the boundary (eq. B.18). When $\lambda_r > 0$ (and

$$\dot{\lambda}_r = 0),$$

$$\begin{aligned} J_{r,r'} &= \frac{\partial \dot{\lambda}_r}{\lambda_{r'}} = -\lambda_r \sum_{\mu} \sigma_{\mu,r} \sigma_{\mu,r'} h_{\mu} \frac{2g_{\mu}^{\max}}{(1+C_{\mu})^3} \\ &= \sum_{\mu} \sigma_{\mu,r} \left[\frac{2\lambda_r h_{\mu} g_{\mu}^{\max}}{(1+C_{\mu})^3} \right]^{\frac{1}{2}} \times \sigma_{\mu,r'} \left[\frac{2\lambda_r h_{\mu} g_{\mu}^{\max}}{(1+C_{\mu})^3} \right]^{\frac{1}{2}} \\ &= (A_r)^{\top} A'_{r'} \end{aligned} \tag{B.31}$$

where the r -th column of A is

$$A_r^{\top} = \left[\left(\frac{2\lambda_r h_{\mu} g_{\mu}^{\max}}{(1+C_{\mu})^3} \right)^{\frac{1}{2}} \sigma_{\mu,r} \right]_{\mu=1}^M,$$

and

$$A = \text{diag} \left(\left[\left(\frac{2\lambda_r h_{\mu} g_{\mu}^{\max}}{(1+C_{\mu})^3} \right)^{\frac{1}{2}} \right]_{\mu=1}^M \right) \times \Sigma, \quad J = A^{\top} A. \tag{B.32}$$

We have relied on the assumption that $h_{\mu}, g_{\mu}^{\max} > 0$, $\mu = 1, \dots, M$. Thus J is positive semi-definite, since

$$v^{\top} J v = v^{\top} A^{\top} A v = (vA)^{\top} A v \geq 0, \tag{B.33}$$

which is a quadratic form. This shows that the function is convex, although it is not strictly convex unless A , and thus Σ , has linearly independent columns. This is only possible when there are fewer receptors than strains, $R \leq M$.

Finally, we show that the cost function (eq. B.13) is a Lyapunov function for the

dynamics:

$$\frac{dC}{dt} = \sum_r \frac{\partial C}{\partial \lambda_r} \frac{d\lambda_r}{dt} = \sum_r -\lambda_r \left[\sum_\mu \sigma_{\mu,r} h_\mu \frac{\partial \tilde{g}_\mu}{\partial C_\mu}(\vec{\lambda}) + \beta \right]^2 \leq 0. \quad (\text{B.34})$$

One strain, one receptor

The immune coverage from the single receptor is given by $C = \sigma\lambda$, and the Lagrangian (eq. B.11) is,

$$L(\lambda, \delta) = hg(\lambda) + \beta\lambda + \delta \cdot (l^2 - \lambda), \quad (\text{B.35})$$

with

$$\tilde{g}(C) = \frac{g^{\max}}{1 + \sigma\lambda}.$$

From the KKT conditions, the optimal solution will satisfies,

$$0 = \frac{\partial L}{\partial \lambda} = h \frac{\partial g}{\partial \lambda} + \beta - \delta = -\frac{g^{\max} \sigma}{(1 + \sigma\lambda)^2} + \beta - \delta \quad (\text{B.36})$$

$$0 = \frac{\partial L}{\partial \delta} = l^2 - \lambda. \quad (\text{B.37})$$

The second equation imposes the non-negativity constraint as it can only be satisfied when $\lambda \geq 0$. We have two cases to consider.

Case 1: $\lambda = 0$. Here the inequality constraint is active and $l = 0$ (and δ may be nonzero), so we check the value of the Lagrangian at this boundary:

$$L(\lambda = 0) = hg^{\max}. \quad (\text{B.38})$$

Note that if $h \leq 0$, this is the only viable solution.

Case 2: $\lambda > 0$. Here the inequality constraint is inactive (i.e. satisfied) and $l^2 > 0$ and $\delta = 0$. From the KKT conditions we have:

$$\beta + h \frac{\partial g}{\partial \lambda} = 0, \quad (\text{B.39})$$

$$\Rightarrow \beta - \frac{g^{\max} h \sigma}{(1 + \sigma \lambda)^2} = 0. \quad (\text{B.40})$$

From this we see that the magnitude of the second term is always less than $g^{\max} h \sigma$, and since $\beta > 0$, a finite positive solution exists only when $\sigma > \frac{\beta}{g^{\max} h} > 0$. These equations are solved to find

$$\tilde{\lambda} = \sigma^{-\frac{1}{2}} \left[\left(\frac{g^{\max} h}{\beta} \right)^{\frac{1}{2}} - \sigma^{-\frac{1}{2}} \right]. \quad (\text{B.41})$$

The full solution is then,

$$\lambda^* = \max \left\{ \sigma^{-\frac{1}{2}} \left[\left(\frac{g^{\max} h}{\beta} \right)^{\frac{1}{2}} - \sigma^{-\frac{1}{2}} \right], 0 \right\} \quad (\text{B.42})$$

and the explicit σ dependence of the solution can be determined by setting $\lambda = 0$ in eq. B.39:

$$\text{receptor reactivity threshold: } \sigma^* = \frac{\beta}{g^{\max} h}, \quad (\text{B.43})$$

such that $\tilde{\lambda}|_{\sigma=\sigma^*} = 0$, and $\lambda^* = 0$ if $\sigma \leq \sigma^*$. This means that for a receptor to be effective, there is a threshold level of activity, $\sigma \geq \sigma^*$, that must be met.

Two strains, two broadly reactive receptors.

We consider the simplest case of a “cluster” of immune-interacting strains: two strains and two receptors. The cross-responsiveness matrix is assumed to be full-rank³ and is given by

$$\Sigma = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix}, \quad (\text{B.44})$$

$$\Sigma^{-1} = \frac{1}{\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2}} \times \begin{bmatrix} \sigma_{2,2} & -\sigma_{1,2} \\ -\sigma_{2,1} & \sigma_{1,1} \end{bmatrix}. \quad (\text{B.45})$$

To find the optimal solution, we must solve, as in eq. B.20, to equation B.22:

$$\Sigma^\top \vec{y}(\vec{\lambda}) = \vec{\delta} - \beta \mathbf{1}.$$

Case 1: $\lambda_1 = \lambda_2 = 0$. This is always the optimal solution when $\sum_{\mu \in \{1,2\}} g_\mu^{\max} h_\mu \sigma_{\mu,r} < \beta$ for $r \in \{1,2\}$. Using eq. B.14:

$$L_\beta(\vec{\lambda} = 0, \vec{\delta}) = \sum_{\mu \in \{1,2\}} h_\mu g_\mu^{\max} \quad (\text{B.46})$$

with $l_r = 0$, $r \in \{1,2\}$.

Cases 2 and 3: $\lambda_1 > 0, \lambda_2 = 0$, or $\lambda_1 = 0, \lambda_2 > 0$. Without loss of generality, consider $\lambda_1 > 0, \lambda_2 = 0$ (i.e $\delta_1 = 0, \delta_2 \neq 0$), after verifying that the reactivity threshold is satisfied (eq. B.65), we must solve

$$\beta = \sum_{\mu \in \{1,2\}} \sigma_{\mu,1} \frac{h_\mu g_\mu^{\max}}{(1 + \sigma_{\mu,1} \lambda_1)^2},$$

³This excludes certain degenerate solutions that arise when the augmented matrix $[\Sigma^\top \mid \vec{\beta}]$ contains redundant rows.

which reduces to a quartic polynomial equation.

If we assume that $\sigma_{\mu,1} = \sigma > 0$ for all strains μ , the equation simplifies to

$$\lambda_1 = \frac{1}{\sigma} \left(\sqrt{\frac{\sigma}{\beta} \times (h_1 g_1^{\max} + h_2 g_2^{\max})} - 1 \right) \quad (\text{B.47})$$

Analogous results hold for the case $\lambda_1 = 0, \lambda_2 > 0$ (i.e. $\delta_1 \neq 0, \delta_2 = 0$).

Binary binding: Consider a simpler situation in which, $\sigma_{1,r} = \sigma_{2,r} = \sigma_r \geq 0$ for $r \in \{1, 2\}$. That is, strains express the same epitopes. Then, provided $\sigma_1 \neq \sigma_2$, the receptor $j = \arg \max\{\sigma_1, \sigma_2\}$ will out-compete and exclude the other.

Case 4: $\lambda_1, \lambda_2 > 0$. If, without loss of generality, $\sigma_{2,2} = 0$, and $\sigma_{1,1} = \sigma_{2,1}, \sigma_{1,2} > 0$, we obtain the following simplifications to the previous conditions for co-existence, with

$$\bar{y}^* = -(\Sigma^\top)^{-1} \beta \mathbf{1} = -\frac{\beta}{\sigma_1 \sigma_2} \times \begin{bmatrix} \sigma_1 \\ \sigma_2 - \sigma_1 \end{bmatrix}. \quad (\text{B.48})$$

Using eq. B.27, we seek to compute

$$C_\mu^* = \left[\frac{g_\mu^{\max} h_\mu}{|y_\mu^*|} \right]^{\frac{1}{2}} - 1, \quad (\text{B.49})$$

and using eq. B.8,

$$g_\mu^* = \frac{g_\mu^{\max}}{1 + C_\mu^*} = \left[\frac{g_\mu^{\max} |y_\mu^*|}{h_\mu} \right]^{\frac{1}{2}}. \quad (\text{B.50})$$

We need to check that the solution is feasible:

1. Is harm decreased, i.e. $\bar{y}^* < 0$? For $h_\mu > 0$, this requires that $\sigma_1 > 0$ and $\sigma_2 > \sigma_1$.
2. Is the immune coverage non-negative, $C_\mu \geq 0$? For strains $\mu \in \{1, 2\}$ and receptors

$r \in \{1, 2\}$, assuming $\sigma_2 > \sigma_1$, we require

$$g_1^{\max} h_1 \geq |y_1^*| = \frac{\beta}{\sigma_2}, \quad g_2^{\max} h_2 \geq |y_2^*| = \beta \frac{\sigma_2 - \sigma_1}{\sigma_1 \sigma_2} = \beta \left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2} \right).$$

The first inequality can be rewritten as

$$\sigma_2 \geq \frac{\beta}{h_1 g_1^{\max}},$$

which says that for receptor 2 to be in the repertoire it must be a cost-effective strategy for controlling strain 1, i.e. the potential reduction in harm is greater than the cost.

The second inequality can be rearranged to

$$\sigma_2 \leq \frac{1}{\frac{1}{\sigma_1} - \frac{h_2 g_2^{\max}}{\beta}}, \quad \text{if } \sigma_1 < \frac{\beta}{h_2 g_2^{\max}},$$

which says that if receptor 2 is sufficiently reactive it can exclude receptor 1. However, when $\sigma_1 > \frac{\beta}{h_2 g_2^{\max}}$, the condition holds for any value of σ_2 and receptor 1 cannot be excluded because it is guaranteed to cost-effectively decrease harm from strain 2. Finally, when $\sigma_1 = \frac{\beta}{h_2 g_2^{\max}}$ this condition is inconclusive and strain 2 does not matter.

If these conditions are satisfied, then

$$C_1^* = \left[\frac{g_1^{\max} h_1 \sigma_2}{\beta} \right]^{\frac{1}{2}} - 1, \quad C_2^* = \left[\frac{g_2^{\max} h_2 \sigma_1 \sigma_2}{\beta(\sigma_2 - \sigma_1)} \right]^{\frac{1}{2}} - 1. \quad (\text{B.51})$$

3. Attainable: Next, we use eq. B.28 to find

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \Sigma^{-1} \begin{bmatrix} C_1^* \\ C_2^* \end{bmatrix} = \frac{1}{\sigma_1 \sigma_2} \times \begin{bmatrix} \sigma_2 C_2^* \\ \sigma_1 C_1^* - \sigma_1 C_2^* \end{bmatrix}. \quad (\text{B.52})$$

Finally, we must check that $\lambda_1, \lambda_2 > 0$ (per the KKT conditions). For this to be true,

we need to satisfy conditions on the immune coverages:

$$C_2^* > 0, \quad C_1^* > C_2^*. \quad (\text{B.53})$$

which simplifies to two strict inequalities

$$\begin{aligned} \sigma_2 &< \frac{\sigma_1}{1 - \frac{h_2 g_2^{\max}}{\beta} \sigma_1}, \quad \text{if } \sigma_1 < \frac{\beta}{h_2 g_2^{\max}}, \\ \sigma_2 &> \sigma_1 \frac{(h_1 g_1^{\max} + h_2 g_2^{\max})}{h_1 g_1^{\max}}, \end{aligned}$$

where the second condition is required for receptor 2 to be recruited.

Collectively, there are 3 conditions that must be satisfied for coexistence:

$$\begin{aligned} \sigma_2 &\geq \frac{\beta}{h_1 g_1^{\max}}, \\ \sigma_2 &> \sigma_1 \frac{(h_1 g_1^{\max} + h_2 g_2^{\max})}{h_1 g_1^{\max}}, \\ \sigma_2 &< \frac{\sigma_1}{1 - \frac{h_2 g_2^{\max}}{\beta} \sigma_1}, \quad \text{if } \sigma_1 < \frac{\beta}{h_2 g_2^{\max}}, \end{aligned} \quad (\text{B.54})$$

Unconstrained binding: If $\sigma_{\mu,r}$ are allowed to take any non-negative, possibly bounded value, case 4 becomes more complicated.

Case 4: $\lambda_1, \lambda_2 > 0$. Solving equation B.22 for the case of two and two strains receptors,

$$y^* = -(\Sigma^\top)^{-1} \beta \mathbf{1} = \frac{-\beta}{\sigma_{1,1} \sigma_{2,2} - \sigma_{2,1} \sigma_{1,2}} \times \begin{bmatrix} \sigma_{2,2} - \sigma_{2,1} \\ \sigma_{1,1} - \sigma_{1,2} \end{bmatrix}. \quad (\text{B.55})$$

Recall that $y_\mu(C_\mu)$ is the rate at which harm is decreased by mitigating strain μ , C_μ is the immune coverage of strain μ , and $\|y_\mu\| \propto \frac{1}{C_\mu}$. Next we check that the solution is feasible:

1. Is harm decreased? For $h_\mu > 0$, this requires that $\sigma_{2,2} \neq \sigma_{2,1}$ and $\sigma_{1,1} \neq \sigma_{1,2}$ and

$$\text{sgn}(\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2}) = \text{sgn}(\sigma_{2,2} - \sigma_{2,1}),$$

$$\text{sgn}(\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2}) = \text{sgn}(\sigma_{1,1} - \sigma_{1,2}).$$

By fixing the sign of the determinant and checking cases, we find that this requires either

$$\sigma_{2,2} > \sigma_{2,1}, \sigma_{1,1} > \sigma_{1,2} \quad \text{or} \quad \sigma_{2,2} < \sigma_{2,1}, \sigma_{1,1} < \sigma_{1,2},$$

that is, no single receptor is more reactive than the other to both strains.

2. Is the required rate of decrease achievable? According to equation B.20, for strains $\{\alpha, \mu\}$ and receptors $\{r, j\}$, we require

$$g_\alpha^{\max} h_\alpha \geq |y_\alpha^*| = \beta \frac{\sigma_{\mu,j} - \sigma_{\mu,r}}{\sigma_{\mu,j}\sigma_{\alpha,r} - \sigma_{\alpha,j}\sigma_{\mu,r}}, \quad g_\mu^{\max} h_\mu \geq |y_\mu^*| = \beta \frac{\sigma_{\alpha,r} - \sigma_{\alpha,j}}{\sigma_{\mu,j}\sigma_{\alpha,r} - \sigma_{\alpha,j}\sigma_{\mu,r}}$$

Assuming $\sigma_{\mu,j} > \sigma_{\mu,r}$ and $\sigma_{\alpha,r} > \sigma_{\alpha,j}$,

$$\begin{aligned} \beta \frac{\sigma_{\mu,j} - \sigma_{\mu,r}}{\sigma_{\mu,j}\sigma_{\alpha,r} - \sigma_{\alpha,j}\sigma_{\mu,r}} &= \beta \frac{\sigma_{\mu,j} - \sigma_{\mu,r}}{\sigma_{\mu,j}\sigma_{\alpha,r} - \sigma_{\mu,r}\sigma_{\alpha,r} + \sigma_{\mu,r}\sigma_{\alpha,r} - \sigma_{\alpha,j}\sigma_{\mu,r}} \\ &= \frac{\beta}{\sigma_{\alpha,r} + \sigma_{\mu,r} \frac{\sigma_{\alpha,r} - \sigma_{\alpha,j}}{\sigma_{\mu,j} - \sigma_{\mu,r}}}, \end{aligned}$$

which implies

$$\sigma_{\alpha,r} + \sigma_{\mu,r} \frac{\sigma_{\alpha,r} - \sigma_{\alpha,j}}{\sigma_{\mu,j} - \sigma_{\mu,r}} \geq \frac{\beta}{g_\alpha^{\max} h_\alpha}. \quad (\text{B.56})$$

This sets a threshold on the weighted reactivity of the receptor most reactive to strain α , accounting for cross-responsiveness. A similar expression holds for $|y_\mu^*|$. If $\sigma_{\mu,r} = 0$,

the expression simplifies to

$$\sigma_{\alpha,r} \geq \frac{\beta}{g_{\alpha}^{\max} h_{\alpha}}.$$

3. Attainable: If the last two conditions are satisfied, we may use eq. B.27 to solve for C_{μ}^* , and find

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \Sigma^{-1} \begin{bmatrix} C_1^* \\ C_2^* \end{bmatrix} = \frac{1}{\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2}} \times \begin{bmatrix} \sigma_{2,2}C_1^* - \sigma_{1,2}C_2^* \\ \sigma_{1,1}C_2^* - \sigma_{2,1}C_1^* \end{bmatrix}. \quad (\text{B.57})$$

Finally, we must check that $\lambda_1, \lambda_2 > 0$ (per the KKT conditions). For this to be true, we need to satisfy conditions on the immune coverages:

- If $\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2} > 0$:

$$\frac{C_2^*}{C_1^*} < \frac{\sigma_{2,2}}{\sigma_{1,2}}, \quad \frac{C_2^*}{C_1^*} > \frac{\sigma_{2,1}}{\sigma_{1,1}}, \quad (\text{B.58})$$

- If $\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2} < 0$:

$$\frac{C_2^*}{C_1^*} > \frac{\sigma_{2,2}}{\sigma_{1,2}}, \quad \frac{C_2^*}{C_1^*} < \frac{\sigma_{2,1}}{\sigma_{1,1}}. \quad (\text{B.59})$$

In two dimensions, this defines a subset of vectors. Going forward, without loss of generality, assume $\sigma_{1,1}\sigma_{2,2} - \sigma_{2,1}\sigma_{1,2} > 0$, then this subset gets bigger as

$$\frac{\sigma_{1,2}}{\sigma_{2,2}}, \frac{\sigma_{2,1}}{\sigma_{1,1}} \rightarrow 0,$$

that is, as the receptors become less broadly reactive.

Towards understanding this condition, supposes that $\frac{g_{\mu}^{\max} h_{\mu}}{y_{\mu}^*} \gg 1$ (i.e. immune coverage

$C_\mu^* \gg 1$), then from eq. B.27

$$C_\mu^* = \left[\frac{g_\mu^{\max} h_\mu}{|y_\mu^*|} \right]^{\frac{1}{2}} - 1 \approx \left[\frac{g_\mu^{\max} h_\mu}{|y_\mu^*|} \right]^{\frac{1}{2}}, \quad (\text{B.60})$$

and from Equation B.8,

$$g_\mu^* = \frac{g_\mu^{\max}}{1 + C_\mu^*} = \left[\frac{g_\mu^{\max} |y_\mu^*|}{h_\mu} \right]^{\frac{1}{2}}. \quad (\text{B.61})$$

Then

$$\frac{C_2^*}{C_1^*} \sim \left[\frac{g_2^{\max} h_2 (\sigma_{2,2} - \sigma_{2,1})}{g_1^{\max} h_1 (\sigma_{1,1} - \sigma_{1,2})} \right]^{\frac{1}{2}}, \quad (\text{B.62})$$

and we see that the repertoire is sensitive to the fine details of receptor affinity.

In the most general case, the optimal repertoire is determined by comparing the solutions from all four cases to find the global minimizer. Since convexity is guaranteed when Σ is full rank, the optimal solution is case 4 when that solution exists.

Many strains and receptors

If strains have private receptors: In general, each receptor may bind to many different strains; however as a simplification, we assume that each strain α is recognized exclusively by a distinct set of receptors $\kappa_\alpha = \{r_{\alpha_1}, \dots, r_{\alpha_l}\} \subset \{1, \dots, R\} = \bigcup_\alpha \kappa_\alpha$. The population of each strain can thus be controlled independently of all others, and the KKT conditions for its associated receptors will be decoupled from all others, such that

$$\begin{aligned} \frac{\partial L}{\partial \lambda_r} = \sigma_{\alpha,r} h_\alpha \frac{\partial \tilde{g}_\alpha}{\partial C_\alpha}(\vec{\lambda}) + \beta - \delta_r = 0 &\Rightarrow h_\alpha \frac{\partial \tilde{g}_\alpha}{\partial C_\alpha}(\vec{\lambda}) = \frac{\delta_r - \beta}{\sigma_{\alpha,r}}, \\ \frac{\partial L}{\partial \delta_r} = l_r^2 - \lambda_r = 0, \end{aligned} \quad (\text{B.63})$$

for $r \in \kappa_\alpha$. The RHS of the implication from first equation is identical for all receptors. For the population of a receptor λ_r to be greater than zero, we must have $\delta_r = 0$. This cannot be the case for any two receptors, λ_r and λ_j except if $\sigma_{\alpha,r} = \sigma_{\alpha,j}$. To see this, first observe that if two receptors, r and j , coexist they must satisfy the KKT conditions with $\delta_r = \delta_j = 0$, and

$$\frac{\partial \tilde{g}_\alpha(\vec{\lambda})}{\partial C_\alpha} = -\frac{\beta}{\sigma_{\mu,r}} = -\frac{\beta}{\sigma_{\mu,j}}.$$

Without loss of generality, suppose $\frac{\partial L}{\partial \lambda_j} = 0$ with $\delta_j = 0$, and $\sigma_{\mu,r} > \sigma_{\mu,j}$ then

$$h_\alpha \frac{\partial \tilde{g}_\alpha(\vec{\lambda})}{\partial C_\alpha} = -\frac{\beta_j}{\sigma_{\mu,j}}$$

and

$$0 = \frac{\partial L}{\partial \lambda_r} = -\sigma_{\alpha,r} \frac{\beta}{\sigma_{\mu,j}} + \beta - \delta_r.$$

and since the sum of the first two terms on the RHS is negative, $\delta_r < 0$. This means that we can decrease harm by increasing λ_r so this cannot be a local minimum, and the repertoire containing λ_j is suboptimal. Thus, we only need to select the receptor(s) that satisfy the leverage cutoff (eq. B.43) with the smallest cost-to-affinity ratio

$$j_\mu \in \arg \min_{r \in \kappa_\mu} \left\{ \frac{\beta}{\sigma_{\mu,r}} \right\}, \quad (\text{B.64})$$

for each of the M strains. If no receptor meets the leverage cutoff then we set all their abundances to zero.

Broadly reactive receptors exclude specialists: From the the KKT conditions (Eq B.18), we can conclude that if there exists a positive solution $\lambda_r^* > 0$ (and $\delta_r = 0$), i.e.,

inactive inequality constraint, then

$$\begin{aligned}
0 < \beta &= \sum_{\mu} h_{\mu} \sigma_{\mu,r} \left[\frac{g_{\mu}^{\max}}{(1 + C_{\mu}(\vec{\lambda}))^2} \right] \\
&\leq \sum_{\mu} g_{\mu}^{\max} h_{\mu} \sigma_{\mu,r},
\end{aligned} \tag{B.65}$$

where the LHS is the marginal cost of maintaining receptor r and the RHS of the inequality is the maximum rate of decrease in harm that can be effected by receptor r . Thus if this condition is not satisfied for receptor r its optimal value must be $\lambda_r^* = 0$. This means that a receptor should only be present if the total maximum marginal protection it confers is greater than its cost.

Next, consider a receptor j that binds a single strain μ' , $\sigma_{\mu,j} = \delta_{\mu',\mu}$ and another receptor r that binds the same strain and possibly others, $m_r = \{\mu\} \ni \mu'$, $\sigma_{\mu,r} = \sigma_r \mathbf{1}_{\mu \in m_r}$. Suppose receptor r belongs to the optimal repertoire then

$$\beta = \sigma_r \sum_{\mu} \mathbf{1}_{\mu \in m_r} \cdot (-y_{\mu}^*) = \sigma_r \sum_{\mu} \mathbf{1}_{\mu \in m_r} \cdot |y_{\mu}^*| \Rightarrow |y_{\mu'}| = \frac{\beta}{\sigma_r} - \sum_{\mu \neq \mu'} \mathbf{1}_{\mu \in m_r} \cdot |y_{\mu}^*| > 0$$

where we have used the fact that $y_{\mu}^* < 0$. For receptor j to also belong to the optimal repertoire, we must have

$$\sigma_{\mu,j} |y_{\mu'}^*| = \beta \Rightarrow \sigma_{\mu,j} = \frac{\beta}{|y_{\mu'}^*|} = \frac{\beta}{\frac{\beta}{\sigma_r} - \sum_{\mu \neq \mu'} \mathbf{1}_{\mu \in m_r} \cdot |y_{\mu}^*|},$$

from which we see that increased broad reactivity of receptor r makes it so that receptor j needs to bind more strongly to be in the optimal repertoire. In this way receptors that recognize many strains tend to exclude receptors that are more selective binders.

For a receptor r , consider another receptor j that only recognizes strains that receptor

r binds to. Suppose $\lambda_r > 0$ is part of the optimal repertoire. Then from Eq B.22, we have

$$\boldsymbol{\sigma}_r^\top \mathbf{y}^* = -\beta$$

for some $\mathbf{y}^* < 0$. If $\boldsymbol{\sigma}_r \geq \boldsymbol{\sigma}_j$ and $\sigma_{\mu,r} > \sigma_{\mu,j}$ for at least one strain μ ,

$$\begin{aligned} 0 &= \frac{\partial L}{\partial \lambda_j} = \boldsymbol{\sigma}_j^\top \mathbf{y}^* + \beta - \delta_j \\ &= (\boldsymbol{\sigma}_j - \boldsymbol{\sigma}_r)^\top \mathbf{y}^* + \boldsymbol{\sigma}_r^\top \mathbf{y}^* + \beta - \delta_j \\ &= (\boldsymbol{\sigma}_j - \boldsymbol{\sigma}_r)^\top \mathbf{y}^* - \beta + \beta - \delta_j \\ &= (\boldsymbol{\sigma}_j - \boldsymbol{\sigma}_r)^\top \mathbf{y}^* - \delta_j \end{aligned}$$

and since the first term is positive, $\delta_j > 0$. This means that we can decrease the cost function by decreasing λ_j , and so receptor j cannot be in the optimal repertoire if $\lambda_r > 0$ is in the repertoire. In this way, broadly reactive receptors tend to exclude narrowly reactive ones.

Limited ecological interactions

We consider a system with two strains and one-sided competitive interactions where immune coverage is identical for both strains,

$$\dot{g}_1 = g_1 \left(\frac{g_1^{\max}}{1+C} - g_1 \right), \quad \dot{g}_2 = g_2 \left(\frac{g_2^{\max}}{1+C+sg_1} - g_2 \right). \quad (\text{B.66})$$

At steady-state, we find:

$$\tilde{g}_1(C) = \frac{g_1^{\max}}{1+C}, \quad \tilde{g}_2(C) = \frac{g_2^{\max}}{1+C+s\frac{g_1^{\max}}{1+C}}. \quad (\text{B.67})$$

For any combination of immune coverage (C) and ecological interaction (s), using eq. B.2, the composition change gap between the two strains induced by immune coverage is given by:

$$\begin{aligned}\Delta\text{CLR}_1 - \Delta\text{CLR}_2 &= \log_{10}(\tilde{g}_1(C)/\tilde{g}_1(0)) - \log_{10}(\tilde{g}_2(C)/\tilde{g}_2(0)) \\ &= -\log_{10}\left(\frac{1 + sg_1^{\max}}{1 + sg_1^{\max}/(1+C)^2}\right).\end{aligned}\tag{B.68}$$

Furthermore, the effect of immune coverage on the abundance of strain 2 will depend on the strength of its ecological interaction with strain 1,

$$\frac{\partial \tilde{g}_2}{\partial C} = \frac{g_2^{\max} \left(s \frac{g_1^{\max}}{(1+C)^2} - 1 \right)}{(1+C + s \frac{g_1^{\max}}{1+C})^2},$$

which for a sufficiently strong ecological interaction, $\sqrt{sg_1^{\max}} - 1 > C$, depends non-monotonically on immune coverage C . In the regime of a strong interaction, increasing immune coverage may actually increase the abundance of strain 2.

B.3 *Estimating cluster harmfulness*

Generalizing eq. B.47 to an arbitrary number of strains within a cluster co-reactive to the same T cell, we obtain

$$C^* \approx \max \left\{ \left[\frac{\sigma \sum_{\mu} h_{\mu} g_{\mu}^{\max}}{\beta} \right]^{\frac{1}{2}} - 1, 0 \right\}.\tag{B.69}$$

Recall that K_C^{-1} is the (unmeasured) susceptibility of microbial abundance to repertoire coverage eq. B.8, and that our expression for C^* is units of this quantity. Then, assuming the observed cluster coverage, $\langle C_{\mu} \rangle_k = K_C C^*$, we can relate it to the abundance-weighted

harm of strains in cluster k :

$$\sum_{\mu \in \text{cluster } k} h_{\mu} g_{\mu}^{\max} / \sum_{\mu \in \text{cluster } k} g_{\mu}^{\max} = \langle h \rangle_k \approx \beta \frac{(K_C^{-1} \langle C_{\mu} \rangle_k)^2 + 1}{\sigma_k \sum_{\mu \in \text{cluster } k} g_{\mu}^{\max}}, \quad (\text{B.70})$$

where β is the (unmeasured) marginal cost of immunity, σ_k is the constant reactivity of strains within the cluster to a common single receptor, and $g_{\mu}^{\max}$ is the maximum niche size of strain μ .

From eqs. B.8 and B.1, the compositional change (ΔCLR_k) from pre- to post-immune coverage follows

$$\Delta \text{CLR}_k = -\log_{10}(1 + K_C^{-1} \langle C_{\mu} \rangle_k) + \sum_k \log_{10}(1 + K_C^{-1} \langle C_{\mu} \rangle_k), \quad (\text{B.71})$$

where the second term can be treated as a constant term in a regression.

Numerical simulations

To obtain optimal repertoire allocations numerically, for a fixed reactivity matrix (Σ), eq. B.9 was solved using the *scikit-learn* package in Python [279]. Reactivity matrices can be specified entry-wise or drawn randomly. In the latter case, ensembles of $M \times R$ reactivity matrices, $\{\Sigma\}_b$, were generated, with each matrix constructed by drawing R columns corresponding to a receptor and its strain reactivities, $\vec{\sigma}_r \sim \text{Unif}[0, 1] \times \text{Binomial}(M, b)$. Solving eq. B.9, properties of the optimal repertoire can be evaluated for each matrix and averaged over the ensemble.

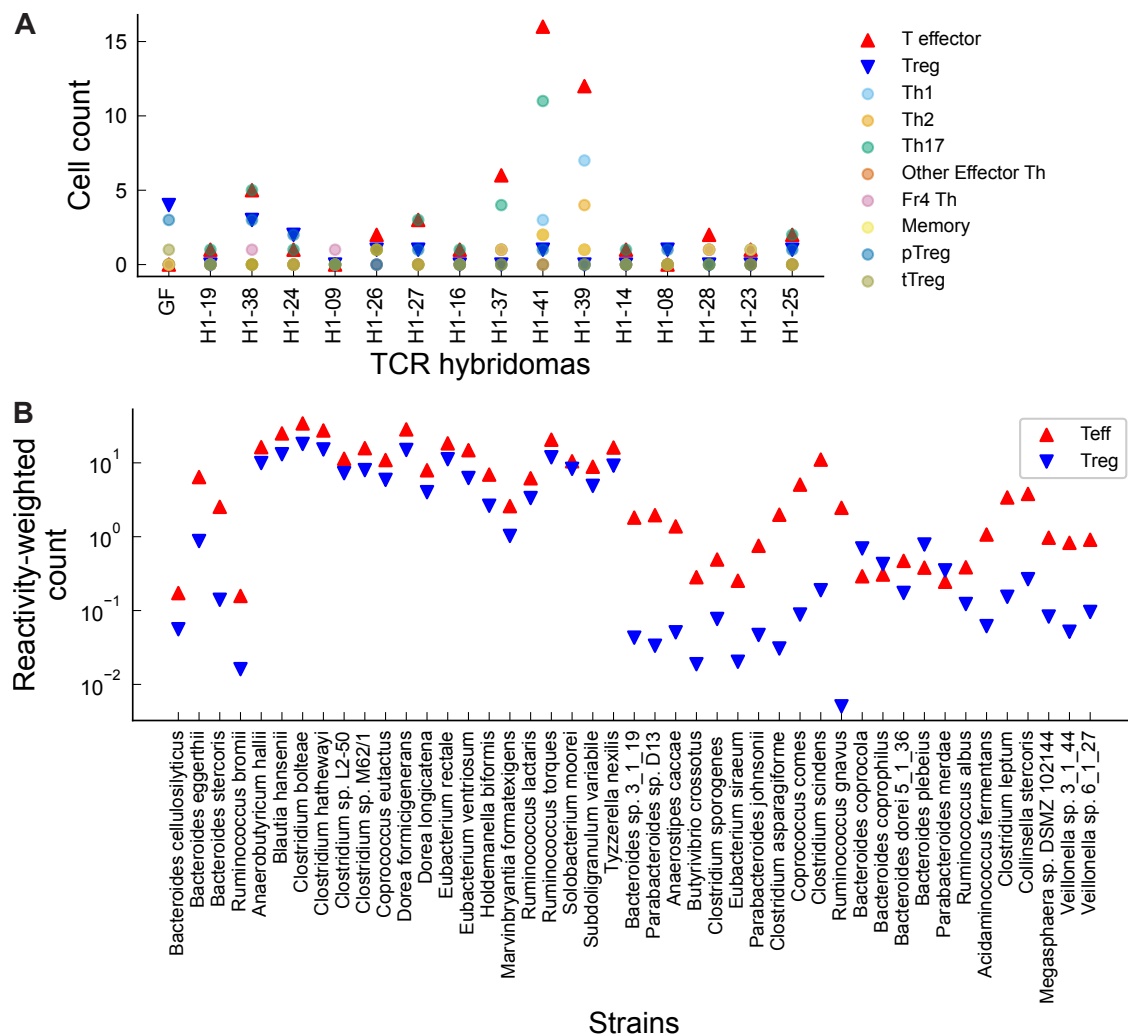


Figure B.1: T cell repertoire composition. (A) Scatter plot showing T cell subtype counts sampled from the colonic lamina propria and mesenteric lymph nodes of inoculated mice. Colors indicate T cell subtypes. (B) Effector (red) and regulatory (blue) T cell reactivity-weighted counts for strains in clusters 1-8.

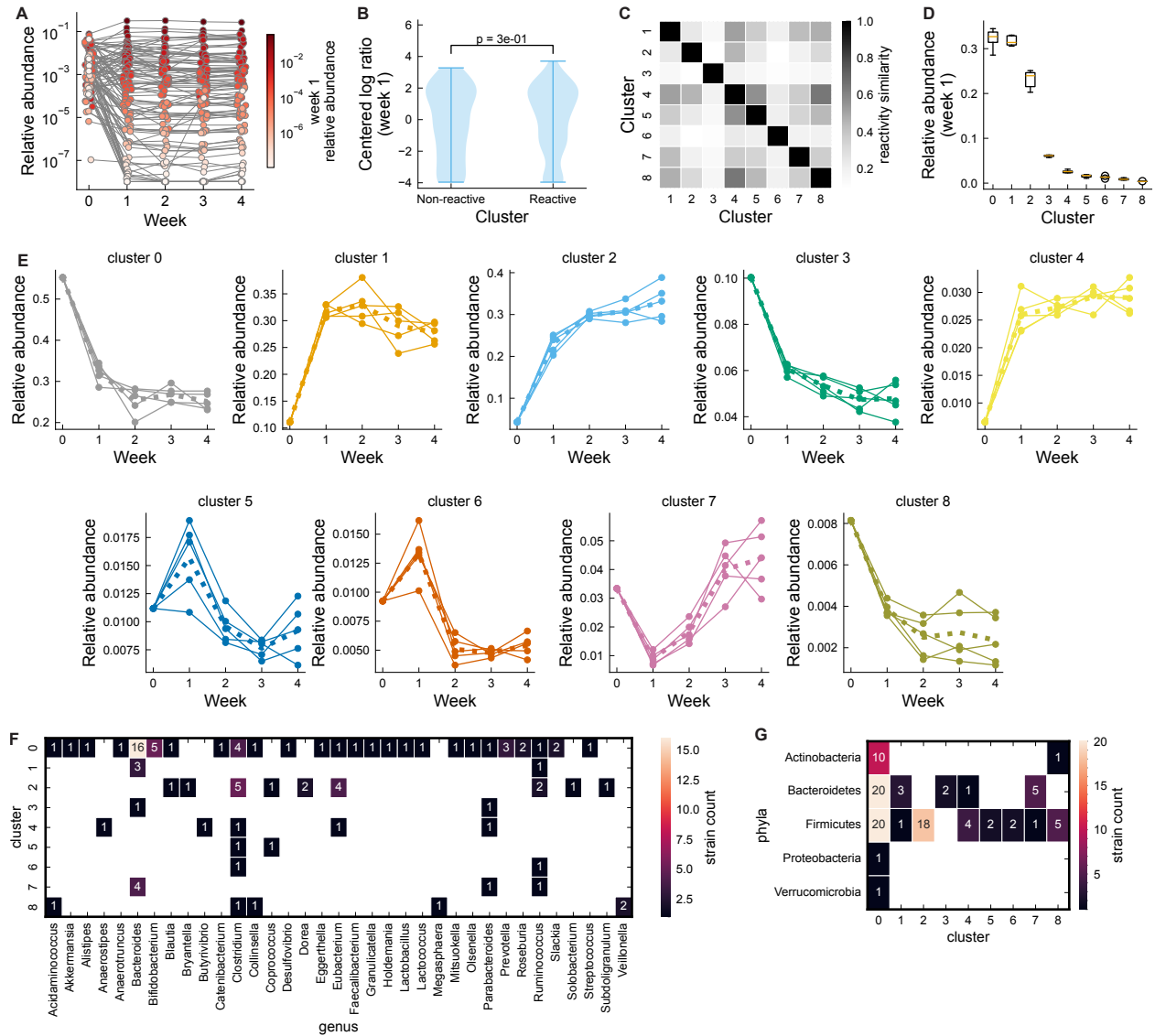


Figure B.2: Composition of microbial clusters. (A) Longitudinal relative abundance of a pre-defined community of 97 microbial strains transplanted into germ free mice [179, 206]. Trajectories are averaged over 5 biological replicates. Five of the 97 strains were not detected at week 1 and were dropped from further analysis. Heat map color indicates strain relative abundance measured at week 1. (B) Distribution of week 1 abundances (centered-log ratio) of non-reactive and reactive strains are statistically indistinguishable, as confirmed by a Mann-Whitney test. (C) Heatmap showing Bray-Curtis similarity score between clusters, measured by the overlap of normalized T cell reactivities (Fig. 4.2) each microbial cluster. (D) Week 1 relative abundance of microbial clusters. Box plots show the mean (orange line) and standard deviation (error bars) across 5 biological replicates. (E) Relative abundance of clusters over time (weeks post-inoculation). Thin lines are biological replicates and the thick dotted line is the mean across replicates. (F) and (G) show strain genera and phylum counts of clusters respectively.

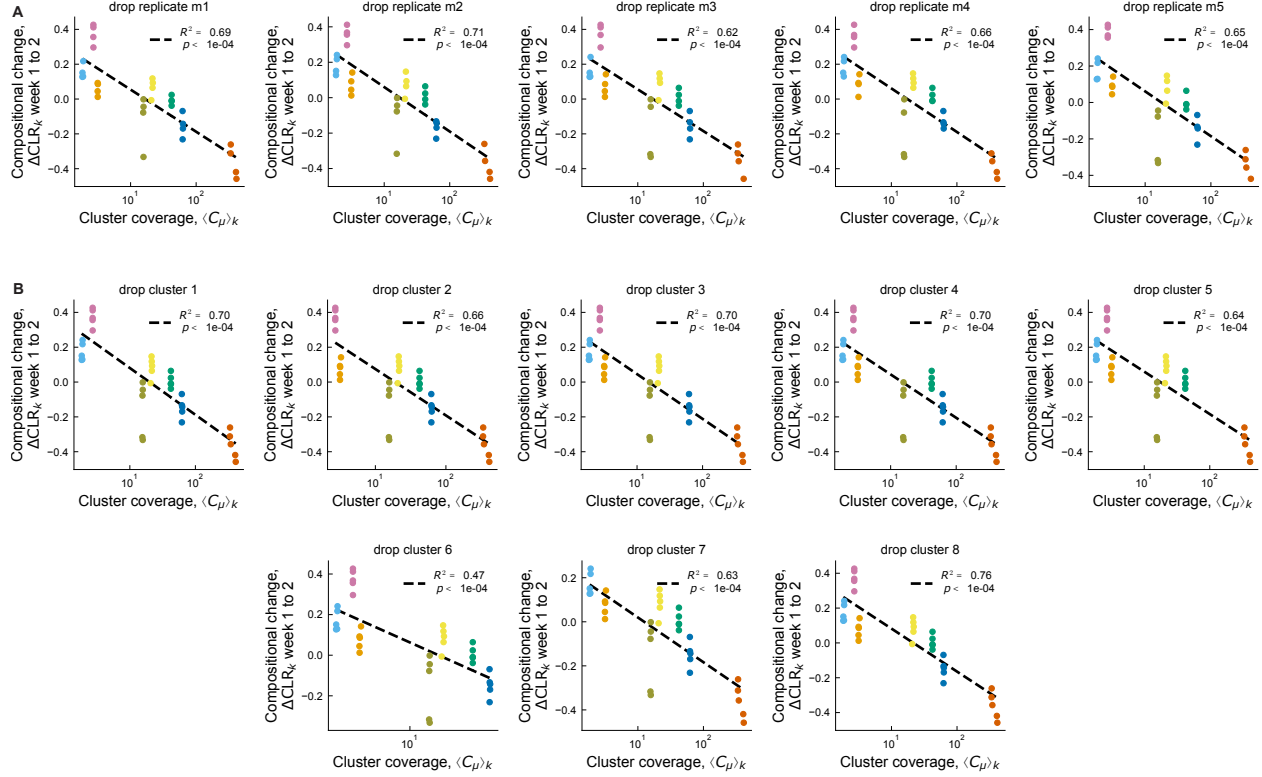


Figure B.3: Robust relationship between microbiota compositional change and repertoire coverage. (A) Scatter plot showing relationship between abundance changes (ΔCLR) and repertoire coverage ($\langle C_\mu \rangle_k$) of microbial clusters. Colors represent clusters and points are biological replicates of the same experiment. In each panel, a single indicated biological replicate was dropped to assess the robustness of the trend. Dotted line represents fit to $\Delta\text{CLR}_k = m \log_{10}(\langle C_\mu \rangle_k) + b$. Statistical significance was quantified through a hierarchical permutation test (i.e., shuffling cluster and replicate labels). (B) Same as (A) but with a single indicated cluster dropped to assess robustness of the trend.

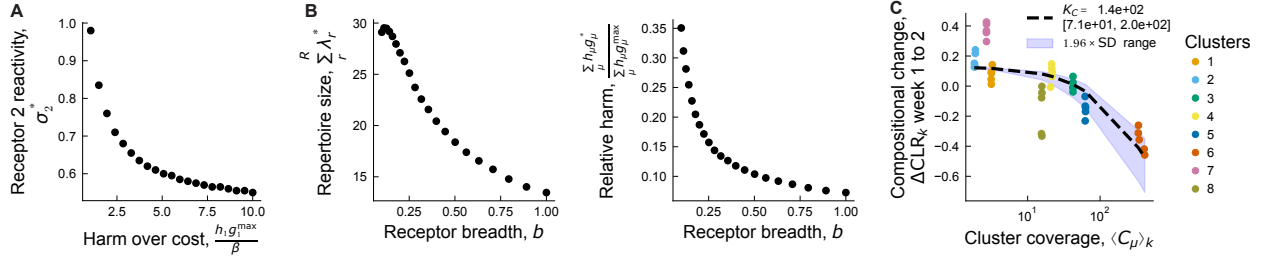


Figure B.4: Properties of the optimal repertoire and inferring K_C . (A) Reactivity of receptor 2 required for repertoire co-existence as a function of the harm-cost ratio ($h_1 g_1^{\max} \beta^{-1}$) for responding to strain 1. As the potential harm from strain 1 increases, threshold for including receptor 2 in the repertoire decreases ($\sigma_1 = 0.5$). (B) Properties of the optimal repertoire as a function of average receptor breadth (b). Left: total number of receptors; right: harm incurred. (C) Scatter plot showing relationship between cluster repertoire coverage ($\langle C_\mu \rangle_k$) and the change in centered log-ratio (ΔCLR_k) from week 1 to week 2. Points of the same color are biological replicates. Dotted line represents fit to $\Delta \text{CLR}_k = -\log_{10}(1 + K_C^{-1} \langle C_\mu \rangle_k) + d$. The blue envelope represents 1.96 times the standard deviation of the inferred constant K_C .

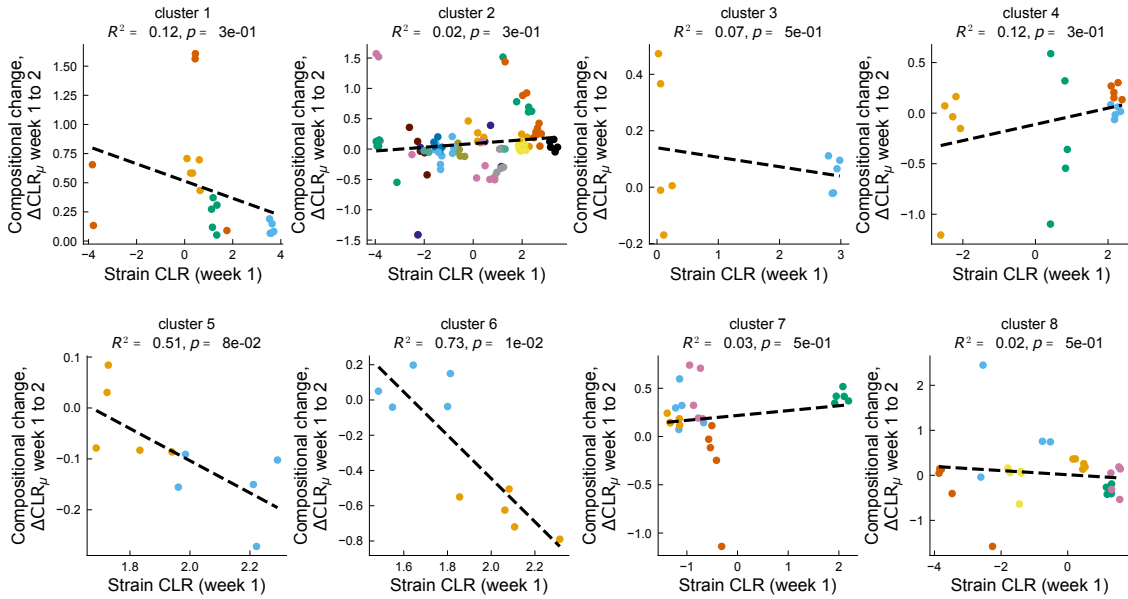


Figure B.5: Intra-cluster trends in strain compositional changes. Relationship between compositional changes (ΔCLR) and week 1 composition of strains within clusters. Dotted line represents a fit to $\Delta \text{CLR}_\mu(1, 2) = -m \cdot \text{CLR}_\mu(1) + b$. Statistical significance (p -values) were corrected for multiple testing (Benjamini-Hochberg).

BIBLIOGRAPHY

- [1] Grégoire Altan-Bonnet, Thierry Mora, and Aleksandra M Walczak. Quantitative immunology for physicists. *Phys. Rep.*, 849:1–83, March 2020.
- [2] Thomas Boehm. Design principles of adaptive immune systems. *Nat. Rev. Immunol.*, 11(5):307–317, May 2011.
- [3] K. Murphy and C. Weaver. *Janeway’s Immunobiology*. W.W. Norton, 2016. ISBN 9781315533247. URL <https://books.google.com/books?id=GmPLCwAAQBAJ>.
- [4] Martin F Flajnik and Masanori Kasahara. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat. Rev. Genet.*, 11(1):47–59, January 2010.
- [5] Andreas Tiffeau-Mayer, Jonathan A Levine, Christopher J Russo, Quentin Marcou, William Bialek, and Benjamin D Greenbaum. How different are self and nonself? *PRX Life*, 4(1):013027, March 2026.
- [6] Mikhail V Pogorelyy, Anastasia A Minervina, Maximilian Puelma Touzel, Anastasiia L Sycheva, Ekaterina A Komech, Elena I Kovalenko, Galina G Karganova, Evgeniy S Egorov, Alexander Yu Komkov, Dmitriy M Chudakov, Ilgar Z Mamedov, Thierry Mora, Aleksandra M Walczak, and Yuri B Lebedev. Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. *Proc. Natl. Acad. Sci. U. S. A.*, 115(50):12704–12709, December 2018.
- [7] Nicole M Gentile, Joseph A Murray, and Darrell S Pardi. Autoimmune enteropathy: a review and update of clinical management. *Curr. Gastroenterol. Rep.*, 14(5):380–385, October 2012.
- [8] Paola Galozzi, Sara Bindoli, Andrea Doria, and Paolo Sfriso. Progress in biological therapies for adult-onset still’s disease. *Biologics*, 16:21–34, April 2022.
- [9] Xian-Pei Xiao, Ming-Yu Wu, Yan Li, Xiao Hu, and Bi-Yuan Qin. Global landscape of autoimmune diseases across different lifespan: A three-decade perspective. *Medicine (Baltimore)*, 105(2):e47140, January 2026.
- [10] Oskar H Schnaack, Luca Peliti, and Armita Nourmohammad. Learning and organization of memory for evolving patterns. *Phys. Rev. X*, 12(2):021063, June 2022.

- [11] Ruslan Medzhitov and Akiko Iwasaki. Exploring new perspectives in immunology. *Cell*, 187(9):2079–2094, April 2024.
- [12] Joy Bergelson, Martin Kreitman, Dmitri A Petrov, Alvaro Sanchez, and Mikhail Tikhonov. Functional biology in its natural context: A search for emergent simplicity. *Elife*, 10, June 2021.
- [13] Rob Phillips. Theory in biology: Figure 1 or figure 7? *Trends Cell Biol.*, 25(12): 723–729, December 2015.
- [14] S M Kaech and R Ahmed. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naïve cells. *Nat. Immunol.*, 2(5):415–422, May 2001.
- [15] Veit R Buchholz, Michael Flossdorf, Inge Hensel, Lorenz Kretschmer, Bianca Weissbrich, Patricia Gräf, Admar Verschoor, Matthias Schiemann, Thomas Höfer, and Dirk H Busch. Disparate individual fates compose robust CD8+ T cell immunity. *Science*, 340(6132):630–635, May 2013.
- [16] Thierry Mora and Aleksandra M Walczak. How many different clonotypes do immune repertoires contain? *Curr. Opin. Syst. Biol.*, 18:104–110, December 2019.
- [17] Adrian Straub, Simon Grassmann, Sebastian Jarosch, Lena Richter, Philipp Hilgen-dorf, Monika Hammel, Karolin I Wagner, Veit R Buchholz, Kilian Schober, and Dirk H Busch. Recruitment of epitope-specific T cell clones with a low-avidity threshold sup-ports efficacy against mutational escape upon re-infection. *Immunity*, May 2023.
- [18] Motonari Kondo, Amy J Wagers, Markus G Manz, Susan S Prohaska, David C Scherer, Georg F Beilhack, Judith A Shizuru, and Irving L Weissman. Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annu. Rev. Immunol.*, 21(1):759–806, April 2003.
- [19] M M Davis and P J Bjorkman. T-cell antigen receptor genes and T-cell recognition. *Nature*, 334(6181):395–402, August 1988.
- [20] Anand Murugan, Thierry Mora, Aleksandra M Walczak, and Curtis G Callan, Jr. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc. Natl. Acad. Sci. U. S. A.*, 109(40):16161–16166, October 2012.
- [21] Yuval Elhanati, Anand Murugan, Curtis G Callan, Jr, Thierry Mora, and Aleksan-dra M Walczak. Quantifying selection in immune receptor repertoires. *Proc. Natl. Acad. Sci. U. S. A.*, 111(27):9875–9880, July 2014.
- [22] K Maude Ashby and Kristin A Hogquist. A guide to thymic selection of T cells. *Nat. Rev. Immunol.*, 24(2):103–117, February 2024.

- [23] Ichiro Taniuchi. CD4 helper and CD8 cytotoxic T cell differentiation. *Annu. Rev. Immunol.*, 36:579–601, April 2018.
- [24] Minyi Lee, Judith N Mandl, Ronald N Germain, and Andrew J Yates. The race for the prize: T-cell trafficking strategies for optimal surveillance. *Blood*, 120(7):1432–1438, August 2012.
- [25] Judith N Mandl, Rachel Liou, Frederick Klauschen, Nienke Vrisekoop, João P Monteiro, Andrew J Yates, Alex Y Huang, and Ronald N Germain. Quantification of lymph node transit times reveals differences in antigen surveillance strategies of naïve CD4⁺ and CD8⁺ T cells. *Proc. Natl. Acad. Sci. U. S. A.*, 109(44):18036–18041, October 2012.
- [26] W Weninger, M A Crowley, N Manjunath, and U H von Andrian. Migratory properties of naive, effector, and memory CD8(+) T cells. *J. Exp. Med.*, 194(7):953–966, October 2001.
- [27] Sarah E Henrickson, Thorsten R Mempel, Irina B Mazo, Bai Liu, Maxim N Artyomov, Huan Zheng, Antonio Peixoto, Michael P Flynn, Balimkiz Senman, Tobias Junt, Hing C Wong, Arup K Chakraborty, and Ulrich H von Andrian. T cell sensing of antigen dose governs interactive behavior with dendritic cells and sets a threshold for T cell activation. *Nat. Immunol.*, 9(3):282–291, March 2008.
- [28] Misty R Jenkins, Justine Mintern, Nicole L La Gruta, Katherine Kedzierska, Peter C Doherty, and Stephen J Turner. Cell cycle-related acquisition of cytotoxic mediators defines the progressive differentiation to effector status for virus-specific CD8⁺ T cells. *J. Immunol.*, 181(6):3818–3822, September 2008.
- [29] Aleksandra J Ozga, Federica Moalli, Jun Abe, Jim Swoger, James Sharpe, Dietmar Zehn, Mario Kreutzfeldt, Doron Merkler, Jorge Ripoll, and Jens V Stein. pMHC affinity controls duration of CD8⁺ T cell-DC interactions and imprints timing of effector differentiation versus expansion. *J. Exp. Med.*, 213(12):2811–2829, November 2016.
- [30] Philippe Bousso and Ellen Robey. Dynamics of CD8⁺ T cell priming by dendritic cells in intact lymph nodes. *Nat. Immunol.*, 4(6):579–585, June 2003.
- [31] Byron B Au-Yeung, Julie Zikherman, James L Mueller, Judith F Ashouri, Mehrdad Matloubian, Debra A Cheng, Yiling Chen, Kevan M Shokat, and Arthur Weiss. A sharp t-cell antigen receptor signaling threshold for t-cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.*, 111(35):E3679–88, September 2014.
- [32] Melissa Lever, Philip K Maini, P Anton van der Merwe, and Omer Dushek. Phenotypic models of T cell activation. *Nat. Rev. Immunol.*, 14(9):619–629, September 2014.
- [33] Raman S Ganti, Wan-Lin Lo, Darren B McAfee, Jay T Groves, Arthur Weiss, and Arup K Chakraborty. How the T cell signaling network processes information to

- discriminate between self and agonist ligands. *Proc. Natl. Acad. Sci. U. S. A.*, 117(42):26020–26030, October 2020.
- [34] Susanne Heinzel, Tran Binh Giang, Andrey Kan, Julia M Marchingo, Bryan K Lye, Lynn M Corcoran, and Philip D Hodgkin. A myc-dependent division timer complements a cell-death timer to regulate T cell and B cell responses. *Nat. Immunol.*, 18(1):96–103, January 2017.
 - [35] J M Marchingo, G Prevedello, A Kan, S Heinzel, P D Hodgkin, and K R Duffy. T-cell stimuli independently sum to regulate an inherited clonal division fate. *Nat. Commun.*, 7:13540, November 2016.
 - [36] Susanne Heinzel, Julia M Marchingo, Miles B Horton, and Philip D Hodgkin. The regulation of lymphocyte activation and proliferation. *Curr. Opin. Immunol.*, 51:32–38, April 2018.
 - [37] E John Wherry, Volker Teichgräber, Todd C Becker, David Masopust, Susan M Kaech, Rustom Antia, Ulrich H von Andrian, and Rafi Ahmed. Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat. Immunol.*, 4(3):225–234, March 2003.
 - [38] Matthew D Martin and Vladimir P Badovinac. Defining memory CD8 T cell. *Front. Immunol.*, 9:2692, November 2018.
 - [39] Shannon M Kahan, Rakesh K Bakshi, Jennifer T Ingram, R Curtis Hendrickson, Elliot J Lefkowitz, David K Crossman, Laurie E Harrington, Casey T Weaver, and Allan J Zajac. Intrinsic IL-2 production by effector CD8 T cells affects IL-2 signaling and promotes fate decisions, stemness, and protection. *Sci Immunol*, 7(68):eabl6322, February 2022.
 - [40] Vandana Kalia and Surojit Sarkar. Regulation of effector and memory CD8 T cell differentiation by IL-2-A balancing act. *Front. Immunol.*, 9:2987, December 2018.
 - [41] Sung-Woo Lee, Gil-Woo Lee, Hee-Ok Kim, and Jae-Ho Cho. Shaping heterogeneity of naive CD8+ T cell pools. *Immune Netw.*, 23(1):e2, February 2023.
 - [42] M J van Stipdonk, E E Lemmens, and S P Schoenberger. Naïve CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nat. Immunol.*, 2(5):423–429, May 2001.
 - [43] Zhengguo Xiao, Matthew F Mescher, and Stephen C Jameson. Detuning CD8 T cells: down-regulation of CD8 expression, tetramer binding, and response during CTL activation. *J. Exp. Med.*, 204(11):2667–2677, October 2007.
 - [44] Lorenz Kretschmer, Michael Flossdorf, Jonas Mir, Yi-Li Cho, Marten Plambeck, Irina Treise, Albulena Toska, Susanne Heinzel, Matthias Schiemann, Dirk H Busch, and

- Veit R Buchholz. Differential expansion of T central memory precursor and effector subsets is regulated by division speed. *Nat. Commun.*, 11(1):113, January 2020.
- [45] Susan M Kaech and Weiguo Cui. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat. Rev. Immunol.*, 12(11):749–761, November 2012.
- [46] Mikel Ruterbusch, Kurt B Pruner, Laila Shehata, and Marion Pepper. In vivo CD4+ T cell differentiation and function: Revisiting the Th1/Th2 paradigm. *Annu. Rev. Immunol.*, 38(1):705–725, April 2020.
- [47] Shimon Sakaguchi, Norihisa Mikami, James B Wing, Atsushi Tanaka, Kenji Ichiyama, and Naganari Ohkura. Regulatory T cells and human disease. *Annu. Rev. Immunol.*, 38(1):541–566, April 2020.
- [48] Christophe Benoist and Diane Mathis. Treg cells, life history, and diversity. *Cold Spring Harb. Perspect. Biol.*, 4(9):a007021, September 2012.
- [49] Kathleen Abadie, Elisa C Clark, Rajesh M Valanparambil, Obinna Ukogu, Wei Yang, Riza M Daza, Kenneth K H Ng, Jumana Fathima, Allan L Wang, Judong Lee, Tahseen H Nasti, Avinash Bhandoola, Armita Nourmohammad, Rafi Ahmed, Jay Shendure, Junyue Cao, and Hao Yuan Kueh. Reversible, tunable epigenetic silencing of TCF1 generates flexibility in the T cell memory decision. *Immunity*, 57(2):271–286.e13, February 2024.
- [50] Ben Youngblood, J Scott Hale, Haydn T Kissick, Eunseon Ahn, Xiaojin Xu, Andreas Wieland, Koichi Araki, Erin E West, Hazem E Ghoneim, Yiping Fan, Pranay Dogra, Carl W Davis, Bogumila T Konieczny, Rustom Antia, Xiaodong Cheng, and Rafi Ahmed. Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature*, 552(7685):404–409, December 2017.
- [51] Marten Plambeck, Atefeh Kazeroonian, Dirk Loeffler, Lorenz Kretschmer, Ciro Salinno, Timm Schroeder, Dirk H Busch, Michael Flossdorf, and Veit R Buchholz. Heritable changes in division speed accompany the diversification of single T cell fate. *Proceedings of the National Academy of Sciences*, 119(9):e2116260119, 2022.
- [52] Derek J Bangs, Brian Hondowicz, Kathleen Abadie, Remi Savard, Madeleine Weiss, Triantafyllia Karakousi, Peter J Skene, ImmGenT Consortium, Amanda W Lund, Ananda W Goldrath, Maximilian Heeg, and Marion Pepper. Tertiary lymphoid structures support the development of allergen-specific progenitor CD4+ T cells. *bioRxiv*, page 2025.10.03.680350, October 2025.
- [53] Julia M Marchingo, Andrey Kan, Robyn M Sutherland, Ken R Duffy, Cameron J Wellard, Gabrielle T Belz, Andrew M Lew, Mark R Dowling, Susanne Heinzl, and Philip D Hodgkin. T cell signaling. antigen affinity, costimulation, and cytokine inputs sum linearly to amplify T cell expansion. *Science*, 346(6213):1123–1127, November 2014.

- [54] Julie Zikherman and Byron Au-Yeung. The role of T cell receptor signaling thresholds in guiding T cell fate decisions. *Curr. Opin. Immunol.*, 33:43–48, April 2015.
- [55] Inbal Eizenberg-Magar, Jacob Rimer, Irina Zaretsky, David Lara-Astiaso, Shlomit Reich-Zeliger, and Nir Friedman. Diverse continuum of CD4+ t-cell states is determined by hierarchical additive integration of cytokine signals. *Proc. Natl. Acad. Sci. U. S. A.*, 114(31):E6447–E6456, August 2017.
- [56] F M Burnet. A modification of jerne’s theory of antibody production using the concept of clonal selection. *CA Cancer J. Clin.*, 26(2):119–121, March 1976.
- [57] Alan S Perelson and Gérard Weisbuch. Immunology for physicists. *Rev. Mod. Phys.*, 69(4):1219–1268, October 1997.
- [58] Arup K Chakraborty. A perspective on the role of computational models in immunology. *Annu. Rev. Immunol.*, 35(1):403–439, April 2017.
- [59] M.W. Hirsch, S. Smale, and R.L. Devaney. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Differential Equations, Dynamical Systems, and an Introduction to Chaos. Elsevier Science, 2013. ISBN 9780123820105. URL https://books.google.com/books?id=csYhsrOEh_MC.
- [60] R J De Boer, M Oprea, R Antia, K Murali-Krishna, R Ahmed, and A S Perelson. Recruitment times, proliferation, and apoptosis rates during the CD8(+) t-cell response to lymphocytic choriomeningitis virus. *J. Virol.*, 75(22):10663–10669, November 2001.
- [61] Yuval Hart, Shlomit Reich-Zeliger, Yaron E Antebi, Irina Zaretsky, Avraham E Mayo, Uri Alon, and Nir Friedman. Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels. *Cell*, 158(5):1022–1032, August 2014.
- [62] Andreas Mayer, Vijay Balasubramanian, Thierry Mora, and Aleksandra M Walczak. How a well-adapted immune system is organized. *Proc. Natl. Acad. Sci. U. S. A.*, 112(19):5950–5955, May 2015.
- [63] Andreas Mayer, Yaojun Zhang, Alan S Perelson, and Ned S Wingreen. Regulation of T cell expansion by antigen presentation dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, 116(13):5914–5919, March 2019.
- [64] Hugo Antonius van den Berg and Yuri Nikolaevich Kiselëv. Expansion and contraction of the cytotoxic T lymphocyte response—an optimal control approach. *Bull. Math. Biol.*, 66(5):1345–1369, September 2004.
- [65] Wenping Cui, Robert Marsland, and Pankaj Mehta. Diverse communities behave like typical random ecosystems. *Phys Rev E*, 104(3-1):034416, September 2021.

- [66] Robert Marsland, 3rd, Owen Howell, Andreas Mayer, and Pankaj Mehta. Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response. *Proc. Natl. Acad. Sci. U. S. A.*, 118(1), January 2021.
- [67] Yael Lebel, Tomer Milo, Alon Bar, Avi Mayo, and Uri Alon. Excitable dynamics of flares and relapses in autoimmune diseases. *iScience*, 26(11):108084, November 2023.
- [68] Uri Alon. *An introduction to systems biology: Design principles of biological circuits*. Chapman and Hall/CRC, Second edition. | Boca Raton, Fla. : CRC Press, [2019], July 2019.
- [69] Simon Grassmann, Lorenz Mihatsch, Jonas Mir, Atefeh Kazeroonian, Roza Rahimi, Sophie Flommersfeld, Kilian Schober, Inge Hensel, Justin Leube, Ludwig O Pachmayr, Lorenz Kretschmer, Qin Zhang, Adrien Jolly, M Zeeshan Chaudhry, Matthias Schiemann, Luka Cicin-Sain, Thomas Höfer, Dirk H Busch, Michael Flossdorf, and Veit R Buchholz. Early emergence of T central memory precursors programs clonal dominance during chronic viral infection. *Nat. Immunol.*, 21(12):1563–1573, December 2020.
- [70] Alon Oyler-Yaniv, Jennifer Oyler-Yaniv, Benjamin M Whitlock, Zhiduo Liu, Ronald N Germain, Morgan Huse, Grégoire Altan-Bonnet, and Oleg Krichevsky. A tunable Diffusion-Consumption mechanism of cytokine propagation enables plasticity in Cell-to-Cell communication in the immune system. *Immunity*, 46(4):609–620, April 2017.
- [71] Jennifer Oyler-Yaniv, Alon Oyler-Yaniv, Mojdeh Shakiba, Nina K Min, Ying-Han Chen, Sheue-Yann Cheng, Oleg Krichevsky, Nihal Altan-Bonnet, and Grégoire Altan-Bonnet. Catch and release of cytokines mediated by tumor phosphatidylserine converts transient exposure into Long-Lived inflammation. *Mol. Cell*, 66(5):635–647.e7, June 2017.
- [72] Grégoire Altan-Bonnet and Ronald N Germain. Modeling T cell antigen discrimination based on feedback control of digital ERK responses. *PLoS Biol.*, 3(11):e356, November 2005.
- [73] Grégoire Altan-Bonnet and Ratnadeep Mukherjee. Cytokine-mediated communication: a quantitative appraisal of immune complexity. *Nat. Rev. Immunol.*, 19(4):205–217, April 2019.
- [74] Sooraj R Achar, François X P Bourassa, Thomas J Rademaker, Angela Lee, Taisuke Kondo, Emanuel Salazar-Cavazos, John S Davies, Naomi Taylor, Paul François, and Grégoire Altan-Bonnet. Universal antigen encoding of T cell activation from high-dimensional cytokine dynamics. *Science*, 376(6595):880–884, May 2022.
- [75] Aleksandra M Walczak, Gasper Tkacik, and William Bialek. Optimizing information flow in small genetic networks. II. feed-forward interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 81(4 Pt 1):041905, April 2010.

- [76] Gasper Tkacik, Curtis G Callan, Jr, and William Bialek. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. U. S. A.*, 105(34):12265–12270, August 2008.
- [77] Srividya Iyer-Biswas and Anton Zilman. First passage processes in cellular biology. March 2015.
- [78] R. Durrett. *Probability: Theory and Examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. ISBN 9781108473682. URL <https://books.google.com/books?id=b22MDwAAQBAJ>.
- [79] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford Science Publications. Oxford University Press, 2020. ISBN 9780198847601. URL <https://books.google.com/books?id=l-vqDwAAQBAJ>.
- [80] Andreas Mayer, Thierry Mora, Olivier Rivoire, and Aleksandra M Walczak. Diversity of immune strategies explained by adaptation to pathogen statistics. *Proc. Natl. Acad. Sci. U. S. A.*, 113(31):8630–8635, August 2016.
- [81] Oskar H Schnaack and Armita Nourmohammad. Optimal evolutionary decision-making to store immune memory. *Elife*, 10, April 2021.
- [82] Jacopo Marchi, Michael Lässig, Aleksandra M Walczak, and Thierry Mora. Antigenic waves of virus-immune coevolution. *Proc. Natl. Acad. Sci. U. S. A.*, 118(27), July 2021.
- [83] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.
- [84] Alfonso Pérez-Escudero, Marta Rivera-Alba, and Gonzalo G de Polavieja. Structure of deviations from optimality in biological systems. *Proc. Natl. Acad. Sci. U. S. A.*, 106(48):20544–20549, December 2009.
- [85] Marianne Bauer, Mariela D Petkova, Thomas Gregor, Eric F Wieschaus, and William Bialek. Trading bits in the readout from a genetic network. *Proc. Natl. Acad. Sci. U. S. A.*, 118(46), November 2021.
- [86] S J Gould and R C Lewontin. The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B Biol. Sci.*, 205(1161):581–598, September 1979.
- [87] O Shoval, H Sheftel, G Shinar, Y Hart, O Ramote, A Mayo, E Dekel, K Kavanagh, and U Alon. Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. *Science*, 336(6085):1157–1160, June 2012.

- [88] Pablo Szekely, Hila Sheftel, Avi Mayo, and Uri Alon. Evolutionary tradeoffs between economy and effectiveness in biological homeostasis systems. *PLoS Comput. Biol.*, 9(8):e1003163, August 2013.
- [89] Yuval Hart, Hila Sheftel, Jean Hausser, Pablo Szekely, Noa Bossel Ben-Moshe, Yael Korem, Avichai Tendler, Avraham E Mayo, and Uri Alon. Inferring biological tasks using pareto analysis of high-dimensional data. *Nat. Methods*, 12(3):233–5, 3 p following 235, March 2015.
- [90] K Murali-Krishna, J D Altman, M Suresh, D J Sourdive, A J Zajac, J D Miller, J Slansky, and R Ahmed. Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity*, 8(2):177–187, February 1998.
- [91] Wen-Hsuan W Lin, Simone A Nish, Bonnie Yen, Yen-Hua Chen, William C Adams, Radomir Kratchmarov, Nyanza J Rothman, Avinash Bhandoola, Hai-Hui Xue, and Steven L Reiner. CD8+ T lymphocyte Self-Renewal during effector cell determination. *Cell Rep.*, 17(7):1773–1782, November 2016.
- [92] John T Chang, Vikram R Palanivel, Ichiko Kinjyo, Felix Schambach, Andrew M Intlekofer, Arnob Banerjee, Sarah A Longworth, Kristine E Vinup, Paul Mrass, Jane Oliaro, Nigel Killeen, Jordan S Orange, Sarah M Russell, Wolfgang Weninger, and Steven L Reiner. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science*, 315(5819):1687–1691, March 2007.
- [93] Boyko Kakaradov, Janilyn Arsenio, Christella E Widjaja, Zhaoren He, Stefan Aigner, Patrick J Metz, Bingfei Yu, Ellen J Wehrens, Justine Lopez, Stephanie H Kim, Elina I Zuniga, Ananda W Goldrath, John T Chang, and Gene W Yeo. Early transcriptional and epigenetic regulation of CD8+ T cell differentiation revealed by single-cell RNA sequencing. *Nat. Immunol.*, 18(4):422–432, April 2017.
- [94] Oliver Bannard, Matthew Kraman, and Douglas T Fearon. Secondary replicative function of CD8+ T cells that had developed an effector phenotype. *Science*, 323(5913):505–509, January 2009.
- [95] Joshy Jacob and David Baltimore. Modelling t-cell memory by genetic marking of memory T cells in vivo. *Nature*, 399(6736):593–597, June 1999.
- [96] Marten Plambeck, Atefeh Kazeroonian, Dirk Loeffler, Lorenz Kretschmer, Ciro Salinno, Timm Schroeder, Dirk H Busch, Michael Flossdorf, and Veit R Buchholz. Heritable changes in division speed accompany the diversification of single T cell fate. *Proceedings of the National Academy of Sciences*, 119(9):e2116260119, 2022.
- [97] A C Fowler. Atto-Foxes and other minutiae. *Bull. Math. Biol.*, 83(10):104, August 2021.

- [98] Rustom Antia, Carl T Bergstrom, Sergei S Pilyugin, Susan M Kaech, and Rafi Ahmed. Models of CD8+ responses: 1. what is the antigen-independent proliferation program. *J. Theor. Biol.*, 221(4):585–598, April 2003.
- [99] Daniela Pais Ferreira, Joana Gomes Silva, Tania Wyss, Silvia A Fuertes Marraco, Léonardo Scarpellino, Mélanie Charmoy, Roeltje Maas, Imran Siddiqui, Li Tang, Johanna A Joyce, Mauro Delorenzi, Sanjiv A Luther, Daniel E Speiser, and Werner Held. Central memory CD8+ T cells derive from stem-like tcf7hi effector cells in the absence of cytotoxic differentiation. *Immunity*, 53(5):985–1000.e11, November 2020.
- [100] Rui Alves and Michael A Savageau. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics*, 16(9):786–798, September 2000.
- [101] Obinna A Ukogu, Zachary Montague, Grégoire Altan-Bonnet, and Armita Nourmohammad. Design principles of the cytotoxic CD8+ T cell response. *Proc. Natl. Acad. Sci. U. S. A.*, 123(17):e2528202123, April 2026.
- [102] Qian Qi, Yi Liu, Yong Cheng, Jacob Glanville, David Zhang, Ji-Yeun Lee, Richard A Olshen, Cornelia M Weyand, Scott D Boyd, and Jörg J Goronzy. Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci. U. S. A.*, 111(36):13139–13144, September 2014.
- [103] Grant Lythe, Robin E Callard, Rollo L Hoare, and Carmen Molina-París. How many TCR clonotypes does a body maintain? *J. Theor. Biol.*, 389:214–224, January 2016.
- [104] Marianne J B van Stipdonk, Gijs Hardenberg, Martijn S Bijker, Edward E Lemmens, Nathalie M Droin, Douglas R Green, and Stephen P Schoenberger. Dynamic programming of CD8+ T lymphocyte responses. *Nat. Immunol.*, 4(4):361–365, April 2003.
- [105] Nikhil S Joshi, Weiguo Cui, Anmol Chandele, Heung Kyu Lee, David R Urso, James Hagman, Laurent Gapin, and Susan M Kaech. Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of t-bet transcription factor. *Immunity*, 27(2):281–295, August 2007.
- [106] Julia M Marchingo, Andrey Kan, Robyn M Sutherland, Ken R Duffy, Cameron J Wellard, Gabrielle T Belz, Andrew M Lew, Mark R Dowling, Susanne Heinzel, and Philip D Hodgkin. T cell signaling. antigen affinity, costimulation, and cytokine inputs sum linearly to amplify T cell expansion. *Science*, 346(6213):1123–1127, November 2014.
- [107] J M Marchingo, G Prevedello, A Kan, S Heinzel, P D Hodgkin, and K R Duffy. T-cell stimuli independently sum to regulate an inherited clonal division fate. *Nat. Commun.*, 7:13540, November 2016.

- [108] Inbal Eizenberg-Magar, Jacob Rimer, Irina Zaretsky, David Lara-Astiaso, Shlomit Reich-Zeliger, and Nir Friedman. Diverse continuum of CD4⁺ T-cell states is determined by hierarchical additive integration of cytokine signals. *Proc. Natl. Acad. Sci. U. S. A.*, 114(31):E6447–E6456, August 2017.
- [109] Vladimir P Badovinac, Jodie S Haring, and John T Harty. Initial T cell receptor transgenic cell precursor frequency dictates critical aspects of the CD8(+) T cell response to infection. *Immunity*, 26(6):827–841, June 2007.
- [110] Nu Zhang and Michael J Bevan. CD8(+) T cells: foot soldiers of the immune system. *Immunity*, 35(2):161–168, August 2011.
- [111] Veit R Buchholz, Michael Flossdorf, Inge Hensel, Lorenz Kretschmer, Bianca Weissbrich, Patricia Gräf, Admar Verschoor, Matthias Schiemann, Thomas Höfer, and Dirk H Busch. Disparate individual fates compose robust CD8⁺ T cell immunity. *Science*, 340(6132):630–635, May 2013.
- [112] Surojit Sarkar, Vandana Kalia, W Nicholas Haining, Bogumila T Konieczny, Shruti Subramaniam, and Rafi Ahmed. Functional and genomic profiling of effector CD8 T cell subsets with distinct memory fates. *J. Exp. Med.*, 205(3):625–640, March 2008.
- [113] Robert A Seder, Patricia A Darrah, and Mario Roederer. T-cell quality in memory and protection: implications for vaccine design. *Nat. Rev. Immunol.*, 8(4):247–258, April 2008.
- [114] Yi-Li Cho, Michael Flossdorf, Lorenz Kretschmer, Thomas Höfer, Dirk H Busch, and Veit R Buchholz. TCR signal quality modulates fate decisions of single CD4⁺ T cells in a probabilistic manner. *Cell Rep.*, 20(4):806–818, July 2017.
- [115] Annie Xin, Frederick Masson, Yang Liao, Simon Preston, Tianxia Guan, Renee Gloury, Moshe Olshansky, Jian-Xin Lin, Peng Li, Terence P Speed, Gordon K Smyth, Matthias Ernst, Warren J Leonard, Marc Pellegrini, Susan M Kaech, Stephen L Nutt, Wei Shi, Gabrielle T Belz, and Axel Kallies. A molecular threshold for effector CD8(+) T cell differentiation controlled by transcription factors blimp-1 and T-bet. *Nat. Immunol.*, 17(4):422–432, April 2016.
- [116] Lorenz Kretschmer, Michael Flossdorf, Jonas Mir, Yi-Li Cho, Marten Plambeck, Irina Treise, Albulena Toska, Susanne Heinzl, Matthias Schiemann, Dirk H Busch, and Veit R Buchholz. Differential expansion of T central memory precursor and effector subsets is regulated by division speed. *Nat. Commun.*, 11(1):113, January 2020.
- [117] Alexander E Downie, Andreas Mayer, C Jessica E Metcalf, and Andrea L Graham. Optimal immune specificity at the intersection of host life history and parasite epidemiology. *PLoS Comput. Biol.*, 17(12):e1009714, December 2021.

- [118] Andreas Mayer, Thierry Mora, Olivier Rivoire, and Aleksandra M Walczak. Diversity of immune strategies explained by adaptation to pathogen statistics. *Proc. Natl. Acad. Sci. U. S. A.*, 113(31):8630–8635, August 2016.
- [119] Douglas R Green, Thomas Ferguson, Laurence Zitvogel, and Guido Kroemer. Immunogenic and tolerogenic cell death. *Nat. Rev. Immunol.*, 9(5):353–363, May 2009.
- [120] K Labbé and M Saleh. Cell death in the host response to infection. *Cell Death Differ.*, 15(9):1339–1349, September 2008.
- [121] Thomas Charles Butler, Mehran Kardar, and Arup K Chakraborty. Quorum sensing allows T cells to discriminate between self and nonself. *Proc. Natl. Acad. Sci. U. S. A.*, 110(29):11833–11838, July 2013.
- [122] Joseph M Cicchese, Stephanie Evans, Caitlin Hult, Louis R Joslyn, Timothy Wessler, Jess A Millar, Simeone Marino, Nicholas A Cilfone, Joshua T Mattila, Jennifer J Linderman, and Denise E Kirschner. Dynamic balance of pro- and anti-inflammatory signals controls disease and limits pathology. *Immunol. Rev.*, 285(1):147–167, September 2018.
- [123] Ruslan Medzhitov. The spectrum of inflammatory responses. *Science*, 374(6571):1070–1075, November 2021.
- [124] Paul François and Vincent Hakim. Design of genetic networks with specified functions by evolution in silico. *Proc. Natl. Acad. Sci. U. S. A.*, 101(2):580–585, January 2004.
- [125] Guy Shinar, Ron Milo, María Rodríguez Martínez, and Uri Alon. Input output robustness in simple bacterial signaling systems. *Proc. Natl. Acad. Sci. U. S. A.*, 104(50):19931–19935, December 2007.
- [126] Jean-Benoît Lalanne and Paul François. Principles of adaptive sorting revealed by in silico evolution. *Phys. Rev. Lett.*, 110(21):218102, May 2013.
- [127] Thomas R Sokolowski, Thomas Gregor, William Bialek, and Gašper Tkačik. Deriving a genetic regulatory network from an optimization principle. February 2023.
- [128] A S Perelson and G F Oster. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. *J. Theor. Biol.*, 81(4):645–670, December 1979.
- [129] Andreas Mayer, Vijay Balasubramanian, Thierry Mora, and Aleksandra M Walczak. How a well-adapted immune system is organized. *Proc. Natl. Acad. Sci. U. S. A.*, 112(19):5950–5955, May 2015.
- [130] Serena Bradde, Armita Nourmohammad, Sidhartha Goyal, and Vijay Balasubramanian. The size of the immune repertoire of bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, 117(10):5144–5151, March 2020.

- [131] Andreas Mayer, Vijay Balasubramanian, Aleksandra M Walczak, and Thierry Mora. How a well-adapting immune system remembers. *Proc. Natl. Acad. Sci. U. S. A.*, 116(18):8815–8823, April 2019.
- [132] Victor Chardès, Massimo Vergassola, Aleksandra M Walczak, and Thierry Mora. Affinity maturation for an optimal balance between long-term immune coverage and short-term resource constraints. *arXiv [q-bio.PE]*, page 2021.07.26.453765, July 2021.
- [133] Yawei Qin, Emily M Mace, and John P Barton. An inference model gives insights into innate immune adaptation and repertoire diversity. *Proc. Natl. Acad. Sci. U. S. A.*, 120(38):e2305859120, September 2023.
- [134] Gregg N Milligan and Alan D T Barrett. *Vaccinology: An Essential Guide*. Wiley-Blackwell, Hoboken, NJ, January 2015.
- [135] Pablo Abuin, Alejandro Anderson, Antonio Ferramosca, Esteban A Hernandez-Vargas, and Alejandro H Gonzalez. Characterization of SARS-CoV-2 dynamics in the host. *Annu. Rev. Control*, 50:457–468, October 2020.
- [136] Paul François, Guillaume Voisinne, Eric D Siggia, Grégoire Altan-Bonnet, and Massimo Vergassola. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism. *Proc. Natl. Acad. Sci. U. S. A.*, 110(10):E888–97, March 2013.
- [137] Luca Gattinoni, Enrico Lugli, Yun Ji, Zoltan Pos, Chrystal M Paulos, Máire F Quigley, Jorge R Almeida, Emma Gostick, Zhiya Yu, Carmine Carpenito, Ena Wang, Daniel C Douek, David A Price, Carl H June, Francesco M Marincola, Mario Roederer, and Nicholas P Restifo. A human memory T cell subset with stem cell-like properties. *Nat. Med.*, 17(10):1290–1297, September 2011.
- [138] Weiguo Cui and Susan M Kaech. Generation of effector CD8+ T cells and their conversion to memory T cells: Effector and memory T-cell differentiation. *Immunol. Rev.*, 236(1):151–166, July 2010. ISSN 0105-2896,1600-065X. doi: 10.1111/j.1600-065X.2010.00926.x.
- [139] Wiet de Ronde, Pieter Rein ten Wolde, and Andrew Mugler. Protein logic: a statistical mechanical study of signal integration at the single-molecule level. *Biophys. J.*, 103(5):1097–1107, September 2012.
- [140] Alon Oyler-Yaniv, Jennifer Oyler-Yaniv, Benjamin M Whitlock, Zhiduo Liu, Ronald N Germain, Morgan Huse, Grégoire Altan-Bonnet, and Oleg Krichevsky. A tunable diffusion-consumption mechanism of cytokine propagation enables plasticity in cell-to-cell communication in the immune system. *Immunity*, 46(4):609–620, April 2017.
- [141] Guillaume Voisinne, Briana G Nixon, Anna Melbinger, Georg Gasteiger, Massimo Vergassola, and Grégoire Altan-Bonnet. T cells integrate local and global cues to

- discriminate between structurally similar antigens. *Cell Rep.*, 11(8):1208–1219, May 2015.
- [142] Miri Adler, Avi Mayo, and Uri Alon. Logarithmic and power law input-output relations in sensory systems with fold-change detection. *PLoS Comput. Biol.*, 10(8):e1003781, August 2014.
 - [143] Avni S Gupta, Victoria Chevé, Adam S Kirosingh, Nicole M Davis, and David S Schneider. Variation in the response to malaria in diversity outbred mice. *bioRxiv*, page 2021.09.29.462483, September 2021.
 - [144] Yael Lebel, Avni S Gupta, Victoria Chevé, Uri Alon, and David S Schneider. An immunologist, ecologist and clinician walk into plato’s cave to discuss infections. *bioRxiv*, page 2025.03.04.641508, March 2025.
 - [145] Philip L F Johnson, Beth F Kochin, Megan S McAfee, Ingunn M Stromnes, Roland R Regoes, Rafi Ahmed, Joseph N Blattman, and Rustom Antia. Vaccination alters the balance between protective immunity, exhaustion, escape, and death in chronic infections. *J. Virol.*, 85(11):5565–5570, June 2011.
 - [146] Akiko Iwasaki and Ruslan Medzhitov. Control of adaptive immunity by the innate immune system. *Nat. Immunol.*, 16(4):343–353, April 2015. ISSN 1529-2908,1529-2916. doi: 10.1038/ni.3123.
 - [147] Loren Koçillari, Piero Fariselli, Antonio Trovato, Flavio Seno, and Amos Maritan. Signature of pareto optimization in the escherichia coli proteome. *Sci. Rep.*, 8(1):9141, June 2018.
 - [148] Dietmar Zehn, Sarah Y Lee, and Michael J Bevan. Complete but curtailed t-cell response to very low-affinity antigen. *Nature*, 458(7235):211–214, March 2009.
 - [149] Joseph N Blattman, Jason M Grayson, E John Wherry, Susan M Kaech, Kendall A Smith, and Rafi Ahmed. Therapeutic use of IL-2 to enhance antiviral t-cell responses in vivo. *Nat. Med.*, 9(5):540–547, May 2003.
 - [150] Estelle Baulu, Célia Gardet, Nicolas Chuvin, and Stéphane Depil. TCR-engineered T cell therapy in solid tumors: State of the art and perspectives. *Sci. Adv.*, 9(7):eadf3700, February 2023.
 - [151] Inés Zugasti, Lady Espinosa-Aroca, Klaudyna Fidyt, Vladimir Mulens-Arias, Marina Diaz-Beya, Manel Juan, Álvaro Urbano-Ispizua, Jordi Esteve, Talia Velasco-Hernandez, and Pablo Menéndez. CAR-T cell therapy for cancer: current challenges and future directions. *Signal Transduct. Target. Ther.*, 10(1):210, July 2025.
 - [152] Anand Rotte, Matthew J Frigault, Ayub Ansari, Brad Gliner, Christopher Heery, and Bijal Shah. Dose-response correlation for CAR-T cells: a systematic review of clinical studies. *J. Immunother. Cancer*, 10(12):e005678, December 2022.

- [153] Michele M Hoffmann and Jill E Slansky. T-cell receptor affinity in the age of cancer immunotherapy. *Mol. Carcinog.*, 59(7):862–870, July 2020.
- [154] Bingxu Liu, Nathan F Greenwood, Julia E Bonzanini, Amir Motmaen, Jeremy Meyerberg, Tao Dao, Xinyu Xiang, Russell Ault, Jazmin Sharp, Chunyu Wang, Gian Marco Visani, Dionne K Vafeados, Nicole Roullier, Armita Nourmohammad, David A Scheinberg, K Christopher Garcia, and David Baker. Design of high-specificity binders for peptide-MHC-I complexes. *Science*, 389(6758):386–391, July 2025.
- [155] Taisuke Kondo, François X P Bourassa, Sooraj Achar, Justyn DuSold, Pablo F Céspedes, Makoto Ando, Alka Dwivedi, Josquin Moraly, Christopher Chien, Salih Majdoul, Adam L Kenet, Madison Wahlsten, Audun Kvalvaag, Edward Jenkins, Sanghyun P Kim, Catherine M Ade, Zhiya Yu, Guillaume Gaud, Marco Davila, Paul Love, James C Yang, Michael L Dustin, Grégoire Altan-Bonnet, Paul François, and Naomi Taylor. Engineering TCR-controlled fuzzy logic into CAR T cells enhances therapeutic specificity. *Cell*, 0(0), April 2025.
- [156] Gian Marco Visani, Michael N Pun, Anastasia A Minervina, Philip Bradley, Paul Thomas, and Armita Nourmohammad. T-cell receptor specificity landscape revealed through de novo peptide design. *bioRxiv.org*, page 2025.02.28.640903, March 2025.
- [157] Giulia Escobar, Katherine Tooley, Joan Pagès Oliveras, Linglin Huang, Hanning Cheng, Michelle L Bookstaver, Camilla Edwards, Eugene Froimchuk, Chang Xue, Davide Mangani, Rajesh K Krishnan, Natanael Hazel, Carola Rutigliani, Christopher M Jewell, Luca Biasco, and Ana C Anderson. Tumor immunogenicity dictates reliance on TCF1 in CD8+ T cells for response to immunotherapy. *Cancer Cell*, 41(9):1662–1679.e7, September 2023.
- [158] Nathiya Muthalagu, Melissa R Junttila, Katrin E Wiese, Elmar Wolf, Jennifer Morton, Barbara Bauer, Gerard I Evan, Martin Eilers, and Daniel J Murphy. BIM is the primary mediator of MYC-induced apoptosis in multiple solid tissues. *Cell Rep.*, 8(5):1347–1353, September 2014.
- [159] Jennifer R Tisoncik, Marcus J Korth, Cameron P Simmons, Jeremy Farrar, Thomas R Martin, and Michael G Katze. Into the eye of the cytokine storm. *Microbiol. Mol. Biol. Rev.*, 76(1):16–32, March 2012.
- [160] Toshiaki Yoshikawa, Zhiwen Wu, Satoshi Inoue, Hitomi Kasuya, Hirokazu Matsushita, Yusuke Takahashi, Hiroaki Kuroda, Waki Hosoda, Shiro Suzuki, and Yuki Kagoya. Genetic ablation of PRDM1 in antitumor T cells enhances therapeutic efficacy of adoptive immunotherapy. *Blood*, 139(14):2156–2172, April 2022.
- [161] Fei Yi, Tal Cohen, Natalie Zimmerman, Friederike Dünder, Paul Zumbo, Razan Eltilib, Erica J Brophy, Hannah Arkin, Judith Feucht, Michael V Gormally, Christopher S Hackett, Korbinian N Kropp, Inaki Etxeberria, Smita S Chandran, Zeguo Zhao, Winsong Cai, Anthony F Daniyan, Jae H Park, Caleb A Lareau, Katharine C Hsu, Michel

- Sadelain, Doron Betel, and Christopher A Klebanoff. CAR-engineered lymphocyte persistence is governed by a FAS ligand-FAS autoregulatory circuit. *Nat. Cancer*, pages 1–18, July 2025.
- [162] Nelson H Knudsen, Giulia Escobar, Felix Korell, Tamina Kienka, Celeste Nobrega, Seth Anderson, Andrew Y Cheng, Maria Zschummel, Alexander Armstrong, Amanda Bouffard, Michael C Kann, Sadie Goncalves, Hans W Pope, Mitra Pezeshki, Alexander Rojas, Juliette S M T Suermondt, Merle Phillips, Trisha R Berger, Sangwoo Park, Diego Salas-Benito, Elijah P Darnell, Filippo Birocchi, Mark B Leick, Rebecca C Larson, John G Doench, Debattama Sen, Kathleen B Yates, Robert T Manguso, and Marcela V Maus. In vivo CRISPR screens identify modifiers of CAR T cell function in myeloma. *Nature*, pages 1–10, September 2025.
- [163] Paul Datlinger, Eugenia V Pankevich, Cosmas D Arnold, Nicole Prankevicius, Jenny Lin, Daria Romanovskaia, Moritz Schaefer, Francesco Piras, Anne-Christine Orts, Amelie Nemc, Paulina N Biesaga, Michelle Chan, Teresa Neuwirth, Artem V Artemov, Wentao Li, Sabrina Ladstätter, Thomas Krausgruber, and Christoph Bock. Systematic discovery of CRISPR-boosted CAR T cell immunotherapies. *Nature*, pages 1–10, September 2025.
- [164] Kiam Heong Ang, G Chong, and Yun Li. PID control system analysis, design, and technology. *IEEE Trans. Control Syst. Technol.*, 13(4):559–576, July 2005.
- [165] Kaspar Bresser, Lianne Kok, Arpit C Swain, Lisa A King, Laura Jacobs, Tom S Weber, Leïla Perié, Ken R Duffy, Rob J de Boer, Ferenc A Scheeren, and Ton N Schumacher. Replicative history marks transcriptional and functional disparity in the CD8+ T cell memory pool. *Nat. Immunol.*, 23(5):791–801, May 2022.
- [166] Simon Grassmann, Lorenz Mihatsch, Jonas Mir, Atefeh Kazeroonian, Roza Rahimi, Sophie Flommersfeld, Kilian Schober, Inge Hensel, Justin Leube, Ludwig O Pachmayr, Lorenz Kretschmer, Qin Zhang, Adrien Jolly, M Zeeshan Chaudhry, Matthias Schiemann, Luka Cicin-Sain, Thomas Höfer, Dirk H Busch, Michael Flossdorf, and Veit R Buchholz. Early emergence of T central memory precursors programs clonal dominance during chronic viral infection. *Nat. Immunol.*, 21(12):1563–1573, December 2020.
- [167] Shiyu Xia, Andrew C Lu, Victoria Tobin, Kaiwen Luo, Lukas Moeller, D Judy Shon, Rongrong Du, James M Linton, Margaret Sui, Felix Horns, and Michael B Elowitz. Synthetic protein circuits for programmable control of mammalian cell death. *Cell*, 187(11):2785–2800.e16, May 2024.
- [168] Hui-Shan Li, Divya V Israni, Keith A Gagnon, Kok Ann Gan, Michael H Raymond, Jeffry D Sander, Kole T Roybal, J Keith Joung, Wilson W Wong, and Ahmad S Khalil. Multidimensional control of therapeutic human cell function with synthetic gene circuits. *Science*, 378(6625):1227–1234, December 2022.

- [169] Greg M Allen, Nicholas W Frankel, Nishith R Reddy, Hersh K Bhargava, Maia A Yoshida, Sierra R Stark, Megan Purl, Jungmin Lee, Jacqueline L Yee, Wei Yu, Aileen W Li, K Christopher Garcia, Hana El-Samad, Kole T Roybal, Matthew H Spitzer, and Wendell A Lim. Synthetic cytokine circuits that drive T cells into immune-excluded tumors. *Science*, 378(6625):eaba1624, December 2022.
- [170] Michael Lässig, Ville Mustonen, and Armita Nourmohammad. Steering and controlling evolution - from bioengineering to fighting pathogens. *Nat. Rev. Genet.*, 24(12):851–867, December 2023.
- [171] Dennis L Chao, Miles P Davenport, Stephanie Forrest, and Alan S Perelson. A stochastic model of cytotoxic T cell responses. *J. Theor. Biol.*, 228(2):227–240, May 2004.
- [172] Vitaly V Ganusov, Daniel L Barber, and Rob J De Boer. Killing of targets by CD8 T cells in the mouse spleen follows the law of mass action. *PLoS One*, 6(1):e15959, January 2011.
- [173] Stephan Halle, Kirsten Anja Keyser, Felix Rolf Stahl, Andreas Busche, Anja Marquardt, Xiang Zheng, Melanie Galla, Vigo Heissmeyer, Katrin Heller, Jasmin Boelter, Karen Wagner, Yvonne Bischoff, Rieke Martens, Asolina Braun, Kathrin Werth, Alexey Uvarovskii, Harald Kempf, Michael Meyer-Hermann, Ramon Arens, Melanie Kremer, Gerd Sutter, Martin Messerle, and Reinhold Förster. In vivo killing capacity of cytotoxic T cells is limited and involves dynamic interactions and T cell cooperativity. *Immunity*, 44(2):233–245, February 2016.
- [174] Lora V Hooper, Dan R Littman, and Andrew J Macpherson. Interactions between the microbiota and the immune system. *Science*, 336(6086):1268–1273, June 2012.
- [175] Eduard Ansaldo, Taylor K Farley, and Yasmine Belkaid. Control of immunity by the microbiota. *Annu. Rev. Immunol.*, 39:449–479, April 2021.
- [176] Markus Arnoldini, Richa Sharma, Claudia Moresi, Griffin Chure, Julien Chabbey, Emma Slack, and Jonas Cremer. Quantifying the varying harvest of fermentation products from the human gut microbiota. *Cell*, 188(19):5332–5342.e16, September 2025.
- [177] Noah W Palm, Marcel R de Zoete, Thomas W Cullen, Natasha A Barry, Jonathan Stefanowski, Liming Hao, Patrick H Degnan, Jianzhong Hu, Inga Peter, Wei Zhang, Elizabeth Ruggiero, Judy H Cho, Andrew L Goodman, and Richard A Flavell. Immunoglobulin a coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell*, 158(5):1000–1010, August 2014.
- [178] Takahiro Yamada and Noah W Palm. A host-centric view of the microbiota metabolome. *Immunity*, 58(12):2940–2956, December 2025.

- [179] Min Wang, Lucas J Osborn, Sunit Jain, Xiandong Meng, Allison Weakley, Jia Yan, William J Massey, Venkateshwari Varadharajan, Anthony Horak, Rakhee Banerjee, Daniela S Allende, E Ricky Chan, Adeline M Hajjar, Zeneng Wang, Alejandra Dimas, Aishan Zhao, Kazuki Nagashima, Alice G Cheng, Steven Higginbottom, Stanley L Hazen, J Mark Brown, and Michael A Fischbach. Strain dropouts reveal interactions that govern the metabolic output of the gut microbiome. *Cell*, 186(13):2839–2852.e21, June 2023.
- [180] Matthew S Schmitt, Kiseok Lee, Freddy Bunbury, Joseph A Landsittel, Vincenzo Vitelli, and Seppe Kuehn. Learning functional groups in complex microbiomes. *bioRxiv*, page 2026.03.03.709366, March 2026.
- [181] Abigail Skwara, Karna Gowda, Mahmoud Yousef, Juan Diaz-Colunga, Arjun S Raman, Alvaro Sanchez, Mikhail Tikhonov, and Seppe Kuehn. Statistically learning the functional landscape of microbial communities. *Nat Ecol Evol*, 7(11):1823–1833, November 2023.
- [182] Richard R Stein, Takeshi Tanoue, Rose L Szabady, Shakti K Bhattarai, Bernat Olle, Jason M Norman, Wataru Suda, Kenshiro Oshima, Masahira Hattori, Georg K Gerber, Chris Sander, Kenya Honda, and Vanni Bucci. Computer-guided design of optimal microbial consortia for immune system modulation. *Elife*, 7, April 2018.
- [183] Marco Tulio Angulo, Claude H Moog, and Yang-Yu Liu. A theoretical framework for controlling complex microbial communities. *Nat. Commun.*, 10(1):1045, March 2019.
- [184] Jacob Moran and Mikhail Tikhonov. Defining coarse-grainability in a model of structured microbial ecosystems. *Phys. Rev. X*, 12(2):021038, May 2022.
- [185] Matthew R Olm, Sean P Spencer, Tadashi Takeuchi, Evelyn Lemus Silva, and Justin L Sonnenburg. Metagenomic immunoglobulin sequencing reveals IgA coating of microbial strains in the healthy human gut. *Nat. Microbiol.*, 10(1):112–125, January 2025.
- [186] Koji Atarashi, Takeshi Tanoue, Kenshiro Oshima, Wataru Suda, Yuji Nagano, Hiroyoshi Nishikawa, Shinji Fukuda, Takuro Saito, Seiko Narushima, Koji Hase, Sangwan Kim, Joëlle V Fritz, Paul Wilmes, Satoshi Ueha, Kouji Matsushima, Hiroshi Ohno, Bernat Olle, Shimon Sakaguchi, Tadatsugu Taniguchi, Hidetoshi Morita, Masahira Hattori, and Kenya Honda. Treg induction by a rationally selected mixture of clostridia strains from the human microbiota. *Nature*, 500(7461):232–236, August 2013.
- [187] David E J Klawon, Nicole Pagane, Matthew T Walker, Nicole K Ganci, Christine H Miller, Eric Gai, Donald M Rodriguez, Bridgett K Ryan-Payseur, Ryan K Duncombe, Erin J Adams, Mark Maienschein-Cline, Nancy E Freitag, Ronald N Germain, Harikesh S Wong, and Peter A Savage. Regulatory T cells constrain T cells of shared specificity to enforce tolerance during infection. *Science*, page eadk3248, February 2025.

- [188] Laura F Su, Daniel Del Alcazar, Erietta Stelekati, E John Wherry, and Mark M Davis. Antigen exposure shapes the ratio between antigen-specific tregs and conventional T cells in human peripheral blood. *Proc. Natl. Acad. Sci. U. S. A.*, 113(41):E6192–E6198, October 2016.
- [189] Thomas Höfer, Oleg Krichevsky, and Grégoire Altan-Bonnet. Competition for IL-2 between regulatory and effector T cells to chisel immune responses. *Front. Immunol.*, 3:268, September 2012.
- [190] Alice McNally, Geoffrey R Hill, Tim Sparwasser, Ranjeny Thomas, and Raymond J Steptoe. CD4+CD25+ regulatory T cells control CD8+ T-cell effector differentiation by modulating IL-2 homeostasis. *Proc. Natl. Acad. Sci. U. S. A.*, 108(18):7529–7534, May 2011.
- [191] Jens Y Humrich, Henner Morbach, Reinmar Undeutsch, Philipp Enghard, Stefan Rosenberger, Olivia Weigert, Lutz Kloke, Juliane Heimann, Timo Gaber, Susan Brandenburg, Alexander Scheffold, Jochen Huehn, Andreas Radbruch, Gerd-Rüdiger Burmester, and Gabriela Riemekasten. Homeostatic imbalance of regulatory and effector T cells due to IL-2 deprivation amplifies murine lupus. *Proc. Natl. Acad. Sci. U. S. A.*, 107(1):204–209, January 2010.
- [192] Eric R Littmann, Jung-Jin Lee, Joshua E Denny, Zahidul Alam, Jeffrey R Maslanka, Isma Zarin, Rina Matsuda, Rebecca A Carter, Bože Susac, Miriam S Saffern, Bryton Fett, Lisa M Mattei, Kyle Bittinger, and Michael C Abt. Host immunity modulates the efficacy of microbiota transplantation for treatment of clostridioides difficile infection. *Nat. Commun.*, 12(1):755, February 2021.
- [193] Melanie Schirmer, Sanne P Smeekens, Hera Vlamakis, Martin Jaeger, Marije Oosting, Eric A Franzosa, Rob Ter Horst, Trees Jansen, Liesbeth Jacobs, Marc Jan Bonder, Alexander Kurilshikov, Jingyuan Fu, Leo A B Joosten, Alexandra Zhernakova, Curtis Huttenhower, Cisca Wijmenga, Mihai G Netea, and Ramnik J Xavier. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell*, 167(4):1125–1136.e8, November 2016.
- [194] Wioletta Lisicka, Zachary M Earley, Joseph J Sifakis, Steven A Erickson, Jonathan R Mattingly, Natalie J Wu-Woods, Siddharth R Krishnamurthy, Yasmine Belkaid, Rustem F Ismagilov, Jason G Cyster, Samantha J Riesenfeld, Albert Bendelac, and Bana Jabri. Immunoglobulin a controls intestinal virus colonization to preserve immune homeostasis. *Cell Host Microbe*, 33(4):498–511.e10, April 2025.
- [195] Ivan Vujkovic-Cvijin, Hugh C Welles, Connie W Y Ha, Lutfi Huq, Shreni Mistry, Jason M Brenchley, Giorgio Trinchieri, Suzanne Devkota, and Yasmine Belkaid. The systemic anti-microbiota IgG repertoire can identify gut bacteria that translocate across gut barrier surfaces. *Sci. Transl. Med.*, 14(658):eabl3927, 2022.

- [196] Kathrin Moor, Médéric Diard, Mikael E Sellin, Boas Felmy, Sandra Y Wotzka, Albulena Toska, Erik Bakkeren, Markus Arnoldini, Florence Bansept, Alma Dal Co, Tom Völler, Andrea Minola, Blanca Fernandez-Rodriguez, Gloria Agatic, Sonia Barbieri, Luca Piccoli, Costanza Casiraghi, Davide Corti, Antonio Lanzavecchia, Roland R Regoes, Claude Loverdo, Roman Stocker, Douglas R Brumley, Wolf-Dietrich Hardt, and Emma Slack. High-avidity IgA protects the intestine by enchainning growing bacteria. *Nature*, 544(7651):498–502, April 2017.
- [197] Yisu Gu, Raquel Bartolomé-Casado, Chuan Xu, Alice Bertocchi, Alina Janney, Cornelia Heuberger, Claire F Pearson, Sarah A Teichmann, Emily E Thornton, and Fiona Powrie. Immune microniches shape intestinal treg function. *Nature*, 628(8009):854–862, April 2024.
- [198] Anne-Katrin Pröbstel, Xiaoyuan Zhou, Ryan Baumann, Sven Wischnewski, Michael Kutza, Olga L Rojas, Katrin Sellrie, Antje Bischof, Kicheol Kim, Akshaya Ramesh, Ravi Dandekar, Ariele L Greenfield, Ryan D Schubert, Jordan E Bisanz, Stephanie Vistnes, Khashayar Khaleghi, James Landefeld, Gina Kirkish, Friederike Liesche-Starnecker, Valeria Ramaglia, Sneha Singh, Edwina B Tran, Patrick Barba, Kelsey Zorn, Johanna Oechtering, Karin Forsberg, Lawrence R Shiow, Roland G Henry, Jennifer Graves, Bruce A C Cree, Stephen L Hauser, Jens Kuhle, Jeffrey M Gelfand, Peter M Andersen, Jürgen Schlegel, Peter J Turnbaugh, Peter H Seeberger, Jennifer L Gommerman, Michael R Wilson, Lucas Schirmer, and Sergio E Baranzini. Gut microbiota-specific IgA+ B cells traffic to the CNS in active multiple sclerosis. *Sci. Immunol.*, 5(53):eabc7191, November 2020.
- [199] Lucia M Kato, Shimpei Kawamoto, Mikako Maruya, and Sidonia Fagarasan. The role of the adaptive immune system in regulation of gut microbiota. *Immunol. Rev.*, 260(1):67–75, July 2014.
- [200] Takeshi Tanoue, Satoru Morita, Damian R Plichta, Ashwin N Skelly, Wataru Suda, Yuki Sugiura, Seiko Narushima, Hera Vlamakis, Iori Motoo, Kayoko Sugita, Atsushi Shiota, Kozue Takeshita, Keiko Yasuma-Mitobe, Dieter Riethmacher, Tsuneyasu Kaisho, Jason M Norman, Daniel Mucida, Makoto Suematsu, Tomonori Yaguchi, Vanni Bucci, Takashi Inoue, Yutaka Kawakami, Bernat Olle, Bruce Roberts, Masahira Hattori, Ramnik J Xavier, Koji Atarashi, and Kenya Honda. A defined commensal consortium elicits CD8 T cells and anti-cancer immunity. *Nature*, 565(7741):600–605, January 2019.
- [201] Naama Geva-Zatorsky, Esen Sefik, Lindsay Kua, Lesley Pasman, Tze Guan Tan, Adriana Ortiz-Lopez, Tsering Bakto Yanortsang, Liang Yang, Ray Jupp, Diane Mathis, Christophe Benoist, and Dennis L Kasper. Mining the human gut microbiota for immunomodulatory organisms. *Cell*, 168(5):928–943.e11, February 2017.
- [202] Lisa Derosa, Bertrand Routy, Andrew Maltez Thomas, Valerio Iebba, Gerard Zalcman, Sylvie Friard, Julien Mazieres, Clarisse Audigier-Valette, Denis Moro-Sibilot,

- François Goldwasser, Carolina Alves Costa Silva, Safae Terrisse, Melodie Bonvalet, Arnaud Scherpereel, Hervé Pegliasco, Corentin Richard, François Ghiringhelli, Arielle Elkrief, Antoine Desilets, Felix Blanc-Durand, Fabio Cumbo, Aitor Blanco, Romain Boidot, Sandy Chevrier, Romain Daillère, Guido Kroemer, Laurie Alla, Nicolas Pons, Emmanuelle Le Chatelier, Nathalie Galleron, Hugo Roume, Agathe Dubuisson, Nicole Bouchard, Meriem Messaoudene, Damien Drubay, Eric Deutsch, Fabrice Barlesi, David Planchard, Nicola Segata, Stéphanie Martinez, Laurence Zitvogel, Jean-Charles Soria, and Benjamin Besse. Intestinal *akkermansia muciniphila* predicts clinical response to PD-1 blockade in patients with advanced non-small-cell lung cancer. *Nat. Med.*, 28(2): 315–324, February 2022.
- [203] Albert C Yeh, Motoko Koyama, Olivia G Waltner, Simone A Minnie, Julie R Boiko, Tamer B Shabaneh, Shuichiro Takahashi, Ping Zhang, Kathleen S Ensbey, Christine R Schmidt, Samuel R W Legg, Tomoko Sekiguchi, Ethan Nelson, Shruti S Bhise, Andrew R Stevens, Tracy Goodpaster, Saranya Chakka, Scott N Furlan, Kate A Markey, Marie E Bleakley, Charles O Elson, Philip H Bradley, and Geoffrey R Hill. Microbiota dictate T cell clonal selection to augment graft-versus-host disease after stem cell transplantation. *Immunity*, 57(7):1648–1664.e9, July 2024.
- [204] Hannah C Wastyk, Gabriela K Fragiadakis, Dalia Perelman, Dylan Dahan, Bryan D Merrill, Feiqiao B Yu, Madeline Topf, Carlos G Gonzalez, William Van Treuren, Shuo Han, Jennifer L Robinson, Joshua E Elias, Erica D Sonnenburg, Christopher D Gardner, and Justin L Sonnenburg. Gut-microbiota-targeted diets modulate human immune status. *Cell*, 184(16):4137–4153.e14, August 2021.
- [205] Elisa Sánchez-Martínez, Liam E Rondeau, Manuel Garrido-Romero, Bruna Barbosa da Luz, Dominic A Haas, Gavin Yuen, Peter Hall, Rebecca Dang, Xuan-Yu Wang, Lucía Moreno-Serna, Celia López-Sanz, Emilio Nuñez-Borque, Maria Garrido-Arandia, Araceli Diaz-Perales, Yolanda R Carrasco, Joshua F E Koenig, Tina D Walker, Manel Jordana, Elena F Verdu, Michael G Surette, Pedro Ojeda, Francisco Vega, Carlos Blanco, Wayne G Shreffler, Sarita U Patil, F Javier Moreno, Rodrigo Jiménez-Saiz, and Alberto Caminero. Microbial metabolism of food allergens determines the severity of IgE-mediated anaphylaxis. *bioRxiv.org*, page 2025.02.17.638013, February 2025.
- [206] Kazuki Nagashima, Aishan Zhao, Katayoon Atabakhsh, Minwoo Bae, Jamie E Blum, Allison Weakley, Sunit Jain, Xiandong Meng, Alice G Cheng, Min Wang, Steven Higginbottom, Alex Dimas, Pallavi Murugkar, Elizabeth S Sattely, James J Moon, Emily P Balskus, and Michael A Fischbach. Mapping the T cell repertoire to a complex gut bacterial community. *Nature*, August 2023.
- [207] Alice G Cheng, Po-Yi Ho, Andrés Aranda-Díaz, Sunit Jain, Feiqiao B Yu, Xiandong Meng, Min Wang, Mikhail Iakiviak, Kazuki Nagashima, Aishan Zhao, Pallavi Murugkar, Advait Patil, Katayoon Atabakhsh, Allison Weakley, Jia Yan, Ariel R Brumbaugh, Steven Higginbottom, Alejandra Dimas, Anthony L Shiver, Adam

- Deutschbauer, Norma Neff, Justin L Sonnenburg, Kerwyn Casey Huang, and Michael A Fischbach. Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell*, 185(19):3617–3636.e19, September 2022.
- [208] Emma Slack, Maria Luisa Balmer, Jörg H Fritz, and Siegfried Hapfelmeier. Functional flexibility of intestinal IgA - broadening the fine line. *Front. Immunol.*, 3:100, May 2012.
- [209] Oliver Pabst. New concepts in the generation and functions of IgA. *Nat. Rev. Immunol.*, 12(12):821–832, December 2012.
- [210] Maya Saleh and Charles O Elson. Experimental inflammatory bowel disease: insights into the host-microbiota dialog. *Immunity*, 34(3):293–302, March 2011.
- [211] Kirthana Ganeshan, Joni Nikkanen, Kevin Man, Yew Ann Leong, Yoshitaka Sogawa, J Alan Maschek, Tyler Van Ry, D Nyasha Chagwedera, James E Cox, and Ajay Chawla. Energetic trade-offs and hypometabolic states promote disease tolerance. *Cell*, 177(2):399–413.e12, April 2019.
- [212] Yingzi Cong, Ting Feng, Kohtaro Fujihashi, Trenton R Schoeb, and Charles O Elson. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.*, 106(46):19256–19261, November 2009.
- [213] Yi Yang, Miriam B Torchinsky, Michael Gobert, Huizhong Xiong, Mo Xu, Jonathan L Linehan, Francis Alonzo, Charles Ng, Alessandra Chen, Xiyao Lin, Andrew Sczesnak, Jia-Jun Liao, Victor J Torres, Marc K Jenkins, Juan J Lafaille, and Dan R Littman. Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature*, 510(7503):152–156, June 2014.
- [214] Zixian Liu, Jiabei Nie, Yimei Li, Maoxin Huang, Ziluo Chen, Shushang Yu, Jiaqi Zheng, Yuyan Tan, and Shengdi Chen. *Bacteroides coprocola* protects dopaminergic neurons in rotenone-induced parkinson’s disease mouse model by modulating gut microbiota dysbiosis and inhibiting the NLRP3 signaling pathway. *Transl. Neurodegener.*, 15(1):7, February 2026.
- [215] Matthew T Henke, Douglas J Kenny, Chelsi D Cassilly, Hera Vlamakis, Ramnik J Xavier, and Jon Clardy. *Ruminococcus gnavus*, a member of the human gut microbiome associated with crohn’s disease, produces an inflammatory polysaccharide. *Proc. Natl. Acad. Sci. U. S. A.*, 116(26):12672–12677, June 2019.
- [216] Karna Gowda, Derek Ping, Madhav Mani, and Seppe Kuehn. Genomic structure predicts metabolite dynamics in microbial communities. *Cell*, 185(3):530–546.e25, February 2022.
- [217] Céline Robert, Christophe Chassard, Paul A Lawson, and Annick Bernalier-Donadille. *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int. J. Syst. Evol. Microbiol.*, 57(Pt 7):1516–1520, July 2007.

- [218] Sameer Elsayed and Kunyan Zhang. Human infection caused by *clostridium hathewayi*. *Emerg. Infect. Dis.*, 10(11):1950–1952, November 2004.
- [219] William H Press. From the cover: Strong profiling is not mathematically optimal for discovering rare malfeasors. *Proc. Natl. Acad. Sci. U. S. A.*, 106(6):1716–1719, February 2009.
- [220] Gary D Wu, Jun Chen, Christian Hoffmann, Kyle Bittinger, Ying-Yu Chen, Sue A Keilbaugh, Meenakshi Bewtra, Dan Knights, William A Walters, Rob Knight, Rohini Sinha, Erin Gilroy, Kernika Gupta, Robert Baldassano, Lisa Nessel, Hongzhe Li, Frederic D Bushman, and James D Lewis. Linking long-term dietary patterns with gut microbial enterotypes. *Science*, 334(6052):105–108, October 2011.
- [221] Amandine Everard, Clara Belzer, Lucie Geurts, Janneke P Ouwerkerk, Céline Druart, Laure B Bindels, Yves Guiot, Muriel Derrien, Giulio G Muccioli, Nathalie M Delzenne, Willem M de Vos, and Patrice D Cani. Cross-talk between *akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci. U. S. A.*, 110(22):9066–9071, May 2013.
- [222] Shishi Luo and Alan S Perelson. Competitive exclusion by autologous antibodies can prevent broad HIV-1 antibodies from arising. *Proc. Natl. Acad. Sci. U. S. A.*, 112(37):11654–11659, September 2015.
- [223] Armita Nourmohammad, Jakub Otwinowski, and Joshua B Plotkin. Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS Genet.*, 12(7):e1006171, July 2016.
- [224] Shimon Sakaguchi. Naturally arising CD4⁺ regulatory t cells for immunologic self-tolerance and negative control of immune responses. *Annu. Rev. Immunol.*, 22(1):531–562, 2004.
- [225] Miguel Reina-Campos, Alexander Monell, Amir Ferry, Vida Luna, Kitty P Cheung, Giovanni Galletti, Nicole E Scharping, Kennedy K Takehara, Sara Quon, Peter P Challita, Brigid Boland, Yun Hsuan Lin, William H Wong, Cynthia S Indralingam, Hayley Neadeau, Suzie Alarcón, Gene W Yeo, John T Chang, Maximilian Heeg, and Ananda W Goldrath. Tissue-resident memory CD8 T cell diversity is spatiotemporally imprinted. *Nature*, 639(8054):483–492, March 2025.
- [226] Sandra Nakandakari-Higa, Sarah Walker, Maria C C Canesso, Verena van der Heide, Aleksey Chudnovskiy, Dong-Yoon Kim, Johanne T Jacobsen, Roham Parsa, Jana Bilanovic, S Martina Parigi, Karol Fiedorczuk, Elaine Fuchs, Angelina M Bilate, Giulia Pasqual, Daniel Mucida, Alice O Kamphorst, Yuri Pritykin, and Gabriel D Victora. Universal recording of immune cell interactions in vivo. *Nature*, 627(8003):399–406, March 2024.

- [227] Joshua G Harrison, W John Calder, Vivaswat Shastry, and C Alex Buerkle. Dirichlet-multinomial modelling outperforms alternatives for analysis of microbiome and other ecological count data. *Mol. Ecol. Resour.*, 20(2):481–497, March 2020.
- [228] Hila Segre, Niv DeMalach, Zalmen Henkin, and Ronen Kadmon. Quantifying competitive exclusion and competitive release in ecological communities: A conceptual framework and a case study. *PLoS One*, 11(8):e0160798, August 2016.
- [229] P K Newton and Y Ma. Nonlinear adaptive control of competitive release and chemotherapeutic resistance. *Phys Rev E*, 99(2-1):022404, February 2019.
- [230] Jacopo Grilli, György Barabás, Matthew J Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210–213, August 2017.
- [231] Daniel P Newton, Po-Yi Ho, and Kerwyn Casey Huang. Modulation of antibiotic effects on microbial communities by resource competition. *Nat. Commun.*, 14(1):2398, April 2023.
- [232] James A Frank. Claudins and alveolar epithelial barrier function in the lung: Claudins in lung injury and repair. *Ann. N. Y. Acad. Sci.*, 1257(1):175–183, June 2012.
- [233] Ruomeng Li, Jing Li, and Xikun Zhou. Lung microbiome: new insights into the pathogenesis of respiratory diseases. *Signal Transduct. Target. Ther.*, 9(1):19, January 2024.
- [234] Margaret Coates, Min Jin Lee, Diana Norton, and Amanda S MacLeod. The skin and intestinal microbiota and their specific innate immune systems. *Front. Immunol.*, 10:2950, December 2019.
- [235] Inta Griboņika, Victor I Band, Liang Chi, Paula Juliana Perez-Chaparro, Verena M Link, Eduard Ansaldi, Cihan Oguz, Djenet Bousbaine, Michael A Fischbach, and Yasmine Belkaid. Skin autonomous antibody production regulates host-microbiota interactions. *Nature*, 638(8052):1043–1053, February 2025.
- [236] Ainsley Lockhart, Daniel Mucida, and Angelina M Bilate. Intraepithelial lymphocytes of the intestine. *Annu. Rev. Immunol.*, 42(1):289–316, June 2024.
- [237] James J Moon, H Hamlet Chu, Marion Pepper, Stephen J McSorley, Stephen C Jameson, Ross M Kedl, and Marc K Jenkins. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity*, 27(2):203–213, August 2007.
- [238] Nicole D Sonnert, Connor E Rosen, Andrew R Ghazi, Eric A Franzosa, Brianna Duncan-Lowey, Jaime A González-Hernández, John D Huck, Yi Yang, Yile Dai, Tyler A Rice, Mytien T Nguyen, Deguang Song, Yiyun Cao, Anjelica L Martin,

- Agata A Bielecka, Suzanne Fischer, Changhui Guan, Julia Oh, Curtis Huttenhower, Aaron M Ring, and Noah W Palm. A host-microbiota interactome reveals extensive transkingdom connectivity. *Nature*, March 2024.
- [239] Roham Parsa, Helder Assis, Tiago B R de Castro, Gabriella Lima Dos Reis, Harald Hartweger, Angelina Bilate, and Daniel Mucida. Tissue-specific clonal selection and differentiation of CD4+ T cells during infection. *bioRxiv*, August 2025.
- [240] Yun Min Chang, Andreas Wieland, Zheng-Rong Li, Se Jin Im, Donald J McGuire, Haydn T Kissick, Rustom Antia, and Rafi Ahmed. T cell receptor diversity and lineage relationship between virus-specific CD8 T cell subsets during chronic lymphocytic choriomeningitis virus infection. *J. Virol.*, 94(20):e00935–20, September 2020.
- [241] Anna Posfai, Thibaud Tallefumier, and Ned S Wingreen. Metabolic Trade-Offs promote diversity in a model ecosystem. *Phys. Rev. Lett.*, 118(2):028103, January 2017.
- [242] Po-Yi Ho, Taylor H Nguyen, Juan M Sanchez, Brian C DeFelice, and Kerwyn Casey Huang. Resource competition predicts assembly of gut bacterial communities in vitro. *Nat Microbiol*, 9(4):1036–1048, April 2024.
- [243] Zhijie Feng, Emmy Blumenthal, Pankaj Mehta, and Akshit Goyal. A theory of ecological invasions and its implications for eco-evolutionary dynamics. *Proc. Natl. Acad. Sci. U. S. A.*, 122(49):e2505850122, December 2025.
- [244] Benjamin H Good and Layton B Rosenfeld. Eco-evolutionary feedbacks in the human gut microbiome. *Nat. Commun.*, 14(1):7146, November 2023.
- [245] Munehiro Furuichi, Takaaki Kawaguchi, Marie-Madlen Pust, Keiko Yasuma-Mitobe, Damian R Plichta, Naomi Hasegawa, Takashi Ohya, Shakti K Bhattarai, Satoshi Sasajima, Yoshimasa Aoto, Timur Tuganbaev, Mizuki Yaginuma, Masahiro Ueda, Nobuyuki Okahashi, Kimiko Amafuji, Yuko Kiridoshi, Kayoko Sugita, Martin Stražar, Julian Avila-Pacheco, Kerry Pierce, Clary B Clish, Ashwin N Skelly, Masahira Hattori, Nobuhiro Nakamoto, Silvia Caballero, Jason M Norman, Bernat Olle, Takeshi Tanoue, Wataru Suda, Makoto Arita, Vanni Bucci, Koji Atarashi, Ramnik J Xavier, and Kenya Honda. Commensal consortia decolonize enterobacteriaceae via ecological control. *Nature*, 633(8031):878–886, September 2024.
- [246] Yasmine Belkaid and Timothy W Hand. Role of the microbiota in immunity and inflammation. *Cell*, 157(1):121–141, March 2014.
- [247] Camille Ameline, Elsa Seixas, Hugo C Barreto, Nelson Frazão, Miguel V Rodrigues, M Rita Ventura, Marta Lourenço, and Isabel Gordo. Evolution of escherichia coli strains under competent or compromised adaptive immunity. *PLoS Pathog.*, 21(4):e1012442, April 2025.

- [248] Husen Zhang, Joshua B Sparks, Saikumar V Karyala, Robert Settlage, and Xin M Luo. Host adaptive immunity alters gut microbiota. *ISME J.*, 9(3):770–781, March 2015.
- [249] Stanley Plotkin. History of vaccination. *Proc. Natl. Acad. Sci. U. S. A.*, 111(34):12283–12287, August 2014.
- [250] Hongbo Chi, Marion Pepper, and Paul G Thomas. Principles and therapeutic applications of adaptive immunity. *Cell*, 187(9):2052–2078, April 2024.
- [251] Georgina Thorborn, Mickaél J Ploquin, Urszula Eksmond, Rebecca Pike, Wibke Bayer, Ulf Dittmer, Kim J Hasenkrug, Marion Pepper, and George Kassiotis. Clonotypic composition of the CD4+ T cell response to a vectored retroviral antigen is determined by its speed. *J. Immunol.*, 193(4):1567–1577, August 2014.
- [252] Andrew C Tremain, Rachel P Wallace, Kristen M Lorentz, Thomas B Thornley, Jennifer T Antane, Michal R Racz, Joseph W Reda, Aaron T Alpar, Anna J Slezak, Elyse A Watkins, Chitavi D Maulloo, Erica Budina, Ani Solanki, Mindy Nguyen, David J Bischoff, Jamie L Harrington, Rabinarayan Mishra, Gregory P Conley, Romain Marlin, Nathalie Dereuddre-Bosquet, Anne-Sophie Gallouët, Roger LeGrand, D Scott Wilson, Stephan Kontos, and Jeffrey A Hubbell. Synthetically glycosylated antigens for the antigen-specific suppression of established immune responses. *Nat. Biomed. Eng.*, 7(9):1142–1155, September 2023.
- [253] Christiaan H van Dorp, Joshua I Gray, Daniel H Paik, Donna L Farber, and Andrew J Yates. A variational deep-learning approach to modeling memory T cell dynamics. *PLoS Comput. Biol.*, 21(7):e1013242, July 2025.
- [254] Elizabeth R Jerison, Nicolas Romeo, and Stephen R Quake. Spatially-structured inflammatory response in the presence of a uniform stimulus. *bioRxiv*, page 2025.01.28.635318, January 2025.
- [255] Thierry Mora and Aleksandra M Walczak. Towards a quantitative theory of tolerance. *Trends Immunol.*, 44(7):512–518, July 2023.
- [256] Shubham Tripathi, John S Tsang, and Kyemyung Park. Systems immunology of regulatory T cells: can one circuit explain it all? *Trends Immunol.*, 44(10):766–781, October 2023.
- [257] Mihai G Netea, Leo A B Joosten, Eicke Latz, Kingston H G Mills, Gioacchino Natoli, Hendrik G Stunnenberg, Luke A J O’Neill, and Ramnik J Xavier. Trained immunity: A program of innate immune memory in health and disease. *Science*, 352(6284):aaf1098, April 2016.
- [258] C Jessica E Metcalf, Andrea L Graham, Andrew J Yates, and Derek A T Cummings. Convergence and divergence of individual immune responses over the life course. *Science*, 389(6760):604–609, August 2025.

- [259] Maeve P Nagle, Gabriela S Tam, Evan Maltz, Zachary Hemminger, and Roy Wollman. Bridging scales: From cell biology to physiology using in situ single-cell technologies. *Cell Syst.*, 12(5):388–400, May 2021.
- [260] Maya E Kotas and Ruslan Medzhitov. Homeostasis, inflammation, and disease susceptibility. *Cell*, 160(5):816–827, February 2015.
- [261] Mason A Porter and James P Gleeson. Dynamical systems on networks: A tutorial. *arXiv [nlin.AO]*, March 2014.
- [262] David S Fischer, Anna C Schaar, and Fabian J Theis. Learning cell communication from spatial graphs of cells. *bioRxiv*, page 2021.07.11.451750, July 2021.
- [263] Tailin Wu, Hongyu Ren, Pan Li, and Jure Leskovec. Graph information bottleneck. *arXiv [cs.LG]*, October 2020.
- [264] Justin Lessler, Nicholas G Reich, Ron Brookmeyer, Trish M Perl, Kenrad E Nelson, and Derek A T Cummings. Incubation periods of acute respiratory viral infections: a systematic review. *Lancet Infect. Dis.*, 9(5):291–300, May 2009.
- [265] Elissa K Deenick, Amanda V Gett, and Philip D Hodgkin. Stochastic model of T cell proliferation: a calculus revealing IL-2 regulation of precursor frequencies, cell cycle time, and survival. *J. Immunol.*, 170(10):4963–4972, May 2003.
- [266] Hochan Cheon, Andrey Kan, Giulio Prevedello, Simone C Oostindie, Simon J Dovedi, Edwin D Hawkins, Julia M Marchingo, Susanne Heinzl, Ken R Duffy, and Philip D Hodgkin. Cyton2: A model of immune cell population dynamics that includes familial instructional inheritance. *Front Bioinform.*, 1:723337, October 2021.
- [267] Anyue Yin, Dirk Jan A R Moes, Johan G C van Hasselt, Jesse J Swen, and Henk-Jan Guchelaar. A review of mathematical models for tumor dynamics and treatment resistance evolution of solid tumors. *CPT Pharmacometrics Syst. Pharmacol.*, 8(10):720–737, October 2019.
- [268] M Uhl, O Kepp, H Jusforgues-Saklani, J-M Vicencio, G Kroemer, and M L Albert. Autophagy within the antigen donor cell facilitates efficient antigen cross-priming of virus-specific CD8⁺ T cells. *Cell Death Differ.*, 16(7):991–1005, July 2009.
- [269] Gabrielle T Belz, Lei Zhang, Matthew D H Lay, Fiona Kupresanin, and Miles P Davenport. Killer T cells regulate antigen presentation for early expansion of memory, but not naive, CD8⁺ T cell. *Proc. Natl. Acad. Sci. U. S. A.*, 104(15):6341–6346, April 2007.
- [270] Jeroen W J van Heijst, Carmen Gerlach, Erwin Swart, Daoud Sie, Cláudio Nunes-Alves, Ron M Kerkhoven, Ramon Arens, Margarida Correia-Neves, Koen Schepers, and Ton N M Schumacher. Recruitment of antigen-specific CD8⁺ T cells in response to infection is markedly efficient. *Science*, 325(5945):1265–1269, September 2009.

- [271] Kamir J Hiam-Galvez, Rachel DeBarge, Caleb A Lareau, Nam Woo Cho, Jacqueline L Yee, Trine Line H Okholm, and Matthew H Spitzer. Transient dendritic cell activation diversifies the T cell response to acute infection. September 2021.
- [272] Steven L Reiner and William C Adams. Lymphocyte fate specification as a deterministic but highly plastic process. *Nat. Rev. Immunol.*, 14(10):699–704, October 2014.
- [273] Wei Liao, Jian-Xin Lin, and Warren J Leonard. Interleukin-2 at the crossroads of effector responses, tolerance, and immunotherapy. *Immunity*, 38(1):13–25, January 2013.
- [274] Anissa Guillemin and Michael P H Stumpf. Noise and the molecular processes underlying cell fate decision-making. *Phys. Biol.*, 18(1):011002, January 2021.
- [275] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.
- [276] Yang Cao, Daniel T Gillespie, and Linda R Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124(4):044109, January 2006.
- [277] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a “knee-dle” in a haystack: Detecting knee points in system behavior. In *2011 31st International Conference on Distributed Computing Systems Workshops*. IEEE, June 2011.
- [278] Kevin Arvai. *Kneed: Knee-point detection in Python*, July 2023. URL <https://doi.org/10.5281/zenodo.8127224>.
- [279] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [280] Hiroki Ikeda, Rob J de Boer, Kei Sato, Satoru Morita, Naoko Misawa, Yoshio Koyanagi, Kazuyuki Aihara, and Shingo Iwami. Improving the estimation of the death rate of infected cells from time course data during the acute phase of virus infections: application to acute HIV-1 infection in a humanized mouse model. *Theor. Biol. Med. Model.*, 11(1):22, May 2014.
- [281] Ruian Ke, Mian-Er Cong, David Li, J Gerardo García-Lerma, and Alan S Perelson. On the death rate of abortively infected cells: Estimation from simian-human immunodeficiency virus infection. *J. Virol.*, 91(18), September 2017.
- [282] W Barchet, S Oehen, P Klenerman, D Wodarz, G Bocharov, A L Lloyd, M A Nowak, H Hengartner, R M Zinkernagel, and S Ehl. Direct quantitation of rapid elimination of viral antigen-positive lymphocytes by antiviral CD8(+) T cells in vivo. *Eur. J. Immunol.*, 30(5):1356–1363, May 2000.

- [283] Hao Hong Yiu, Andrea L Graham, and Robert F Stengel. Dynamics of a cytokine storm. *PLoS One*, 7(10):e45027, October 2012.
- [284] “Estimating a Dirichlet distribution”. `dirichlet`. <https://pypi.org/project/dirichlet/>. Accessed: 2026-6-11.
- [285] Thomas P Minka. Estimating a dirichlet distribution, 2000.