Control of stemness in CD8 T cell differentiation

Kathleen V Abadie

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Hao Yuan Kueh, Chair

Michael Gerner

William Stafford Noble

Program Authorized to Offer Degree:

Bioengineering

University of Washington

Abstract

Control of stemness in CD8 T cell differentiation

Kathleen V Abadie

Chair of the Supervisory Committee:
Hao Yuan Kueh
Department of Bioengineering

T cell stemness enables continual generation of cytotoxic effector cells from a self-renewing pool of multipotent progenitors. After an acute infection, memory T cells persist for years in a quiescent state and upon rechallenge, rapidly reconstitute an entire expanded T cell response. In chronic infection and cancer, characterized by progression of functional cytotoxic T cells to dysfunctional exhausted cells, stem-like progenitors provide a continual source of cytotoxic cells to fuel a persistent response. In both cases, the system must balance the need for differentiation to generate effector cells and for protection of the stem cell population. Different immune challenges manifest over different time scales and with variable pathogenicity, and as such, the size and characteristics of the stem cell pool that forms vary across these different challenges with distinct signaling environments. While the molecular components that control these outcomes have been exhaustively characterized, the regulatory principles by which they act remain unclear. Thus, the goal of my thesis work has been to reveal these regulatory principles using reductionist approaches

that measure T cell differentiation at the clonal level, over time, and with isolation and perturbation of distinct differentiating populations under defined signaling conditions. I first provide evidence for a stochastic epigenetic gene silencing mechanism that enables robust memory formation in acute infection. I then identify cell-state and signaling context-dependent regulators of T cell stemness using a targeted screening approach across multiple phases of T cell differentiation under chronic stimulation.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

A PhD is hard, and it was hard for reasons I did not expect. When I came to graduate school, I was ready to grit my teeth and work 80 hours a week, crank out experiments, analyze data, make discoveries, and publish them. Hard work, on top of the privilege of a good education, had always led me to success in the past, and I expected the same moving forward. When I got here, however, I realized that the challenge of grad school is not the many hours it takes to do experiments or build analysis pipelines. It's not about being clever or having intellectual conversations with illustrious professors and peers, nor is it about having the most organized sample boxes in the freezer or the smoothest running code. As Doug Fowler told me after I shed some tears in his office in my second year, success in grad school is learning to pick up your flag and move it forward, inch by inch, knowing that if you do not move it forward, no one will. I have found that to be consistently true over the past six years, and as I look ahead, I know that the most valuable thing I will take from my PhD is the confidence to move my flag forward, wherever I go.

The only reason I'm still holding onto my flag after six years of grad-school-induced self-doubt is because of my wonderful community of mentors, family, and friends who remind me that I am good and loved.

First, to my mentors. I have been lucky to have at least one incredibly caring, supportive mentor at every stage of my career. My fluid transport professor at Rice University, Sean Hartig, was the first to introduce me to the joys of molecular biology, taking me under his wing in his research lab. He told me that my CV was impressive, and I should apply to grad school and shoot for the best schools. In my first biopharma internship at Celgene in 2013, I had the great fortune to be paired with Ellen Filvaroff as a research mentor and Kristen Hege as a personal mentor, and

these two inspirational women have supported and guided my career ever since. Ellen showed me what it means to be a thoughtful mentor and manager, printing out papers or relevant slide decks and dropping them off at my desk with a post-it note, and she has patiently talked me through the periods of my PhD in which I felt the most uncertainty and shame. Kristen is my biggest role model, providing a steady example of what it means to be a fierce, patient-motivated scientist, a serious athlete and outdoorswoman, a champion of women and minorities in science, and a caring mother and friend. She has opened more doors for me than I can count, and as I have gotten further in my career, she has helped me mentor others in her footsteps. I do not know where I would be in my career without her generous support, and even though she says I can't buy her dinner until I'm VP of a company, I am always looking for ways to repay her. At Genentech, I serendipitously fell into the cycling-to-work group of James Sabry, who has been a cheerleader for my cycling and my science ever since. He sent me and my partner Flo off to grad school with our own copies of Karl Popper's *The Philosophy of Science* and a big *Molecular Biology of the Cell* textbook, and he always makes time in his schedule for a coffee or a ride whenever we are back in San Francisco. I cannot emphasize enough what it has meant to me that these amazing and accomplished people have believed in me, and a big part of what makes me want to be successful in my own career is to be able to believe in other young scientists and have it mean as much to them.

I thank my research advisor, Kueh, for always lighting up my brain and sticking by my side through the ups and downs of scientific research. Kueh has an amazing ability to talk with you for hours about your project, dig into your data with you, and go back and forth on experiment design, while still giving you the freedom and creativity to become an independent-thinking scientist. He has taught me to embrace the unexpected in my data and to treat the moments when

the science does not match my hypotheses as opportunities rather than disappointments. Kueh's joy for science is palpable, but even so, he prioritizes the people in his lab over projects. He listens to our hopes and goals and does everything in his power to help us achieve them. I am so lucky to have had him as an advisor and can't wait to see what the lab does in the coming years.

Thank you to my research advisor at the Weizmann Institute, Ido Amit, for welcoming me into his beautiful lab of warm, hilarious, and creative scientists. Ido has a singular ability to craft a meaningful immunology story that made a lasting impression on me. My year in his lab made me lighter, tougher, and hopefully a little funnier.

To my family, I am grateful for their endless love and confidence in me. My parents have always been so proud of me, enthusiastically cheering me on in science and sports from childhood to adulthood. When I talk to my dad on the phone, he usually either asks me how the mice are doing or if I'm getting closer to getting my Nobel Prize. While this is of course in jest, I know my dad actually thinks I could win a Nobel Prize, and that means the world to me. My three older siblings have provided the best role models over the years, and I am grateful for the close relationship we all have in our adult lives. Thank you to close friends who have always kept me grounded and shared in the joys of academics, sports, and just doing nothing but laughing.

To my partner Flo, I am grateful for many things. Perhaps the most important during grad school is that she has helped me hold onto all the things that make me me when I would otherwise have let work consume me. Flo is the best friend and family member anyone could have and always prioritizes the things in life that matter. She reminds me to be thoughtful, to be adventurous, and that going for a bike ride is usually a good idea. She is my biggest cheerleader, the best wheel, and

my own personal free genomics consultant. I couldn't be prouder to be defending my thesis just five days after her, and I look forward to continuing our scientific careers and lives together.

Lastly, thank you to my fellow lab mates, with whom it has been an honor to go through this journey. Elisa Clark has been my steady project partner and co-author for the second half of my PhD, and looking back, I realize that grad school suddenly became a lot more enjoyable when we started working together. Thank you to the custodial staff in our building who brighten the early mornings in lab and who have welcomed me into their lunchtime potlucks. Thank you to our research scientists, Kenny, Lihua, and Paul, who got me started in the lab and keep things running smoothly. Thank you to the new, motivated batch of Kueh lab grad students who give me great confidence that the lab has a big future ahead. Thank you to the many mice whose lives I have used for science. And please forgive me for the massive amount of laboratory plastic I have used in the past six years.

While a PhD is hard, it is a privilege more than anything. What a joy it has been to follow scientific curiosities and spend my time and effort working on something that feels meaningful. I am endlessly grateful to the people and circumstances that have made it possible for me.

Chapter 1. INTRODUCTION

1.1 THE IMMUNE CELL'S CHALLENGE

Chapter 1.1 is adapted from the following manuscript:

Abadie, K., Pease, N. A., Wither, M. J., & Kueh, H. Y. (2019). Order by chance: Origins and benefits of stochasticity in immune cell fate control. *Current Opinion in Systems Biology*, 18, 95–103.

To protect against diverse challenges, the immune system must continuously generate an arsenal of specialized cell types, each of which can mount myriad effector responses upon detection of potential threats. To do so, it must generate multiple differentiated cell populations with defined sizes and proportions, often from rare starting precursor cells. Upon pathogen detection, these cells must activate and acquire a spectrum of effector functions, with the timing and frequency closely tailored to the nature and magnitude of the threat. Unlike most other cell types, immune cells and progenitors are highly motile and broadly distributed in the body. As a result, they must depend less on pre-defined spatial cues and constraints to control population sizes and fractions, but instead rely on bottom-up self-organization. In the fly embryo, for instance, developing cells generate different fates at defined proportions by reading out levels of external signals in their vicinity (Gregor, Bialek, Steveninck, Tank, & Wieschaus, 2005; Gregor, Wieschaus, McGregor, Bialek, & Tank, 2007), a process that unfolds across a largely fixed spatial domain. In contrast, hematopoietic progenitors and activating T and B cells generate diverse progeny in the presence of apparently uniform signals (V. R. Buchholz et al., 2013; Suda, Suda, & Ogawa, 1984; Taylor, Pape, Steach, & Jenkins, 2015; Weaver et al., 1998), and do so while dramatically expanding in number, often by multiple orders of magnitude. Such autonomous, bottom-up generation of

heterogeneous cell populations imposes distinct functional constraints on the molecular circuits controlling fate decisions in immune cells. An emerging view is that inherently probabilistic events in cells – stemming from the stochastic nature of biochemical reactions – direct cell fates and responses in the immune system and underlie proper population control (Veit R. Buchholz, Schumacher, & Busch, 2016; Hodgkin, 2018). Since early pioneering work in hematopoiesis (Becker, McCulloch, & Till, 1963; Suda et al., 1984), it has been observed at many developmental and effector response decision points that clonally related cells can show highly heterogeneous outcomes (V. R. Buchholz et al., 2013; Cho et al., 2017; Hawkins, Markham, McGuinness, & Hodgkin, 2009), even in a uniform environment (J. T. Chang et al., 2007; Duffy et al., 2012). This heterogeneity has long been suspected to reflect stochastic control at the single cell level; however, it has been unclear how such stochasticity arises on a molecular level. In this dissertation, I will explore the immune cell's challenge with a specific focus on cytotoxic T cell differentiation.

1.2 ADAPTIVE IMMUNITY AND THE IMPORTANCE OF MEMORY

A hallmark of adaptive immunity is the ability to simultaneously respond to an ongoing immune challenge and generate memory to protect from future encounters with that challenge. This feature is enabled by differentiation of progenitor adaptive immune cells into heterogeneous cell types equipped either for rapid killing of target cells or for persistence. Cytotoxic CD8 T cells form a major arm of the adaptive response, responsible for direct killing of cells displaying foreign antigen. Upon encounter with antigen, naive CD8 T cells undergo massive proliferation and differentiation into diverse functional states, including short-lived, cytotoxic effector cells that directly mediate target killing and long-lived, self-renewing memory cells that persist for years after antigen clearance. Each of these differentiated subsets is critical to managing the immune

challenge over time: effector cells control the immediate challenge, while memory cells provide a fast-responding progenitor population to mount a secondary response, if the challenge reoccurs. As both populations derive from the same limited number of progenitor cells, the decision to become one or the other must be appropriately regulated in order to produce adequate numbers of each and to do so robustly across the diverse range of threats that the immune system encounters.

Defining the regulatory mechanisms that enable such robust control of effector and memory T cell differentiation has long been a goal of immunologists. From the latter half of the 1900s to the present, key features of the T cell immune response have come to light hand-in-hand with technological and methodological innovations. Development of transgenic T cell receptor (TCR) mice, whose T cells respond specifically to known antigen either naturally present or engineered into viral or bacterial strains, enabled controlled initiation and measurement of T cell responses to pathogen (Pircher, Bürki, Lang, Hengartner, & Zinkernagel, 1989a). With these model systems, it became possible to isolate a population of T cells with the same TCR specific for a known antigen (such as P14 cells specific for the lymphocytic choriomeningitis virus gp33 antigen or OT-I cells specific for the chicken ovalbumin antigen), transfer this population to a recipient mouse, infect that recipient mouse with a virus or bacterial strain expressing that antigen, and then isolate the expanded and differentiated progeny of the transferred T cell population at defined time points after infection for analysis. In combination with continual advancement of immune cell phenotyping methods using flow cytometry, transcriptomics, and epigenomics, as well as the cleverness of those wielding them, these technologies have revealed key principles in CD8 T cell memory and effector differentiation. Below I highlight conceptual advances and researchers that have motivated my thesis questions, while acknowledging the countless advances from the many individuals and studies that I do not discuss here.

1.2.1 FINDING THE PRECURSORS TO MEMORY

Early work from Rafi Ahmed, Susan Kaech, and others characterized heterogeneous memory and effector differentiation from naive T cell populations. Their pioneering studies established the distinct phases of the T cell response to infection: expansion and effector cell generation, contraction, and stable memory formation (Kaech & Ahmed, 2001). Among the massive population of cells that arise during expansion, they identified distinct memory precursor and short-lived effector cells, characterized by opposite expression of the interleukin 7 receptor alpha-chain (IL-7R) and killer cell lectin-like receptor G1 (KLRG1) (Joshi et al., 2007; Kaech et al., 2003). These definitions of memory precursor and short lived effector cells, easily distinguishable by a few cell surface markers, enabled many subsequent studies to measure memory and effector heterogeneity in different models and in response to perturbation. While we appreciate now, with the advent of single-cell transcriptomics, that these binary definitions of memory and effector fated cells likely oversimply a more nuanced population of differentiating cells, the ability to isolate and measure these cell states has yielded rich insights into control of memory and effector differentiation over the decades since their establishment (Kaech & Cui, 2012).

1.2.2 A SINGLE NAIVE OR MEMORY T CELL CAN DIFFERENTIATE INTO ALL HETEROGENEOUS EFFECTOR AND MEMORY STATES.

Through elegant single-cell adoptive transfer experiments using OT-I transgenic TCR mice and *Listeria monocytogenes* expressing chicken ovalbumin (Lm-Ova), Stemberger and subsequently Graef and colleagues established the stemness of both naive and memory CD8 T cells (Graef et

al., 2014; Stemberger et al., 2007). Prior to Graef's study, it was hypothesized but not proven that immune memory could operate via stem cell populations with self-renewal capabilities (Fearon, Manders, & Wagner, 2001). Graef demonstrated that a single central memory T cell isolated from the initial response of a single naive T cell to Lm-Ova could expand to the same number and diversity of progeny as the initially responding naive cell. Critically, to meet the definition of stemness, a central memory T cell from that secondary response could then be isolated for rechallenge and give rise to an additional fully expanded and diverse response. More recently, Soerens and colleagues massively extended this line of serial transfer interrogation to demonstrate that memory T cells generated and sustained in mice through consecutive prime-boost immunizations and transfers across multiple individual animals for ten years were able to expand seemingly indefinitely (Soerens et al., 2023).

1.2.3 T CELL DIFFERENTIATION IS STOCHASTIC AT THE CLONAL LEVEL BUT REPRODUCIBLE AT THE POPULATION LEVEL

The T cell response to an immune challenge must robustly generate adequate numbers of effector and memory T cells to counter the immediate challenge and protect against future challenges. For any given antigen, the pool of naive T cell clones specific for that antigen ranges from 100 to 1000 (Blattman et al., 2002). Does each of those clones respond to antigen identically, giving rise to the same number of expanded cells and fraction of memory and effector phenotypes? Or alternatively, do distinct subsets of the initial pool give rise to separate memory and effector progeny? These questions were unresolved until two pioneering studies from Gerlach and Buchholz in 2013 demonstrated that highly heterogeneous individual clone outcomes give rise to the observed robust population-level T cell responses (V. R. Buchholz et al., 2013; Gerlach et al., 2013). Plumlee and

colleagues independently demonstrated the heterogeneity in clonal responses to a given infection and furthermore showed that while differentiation at the clonal level is stochastic, it is neither random nor uncontrolled (Plumlee, Sheridan, Cicek, & Lefrançois, 2013). Different infections give rise to distinct environmental cytokine compositions, which in turn drive different effector and memory differentiation outcomes at the clonal and population level.

1.3 Regulatory mechanisms maintaining T cell stemness

The broad goal of my thesis work was to identify regulatory mechanisms that enable the features of the immune system described above. Specifically, I sought to understand how activating T cells control the decision to become memory and effector cells such that differentiated populations are variable between clones but highly reproducible at the population level. Furthermore, I asked what properties of the system enable appropriate scaling of the response to dynamic immune challenges, for which information about pathogen severity conveyed by antigen and cytokine levels is constantly changing. I focused much of my investigation on the transcription factor TCF1, encoded by the gene *Tcf7*, which is essential for the self-renewing properties of memory T cells. In my experiments, I use *Tcf7* as an imperfect but useful marker of stem-like T cells, as well as a model gene for understanding how these heterogeneous cell state decisions may be regulated at the individual locus level. In biological research, we are often forced to choose between high resolution measurement of one gene or lower resolution measurement of many genes. I would have preferred to investigate many key genes involved in T cell differentiation under the same resolution I used to measure *Tcf7*, but it was not experimentally feasible. However, with the rapid pace of genomic and imaging method development, I am optimistic these investigations may soon be within reach.

In **Chapter 2** of this dissertation, we identify a flexible memory decision-making strategy enabled by a stochastic, reversible, epigenetic switch silencing *Tcf7*. In **Chapter 3**, we extend our investigation of T cell stemness to the chronic stimulation setting, under which classical memory cells, defined by the absence of stimulation, cannot form. We characterize epigenetically distinct *Tcf7* silenced states and perturb potential regulators of the silent state across different phases of the response to chronic stimulation. In **Chapter 4**, I look forward to method development that promises to provide temporal and spatial resolution to our understanding of immune cell differentiation processes.

Chapter 2. FLEXIBLE AND SCALABLE CONTROL OF T CELL MEMORY BY A REVERSIBLE EPIGENETIC SWITCH

Chapter 2 is adapted with minimal modifications from the following manuscript:

Abadie, K.*, Clark, E.C.*, Valanparambil, R.M., Ukogu, O., Yang, W., Daza, R.M., Ng, K.H., Fathima, J., Wang, A., Bhandoola, A., Nourmohammad, A., Ahmed, R., Shendure, J., Cao, J., Kueh, H.Y. (2023). Flexible and scalable control of T cell memory by a reversible epigenetic switch. *under revision and resubmission, Immunity.*

The supplementary movies and tables can be found on the Bioarxiv preprint: https://www.biorxiv.org/content/10.1101/2022.12.31.521782v2

2.1 ABSTRACT

The immune system encodes information about the severity of a pathogenic threat in the quantity and type of memory cell populations formed in response. This encoding emerges from the decisions of lymphocytes to maintain or lose self-renewal and memory potential during a challenge. By tracking responding CD8 T cells at the single-cell and clonal lineage level using time-resolved transcriptomics and quantitative live imaging, we identify a remarkably flexible decision-making strategy, whereby T cells initially choose whether to maintain or lose memory

potential early after antigen recognition, but following pathogen clearance may regain memory potential if initially lost. Mechanistically, this flexibility is implemented by a *cis*-epigenetic switch that silences the memory regulator TCF1 in a stochastic and reversible manner in response to stimulatory inputs. Mathematical modeling shows how this strategy allows memory T cell numbers to scale robustly with pathogen virulence and immune response magnitudes. We propose that flexibility and stochasticity in cellular decision making ensures optimal immune responses against diverse threats.

2.2 INTRODUCTION

The immune system keeps a memory of prior infections with information about the inducing threat. This memory is encoded by the numbers and types of memory lymphocytes generated upon challenge. The quantity of memory T cells, in particular, scales with the magnitude of a prior infection, such that the memory population is a fixed fraction of the T cell number at the infection peak, across a range of pathogenic challenges (Busch, Pilip, Vijh, & Pamer, 1998; Murali-Krishna et al., 1998; Vijh & Pamer, 1997). This scaling in memory production is robust across T cell clones with different epitope specificities and allows the body to generate memory proportional to the severity of the pathogenic challenge. The regulatory mechanisms that enable this critical feature of adaptive immunity are not well understood.

The size and characteristics of the memory compartment are determined by the fate decision-making strategies of T cells responding to an acute infection (Kaech, Wherry, & Ahmed, 2002). As naive CD8 T cells respond to antigens, they must decide whether and when to maintain long-term viability and self-renewal potential, and thereby persist to form memory cells as the infection is cleared. One class of models posits that cells make this decision early after antigen encounter, and in a mutually exclusive manner with effector differentiation (**Fig. 2.1A**) (J. T.

Chang et al., 2007; Kakaradov et al., 2017; Lin et al., 2016). Under this model, memory cells form directly from naive cells without first passing through an effector phase, but through an early lineage bifurcation that concurrently gives rise to short-lived effector cells. A second class of models posits that cells decide later, only after they have undergone effector differentiation (**Fig. 2.1A**) (Bannard, Kraman, & Fearon, 2009; Jacob & Baltimore, 1999; Youngblood et al., 2017). In this strategy, cytotoxic effectors that maintain memory potential populate the memory compartment upon infection clearance. However, in contrast with both models, it is also possible that this process is inherently flexible (Chung, McDonald, & Kaech, 2021), such that T cells have multiple opportunities to commit to the memory state. From a social and cognitive sciences perspective (Benjaafar, Morin, & Talavage, 1995; Tello-Ramos, Branch, Kozlovsky, Pitera, & Pravosudov, 2019), flexibility in decision making allows individuals to adapt and better respond to uncertain and dynamic environments; in the immune system, such flexibility may allow T cells to optimize memory formation for threats whose properties might only manifest as they unfold in time. It is unclear whether there exists such flexibility in T cell memory formation and, if so, what its underlying mechanisms and functional roles are.

In this study, we sought to elucidate the memory decision-making dynamics of CD8 T cells by following the regulation of TCF1 (encoded by *Tcf7*), a transcription factor essential for memory cell generation (Zhao, Shan, & Xue, 2021). *Tcf7* is expressed in naive and memory cells, where it is crucial for maintaining self-renewal, and is silenced during effector differentiation, resulting in loss of memory potential and entry into a short-lived state (Jeannet et al., 2010; Lin et al., 2016). To follow *Tcf7* regulation and memory decision-making in a controlled environment where cells can be continuously observed and signaling inputs carefully manipulated, we developed an *ex vivo* system to mimic stimulation of T cells by acute challenge. Using this system and complementary

testing *in vivo*, we uncover a flexible decision-making strategy: T cells can gain or lose memory potential at multiple junctures after antigen encounter, and do so in a stochastic and reversible manner. Mathematical modeling reveals that this flexible decision-making strategy allows for the number of memory cells to scale linearly with total numbers of expanded T cells at the peak of infection, thereby encoding information about the severity of the prior threat.

2.3 RESULTS

2.3.1 A MINIMAL EX VIVO SYSTEM FOR EFFECTOR AND MEMORY DIFFERENTIATION

In our system, naive ($CD44^-CD62L^+$) CD8 T cells with a knock-in YFP reporter for *Tcf7* (Harly et al., 2019) are activated with plate-immobilized anti-CD3 and anti-CD28 antibodies and IL-2, together with additional cytokines present during acute infection (IL-12, IL-7, and IL-15 (Mescher et al., 2006; Rubinstein et al., 2008; Xiao, Casey, Jameson, Curtsinger, & Mescher, 2009)). These conditions minimize variability in the exposure of individual cells to stimulatory signals, enabling cell-intrinsic lineage control mechanisms to be studied apart from environmental heterogeneity.

In this system, all cells begin dividing rapidly after 24 hours and upregulate the transmembrane glycoprotein CD44, indicating uniform activation (**Fig. 2.1B**). Activated cells downregulate *Tcf7* and the lymph node-homing adhesion molecule CD62L, consistent with effector differentiation. The inflammatory cytokines IL-12 and IFN-β1 enhance *Tcf7*-YFP silencing (**Fig. 2.1C**, **Fig. 2.2C-D**), consistent with their roles in driving effector differentiation (Danilo, Chennupati, Silva, Siegert, & Held, 2018; Joshi et al., 2007). When TCR stimulation (anti-CD3/CD28) and inflammation (IL-12) are removed to mimic pathogen clearance, the cells demonstrate a population-level increase in CD62L and *Tcf7*-YFP while continuing to divide, as

previously observed (Kaech et al., 2002). *Tcf7* and CD62L levels are heterogeneous both during stimulation and after removal, suggestive of an early memory and effector differentiation decision. YFP levels closely matched TCF1 protein levels throughout activation, validating use of the reporter in this system (**Fig. 2.2A-B**).

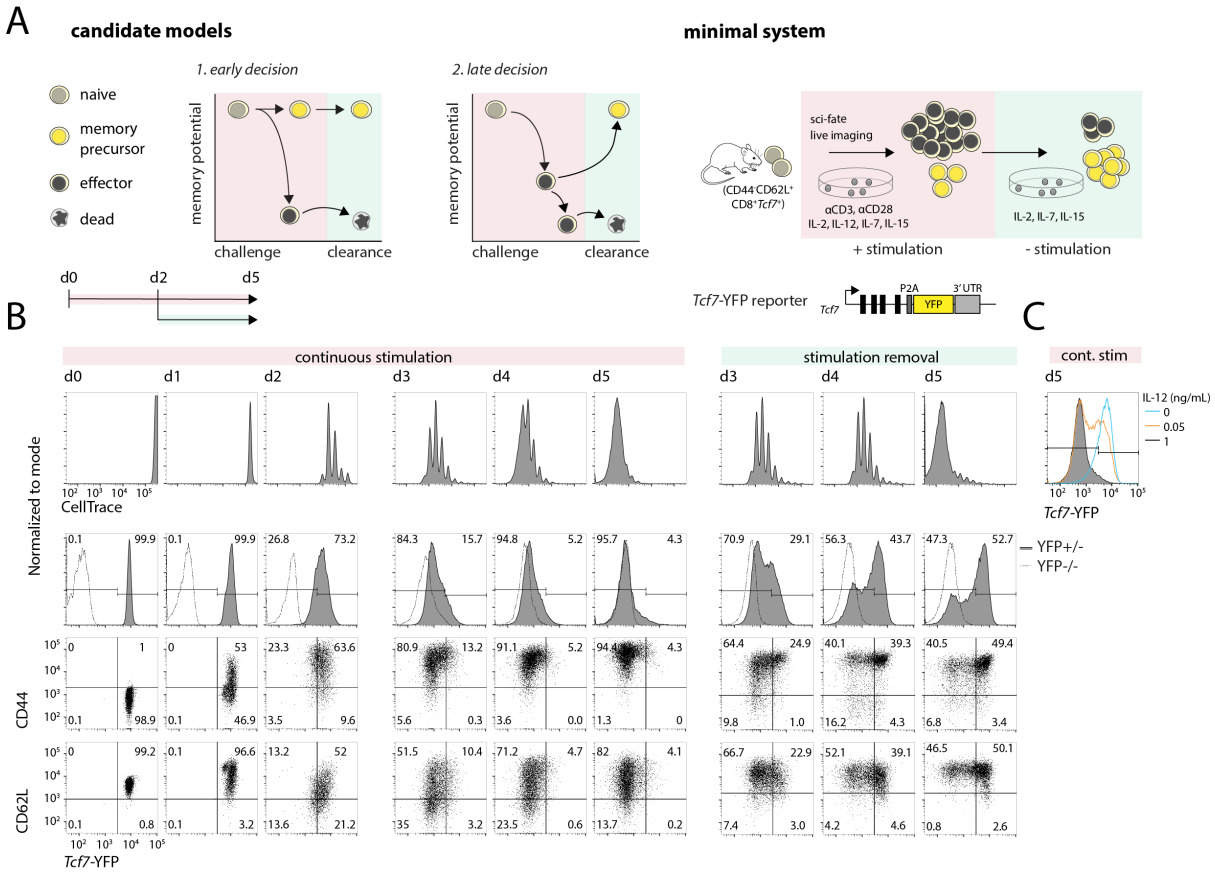


Figure 2.1: A minimal *ex vivo* system to track CD8 T cell effector and memory decision making dynamics. (A) Candidate decision-making strategies for CD8 T cell memory generation (left); a minimal *ex vivo* system for tracking memory decision-making dynamics at the single-cell level. (**B-C**) Naive CD8 T cells were isolated from *Tcf*7-YFP reporter mice, then cultured using this minimal *ex vivo* system. Flow cytometry plots show analysis of cultured CD8 T cells during initial stimulation for 2 days (left) and continued stimulation to day 5 (middle), or after stimulation withdrawal (removal of αCD3/αCD28 after day 2 and IL-12 after day 3) (right). (**C**) *Tcf*7-YFP silencing is tunable by IL-12 levels. Data are from a single experiment representative of at least 3 independent experiments.

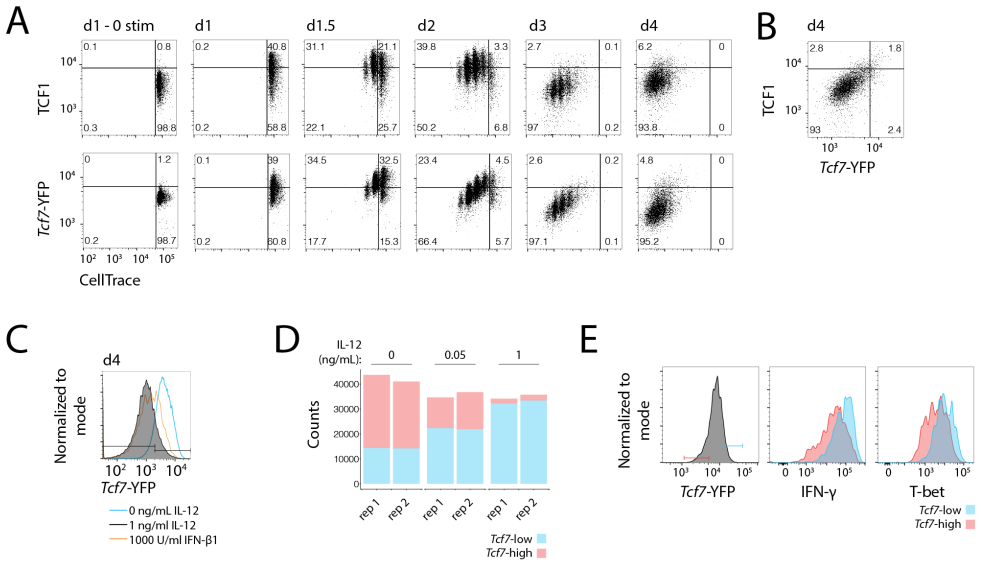


Figure 2.2: Characterization of the *Tcf7*-YFP reporter during *ex vivo* CD8 T cell activation. ***(*A-B**)* TCF1 and *Tcf7*-YFP correlation in cells activated with 1 ng/ml IL-12. Note, *Tcf7*-YFP dynamic range is reduced in fixed and permeabilized samples due to leakage of fluorescent protein out of permeabilized cells. (**C**) *Tcf7* silencing in cells stimulated with IFN-β1 or IL-12. (**D**) Total *Tcf7*-YFP high and low cell counts for samples activated for 4 days with different IL-12 levels. (**E**) IFN-ɣ and T-bet levels in cells stimulated for 2 days with IL-12. Data are from a single experiment representative of [A-B] 2 independent experiments, [C] 1 experiment for IFN-β1 and at least 3 for others, [D] at least 3 independent experiments, [E] 1 experiment.

2.3.2 NAIVE CELLS BIFURCATE EARLY INTO EFFECTORS AND MEMORY PRECURSORS

To determine whether the heterogeneity in *Tcf7* and CD62L downregulation reflects early memory and effector programming (**Fig. 2.1**), we analyzed *ex vivo* activated cells using the temporally-resolved single-cell transcriptome sequencing method, *sci-fate* (Cao, Zhou, Steemers, Trapnell, & Shendure, 2020). Here, metabolic labeling of newly-synthesized transcripts reveals a cell's current activity state apart from its history (Cao et al., 2020; Erhard et al., 2022) (**Fig. 2.3A**). We subjected cells at days 1, 2, and 4 to 4-thiouridine (4sU) pulse-labeling for 2 hrs, followed by sequencing and analysis as previously described (Cao et al., 2020). We obtained old and new transcriptomes

for ~17,000 single cells, with a median of 17,574 total and 2,529 new transcripts detected per cell (**Fig. 2.4A**). To disentangle effector and memory gene programs from other activation-induced programs, we performed an integrative analysis of our temporally-resolved transcriptome data and existing transcription factor (TF) binding data (He et al., 2016) to identify TF modules, consisting of co-regulated groups of TFs and their cognate target genes (see Methods). This analysis revealed two main TF modules, a cell cycle module and a T cell differentiation module, the latter further separable into submodules that included known regulators of effector and memory differentiation (**Fig. 2.3B**; **Fig. 2.4C**).

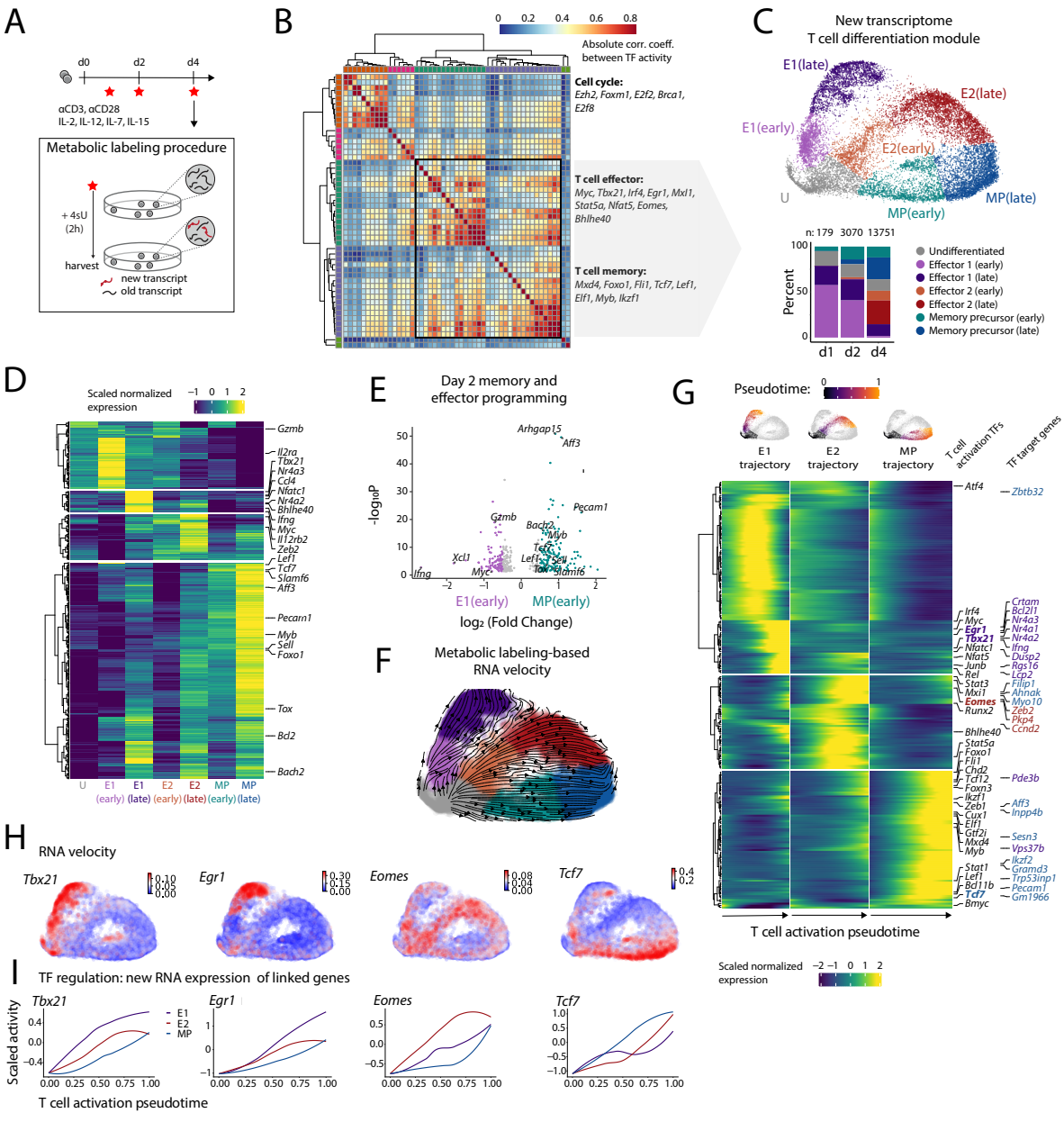

A
d0
d2
d4
αCD3, αCD28
IL-2, IL-12, IL-7, IL-15
Metabolic labeling procedure
+ 4sU
(2h)
harvest
new transcript
old transcript
B
Absolute corr. coeff.
between TF activity
Cell cycle:
Ezh2, Foxm1, E2f2, Brca1, E2f8
T cell effector:
Myc, Tbx21, Irf4, Egr1, Mxl1, Stat5a, Nfat5, Eomes, Bhlhe40
T cell memory:
Mxd4, Foxo1, Fli1, Tcf7, Lef1, Elf1, Myb, Ikzf1
C
New transcriptome
T cell differentiation module
E1(late)
E2(late)
E1(early)
E2(early)
U
MP(late)
MP(early)
n: 179 3070 13751
Percent
d1 d2 d4
Undifferentiated
Effector 1 (early)
Effector 1 (late)
Effector 2 (early)
Effector 2 (late)
Memory precursor (early)
Memory precursor (late)
D
Scaled normalized expression
Gzmb
Il2ra
Tbx21
Nr4a3
Ccl4
Nfatc1
Nr4a2
Bhlhe40
Ifng
Myc
Il12rb2
Zeb2
Lef1
Tcf7
Slamf6
Aff3
Pecam1
Myb
Sell
Foxo1
Tox
Bcl2
Bach2
U E1 E1 E2 E2 MP MP
(early) (late) (early) (late) (early) (late)
E
Day 2 memory and effector programming
Arhgap15
Aff3
Pecam1
Gzmb
Bach2
Myb
Tcf7
Xcl1
Lef1
Sell
Tox
Slamf6
Ifng
Myc
-log10P
E1(early)
MP(early)
log2 (Fold Change)
F
Metabolic labeling-based RNA velocity
G
Pseudotime:
E1 trajectory
E2 trajectory
MP trajectory
T cell activation TFs
TF target genes
Atf4
Zbtb32
Irf4
Myc
Egr1
Tbx21
Nfatc1
Nfat5
Junb
Rel
Stat3
Mxi1
Eomes
Runx2
Bhlhe40
Stat5a
Foxo1
Fli1
Chd2
Tcf12
Foxn3
Ikzf1
Zeb1
Cux1
Elf1
Gtf2i
Mxd4
Myb
Stat1
Lef1
Bcl11b
Tcf7
Bmyc
Crtam
Bcl2l1
Nr4a3
Nr4a1
Nr4a2
Ifng
Dusp2
Rgs16
Lcp2
Filip1
Ahnak
Myo10
Zeb2
Pkp4
Ccnd2
Pde3b
Aff3
Inpp4b
Sesn3
Vps37b
Ikzf2
Gramd3
Trp53inp1
Pecam1
Gm1966
T cell activation pseudotime
Scaled normalized expression
H
RNA velocity
Tbx21
Egr1
Eomes
Tcf7
I
TF regulation: new RNA expression of linked genes
Tbx21
Egr1
Eomes
Tcf7
Scaled activity
E1
E2
MP
T cell activation pseudotime

Figure 2.3: Naive cells diverge into effector and memory states early after activation. (**A**) Naive CD8 T cells were activated with αCD3 and αCD28 and 0.05 ng/mL IL-12, as well as IL-2, IL-7 and IL-15. After 1, 2, and 4 days, cells were treated with 4sU for 2 hours to label new transcripts, then harvested for time-resolved transcriptomics using *sci-fate*. (**B**) Heatmap showing the absolute Pearson's correlation coefficient between the activities of pairs of TFs, generated using *sci-fate*. Key TFs in each module are labeled at right. T cell differentiation module used for subsequent analysis is boxed. (**C**) UMAP visualization of cells based on the activity of T cell differentiation-related TF module, using newly synthesized mRNA, colored by cluster ID (top). Percentage of cells in each T cell activation state cluster after indicated days (bottom). (**D**) Aggregated expression (scaled, $\log_{10}$ normalized) of top 400 differentially expressed (DE) genes between clusters ($q < 3 \times 10^{-45}$ for all genes except for *Ifng*, $q = 7.3 \times 10^{-29}$). (**E**) DE genes between E1(early) and MP(early) at day 2 only; log2FC > 0.5 and adj. $p < 0.05$. (**F**) UMAP visualization as in (C), characterized by labeling-based RNA velocity analysis. Streamlines indicate the integration paths that connect local projections from the observed state to the extrapolated future state(*26*). (**G**) Pseudotemporal ordering of top 200 DE genes and additional genes of interest ($q < 1.4 \times 10^{-17}$) between trajectories. Gene labels correspond to all DE TFs in the T cell differentiation TF module (left text) and DE target genes linked to *Tbx21*, *Egr1*, *Eomes*, and *Tcf7* (right text). (**H**) RNA velocity magnitude and (**I**) Loess smoothed TF activity over pseudotime for four of the most DE genes between trajectories. TF activity is calculated as the normalized aggregation of levels of newly synthesized mRNA for all TF target genes, scaled across all cells. Cells in the undifferentiated (U) cluster are set to pseudotime = 0 for each trajectory.

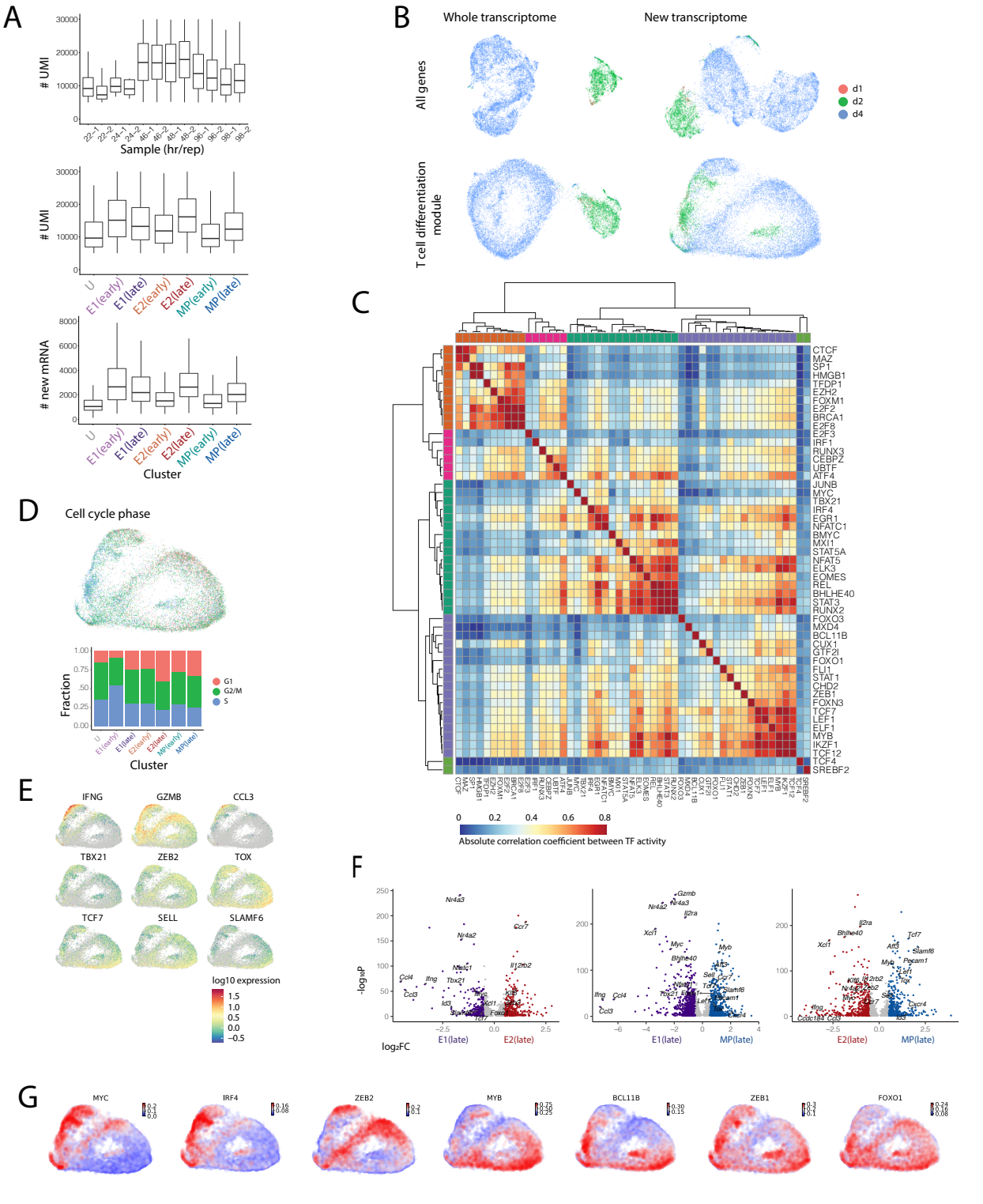

A
UMI
Sample (hr/rep)
UMI
new mRNA
U
E1(early)
E1(late)
E2(early)
E2(late)
MP(early)
MP(late)
Cluster
B
Whole transcriptome
New transcriptome
All genes
T cell differentiation module
d1
d2
d4
C
Absolute correlation coefficient between TF activity
D
Cell cycle phase
Fraction
G1
G2/M
S
Cluster
E
IFNG
GZMB
CCL3
TBX21
ZEB2
TOX
TCF7
SELL
SLAMF6
log10 expression
F
-log10P
log2FC
E1(late)
E2(late)
MP(late)
G
MYC
IRF4
ZEB2
MYB
BCL11B
ZEB1
FOXO1

Figure 2.4: *Sci-fate* metrics and TF module analysis. (**A**) Number of UMI per cell for each time point and replicate sample (top); number of UMI (middle) and newly synthesized reads (bottom) per cell for each cluster. (**B**) UMAP projections of cells using whole or new transcriptome and all genes or only T cell differentiation module genes. (**C**) Transcription factor module analysis as in Fig. 2.3B, enlarged to show direct TF correlations. (**D**) UMAP projection with cells colored by cell cycle phase and fraction of cells in each phase for all clusters. (**E**) Expression of representative genes in UMAP space. **(F)** Differentially expressed genes between indicated clusters, using cells from all timepoints; log2FC > 0.5 and adj. $p < 0.05$. (**G**) RNA velocity magnitude for additional effector and memory genes, similar to Fig. 2.3H.

By visualizing cell states using genes in the T cell differentiation module for Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction, we resolved distinct effector and memory states with coherence between timepoints (**Fig. 2.3C**; **Fig. 2.4B, D**). Unsupervised clustering and differential gene expression analysis revealed distinct early and late (A vs. B) effector (E1 and E2) and memory precursor (MP) states. E1 and E2 cells exhibited higher expression of the effector-associated genes *Gzmb, Ifng, Tbx21, Zeb2,* and *IL12rb2,* while MP cells had higher expression of the stem- and memory-associated factors *Bach2, Lef1, Tcf7, Sell*, and *Slamf6*, and lower expression of effector-associated genes (**Fig. 2.3D-E**; **Fig. 2.4E-F**; Supplementary Table 1-2) (Best et al., 2013). These differential gene expression patterns were present at day 2 and amplified at day 4.

Consistent with an early fate bifurcation, RNA velocity vectors calculated using reads from newly synthesized transcripts originate from the undifferentiated state (U), and flow along separate effector and memory branches (Qiu et al., 2020) (**Fig. 2.3F**). To gain insight into the dynamics of genes differentially regulated between divergent trajectories, we visualized their expression over pseudotime along each trajectory (**Fig. 2.3G**; Supplementary Table 3). This analysis, together with RNA velocity and TF activity analysis (**Fig. 2.3H-I**; **Fig. 2.4G**), identified effector and memory regulators with greatest differential regulation along their respective trajectories. *Tbx21, Egr1,* and *Irf4*, among other effector regulatory genes, were specifically active along the E1 trajectory, while

a distinct set of effector regulators, including *Eomes, Bhlhe40*, *Stat5a* and *Stat3*, characterize the E2 trajectory. This effector heterogeneity and its potential influence on downstream differentiation will be interesting to investigate in future studies but is not further pursued here. Finally, regulators of T cell stemness and survival, including *Tcf7, Myb, Mxd4* and *Fli1*, were active in the MP trajectory. *Tcf7* was the most significantly differentially expressed gene between trajectories, upregulated early along the MP trajectory and absent in both E1 and E2 trajectories. Its expression furthermore coincided with that of target genes identified through TF linkage that promote self-renewal, such as *Ikzf2*, *Sesn3*, *Aff3*, and *Pecam1* (CD31). Thus, *Tcf7* is a critical driver of this early divergent memory trajectory in our system.

2.3.3 THE EARLY EFFECTOR AND MEMORY DECISION OCCURS HETEROGENEOUSLY WITHIN CLONES

The divergence of cells into effector and memory lineages, occurring even under the strong, uniform stimulatory conditions of our *ex vivo* system, is suggestive of a cell-intrinsic regulatory mechanism involving *Tcf7* that generates heterogeneity in fate outcomes. To elucidate the degree to which this decision is heterogeneous within cell lineages amid constant environmental signals, we acquired multi-day time-lapse movies of clonal CD8 T cell lineages during activation with continuous measurement of *Tcf7* reporter levels (**Fig. 2.5**). As T cells are difficult to track with live imaging due to their high mobility, tendency to adhere to one another, and rapid proliferation, we optimized adhesion conditions and computational analyses that allow continuous tracking of a fate regulating TF across clonal CD8 T cell lineages (**Fig. 2.5**; **Fig. 2.6**; see Methods) (Nguyen et al., 2021). Using this method, we tracked 104 clonal lineages over 4 days and an average of 4.4 cell generations.

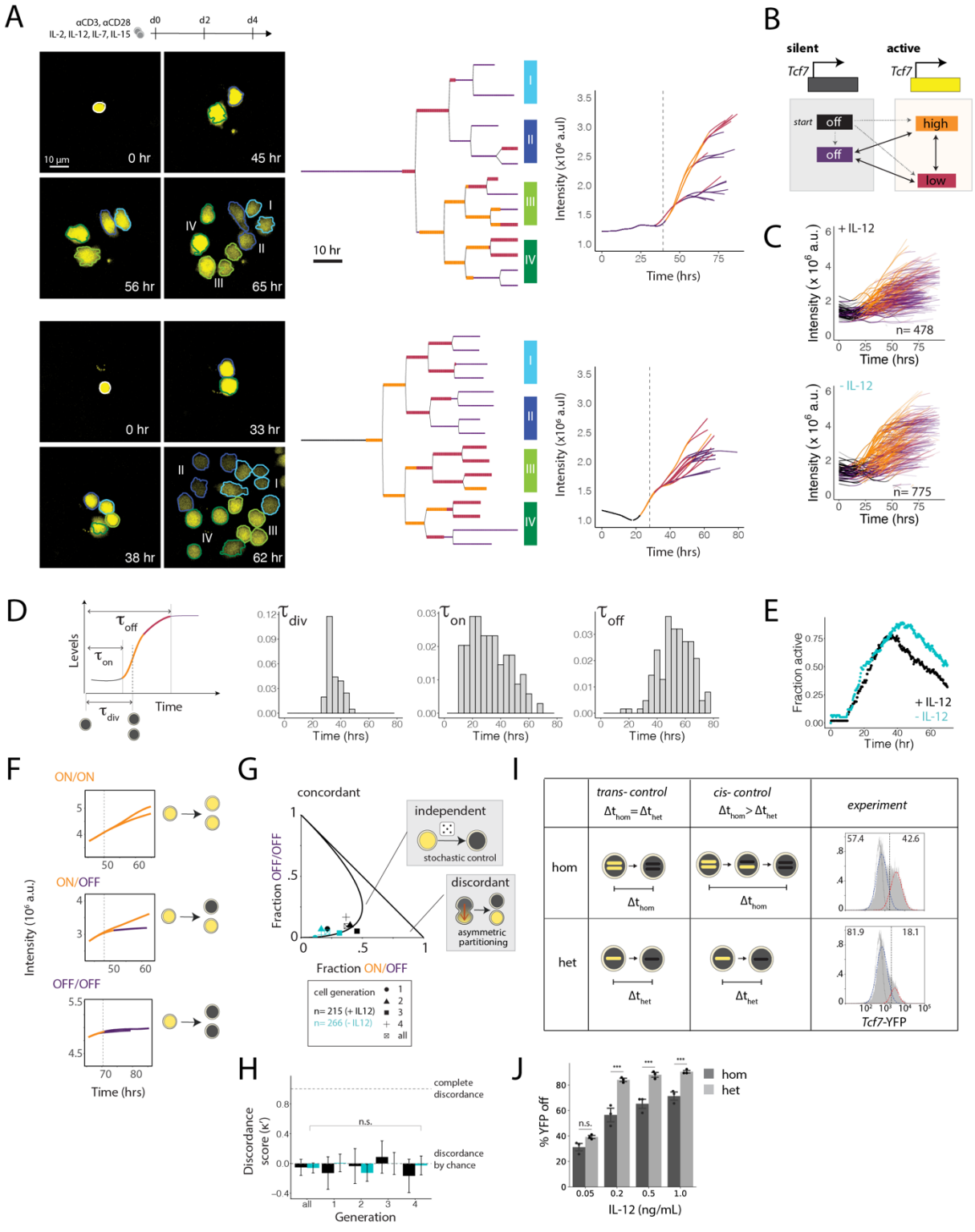

A
αCD3, αCD28
IL-2, IL-12, IL-7, IL-15
d0
d2
d4
10 μm
0 hr
45 hr
56 hr
65 hr
0 hr
33 hr
38 hr
62 hr
10 hr
Intensity (x10⁶ a.u)
Time (hrs)
B
silent
active
Tcf7
start
off
high
low
C
+ IL-12
- IL-12
Intensity (x 10⁶ a.u.)
n= 478
n= 775
Time (hrs)
D
Levels
Time
Time (hrs)
E
Fraction active
+ IL-12
- IL-12
Time (hr)
F
ON/ON
ON/OFF
OFF/OFF
Intensity (10⁶ a.u.)
Time (hrs)
G
concordant
independent
stochastic control
discordant
asymmetric partitioning
Fraction OFF/OFF
Fraction ON/OFF
cell generation
n= 215 (+ IL12)
n= 266 (- IL12)
H
complete discordance
discordance by chance
n.s.
Discordance score (K')
Generation
I
trans- control
cis- control
experiment
hom
het
Tcf7-YFP
J
% YFP off
IL-12 (ng/mL)
hom
het
n.s.

Figure 2.5: Heterogeneous *Tcf7* silencing within clones is controlled by a stochastic epigenetic switch. (**A top**) Naive CD8 T cells were activated with αCD3, αCD28, and IL-12, as well as IL-2, IL-7 and IL-15. IL-12 concentrations used are indicated for each subpanel below. (**A bottom**) Representative lineages demonstrating clonal heterogeneity in *Tcf7*-YFP silencing: image snap shots (left), lineage trees (middle), and reporter intensity (area x median YFP levels) over time for each track (right), with the first cell division marked by a vertical dashed line. Cell borders in snapshots are colored and labeled to match their corresponding leaves in the lineage trees. Lineage trees and tracks are colored by HMM-derived promoter state, outlined in (**B**). Cells are cultured with 1 ng/ml IL-12 unless otherwise indicated. (**C**) Reporter intensity for all overlaid tracks, colored by promoter state. (**D**) For each track, from left to right: time of first division, time of first transition to a stable active state, time of first transition to a stable silent state (stable state ≥ 10 hrs). (**E**) For all lineages combined, fraction of cells in an active promoter state over time, +/- 1 ng/ml IL-12. (**F-H**) Each division of a parent cell with the *Tcf7* promoter ON was categorized as giving rise to zero, one, or two daughters that transition to an OFF state. (**F**) Examples of each division category. (**G**) The OFF/OFF fraction by ON/OFF fraction is plotted separately for each generation, +/- IL-12, to distinguish concordant, independent, and asymmetric silencing mechanisms. (**H**) Modified Cohen's kappa test for division events in (G). (**I**) Comparison of YFP/YFP and YFP/+ reporters to distinguish *cis* and *trans* regulation of *Tcf7* silencing (left). YFP distributions for YFP/YFP and YFP/+ reporters cultured for 5 days with 0.2 ng/ml IL-12 (right). YFP off fractions are calculated from gaussian fits to distributions. (**J**) YFP off percentages as in (I), over a range of IL-12 concentrations. Mean ± s.d. Statistical significance was calculated with an unpaired two-tailed t test; n.s. $p=0.05$, ***$p<0.005$. Individual data points are from a single experiment representative of 2 independent experiments (I-J).

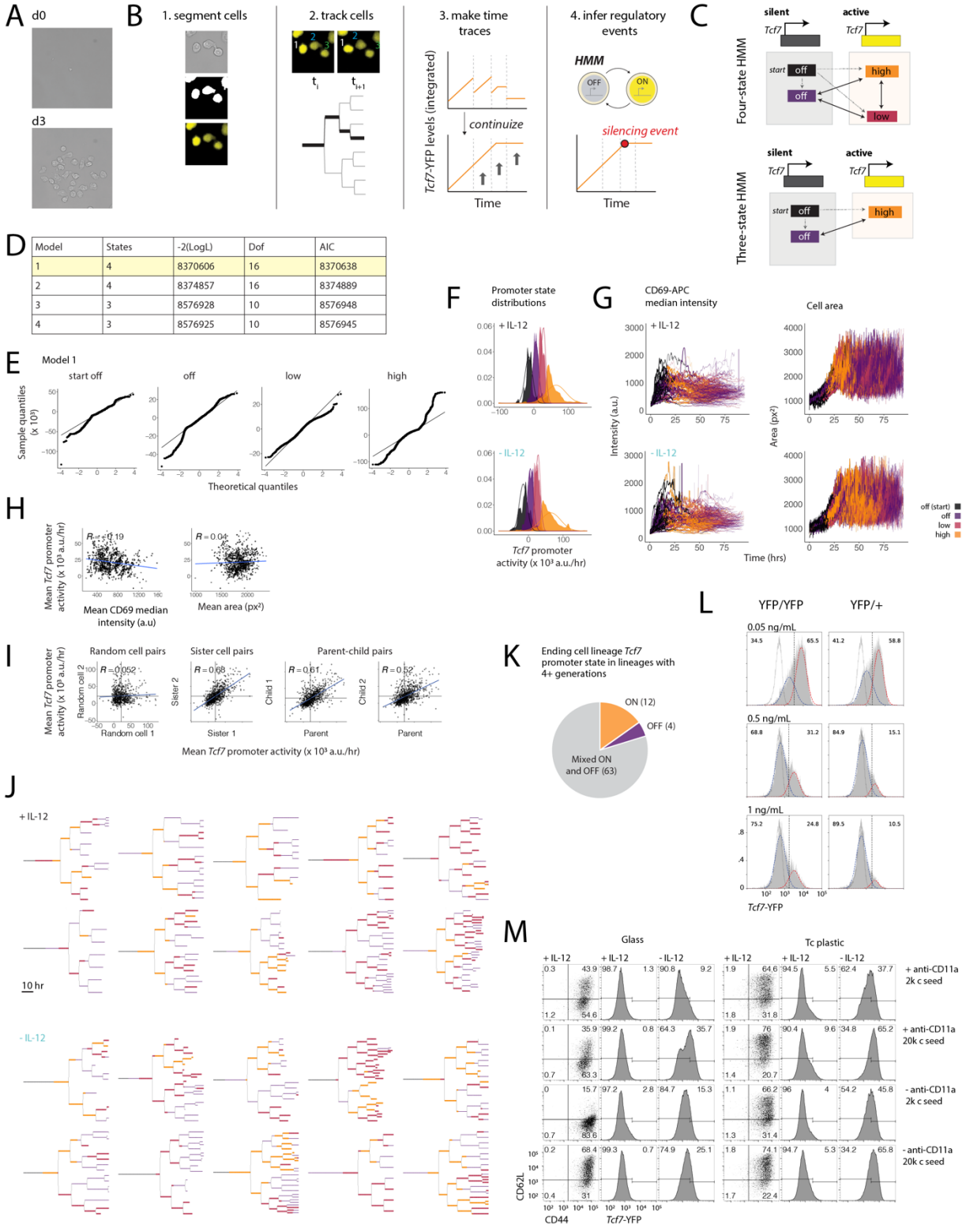

A
d0
d3
B
1. segment cells
2. track cells
3. make time traces
4. infer regulatory events
Tcf7-YFP levels (integrated)
continuize
Time
HMM
OFF
ON
silencing event
Time
C
silent
active
Tcf7
Four-state HMM
start
off
high
low
Three-state HMM
D
Model | States | -2(LogL) | Dof | AIC
1 | 4 | 8370606 | 16 | 8370638
2 | 4 | 8374857 | 16 | 8374889
3 | 3 | 8576928 | 10 | 8576948
4 | 3 | 8576925 | 10 | 8576945
E
Model 1
start off
off
low
high
Sample quantiles (x 10³)
Theoretical quantiles
F
Promoter state distributions
+ IL-12
- IL-12
Tcf7 promoter activity (x 10³ a.u./hr)
G
CD69-APC median intensity
Intensity (a.u.)
Cell area
Area (px²)
Time (hrs)
off (start)
off
low
high
H
Mean Tcf7 promoter activity (x 10³ a.u./hr)
R = 0.19
Mean CD69 median intensity (a.u)
R = 0.04
Mean area (px²)
I
Random cell pairs
Sister cell pairs
Parent-child pairs
R = 0.052
R = 0.68
R = 0.61
R = 0.52
Random cell 1
Random cell 2
Sister 1
Sister 2
Parent
Child 1
Child 2
Mean Tcf7 promoter activity (x 10³ a.u./hr)
J
+ IL-12
10 hr
- IL-12
K
Ending cell lineage Tcf7 promoter state in lineages with 4+ generations
ON (12)
OFF (4)
Mixed ON and OFF (63)
L
YFP/YFP
YFP/+
0.05 ng/mL
0.5 ng/mL
1 ng/mL
Tcf7-YFP
M
Glass
Tc plastic
+ IL-12
- IL-12
+ anti-CD11a 2k c seed
+ anti-CD11a 20k c seed
- anti-CD11a 2k c seed
- anti-CD11a 20k c seed
CD62L
CD44
Tcf7-YFP

Figure 2.6: Quantitative live imaging reveals dynamics of epigenetic Tcf7 silencing in clonal lineages. (**A**) Representative images of a single clone adhered to the imaging plate at day 0 and day 3. (**B**) Quantitative image analysis procedure. (**C**) Four and three-state HMMs were compared and (**D**) evaluated by log-likelihood and AIC. Models 2 and 3 compared to models 1 and 4 have tighter constraints on the promoter activity range of the off states. (**E**) For each state in the selected HMM (model 1), quantiles of the theoretical gaussian distribution versus the quantiles of the observed residuals. The observed residuals would fall on the straight line in a perfect fit[2]. (**F**) *Tcf7*-YFP derivative, or promoter activity, of cells assigned to each HMM-derived promoter state. Theoretical distributions for each are overlaid as lines. (**G**) Median CD69 intensity and cell mask area for all overlaid tracks, colored by promoter state. (**H**) Average *Tcf7* promoter activity versus average CD69 median intensity and area for each cell trace, where a cell trace is an ending cell tracked from its naive progenitor. (**I**) Promoter activity averaged over entire cell cycle for related cells. (**J**) Additional lineage trees, as in Fig. 3A. (**K**) Classification of lineages with cells tracked to at least 4 generations as having either all ON, all OFF, or mixed promoter states of terminal progeny. (**L**) YFP distributions for homozygous and heterozygous reporters cultured for 5 days. (**M**) CD62L x CD44 and *Tcf7*-YFP distributions for varied activation conditions after 4 days of stimulation. [L] Data are from a single experiment representative of 2 independent experiments. [M] Data are representative of one experiment.

Naive cells in these time-lapse movies start small, adhere to the antibody-bound plate, acquire CD69 expression, increase dramatically in size, and divide rapidly after 1-2 days (**Fig. 2.6G; Supplemental movie 1**). Strikingly, individual activating T cell clones often gave rise to *Tcf7* high and low subpopulations (**Fig. 2.5A**; **Fig. 2.6J, Supplementary Movie 1**), indicating that the effector and memory decision is made heterogeneously within clones. Of note, *Tcf7* low and high cells showed similar degrees of attachment to the surface, indicating that these intra-clonal differences are not due to differences in TCR stimulation, but more likely due to cell-intrinsic mechanisms generating heterogeneity in *Tcf7* silencing.

Differences in *Tcf7*-YFP levels after multiple cell divisions likely stem from earlier *Tcf7* silencing events propagated through dilution of the stable fluorescent protein by cell division. To pinpoint the timing of early regulatory events that give rise to these differences in *Tcf7*-YFP levels, we calculated the *Tcf7* promoter activity over time in single cells, defined as the rate at which total *Tcf7*-YFP levels increase over time, using a Hidden Markov Model (HMM) to assign *Tcf7*

promoter activity states to each cell at each timepoint and identify switching points between those states (**Fig. 2.5A-C**; **Fig. 2.6A-F**; see **Methods**).

This analysis revealed that cells silence *Tcf7* expression at variable times after the onset of stimulation, and can do so as early as the first cell division as well as at later generations. Cells activated the *Tcf7* promoter prior to the first cell division, reflecting exit from quiescence, and then proceeded to switch the *Tcf7* promoter to a silent state. The timing at which the *Tcf7* promoter transitioned to the silent state varied between cell tracks both within and between cell lineages, consistent with observed heterogeneity in *Tcf7*-YFP levels within clones (**Fig. 2.5A-D**). Lineages with variability in *Tcf7* promoter states in their terminal progeny were more prevalent than those having a uniform promoter state (either all ON or OFF), at least over timescales of observation, indicating that intraclonal heterogeneity in *Tcf7* regulation is the norm (**Fig. 2.6J-K**). Removing IL-12 increased the fraction of cells in an active promoter state (**Fig. 2.5C,E**). Silent *Tcf7* promoter states persisted across multiple cell divisions (**Fig. 2.5A**; **Fig. 2.6I-J**) and thus represent heritable regulatory changes as opposed to more transient dynamics such as transcriptional bursting. These results provide evidence that a cell-intrinsic *Tcf7* silencing event, occurring heterogeneously within clones, underlies the early divergence in effector and memory states.

2.3.4 A STOCHASTIC EPIGENETIC SWITCH CONTROLLING *TCF7* SILENCING UNDERLIES THE EARLY EFFECTOR AND MEMORY DECISION

Heterogeneity in *Tcf7* silencing could derive from asymmetric cell division (J. T. Chang et al., 2007; Verbist et al., 2016), whereby cell fate determinants partition unequally, giving rise to

discordant behavior between two sister cells. Alternatively, this heterogeneity could result from stochastic control (V. R. Buchholz et al., 2013; Duffy et al., 2012; Gerlach et al., 2013), whereby two sisters would make *Tcf7* silencing decisions independently. While two sister cells could still make different decisions, they would silence *Tcf7* discordantly no more frequently than expected by chance. To test these predictions, we analyzed the fractions of daughter cell pairs that silence *Tcf7* either discordantly (ON/OFF) or concordantly (OFF/OFF), doing so for cell pairs across all cell generations, with or without IL-12 (**Fig. 2.5F**). By plotting concordant (OFF/OFF) versus discordant (ON/OFF) sister pair fractions, we found that all data points adhered to a theoretical curve representing the expected relationship between sister pair fractions for independent regulation (**Fig. 2.5G**). Consistently, by statistical analysis using a modified Cohen's kappa coefficient (κ'), we found that daughter cells were no more likely to make discordant decisions than expected by chance (**Fig. 2.5H**; Supplementary Table 4). These findings support the view that *Tcf7* silences in a stochastic manner to drive divergent decision making within clones.

Epigenetic switching mechanisms, involving changes in chromatin modifications or conformation at gene loci, can introduce stochastic rate-limiting steps to gene activation or silencing (Bintu et al., 2016; Lövkvist et al., 2021; Pease et al., 2021). As *Tcf7* silencing involves repressive DNA or histone methylation (Danilo et al., 2018; Gray, Amezquita, Guan, Kleinstein, & Kaech, 2017; Ladle et al., 2016; Zhao et al., 2021), it could be gated by such a mechanism. Epigenetic mechanisms act in *cis* at individual gene loci and therefore would silence each *Tcf7* locus independently. To test for this mechanism, we compared *Tcf7*-YFP silencing kinetics in cells from mice homozygous and heterozygous for the reporter, with the prediction that homozygous reporter cells would yield a smaller population of YFP-low cells, since both loci need to silence for loss of reporter expression (**Fig. 2.5I-J**, **Fig. 2.6L**). Indeed, the *Tcf7*-YFP silent population was

significantly smaller in homozygous reporter cells and increased with IL-12, consistent with a *cis*-epigenetic silencing mechanism modulated by inflammation. Together, these results provide evidence that a stochastic *cis*-epigenetic switch, tunable by external stimuli, controls the early decision of naive cells to silence *Tcf7* expression and memory potential.

2.3.5 REVERSIBILITY OF *TCF7* SILENCING ENABLES A LATE MEMORY DECISION

Tcf7 silencing has been proposed to be an irreversible event that marks a 'point of no return' for effector differentiation and loss of memory potential (Lin et al., 2016; Pais Ferreira et al., 2020). Conversely, various studies have demonstrated that cells that acquire cytotoxic effector function are able to populate memory compartments after an infection is resolved (Bannard et al., 2009; Herndler-Brandstetter et al., 2018; Youngblood et al., 2017), suggesting that *Tcf7*-silenced effectors may still be able to reactivate *Tcf7* and reacquire memory potential. Our data thus far provide evidence for an early T cell decision to abandon or maintain memory potential, driven by stochasticity in antigen-driven *Tcf7* silencing, but do not exclude the possibility that effector cells can reverse their decisions and regain memory potential later after withdrawal of stimulation.

To test this possibility, we sorted *Tcf7*-YFP low and *Tcf7*-YFP high cells after initial culture and subjected them to reculture with variable stimulation conditions *ex vivo* (**Fig. 2.7A**). As expected, sorted *Tcf7*-YFP high cells maintained *Tcf7*-YFP expression without stimulation but underwent heterogeneous silencing under continuing stimulation (**Fig. 2.7B-C**, **Fig. 2.8A-B**). Furthermore, *Tcf7*-YFP low cells maintained a silent state upon continued stimulation, as observed. Strikingly, however, upon stimulation withdrawal, *Tcf7* reactivated, with the fraction of *Tcf7* expressing cells increasing over 6 days. *Tcf7* reactivation upon stimulation withdrawal

coincided with CD25 downregulation and CD62L upregulation, suggesting re-entry into a memory state (**Fig. 2.7D**).

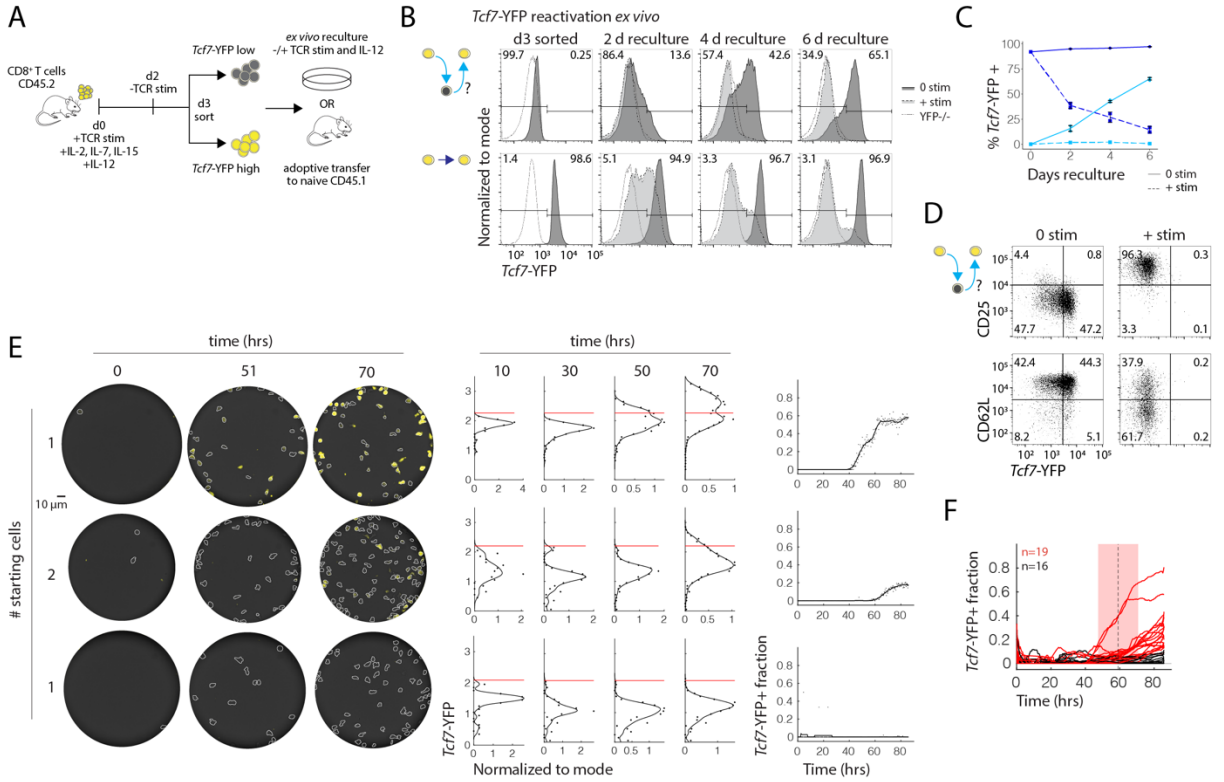


Figure 2.7: Effector cells reverse *Tcf7* silencing and regain memory potential upon stimulation withdrawal. (A) Naive cells from *Tcf7*-YFP mice were stimulated as indicated, sorted for *Tcf7*-YFP low and high populations after 3 days, and recultured either *ex vivo* (B-D, H-I) or adoptively transferred to naive recipients (E-G). Light and dark blue coloring throughout correspond to sorted YFP low and high populations, respectively. (**B-C**) *Tcf7*-YFP levels during reculture with or without αCD3/αCD28 and IL-12 (+/- stim) compared to non-fluorescent controls. (**D**) CD25, CD62L, and *Tcf7*-YFP expression in *Tcf7*-low sorted cells recultured +/- stimulation. (**E**) Representative microwells of sorted *Tcf7*-low cells recultured without stimulation: snap shots (left), top and bottom wells represent single clones; corresponding histograms (middle) with binned cell data for each time point, with YFP +/- gate drawn at 2 standard deviations above the mean YFP intensity from the first 25 hrs; corresponding YFP^+ fractions over time (right). (**F**) YFP^+ fraction for all wells overlaid. Mean activation time = 59.1 hr. [C, F] Mean ± s.d. [B-D] Data are from a single experiment representative of 1 and 3 independent experiments for +stim and 0 stim, respectively. [E-G] Data are from a single experiment with n=3-4 biological replicates.

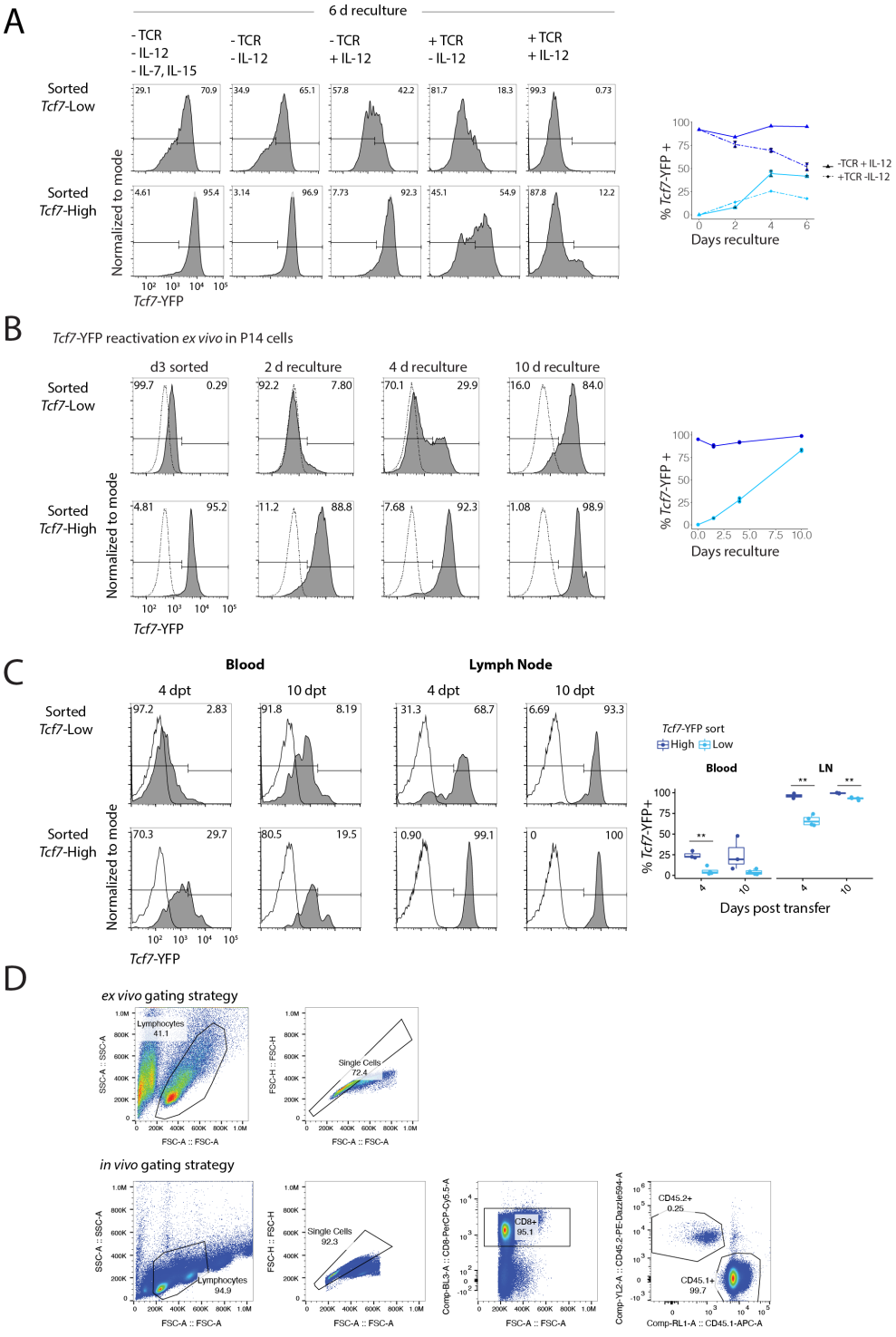

A
6 d reculture
- TCR
- IL-12
- IL-7, IL-15
- TCR
- IL-12
- TCR
+ IL-12
+ TCR
- IL-12
+ TCR
+ IL-12
Sorted Tcf7-Low
Sorted Tcf7-High
Normalized to mode
Tcf7-YFP
% Tcf7-YFP +
Days reculture
-TCR + IL-12
+TCR -IL-12
B
Tcf7-YFP reactivation ex vivo in P14 cells
d3 sorted
2 d reculture
4 d reculture
10 d reculture
C
Blood
Lymph Node
4 dpt
10 dpt
Tcf7-YFP sort
High
Low
LN
Days post transfer
D
ex vivo gating strategy
Lymphocytes
Single Cells
in vivo gating strategy
CD8+
CD45.2+
CD45.1+

Figure 2.8: *Tcf7* reactivation upon signal withdrawal. (**A**) Naive *Tcf7*-YFP low and high cells were sorted after 2 days of TCR stimulation with IL-12 and one day with removal of TCR stimulation. Sorted populations were recultured *ex vivo* under different conditions, with or without TCR stimulation and IL-12. All conditions are with continued IL-7 and IL-15 except where indicated. Blue and light blue represent NM and NEM samples, respectively. (**B**) *Tcf7*-YFPxP14 cells were sorted on day 3 as described for (**Fig. 2.7A**) and recultured *ex vivo* in the presence of IL-2, IL-7, and IL-15 for 10 days. (**C**) *Tcf7*-YFP levels in blood and inguinal lymph nodes at days 4 and 10 after transfer. Clear histograms represent CD45.1$^+$ host cells and filled histograms represent CD45.2$^+$ donor cells. Quantification of percentage of *Tcf7*-YFP$^+$ cells in each tissue at each time point. (**D**) Representative gating strategy for *ex vivo* and *in vivo* experiments. [A-B] Mean ± s.d., data are from a single experiment. [C] Data are from a single experiment with n=2-4 biological replicates. Statistical analysis performed using two-tailed unpaired t tests between *Tcf7* low and *Tcf7* high cells for each tissue at each time point. **p<0.01.

We next used clonal live imaging of sorted *Tcf7*-YFP low cells confined in microwells to test if *Tcf7* reactivation is heterogeneous within individual effector clones, as would be expected if reactivation occurs via reversal of stochastic *cis*-epigenetic silencing (**Fig. 2.5**). Consistent with reactivation observed from bulk starting populations, individual *Tcf7* silenced cells gave rise to *Tcf7* high cells (**Fig 1.7E-F**; **Supplementary Movies 2** and **3**; **Supplementary Table 5**). Similar to the initial *Tcf7* silencing event, reactivation was heterogeneous within clones. Reactivation occurred only after multiple divisions, which may reflect the need for cell division for permissive chromatin state changes. Overall, these results indicate that cells that have silenced *Tcf7* and relinquished memory potential can reverse this decision later, after resolution of an immune challenge.

2.3.6 *TCF7* HIGH CELLS FORMED THROUGH EARLY AND LATE DECISIONS ACQUIRE A COMMON MEMORY PROGRAM

Our results show that memory cells can arise through two pathways: a “naive to memory” (NM) pathway, whereby some cells maintain *Tcf7* expression during initial antigen stimulation, and a “naive to effector to memory” (NEM) pathway, by which cells that have silenced *Tcf7* and entered

an effector state can turn expression back on after stimulation removal. To determine whether *Tcf7* high cells emerging through these two pathways both have genomic and functional memory programs, we subjected them to transcriptomic, epigenomic, and cytokine secretion analysis, alongside control *in vivo* naive ($CD44^{-}CD62L^{+}$), memory ($CD44^{+}CD62L^{+}$), and *ex vivo* generated effector cells (**Fig. 2.9**).

Remarkably, NM and NEM cells showed similar memory characteristics, despite different *Tcf7* regulatory history. They were both more similar to naive and memory *in vivo* controls compared to *ex vivo* generated effector cells in their shared expression of memory-defining genes, though they also maintained some effector characteristics, in line with their recent stimulation (**Fig. 2.9A-C**; **Fig. 2.10A**). Similar to memory controls, NM and NEM cells demonstrated greater TNF-α and IFN-ɣ secretion upon re-stimulation compared to naive cells (**Fig. 2.9D-E**). NM and NEM cells were most similar in global chromatin accessibility to memory controls (**Fig. 2.9F**; **Fig. 2.10B-C**). NEM cells recovered similar *Tcf7* accessibility to NM cells (**Fig. 2.9G**). At the *Ifng* locus, intermediate accessibility of NM and NEM cells between naive and effector controls suggests that both were poised for rapid recall response, and accessibility at other memory- and effector- associated loci support this conclusion (**Fig. 2.10D**).

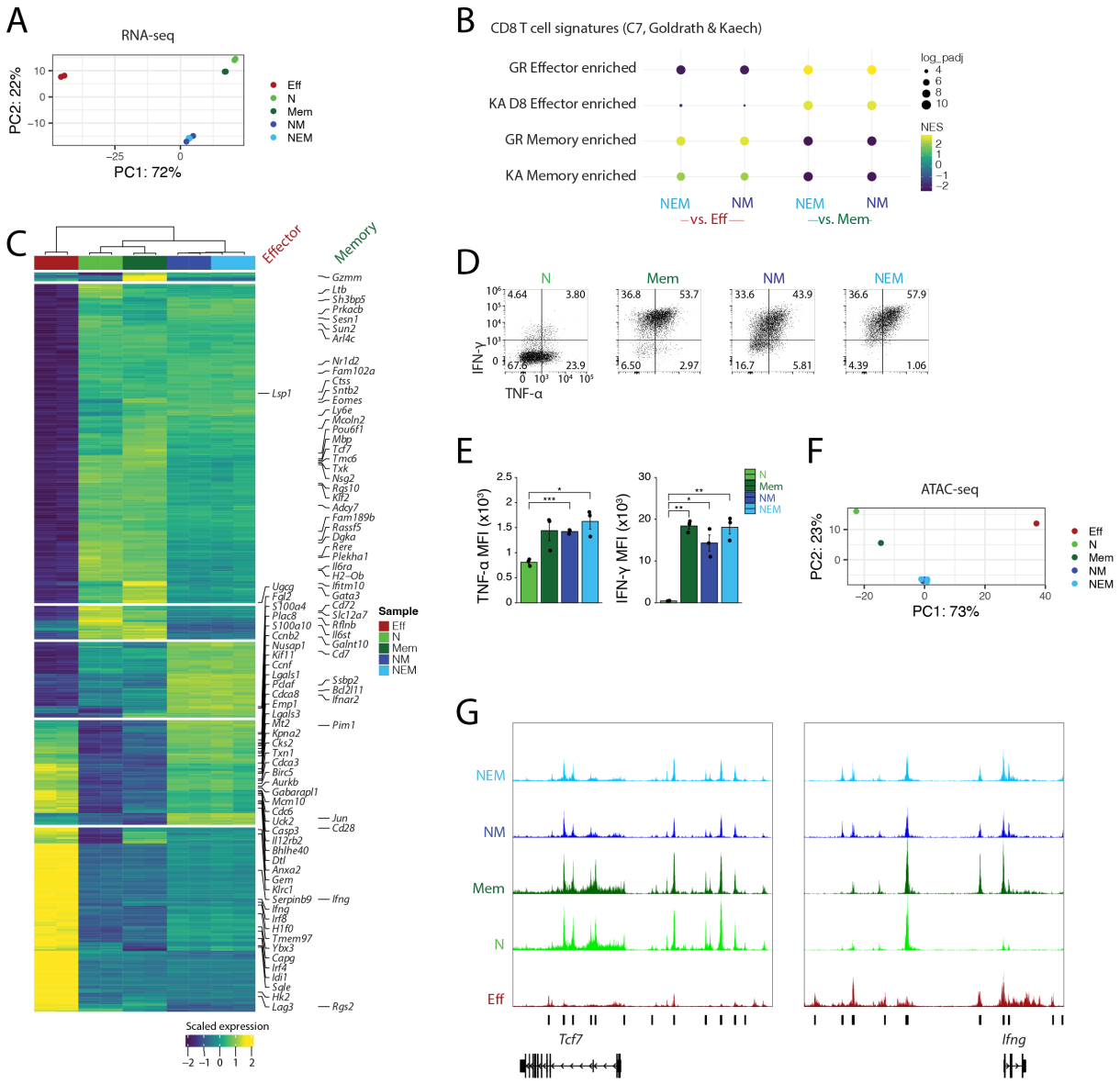


Figure 2.9: *Tcf7* high cells emerging from different routes acquire memory programming and functions. *Tcf7*-YFP low and high cells were sorted after 2 days of stimulation followed by one day of rest, recultured without TCR stimulation or IL-12 for an additional 6 days *ex vivo,* and then sorted for high *Tcf7*-YFP expression and subjected to genomic and functional analyses.
(**A**) PCA of RNA-seq profiles (top 500 DE genes) for recultured cells compared to day 3 effector (Eff) and day 0 naive ($CD44^{-}CD62L^{+}$, N) and memory ($CD44^{+}CD62L^{+}$, Mem) controls. NM and NEM cells were sorted as YFP-high and YFP-low on day 3, respectively. (**B**) GSEA of gene signatures from MSigDB (C7, collections deposited by Goldrath (GR) and Kaech (KA) comparing *ex vivo* recultured populations to Eff and Mem controls. (**C**) Heatmap displaying top 500 DE genes (lfc ≥ 2, Bonferroni-adjusted p value <0.05) between recultured populations and Eff, N, and Mem controls. Scale bar indicates row z-scores of regularized log transformed count data. Memory and

effector associated genes from MSigDB Goldrath and Kaech collections are highlighted. (**D-E**) Cytokine secretion of recultured cells compared to N and Mem controls after PMA/Ionomcyin restimulation. (**F**) PCA of ATAC-seq counts of top 500 differentially accessible peaks between recultured cells and controls. (**G**) ATAC-seq read coverage tracks; vertical bars annotate differentially accessible peaks between recultured cells and controls. [A-C] n = 2 biological replicates for each sample. [D-E] Mean ± s.d. Statistical significance was calculated with an unpaired two-tailed t test performed between groups. *p<0.5, **p<0.01, ***p<0.001. Data are n=3 biological replicates from a single experiment. [F-G] n = 1 biological replicate for Eff, N, Mem, n = 2 for NM, n = 3 for NEM.

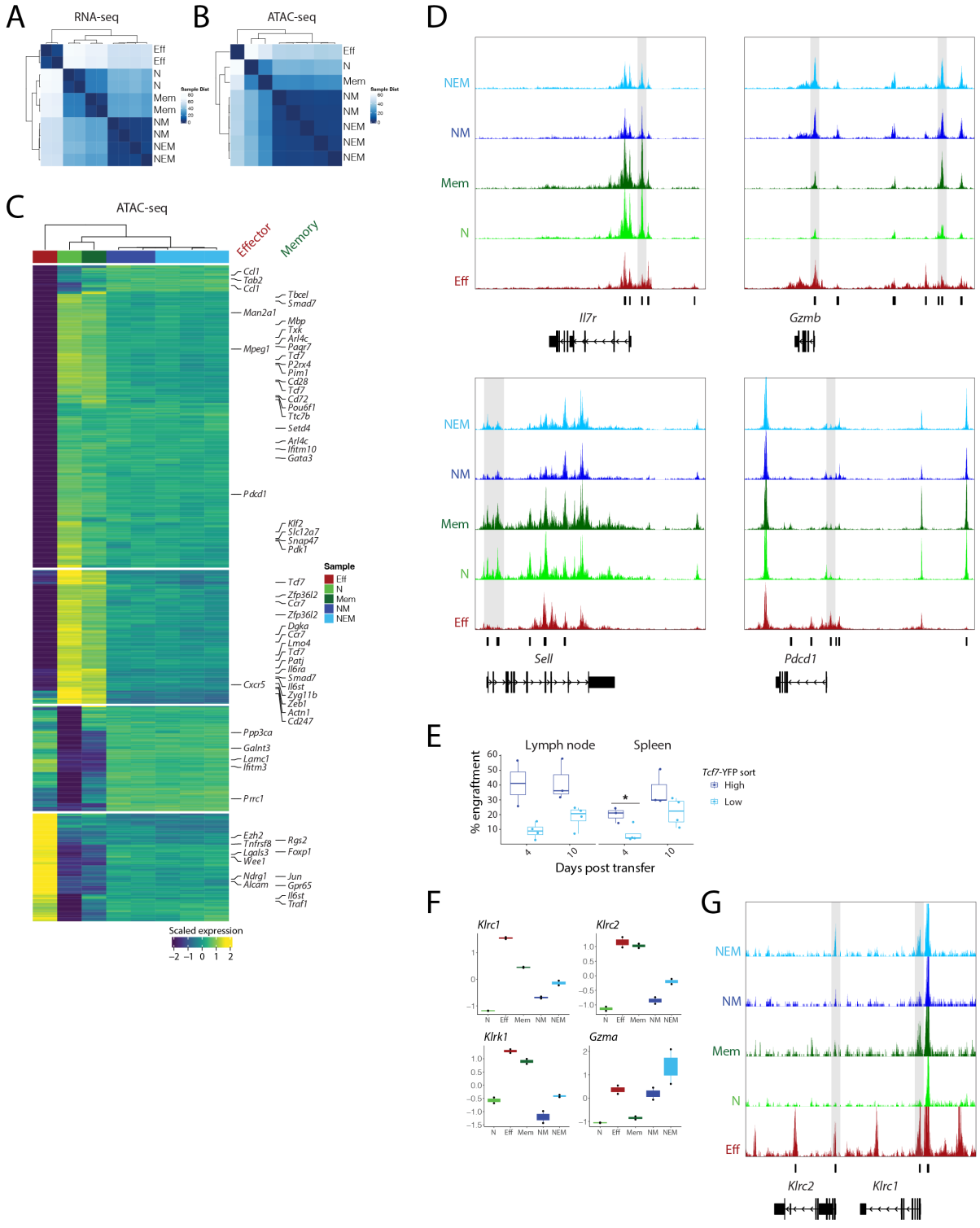

A
RNA-seq
Eff
Eff
N
N
Mem
Mem
NM
NM
NEM
NEM
B
ATAC-seq
Eff
N
Mem
NM
NM
NEM
NEM
NEM
C
ATAC-seq
Effector
Memory
Ccl1
Tab2
Ccl1
Tbcel
Smad7
Man2a1
Mbp
Txk
Arl4c
Pqgr7
Mpeg1
Tcf7
P2rx4
Pim1
Cd28
Tcf7
Cd72
Pou6f1
Ttc7b
Setd4
Arl4c
Ifitm10
Gata3
Pdcd1
Klf2
Slc12a7
Snap47
Pdk1
Sample
Eff
N
Mem
NM
NEM
Tcf7
Zfp36l2
Ccr7
Zfp36l2
Dgka
Ccr7
Lmo4
Tcf7
Patj
Il6ra
Smad7
Cxcr5
Il6st
Zyg11b
Zeb1
Actn1
Cd247
Ppp3ca
Galnt3
Lamc1
Ifitm3
Prrc1
Ezh2
Tnfrsf8
Lgals3
Wee1
Rgs2
Foxp1
Ndrg1
Alcam
Jun
Gpr65
Il6st
Traf1
Scaled expression
−2 −1 0 1 2
D
NEM
NM
Mem
N
Eff
Il7r
Gzmb
Sell
Pdcd1
E
Lymph node
Spleen
% engraftment
Tcf7-YFP sort
High
Low
Days post transfer
F
Klrc1
Klrc2
Klrk1
Gzma
N Eff Mem NM NEM
G
NEM
NM
Mem
N
Eff
Klrc2
Klrc1

Figure 2.10: Memory programming in cells that decide to maintain *Tcf7* expression early (NM) or reactivate *Tcf7* later (NEM). (A) Alternative representation to **Fig. 2.9A**: correlation matrix of RNA-seq profiles (top 500 DEG) for recultured cells compared to effector (Eff) and day 0 naive (N) and memory (Mem) controls. (B) Alternative representation to **Fig. 2.9F**: PCA of ATAC-seq counts of top 500 differential peaks between recultured cells and controls. (C) Heatmap displaying top 500 differentially accessible peaks (lfc ≥ 2, Bonferroni-adjusted p value < 0.05) between recultured populations and Eff, N, and Mem controls. Color legend indicates row z-scores of regularized log transformed count data. Memory and effector associated genes from MSigDB Goldrath and Kaech collections are highlighted in green and red, respectively. (**D**) ATAC-seq read coverage tracks. Vertical bars annotate differentially accessible peaks between recultured cells and controls, and grey shading highlights peaks of interest. (**E**) Engraftment of *Tcf7*-YFP sorted low and high cells in secondary lymphoid organs after adoptive transfer (see **Fig. 2.7A**), quantified as the percentage of transferred $CD45.2^+$ cells in each organ divided by the percentage of $CD45.2^+$ cells in the blood. (**F**) Scaled gene expression from bulk RNA-seq for DE genes (lfc ≥ 2, Bonferroni-adjusted p value < 0.05 except for *Gzma*: non-adjusted p = 0.0175) between D9 recultured NM and NEM cells (see **Fig. 2.9**). (**G**) ATAC-seq read coverage tracks. Vertical bars annotate differentially accessible peaks between recultured cells and controls, and grey shading highlights peaks of interest.[A, F] n = 2 biological replicates for each sample. [B-D, G] n = 1 biological replicate for Act, N, Mem, n = 2 for NM, n = 3 for NEM. [E] Data are from a single experiment with n=2-4 biological replicates. Statistical analysis performed using two-tailed unpaired t tests between *Tcf7* low and *Tcf7* high cells for each tissue at each time point. *p<0.05.

While NM and NEM cells were largely similar, notable differences in tissue localization and gene expression suggest they may be primed for different functional memory properties *in vivo*. When transferred to naive recipient mice, *Tcf7* high cells showed greater engraftment in secondary lymphoid organs than *Tcf7* low sorted cells, suggesting different homing capabilities (**Fig. 2.10E**). NEM cells also had higher expression and accessibility of some effector-associated genes compared to NM, possibly indicative of enhanced effector capabilities or an effector memory state (Herndler-Brandstetter et al., 2018; Pais Ferreira et al., 2020) (**Fig. 2.10F-G**). Overall, both NM and NEM decision strategies give rise to cells with genomic and functional characteristics of memory, suggesting that memory formation may proceed through a flexible decision-making strategy, allowing both for memory and effector divergence during the initial immune challenge and for effector reacquisition of memory potential after the challenge is resolved.

2.3.7 EARLY AND LATE MEMORY DECISIONS CAN BOTH LEAD TO CENTRAL MEMORY FORMATION IN RESPONSE TO ACUTE LCMV INFECTION

We next asked if the flexible memory decision making we observe in our *ex vivo* system also occurs in an acute infection *in vivo*. To test for presence of the early memory decision, we transferred splenocytes containing $1x10^6$ Cell Trace Violet labeled CD8+ T cells from *Tcf7*-YFP / P14 donor mice to CD45.1 recipient mice, which were then infected one day later with 2×10^6 p.f.u. of LCMV Armstrong (**Fig. 2.11A**). From flow cytometry analysis of splenocytes at 1, 2, and 3 days post-infection, we observed early heterogeneity in both TCF1 protein and *Tcf7*-YFP reporter levels at day 1, which developed into distinct TCF1 high and low populations at days 2 and 3 (**Fig. 2.11B**). Cells in both populations divided many times and upregulated PD-1 uniformly by day 2, demonstrating that the divergence in *Tcf7* regulation was not explained by variable exposure to antigen. These dynamics are consistent with the early heterogeneity in *Tcf7* silencing we observe *ex vivo* in response to homogenous strong signaling. To test if cells that have silenced *Tcf7 in vivo* are capable of reactivating it to make a late memory decision, we sorted *Tcf7*-YFP high and low cells three days post-infection and transferred $1x10^6$ cells each to CD45.1 recipient mice that had been infected five days prior with $2x10^5$ p.f.u of LCMV Armstrong (**Fig. 2.11C-D**). While the *Tcf7*-YFP low transferred population remained YFP low in mice harvested 2 days after secondary transfer, they became progressively YFP high at 9 and 17 days after transfer (**Fig. 2.11E**). This upregulation of *Tcf7 in vivo* coincided with upregulation of CD62L and CD127, consistent with reacquisition of a stem-like memory phenotype. The cells that had not reactivated

Tcf7 at days 9 and 17 were enriched in the KLRG1+CD127- population, suggesting that these cells retained more of an effector phenotype (**Fig. 2.11F**). Taken together, these results provide evidence for flexible memory decision making *in vivo*, whereby cells can decide early to maintain *Tcf7* expression or reactivate it later after initial silencing.

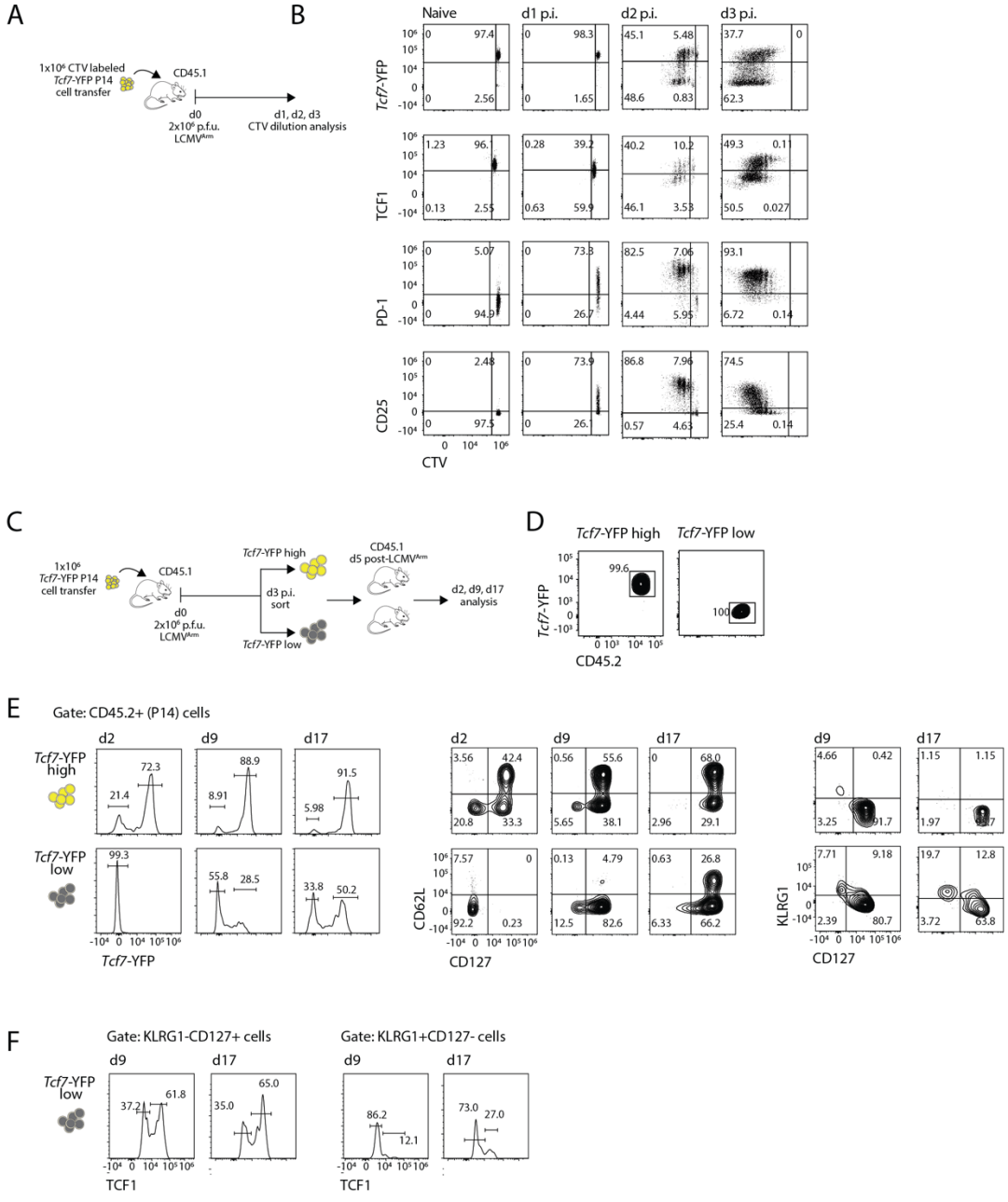


Figure 2.11: Flexible memory decision-making in acute LCMV infection. (A) Experiment design for early decision analysis. $1x10^6$ P14 cells from *Tcf7*-YFP reporter mice were transferred into congenically marked recipient B6 mice after Cell Trace Violet (CTV) staining. The recipient mice were infected with $2x10^6$ p.f.u of LCMV Armstrong. Phenotypical analysis of transferred CD8 cells was performed 24, 48, and 72h p.i. **(B)** CTV, YFP, TCF1, PD-1, and CD25 levels were analyzed at each time point. **(C)** Experiment design for late decision analysis. $1x10^6$ P14 cells from *Tcf7*-YFP reporter mice were transferred into congenically marked recipient B6 mice. The recipient mice were infected with $2x10^6$ p.f.u of LCMV Armstrong. On day 3 p.i., YFP high and low subsets were sort purified and re-transferred into congenically marked B6 recipient mice which were infected with $2x10^5$ p.f.u of LCMV Armstrong 5 days prior to re-transfer, with each mouse receiving $1x10^6$ sorted cells. Phenotypical analysis of the transferred cells was performed on day 2, 9 and 17 after re-transfer. **(D)** On day 3 p.i., splenocytes were isolated from the recipient mice and the P14 cells were sorted into YFP high and low subsets. (**E)** *Tcf7*-YFP, CD62L, KLRG1, and CD127 levels in each subset at each analysis time point, gated on P14 cells. (**F)** TCF1 levels in KLRG1-CD127+ and KLRG1+CD127- cells from the YFP low starting population at days 9 and 17.

2.3.8 MULTIPLE PATHS TO MEMORY ENABLE ROBUST ENCODING OF PATHOGEN EXPERIENCE THROUGH MEMORY POPULATION SIZE

Flexibility in memory decision-making may have functional benefits, and may in particular allow for scaling of memory population sizes with immune response magnitudes. To test this idea, we used mathematical modeling to evaluate different T cell decision-making strategies in their memory outcomes in response to pathogens of different virulence, modeled as having different rates of replication (see Mathematical Appendix). In our first model, we consider the flexible strategy we observed (**Fig. 2.12A**). Here, naive T cells (T_n) initially transition to a *Tcf7*-expressing memory-competent state (MC, T_m) that divides upon exposure to pathogen (v), but stops dividing and persists upon pathogen clearance. These cells can either maintain memory competence upon continuing stimulation, or transition to *Tcf7*-silent effector state (T_e), where they control pathogen growth, but are short-lived. Based on our findings (**Fig. 2.5**), this transition to an effector state is

stochastic, with a probabilistic rate that increases with pathogen. Effector cells can reverse *Tcf7* silencing and re-enter the memory-competent state in the absence of pathogen, as we observe (**Fig. 2.7**).

Mathematical simulations of this flexible decision model recapitulate the canonical features of the T cell response to acute infection (**Fig. 2.12B**; **Fig. 2.13A-B**). T cells expand rapidly in response to pathogen, reaching a peak 4-8 days after infection onset that consists mostly of effector cells, followed by a contraction to a stable, lower level of memory-competent cells (T_m). Consistent with known studies (S. Hou, Hyland, Ryan, Portner, & Doherty, 1994; Murali-Krishna et al., 1998), the quantity of memory cells is ~5% of the peak cell number.

In response to pathogens with varying replication rates, this flexible decision model allows memory cells to form robustly and scale linearly with peak cell expansion numbers. Increasing effector expansion with faster pathogen replication was accompanied by a proportional increase in memory cells, such that the memory fraction relative to the peak T cell number remained constant (**Fig. 2.12B and C** – top, yellow shading, γ_v>0.02/hr). This relation is given by:

$$f_{T_m} = \frac{\beta_{e,m}}{\beta_{e,m} + \delta_e}$$

where $\beta_{e,m}$ is the maximum effector to memory conversion rate and δ_e is the effector death rate. This scaling breaks down when pathogen replication is slow (γ_v<0.02/hr): reduced antigen encounter decreases the probability of the early effector cell decision, such that the number of memory cells generated converges to the starting naive cell number rather than increasing with pathogen replication rate. This ensures a baseline level of memory amid weak challenges that do not elicit a full effector response (Busch et al., 1998).

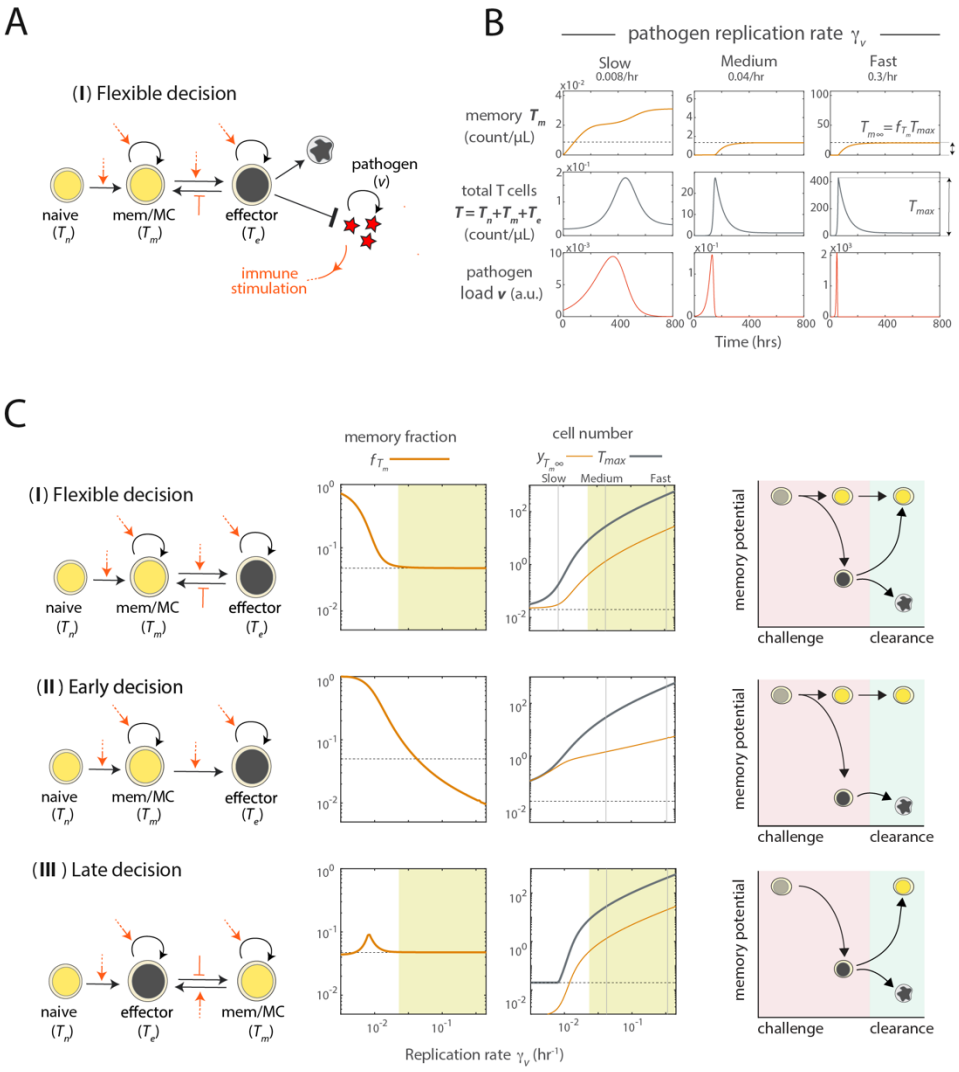
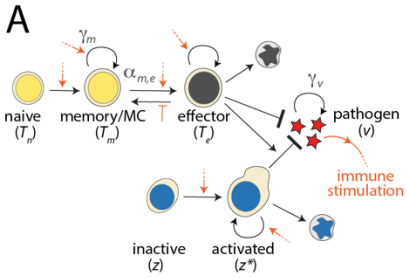
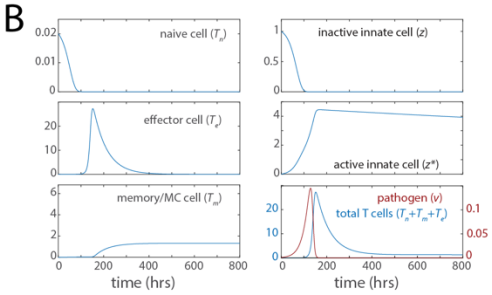
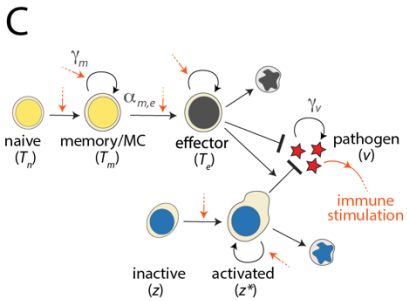
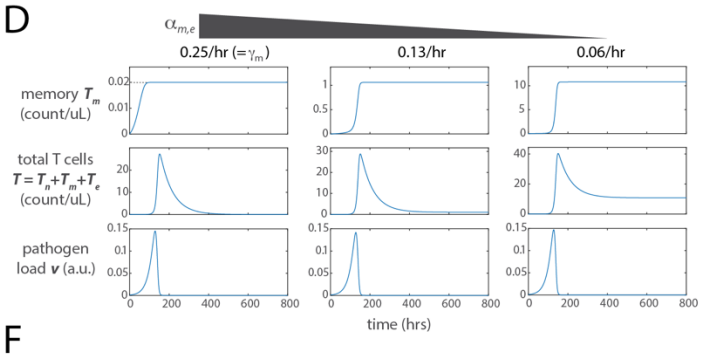


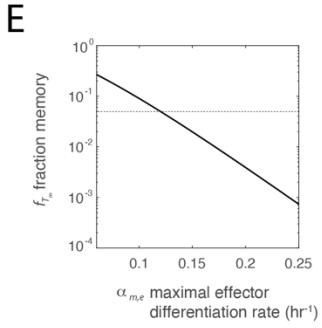
Figure 2.12: Flexible decision making enables quantitative encoding of pathogen experience during T cell memory formation. (A) Model incorporates pathogen proliferation, T cell memory decision making through reversible epigenetic switching. Orange arrows indicate modulation of T cell state transitions by pathogen load. (**B**) Time traces show memory T cell levels (top), total T cell levels (middle) and pathogen load (bottom), for different rates of pathogen replication (left to right). Dotted line shows the number of memory T cells formed in the case when this number is a defined fraction of the peak total T cell number, f_{Tm}. (**C**) Distinct strategies for memory decision making: flexible (top), early (middle) or late (bottom); the fraction of T cells at the response peak that become memory cells f_{Tm}; the peak cell number (black) and memory cell number (orange), both plotted against pathogen replication rate γ_v. The dotted line indicates the number of starting naive cells, and the yellow shading marks scalable memory.

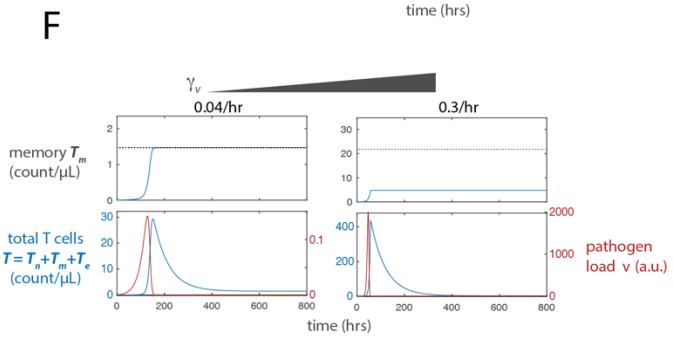

A
naive
(T_n)
memory/MC
(T_m)
effector
(T_e)
pathogen
(v)
immune
stimulation
inactive
(z)
activated
(z^*)

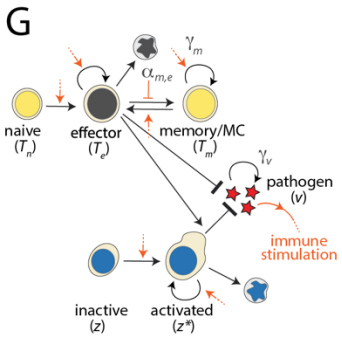

B
naive cell (T_n)
inactive innate cell (z)
effector cell (T_e)
active innate cell (z^*)
memory/MC cell (T_m)
pathogen (v)
total T cells (T_n+T_m+T_e)
time (hrs)
time (hrs)


C
naive
(T_n)
memory/MC
(T_m)
effector
(T_e)
pathogen
(v)
immune
stimulation
inactive
(z)
activated
(z^*)


D
0.25/hr (=γ_m)
0.13/hr
0.06/hr
memory T_m
(count/uL)
total T cells
$T=T_n+T_m+T_e$
(count/uL)
pathogen
load v (a.u.)
time (hrs)


E
f_{T_m} fraction memory
$\alpha_{m,e}$ maximal effector
differentiation rate (hr^{-1})


F
0.04/hr
0.3/hr
memory T_m
(count/μL)
total T cells
$T=T_n+T_m+T_e$
(count/μL)
pathogen
load v (a.u.)
time (hrs)


G
naive
(T_n)
effector
(T_e)
memory/MC
(T_m)
pathogen
(v)
immune
stimulation
inactive
(z)
activated
(z^*)

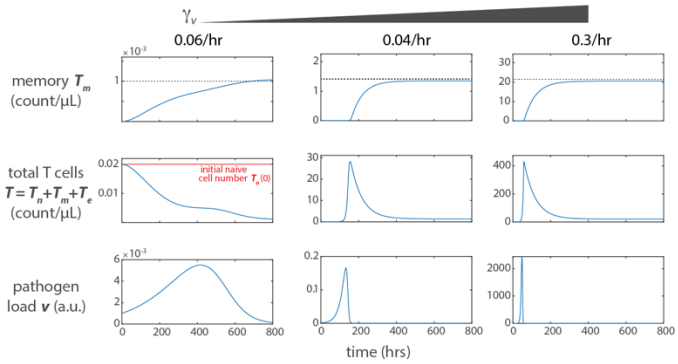

H
0.06/hr
0.04/hr
0.3/hr
memory T_m
(count/μL)
total T cells
$T=T_n+T_m+T_e$
(count/μL)
initial naive
cell number $T_n(0)$
pathogen
load v (a.u.)
time (hrs)

Figure 2.13: Immune response dynamics for flexible, early, and late decision models.
The flexible T cell decision making model (A-B) recapitulates the canonical immune response. **(A)** Schematic for the flexible decision model, including the rate constants for effector differentiation ($\alpha_{m,e}$), memory precursor proliferation (γ_m), and pathogen replication (γ_v). (**B**) Time traces show the time evolution of the levels of naive T cells (T_n), effector T cells (T_e), memory/MC T cells (T_m), inactive innate immune cells (z), active innate immune cells (z^*), and pathogen load (v). Following pathogen emergence, there is an increase in the numbers of activated innate immune cells, followed by an expansion in T cell number that occurs concomitantly with a decrease in pathogen load. Following pathogen clearance, there is a decline in T cell numbers to a stable level, reflecting a rise in the number of long-lived memory cells. *In the early (irreversible) decision model (C-F), the balance between effector differentiation and memory precursor proliferation sets the size of the memory population.* (**C**) Schematic for the early decision model, indicating the rate constants for effector differentiation ($\alpha_{m,e}$), memory precursor proliferation (γ_m), and pathogen replication (γ_v). (**D**) Simulation traces showing time evolution of memory cell numbers (top), total T cell numbers (middle), and pathogen load (bottom, v), obtained for different values of $\alpha_{m,e}$. Dotted line (top left) shows the initial number of naive cells in the system. Reducing the rate of effector differentiation relative to memory precursor proliferation increases the size of the resultant memory population. (**E**) Plot showing the fraction of cells at the peak of expansion that become memory cells, f_{Tm}, as a function of the effector differentiation rate $\alpha_{m,e}$. Dotted line shows a memory fraction $f_{Tm} = 0.05$, used to set the differentiation rate for comparison with the reversible switching model. (**F**) Simulation traces showing time evolution of memory cell numbers (top), total T cell numbers and pathogen load (bottom), for different pathogen replication rates. *A late (obligate reversible) decision strategy (G-H) enables scalable memory generation but fails to generate memory upon a mild challenge.* (**G**) Schematic for the late decision model, indicating the rate constants for effector differentiation ($\alpha_{m,e}$), memory precursor proliferation (γ_m), and pathogen replication (γ_v). (**H**) Simulation traces showing time evolution of memory cell numbers (top), total T cell numbers (middle), and pathogen load (bottom, v), obtained for different rates of pathogen replication γ_v. Dotted line (top) shows the final memory cell number number in the system. Red line (middle left) indicates the initial number of naive cells in the system. At low rates of pathogen proliferation, the number of memory cells formed are only a small fraction of the initial number of naive cells; the system is unable to store substantial amounts of memory in this regime. See Mathematical Appendix Table 2 for parameter values used for simulations of all models.

To ask whether flexibility is necessary for scalable memory encoding, we analyzed two alternative decision models, where memory decisions are made at only one juncture. The early decision model, where naive cells irreversibly commit to the *Tcf7*-silent effector state, generated robust memory upon challenge with slow-dividing pathogens but cannot reproduce the linear scaling of the memory population to the peak population in response to faster-replicating pathogens (**Fig. 2.12C**, middle; **Fig. 2.13C-F**; see also **Mathematical Appendix**). Conversely, the

late decision model, where naive cells transition obligatorily to the effector state and decide later whether to regain memory competence, generated constant memory fractions upon stronger challenges but attenuated memory populations in response to weaker challenges (**Fig. 2.12C**, bottom; **Fig. 2.13G-H**). These analyses underscore the importance of flexibility in memory decision making for optimal long-term immunity against variable threats.

2.4 DISCUSSION

Our finding that reversible epigenetic silencing of *Tcf7* generates inherent flexibility in the T cell memory decision reconciles two prevailing models for memory development that have often been regarded as mutually opposed. While there is evidence that memory cells can form both directly from naive cells with little or no effector differentiation and from effector cells that dedifferentiate upon infection clearance (Bannard et al., 2009; Youngblood et al., 2017), no model has explained how both pathways can coexist. In this mechanism, stochastic control of *Tcf7* silencing enables early divergent memory and effector decision making, and its reversibility enables late effector dedifferentiation. Antigen and inflammatory signals tune the decision-making probabilities at both junctures and would thereby influence which pathway would predominate across challenges that differ in signal duration and intensity (Plumlee et al., 2013). This study, together with others (Pease et al., 2021), implicate stochastic epigenetic switches as drivers of cellular diversification in the immune system. Through regulatory events that initiate over timescales spanning cell generations, these switches allow multiple cell populations to emerge in defined numbers without strict spatially-organized cues (Abadie, Pease, Wither, & Kueh, 2019), facilitating division of labor for optimal pathogen defense.

Our modeling results lay the groundwork for understanding how the adaptive immune system can encode information about the nature and severity of a pathogen in its memory cell

population (**Fig. 2.12**). In future work, it will be interesting to determine whether other pathogen features, such as antigenicity or latency, may also be encoded quantitatively. Our findings that memory cells emerging from different decision points may differ in their functional properties (**Fig. 2.10E-G**) raise the possibility that flexible decision making could underlie qualitative encoding of pathogen information through the generation of heterogeneous memory subsets (Herndler-Brandstetter et al., 2018; Pais Ferreira et al., 2020). In future work, it will be interesting to investigate the extent to which each decision pathway is utilized under various threats *in vivo* and whether cells emerging from different pathways are functionally heterogeneous (Wagner & Klein, 2020).

Overall, our study highlights the utility of plasticity in cell fate decision making in biological systems. From a strategic standpoint, flexibility enables decision makers to change their minds with new information, allowing them to mount optimal responses amid uncertain circumstances (Benjaafar et al., 1995). For immune cells responding to a pathogen, an ability to reassess prior decisions, as opposed to making early commitments, may enable bet-hedging and greater responsiveness as an immune challenge evolves. Observed plasticity in mammalian stem cell fate decision making (Weinreb & Klein, 2020; Xin, Gonzalez, Rompolas, & Greco, 2018) may similarly allow the body to rapidly adapt its regenerative output to changing physiological needs (Manz & Boettcher, 2014). A fuller consideration of flexibility in cellular decision making, along with its mechanisms and roles, will shed light into design principles of these systems and provide valuable insight for harnessing cells as environmentally-responsive therapeutic agents.

2.5 METHODS

Mice

Tcf7-YFP mice have been described (Harly et al., 2019). We note that a small number of experiments utilized mice harboring an additional non-perturbing *Tbx21*-CFP BAC transgene reporter allele (Zhu et al., 2012), though this reporter was not further analyzed for this study. All mice used in experiments were heterozygous for the *Tcf7*-YFP reporter except where specified. WT C57BL/6 mice (Jackson Laboratory) were utilized as reporter negative controls, where applicable. Both male and female mice were used for *ex vivo* experiments, aged 8 to 12 weeks. Female CD45.1 mice, 8-12 weeks of age, were purchased from the Jackson Laboratory for use as recipients for adoptive transfer experiments. For donors for adoptive transfer experiments, homozygous *Tcf7*-YFP mice were crossed with an LCMV specific TCR transgenic strain (Pircher, Bürki, Lang, Hengartner, & Zinkernagel, 1989b) (P14) (Jackson Laboratory) and heterozygous offspring were used. P14 homozygous mice without *Tcf7*-YFP were utilized as non-fluorescent controls for sort gate setting. All mice were used in accordance with Institutional Animal Care and Use Committee guidelines for the University of Washington.

Naive T cell extraction

Spleens were harvested from mice, massaged between rough glass slides to generate a single-cell suspension, and filtered through 40 μm nylon mesh into HBH (HBSS, 10 mM HEPES, 0.5% BSA, pH 7.4). Cells were spun down for 5 min at 300g, resuspended in 3 ml red blood cell (RBC) lysis buffer (150 mM NH_4Cl, 10 mM $NaHCO_3$, 1 mM EDTA) for 3-5 min, and quenched with HBH. Cells were spun down for 5 min at 300g and resuspended in HBH with 2.4G2 blocking solution and incubated for 30 min on ice. Cells were counted, spun down again, and then enriched for CD8

T cells using a CD8a$^+$ T Cell Isolation Kit, mouse (Miltenyi, #130-104-075), with the volume and amount of antibodies and microbeads used scaled down to 70% of that specified by the manufacturer. One LS column was used per spleen (Miltenyi, # 130-042-401). To obtain a pure population of naive CD8 T cells, the cell suspension was stained with anti-CD8 (PerCP/Cyanine5.5, eBioscience, # 45-0081-82 or Biolegend, #100734), anti-CD44 (APC or PE, Invitrogen, #17-0441-82, or #12-0331-82), and anti-CD62L (APC/eFluor780, Invitrogen, #47-0621-82) at 1:600 antibody to cell suspension volume ratio in 30x10^6 cell/ml HBH with Fc block for 15-30 min on ice and then sorted with a BD FACS Aria III (BD Biosciences) with assistance from the University of Washington Pathology Flow Cytometry Core Facility. The naive population was gated as CD8$^+$CD44$^-$CD62L$^+$*Tcf7*-YFP$^+$. Memory cells were gated as CD8$^+$CD44$^+$CD62L$^+$*Tcf7*-YFP$^+$. The cells were sorted into HBH and kept on ice until plating.

***Ex vivo* T cell differentiation**

One day prior to T cell harvest and activation (day -1), plates were prepared by coating with anti-CD3e (Tonbo, #40-0031-U100), anti-CD28 (Tonbo, #40-0281-U100), RetroNectin (Takara, #T100B), and when specified, anti-CD11a (Biolegend, #101117). Unless otherwise specified, each well of a 96-well plate received 0.2 µg anti-CD3, 0.1 µg anti-CD28, 1 µg Retronectin, and (when specified) 1 µg anti-CD11a in 50 µL of PBS. For differentiation in larger wells, these amounts were scaled up by well surface area. Plates were sealed with parafilm and incubated at 4°C overnight. On day 0, plates were allowed to come to room temperature for at least 30 min and washed 2x with PBS. Purified cells were added to wells in T cell media [85% RPMI 1640 with L-glutamine, 10% Fetal Bovine Serum, Pen-Strep-Glutamine, 20 mM HEPES, 1 mM Sodium Pyruvate, 0.1 mM NEAA, 50 µM BME] with indicated cytokine concentrations, mixed, and spun

down for 1 min at 150g to ensure initial contact for all cells with the coated plate surface. Cytokines added to the media were 100 U/mL IL-2 (PeproTech, # 200-02), 0.5 ng/mL IL-7 (PeproTech, # 200-07), 50 ng/mL IL-15 (PeproTech, # 210-15), and 1 ng/mL IL-12 (PeproTech, #210-12) unless otherwise specified. Where specified, IFN-β1 (Biolegend, #581302) was added at 1000 U/mL. The cell seeding concentration was 0.1 - 2.5 million cells / ml unless otherwise indicated. Cells were incubated at 37°C in 5% CO_2 and split every two days by mixing, removing half of the well volume, and topping off the volume with TCM and respective cytokines. Where applicable, prior to seeding, cells were stained with 5 μM CellTrace Violet (CTV) (Invitrogen, #C34557) following the manufacturer's instructions.

Flow cytometry analysis

For timecourse analyses with cell surface protein staining, cells were spun down in round-bottom 96-well plates or 1.5 ml eppendorf tubes, resuspended in 2.4G2 blocking solution for 15-30 min on ice, stained with cell surface antibodies at 1:1200 (anti-CD8: PerCP-Cyanine5.5, eBioscience, # 45-0081-82, or Biolegend, #100734, anti-CD44: APC, Invitrogen, #17-0441-82, anti-CD62L: APC-e780, Invitrogen #47-0621-82 , anti-CD25: APC, #17-0251-82), 1:400 (anti-CD45.1: APC, Biolegend, #110714), 1:200 (anti-CD45.2: PE/Dazzle594, Biolegend, #109846) or 1:100 (anti-CD127/IL7Rα: PE, Invitrogen, #12-1271-83) antibody to cell suspension volume ratio for an additional 15-30 min on ice, and spun down again for a final resuspension in HBH prior to acquisition. For samples that required intracellular protein staining, cells were fixed and permeabilized using Cytofix/Cytoperm Fixation and Permeabilization kit (BD #554714) according to manufacturer instructions and incubated with antibody for 30 min on ice. The TCF1 antibody

(PE, BD Biosciences, #564271) and T-bet antibody (PE, Biolegend, #644809) were used at 1:50 and 1:200, respectively. For samples that required intracellular cytokine staining, cells were restimulated for 5 hr with PMA/Ionomycin (1x in 100 µL per sample Thermofisher, #00-4970-93) in round-bottom 96-well plates, with a protein transport inhibitor (1x Thermofisher, #00-4980-93) added after 1 hr. For cytokine secretion after sorting (for Naive, Mem, and NM/NEM) cells were stained with Zombie Near IR at a 1:1000 dilution in PBS following the manufacturer's instructions (Biolegend, #423117). Cells were then fixed, permeabilized, and stained with antibodies for cytokine and other intracellular protein antibodies as described above. All cytokine antibodies were used at 1:100 dilution in 1x BD Perm/Wash buffer (anti-IFN-ɣ (APC/Cyanine7 or PE, Biolegend, #505849, #505808) and anti-TNF-α (BV711 Biolegend, #506349). Data were acquired using an Attune Nxt Flow Cytometer (ThermoFisher Scientific) and analyzed using FlowJo (BD) software.

Sample processing for sci-fate-seq

Naive CD8 T cells were activated *ex vivo*, as described. For this experiment, media was supplemented with 100 U/mL IL-2, 0.5 ng/mL IL-7, 50 ng/mL IL-15, and 0.05 ng/mL IL-12. The moderate level of IL-12, 0.05 ng/ml, was chosen for this experiment to produce a relatively even representation of *Tcf7* high and low cells (see **Fig. 2.1C**). At days 1, 2, and 4 of activation, two subsequent sci-fate time points were taken as follows: cells were mixed and split into two wells, which had been coated with anti-CD3 and anti-CD28 at day -1 and remained in the incubator with TCM; 4sU was added to one well for a final concentration of 200 µM, and that well was harvested 2 hr later. At that time, 4sU was similarly added to the second well, and that well was harvested 2 hr later. After each 4sU addition, cells were mixed and spun down at 150g for 1 min. Harvested

cells were prepared for sci-RNA-seq as described for the sci-fate protocol (Cao et al., 2020). Briefly, cells were fixed with ice-cold 4% PFA for 15 min, washed and flash frozen with PBSR [PBS, pH 7.4, 0.2 mg/ml bovine serum albumin (Fisher), 1% SuperRnaseIn (Thermofisher) and 10 mM dithiothreitol (DTT)]. PFA-fixed cells were thawed, washed, and treated with iodoacetamide (IAA) to attach a carboxyamidomethyl group to 4sU. Following these steps, a single-cell RNA sequencing library was prepared using the sci-RNA-seq protocol (Cao et al., 2018, 2017). The library was sequenced on the Illumina Next-seq system.

Computational analysis for sci-fate-seq

Read alignment, downstream processing, and TF module construction

Read alignment and downstream processing, linking of TFs to regulated genes, and construction of TF modules was performed as described in Cao et al., 2020, with minor modifications. Briefly, for each gene, across all cells, the correlation between mRNA levels of each expressed TF and that gene was computed using LASSO (least absolute shrinkage and selection operator) regression. We sought to comprehensively define gene programs with distinct dynamics by doing this correlation separately both using only newly synthesized transcript levels for potential target genes and using overall transcript levels, expecting that target genes with more stable transcripts would be more readily identified using newly synthesized transcripts, while less abundant, more lowly detected target genes would be more readily identified in the overall transcriptome. After filtering out the resultant covariance links with a correlation coefficient less than 0.03, we identified 2,117 putative TF - target gene covariance links using newly synthesized transcriptome levels and 9,927 using overall transcriptome levels, resulting in a total of 10,405 unique links after aggregation. These

were further filtered to retain only links supported by ChiP-seq binding, motif enrichment, or predicted enhancer binding (He et al., 2016), resulting in 1065 links between 51 TFs and 632 genes. Of these 1065 links, 147 were identified using the newly synthesized transcriptome levels, 649 were identified using the overall transcriptome levels, and 269 were identified by both. To calculate TF activity scores in each cell, newly synthesized unique molecular identifier (UMI) counts for all linked target genes were scaled by library size, log transformed, aggregated, and normalized. The absolute correlation coefficient was computed between all TF pairs with respect to their activity across all cells. Pairwise correlations were hierarchically clustered using the ward D2 method to identify TF modules, with the reasoning that co-regulatory TFs must be simultaneously active within the same cell.

Cell ordering, clustering, and differential gene expression analysis between clusters

We initially attempted to resolve T cell differentiation states by performing dimensionality reduction with Uniform Manifold Approximation and Projection (UMAP) on whole or new transcriptomes using all detected genes. This analysis largely separated cells by the time point at which they were sampled (**Fig. 2.4B**), as previously observed (Arsenio et al., 2014; Kurd et al., 2020), likely a consequence of the host of other temporal changes occurring during activation apart from differentiation, such as cell cycle control and metabolic programming. To characterize T cell differentiation dynamics apart from other regulatory processes, cells were represented in UMAP space using newly synthesized reads for all genes within the T cell differentiation TF module with monocle3 (v.0.2.3.0) (reduction_method = ‘UMAP’, umap.n_neighbors = 15L, umap.min_dist = 0.001) (Cao et al., 2019) using the function align_cds (Haghverdi, Lun, Morgan, & Marioni, 2018) to remove effects of cell cycle phase (preprocess_method = 'PCA', alignment_group = 'Phase').

The resultant UMAP was clustered using density peak clustering (Rodriguez & Laio, 2014), which resulted in 5 main clusters (**Fig. 2.3C**, U and E2 (A) combined, E1 (A), E1 (B), E2 (B), and MP (A) and MP (B) combined). To further resolve observed variable T cell differentiation marker expression within two of these clusters, *k*-means clustering was used to further divide U and E2 (A) into separate states and MP (A) and MP (B) into separate states ($k = 2$ and 2.5, respectively). Cells in different cell cycle phases were relatively evenly distributed across this UMAP, with S phase representation highest in E1 (A) (**Fig. 2.4D**). Differential gene expression testing was performed between clusters using the monocle3 fit_models function.

RNA velocity analysis

RNA velocity analysis and visualization of velocity streamlines was performed using Dynamo (v.0.95.2.dev) (Qiu et al., 2020) using expression matrices from the full and new transcriptome. The dataset was subsetted to include only the T cell differentiation module genes prior to analysis, but the resultant streamlines were similar when the analysis was performed with all genes. The streamline results were also similar when scVelo (v.0.2.2) (La Manno et al., 2018) was used for velocity analysis (data not shown), with the full and new transcriptome used as the unspliced/spliced expression matrices, indicating that the streamline results are consistent between multiple analysis methods. The scVelo results were also similar with or without subsetting to include only the T cell differentiation module genes.

Trajectory analysis

Cells in each putative trajectory (E1, E2, MP) were ordered in pseudotime based on the point position on the principal curve estimated using the princurve package (Hastie & Stuetzle, 1989).

To align the precursor cells between trajectories, cells in the undifferentiated (U) cluster were set to pseudotime = 0. To identify genes that distinguish the trajectories, differentially expressed genes were identified using the monocle3 fit_models function with the model formula as the trajectory and pseudotime terms. Only resulting DEG associated with the trajectory term were selected.

Time-lapse imaging

Long-term time-lapse imaging of cultured cells, both to track *Tcf7* regulation during initial activation in naive cells and to track *Tcf7* reactivation in sorted *Tcf7*-low cells, was performed as previously described with some modifications (H. Y. Kueh, Champhekar, Nutt, Elowitz, & Rothenberg, 2013; Hao Yuan Kueh et al., 2016). Images were acquired with an inverted widefield fluorescence microscope (Leica DMi8) fit with an incubator to maintain a constant humidified environment at 37°C and 5% CO2, using a 40X dry objective. For imaging of the initial 4 days of activation (**Fig. 2.5**), cells were seeded at low density (2-5k c/well) in wells of a 96-well glass bottom plate (Mattek) coated with anti-CD3, anti-CD28, anti-CD11a, and RetroNectin, as described above. For *Tcf7* reactivation imaging experiments (**Fig. 2.7**), *Tcf7*-low cells were sorted on day 3 after 2 days of initial culture with anti-CD3 and anti-CD28 in media with IL-2, IL-7, IL-15, and IL-12 and one additional day of culture with anti-CD3 and anti-CD28 removed. These cells were seeded onto PDMS micromesh (250 μm hole diameter, Microsurfaces) mounted on top of a 24-well glass bottom plate (Mattek) to enable clonal tracking, as seeded cells show considerably enhanced motility in the absence of TCR stimulation. To prepare the micromesh for imaging, the surface was first coated with BSA while mounted on top of a 24-well plate overnight at 4°C and then transferred to a new glass well and coated with anti-CD11a and RetroNectin for

improved adhesion but without anti-CD3 and anti-CD28. For reactivation experiments, cells were cultured in TCM with IL-2, IL-7, and IL-15, but without IL-12.

To determine if the experimental conditions required for imaging affect differentiation, we systematically compared expression of CD44, CD62L, and *Tcf7*-YFP in cells activated on glass or tissue culture plates, at high or low seeding density, and with or without presence of anti-CD11a (Fig. 2.6M). CD44 levels were comparable across all conditions, confirming that all cells activated in all conditions. In tissue culture plates, CD62L and *Tcf7*-YFP levels were also comparable, though the *Tcf7*-YFP levels were slightly reduced at lower cell density, particularly in the condition without IL-12, consistent with previous findings that memory differentiation occurs less efficiently at lower cell densities (Polonsky et al., 2018). On glass plates, the fraction of CD62L low cells was increased compared to on tissue culture plates. *Tcf7* levels were similarly low for the condition with IL-12, but the combination of low seeding density and presence of anti-CD11a on the glass plate resulted in a lower *Tcf7* distribution in the no IL-12 condition than was otherwise observed. This analysis shows that the specific conditions used for imaging do not affect overall differentiation trends but may underestimate the differences in differentiation between conditions with and without IL-12.

Computational analysis for time-lapse imaging

Image segmentation and tracking

Image pre-processing, cell segmentation, and tracking was performed in MATLAB (Mathworks, Natick, MA) using the ictrack movie analysis pipeline we described previously (Hao Yuan Kueh et al., 2016; Ng et al., 2018) (**Fig. 2.6A-B**), modified to enable segmentation of cells from

brightfield movies. Importantly, to segment cells without additional fluorescent labels besides *Tcf7*-YFP, we first trained a convolutional neural network (CNN) with a U-net architecture (Ounkomol, Seshamani, Maleckar, Collman, & Johnson, 2018) to predict fluorescence images of whole cells from brightfield images, using images of cell-trace violet labeled T cells as a training data set (Nguyen et al., 2021). We trained separate CNNs for the images acquired in 96-well plates (**Fig. 2.5**) and in microwells (**Fig. 2.7**), as predictions are optimal when images for training and prediction have similar features. For each training dataset, hundreds of images of CTV-stained cells were acquired at multiple timepoints during the process of interest (e.g. initial T cell activation or culture after stimulation removal). Using the trained CNN, we then generated predicted whole-cell fluorescence images from acquired brightfield movies, which were used for cell segmentation (**Fig. 2.6B, 1.).** Briefly, in the ictrack analysis pipeline, images underwent (1) correction by subtraction of uneven background signal stemming from the bottom of the glass plate or the side of the PDMS microwells (2) Gaussian blur followed by pixel value saturation to fix uneven signal intensity within the nucleus of the cell and (3) Laplacian edge detection algorithm to identify the nucleus boundary. Non-cell objects were excluded via size and shape limit exclusions. To generate clonal lineage trees, cells were tracked automatically between adjacent movie frames using the Munkres assignment algorithm, and the resulting cell tracks were manually checked for errors and to annotate cell divisions (**Fig. 2.6B, 2.**).

Tcf7 *promoter state assignment and analysis*

To enable quantitative analysis of *Tcf7* promoter activity in clonal cell lineages, we assembled separate full tracks of total *Tcf7*-YFP fluorescence levels from the starting cell to each ending cell within a lineage tree, for all lineage trees analyzed (**Fig. 2.6B, 3.**). Fluorescence levels are halved

at each cell division; thus, to ensure continuity in *Tcf7*-YFP levels in these tracks, we calculated for each parent-daughter cell pair an offset in *Tcf7*-YFP levels, that we added to the daughter cells and their progeny, as previously implemented (Bintu et al., 2016). These 'continuized' tracks were then smoothed using MATLAB medfilt1 (N=5) and smooth (span = 80 time points, equivalent to 20 hours, method = lowess), and their first derivatives with respect to time were calculated to generate single-cell tracks of *Tcf7* promoter activity for downstream HMM analysis (**Fig. 2.6B, 4.**).

Cell tracks were exported from MATLAB to R for downstream processing. *Tcf7* promoter states for each cell and time point were called from tracks of *Tcf7*-YFP level derivatives using Hidden Markov Model (HMM) modeling, implemented with the msm Package for R (v1.6.9) (Jackson, 2011). We initially tested four candidate HMM models with either three or four promoter states and variable constraints on the derivative ranges within each state (**Fig. 2.6C-D**). For each model, we constrained the mean and variance in *Tcf7* promoter activities of each state by fitting Gaussian distributions to the *Tcf7*-YFP derivatives at different time windows, to reflect our observations that cells are expected to be mostly in an inactive, active, or attenuated state at different times.

We then compared the performance of these four models by calculating their log-likelihood and corresponding AIC (Akaike information criterion) scores. We also checked the quality of each model's fit to the data by assessing whether residuals of the fit follow a Gaussian distribution (Lian et al., 2008) (**Fig. 2.6E**). Based on this analysis, we chose a model in which cells transition between 4 states: off (initial), low active, high active, and off (**Fig. 2.5B**, **Fig. 2.6F**), and all start in the off-initial state at the beginning of the track. This four-state model performed favorably compared to

other models, likely because it better accounts for the distinct distributions of promoter activity of silent and active cells at initial and later time points.

Using this four-state model, we assigned promoter activity states at each time point for each cell, removing potentially spurious transient promoter states by finding all promoter states lasting less than 8 hours and replacing them with the previously assigned promoter state. From these states, we then identified promoter silencing events as those involving a switch from active (high or low) to an inactive (off) state, and activation events as those involving a switch from inactive (off-initial or off) to active (high or low) states. We did not allow transitions back to the starting inactive (off-initial) state, as this state has a distinct *Tcf7* promoter activity distribution from the later silent state (off), likely reflecting the distinct noise characteristics of *Tcf7*-YFP levels at different stages after activation.

For analysis of *Tcf7* silencing between sister cells, we first assigned an ending cell state to all cells in the dataset, representing the final promoter state of the cell prior to division or the end of the cell track. Cells with a tracked duration of less than 3 hours and parents with ending cell state durations of less than 10 hours were also excluded, to ensure the analysis only includes sufficiently tracked cells and durable promoter states. We then collected all division events for which the parent cell was in an ON promoter state prior to division and asked whether the daughter cell tracks ended in an ON or OFF promoter state. We thus calculated the number of division events that lead to no (ON/ON), unequal (ON/OFF), or concordant (OFF/OFF) daughter silencing and then calculated the fractions of each category in the entire dataset and within each generation. To statistically analyze the degree of discordance in *Tcf7* silencing decisions between sister pairs by modifying Cohen's kappa statistical test for inter-rater reliability as follows: division events were categorized as concordant (ON/ON or OFF/OFF) or discordant (ON/OFF) between sisters.

The modified Cohen's kappa coefficient, κ', was calculated as the observed percentage of discordant events minus the percentage of discordant events expected by chance, divided by 1 minus the percentage of discordant events expected by chance (Cohen, 1960) (Supplementary Table 4).

Analysis of *Tcf7*-YFP negative fractions in homozygous and heterozygous reporter cells

For analysis in **Fig. 2.5I-J** and **Fig. 2.6L**, YFP distributions were exported from FlowJo as csvs, imported to Python, and represented as histograms. The positive and negative populations were fit simultaneously as two gaussian distributions using the scipy.optimize.least_squares function (scipy v1.5.2), and the gate between YFP positive and negative populations was identified as the intersection between the gaussian curves. The silent fraction was then calculated as the sum of the histogram below the gate divided by the sum of the entire histogram. Two-tailed unpaired t tests between homozygous and heterozygous YFP silent fractions were performed using scipy.stats.

Sort and adoptive transfer or *ex vivo* reculture of activated cells

Cells were cultured *ex vivo* in the presence of anti-CD3/28 (+TCR stim) and IL-2, IL-7, IL-12, and IL-15 as described. On day 2, cells were transferred to a non-antibody-coated plate (-TCR stim) but kept in the same cytokine environment until day 3 for sorting. For adoptive transfer experiments only, CD8 T cells were activated directly after purifying with the Miltenyi $CD8a^+$ T Cell Isolation Kit using 100% recommended reagent amounts, without further purifying naive starting cells by sorting, and RetroNectin was not used during anti-CD3/anti-CD28 stimulation. Prior to *ex vivo* activation, cells were stained with 2 or 5 μM CellTrace Violet (Invitrogen, #C34557). For *ex vivo* reculture experiments, cells were sorted from a single CellTrace peak

representing cells that had undergone the same number of divisions over the 3 day culture period, to ensure YFP differences were due entirely to *Tcf7* regulation differences and not cell division differences. Cells were recultured with and without TCR stimulation and IL-12 (with IL-2, IL-7, and IL-15 maintained except where specified), as labeled in each figure, for an additional 6-10 days. For genomics experiments, effector controls (Eff) were activated with TCR stimulation and cytokines for 3 full days. For adoptive transfer, cells were sorted that had undergone at least 4 divisions. Cells were sorted on CellTrace Violet and *Tcf7*-YFP levels. The *Tcf7* low gate was set using wild type non-fluorescent control cells that were stimulated identically *ex vivo.* Using this negative gate, the top and bottom 20% of the YFP distribution were selected as *Tcf7* high and low. Sorted cells were resuspended in PBS and transferred retro-orbitally (1 million cells transferred per recipient) to naive CD45.1 mice. On days 1.5, 4, and 10 after adoptive transfer, mice were euthanized, and blood, spleens, and lymph nodes were collected for flow cytometry.

Blood and lymph node processing

Blood was collected from euthanized mice by cardiac puncture. Red blood cells were lysed 2x for 5 minutes at room temperature using 1x RBC lysis buffer (described in naive T cell extraction), prior to proceeding with cell staining as described in Flow Cytometry Analysis. Inguinal lymph nodes were harvested, and massaged over a 40 μm cell strainer and resuspended for flow cytometry staining as described in Flow Cytometry Analysis.

Sample processing for RNA-seq

Cells were centrifuged at 500g for 5 minutes, resuspended in 350 μL of Trizol (Ambion), mixed well, and frozen at -80°C for processing, starting from step 2 of the RNeasy micro kit (Qiagen,

#74004) following the manufacturer's instructions. After processing, RNA was resuspended in RNase free water, quantified using a NanoDrop 2000c (Thermo Scientific), and shipped on dry ice to Novogene Corporation Inc. (Sacramento, CA) for library preparation and sequencing.

Computational analysis for RNA-seq

Raw FASTQ files from RNA-seq paired-end sequencing were aligned to the GRCm38/mm10 reference genome using Kallisto (v0.46.1) (Bray, Pimentel, Melsted, & Pachter, 2016), and the resultant transcript-level abundance estimates were imported to genes by cells matrices using tximport (v1.18.0) for downstream analysis. Transcripts with low counts (<10) were removed. Differentially expressed genes were identified with DESeq2 (v1.30.1) (Love, Huber, & Anders, 2014). PCA plots were generated using the top 500 differentially expressed genes between NM and NEM samples and naive, memory, and effector controls. Significantly differentially expressed genes were also used for gene set enrichment analysis, performed with fgsea (v1.16.0) (Korotkevich et al., 2021) and using gene sets from the C7 immunologic or the H Hallmark gene-sets from Molecular Signatures Database deposited by Goldrath and Kaech.

Sample processing for ATAC-seq

After sorting, cells were centrifuged at 500g for 5 minutes then supernatant was aspirated without disturbing the pellet. The pellets were resuspended in 100 µL of ATAC freezing buffer (Saunders, Core, Sutcliffe, Lis, & Ashe, 2013) (50 mM Tris at pH 8.0, 25% glycerol, 5 mM Mg $(OAc)_2$, 0.1 mM EDTA, 5 mM DTT, 1× protease inhibitor cocktail (Roche-noEDTA tablet), 1:2,500 superasin (Ambion)), flash frozen in liquid nitrogen and stored at -80°C. On the day of processing, samples were thawed, centrifuged at 4°C 500g for 5 minutes, and washed with 100 µL of cold 1X

PBS. Cells were again centrifuged and resuspended in 100 µL Omni lysis buffer (Corces et al., 2017) (RSB with 0.1% NP40, 0.1% Tween 20 and 0.01% Digitonin) and incubated on ice for 3 minutes, then quenched with 500 mL of RSB + 0.1% Tween 20. Nuclei were pelleted at 500g for 5 minutes at 4°C, resuspended in 100 µL cold PBS and counted. 50,000 nuclei were used per reaction, pelleted (500g for 5 min at 4°C), resuspended in tagmentation master mix (Corces et al., 2017) (50 µL total: 25 µL 2X TD buffer, 16.5 µL 1x DPBS, 0.5 µL 1% Digitonin, 0.5 µL 10% Tween 20, 5 µL water, 2.5 µL Tn5 enzyme), and incubated at 55°C for 30 minutes. Samples were purified using DNA Clean and Concentrate-5 (Zymo Research) and eluted in EB buffer (10 mM Tris) for amplification of tagmented DNA. PCR amplifications were performed using Illumina indexed primers and NEBNext High-Fidelity 2X PCR Master Mix. SYBR green was added to each PCR reaction to monitor amplification before it reached saturation. Samples in this study were amplified between 11-15 cycles using recommended conditions (Cusanovich et al., 2018). Unpurified products were run on a 6% TBE gel for quality control. PCR product/library were purified using DNA Clean and Concentrate-5 (Zymo Research) then ran on a tapestation to visualize nucleosome distribution. The libraries were normalized to 2nM then pooled equimolar for sequencing. Pooled libraries were loaded onto a NextSeq 500 High150 cycle kit at 1.5 pM loading concentration with paired ends sequencing (read 1: 74 cycles, read 2: 74 cycles, index 1: 10 cycles, index 2: 10 cycles).

Computational analysis for ATAC-seq

Raw ATAC-seq FASTQ files from paired-end sequencing were processed and aligned to the mm10 mouse genome using the PEPATAC (v0.10.3) (Smith et al., 2021) pipeline, which uses bowtie2 (Langmead & Salzberg, 2012) for alignment. Unmapped, unpaired, and mitochondrial

reads were removed. Following alignment, peak calling, merging across all samples, and annotation was performed using HOMER (v4.10) (Heinz et al., 2010). Differentially accessible regions were identified using DESeq2. PCA plots were generated using the top 500 differentially accessible regions between recultured samples and naive, memory, and effector controls. Coverage tracks were generated from bigwig read alignment files using karyoploteR (v1.14.1).

Statistical Analysis

All analyses and p or adjusted p value significance are listed with each figure caption. Statistics were performed in R using the rstatix package (v0.7.0) or Python using scipy (v1.5.2).

Chapter 3. REGULATION OF T CELL STEMNESS UNDER CHRONIC STIMULATION

3.1 INTRODUCTION

Activity of cytotoxic T cells is a critical component of the immune response to chronic infection and cancer. However, under the chronic stimulation conditions that define these challenges, T cells transition to an exhausted or dysfunctional state that loses target cell killing capabilities (Wherry et al., 2007). Therefore, shifting the balance between functional and dysfunctional T cell states in chronic infection and cancer represents a major therapeutic opportunity for improving patient outcomes in these diseases. To realize this opportunity requires understanding the transitions between these functional and dysfunctional states and the signals and regulatory mechanisms that control them.

Recent studies demonstrate that self-renewing TCF1+ stem-like CD8 T cells fuel persistent T cell responses in chronic viral infections and cancer (Hudson et al., 2019; Sade-Feldman et al., 2018; Utzschneider et al., 2016). Similar to the role of TCF1+ naive or memory cells in the context of acute infection, the TCF1+ cells present during chronic infection and in tumors can differentiate into functional cytotoxic cells that mediate target killing. Indeed, multiple studies have shown that it is these TCF1+ cells that provide a proliferative burst of cytotoxic cells in response to PD-1 blockade immunotherapy (Im et al., 2016). Despite the importance of TCF1+ stem-like cells to persistent T cell responses, it remains unclear how this population is maintained under chronic stimulation. As the persistent signaling present in chronic challenges generally drives TCF1 silencing and effector differentiation, there must exist some mechanism by which TCF1+ cells are protected from differentiation or replenished from additional sources. One possible mechanism for maintaining the TCF1+ population is continual migration of stem-like cells from lymph nodes to the site of infection or tumor. As methods to track T cell clonality and migration in vivo through TCR sequencing and cell labeling have become more facile, multiple groups have revealed evidence for TCF1+ T cell trafficking from lymph nodes to tumor tissues to fuel persistent responses (Connolly et al., 2021; Li et al., 2022; Pai et al., 2023). Another possible mechanism for maintaining a pool of TCF1+ cells is by reactivation of TCF1 in cells that have silenced it and further differentiated. While the role of lymph node extravasation in maintaining the TCF1+ population in cancer is becoming clear, whether or not tumor infiltrating T cells that have silenced TCF1 and lost self-renewal potential are capable of reactivating TCF1 and replenishing the stem-like pool is still an open question.

In acute infection, as described in **Chapter 2**, we discovered that activated T cells have a remarkable degree of plasticity in their control of self-renewal. Cells can silence TCF1 and lose

self-renewal potential initially in response to TCR stimulation and inflammation during an acute challenge but are able to reactivate it and regain self-renewal potential upon stimulation withdrawal, when the challenge is resolved. Thus, T cells have multiple opportunities to decide whether to maintain or lose self-renewal potential as an immune challenge develops, enabling robust generation of effector and memory populations across dynamic challenges with ranging pathogenicity. This flexible system is enabled by stochastic epigenetic silencing of TCF1, whereby probabilistic silencing of TCF1 initially in response to stimulation enables early memory and effector heterogeneity, while reversibility of the TCF1 silent state enables later reacquisition of self-renewal and memory potential in effector populations.

These results implicate plasticity of the TCF1 silent state as a critical feature of robust immunity to acute challenge. However, it is unclear whether cells can similarly maintain this plasticity and TCF1 reactivation potential upon prolonged stimulation characteristic of chronic infection and cancer. We therefore sought to ask whether cells stimulated for longer durations maintain this plasticity indefinitely, or if they eventually lock down the TCF1 silent state permanently. If cells do progress from an initially plastic to a terminally silent TCF1 state during chronic stimulation, we would then ask what epigenetic mechanisms regulate the transitions between these distinct states. Addressing these questions will provide insight into the maintenance of the TCF1+ state in chronic infection and cancer, potentially revealing design principles for increasing T cell function in these disease contexts through immunotherapy.

3.2 RESULTS

3.2.1 CHRONIC STIMULATION *EX VIVO* DRIVES LOSS OF STEMNESS AND ACQUISITION OF T CELL EXHAUSTION CHARACTERISTICS

To understand whether T cells within tumors or at the site of chronic infection can regain stemness after differentiating to a cytotoxic effector or exhausted state, the ability to capture linear progression of cell states in response to known signals, and to perturb those signals to change cell state progression, is critical. While recent advances in single-cell RNA sequencing have intricately characterized the composition of T cell states present in tumors or chronic infection, these snapshot measurements fail to capture the temporal relationships between states (Pritykin et al., 2021). Ongoing efforts to develop methods that measure both cell state and environmental signals *in vivo* address this limitation and will clarify the paths that progenitor cells take to the spectrum of function and dysfunctional states at the site of tumor or infection (Erhard et al., 2022; Sheth & Wang, 2018). A complementary approach is to remove the cells of interest from their complex *in vivo* environment and measure their differentiation in response to controlled signals that recapitulate the key features of that environment. This reductionist approach can reveal nuanced, context-dependent effects of signals that may be overlooked in an *in vivo* system, where signaling contexts are often mixed.

To execute this reductionist approach to study regulation of stemness under chronic stimulation, we adapted our *ex vivo* acute T cell stimulation approach (**Chapter 2**) to a chronic setting. Multiple studies of T cell exhaustion in various contexts have converged on a hypothesized progression of states from the stem-like, TCF1+ T_{PEX} state, to a TCF1- functional intermediate

T_{INT} state, to a dysfunctional T_{EX} state marked by upregulation of inhibitory receptors and loss of cytokine secretion potential (Guo et al., 2021; Tsui et al., 2022; Utzschneider et al., 2020). Therefore, we sought to recapitulate this progression in our *ex vivo* assay. In this assay, we stimulate CD8 T cells isolated from *Tcf7*-YFP reporter mice with plate-bound anti-CD3 and anti-CD28 antibodies in the presence of either of the inflammatory cytokines IL-12 or IFNα (**Fig. 3.1A**). Cells are restimulated on fresh antibody-coated plates at days 2 and 4, with removal of anti-CD28 after 2 days to reflect the loss of costimulation characteristic of T cell interaction with antigen in tumors. Over these 6 days of progressive stimulation and in an inflammation-dependent manner, these cells progressively downregulate *Tcf7* and upregulate Tim3, an inhibitory receptor commonly used to mark exhausted cells (**Fig. 3.1B-C**). In parallel, prolonged stimulation in this assay drives loss of TNFα, IFNɣ, and IL-2 secretion potential and in the context of IFNα stimulation, upregulation of the exhaustion-associated transcription factor, Tox (**Fig 3.1D-G**). Thus, this system recapitulates key features of exhaustion differentiation and provides a platform for controlled interrogation of its regulation.

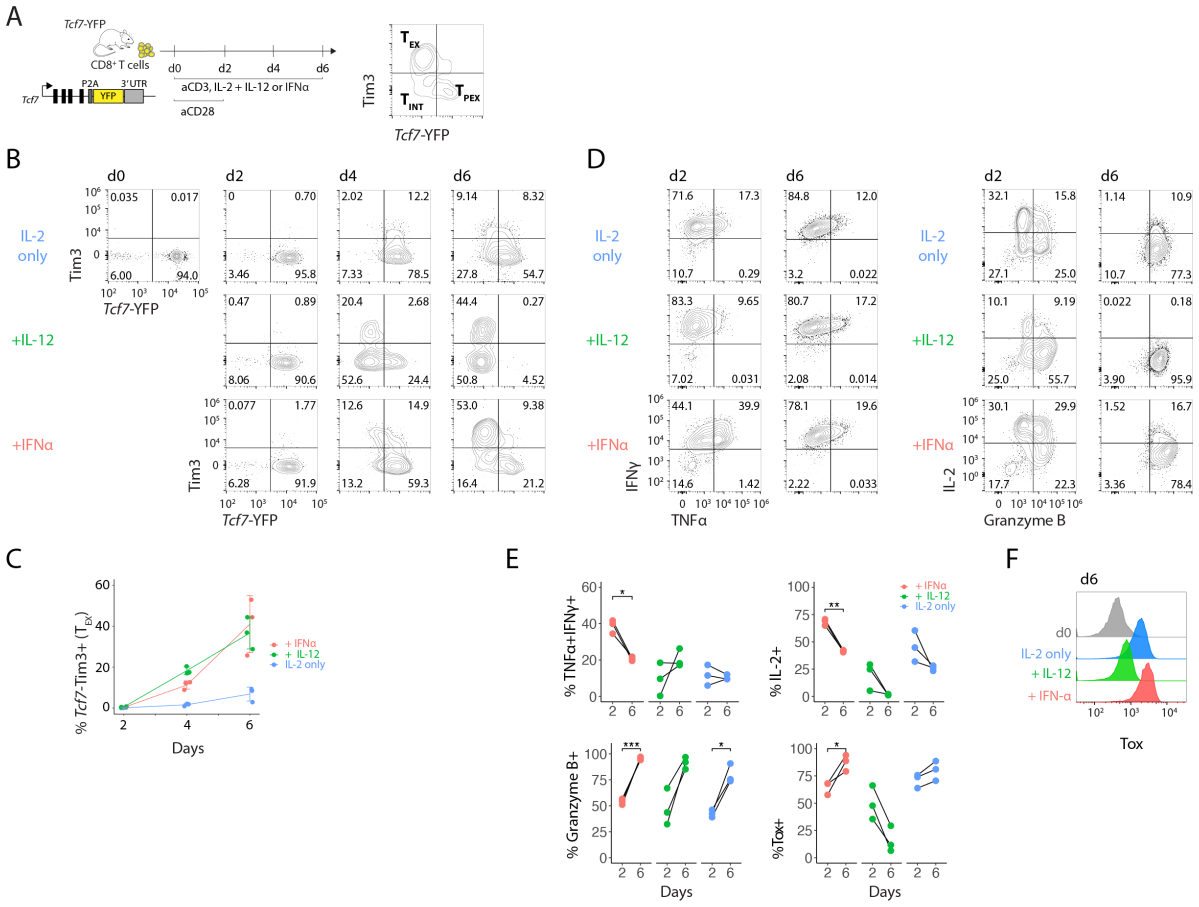


Figure 3.1: An *ex vivo* chronic stimulation assay recapitulates key features T cell exhaustion differentiation. (A) Chronic stimulation assay schematic (left); distinction of T_{PEX}, T_{INT}, and T_{EX} populations by *Tcf7* and Tim3 (right). **(B)** *Tcf7*-YFP and Tim3 expression over 6 days of chronic stimuulation under different cytokine conditions. **(C)** Quantitation of *Tcf7*-Tim3+ T_{EX} population fraction over time in (B). **(D-E)** Cytokine secretion and granzyme B levels over time on chronic stimulation. **(F)** Tox levels at day 6 compared to day 0.

3.2.2 ACTIVATED T CELLS TRANSITION FROM A PLASTIC TO TERMINAL *TCF7* SILENT STATE DURING CHRONIC STIMULATION

To understand how chronic stimulation affects the plasticity of *Tcf7*- T cells to return to the *Tcf7*+ stem-like state, we isolated *Tcf7*- cells at multiple stages during this process and measured their *Tcf7* reactivation potential. Cells were stimulated for either 2, 4, or 6 days with either IL-12 or IFNα as the inflammatory cytokine (**Fig. 3.2A**). In the IL-12 context, aCD3 and aCD28 were both maintained for the full course of stimulation, and aCD28 was removed after day 2 in the IFNα context. Cells were rested in IL-2 only for 1 day after stimulation, sorted to obtain a pure population of *Tcf7*- cells, and then rested for an additional 7 days. We observed a progressive loss of *Tcf7* reactivation potential with longer stimulation in both the IL-12 and IFNα contexts, with the *Tcf7*+ population after 7 days of rest decreasing from 54.9 (±2.7, n=3) to 18.6 (±1.5%, n=3) percent between cells stimulated for 2 and 6 days with IL-12, respectively (**Fig. 3.2B-C**). As the Tim3+ T_{EX} cell percentage of the *Tcf7*- population increases over time in our assay (**Fig. 3.2D**), we hypothesized that the observed loss of plasticity could be associated with the increasing T_{EX} population. To test this, we additionally sorted the *Tcf7*- population after 4 and 6 days of stimulation into T_{INT} and T_{EX} subsets, distinguished by Tim3 expression. Indeed, after a given stimulation duration, T_{EX} cells were significantly less able to reactivate *Tcf7* than T_{INT} cells (**Fig 3.2E-F**). Taken together, these data are consistent with the model that T cells first enter a plastic *Tcf7* off state upon activation, but upon continued stimulation transition to a more terminal, stably silent *Tcf7* off state; the former can readily reactivate upon stimulation withdrawal, while the latter cannot.

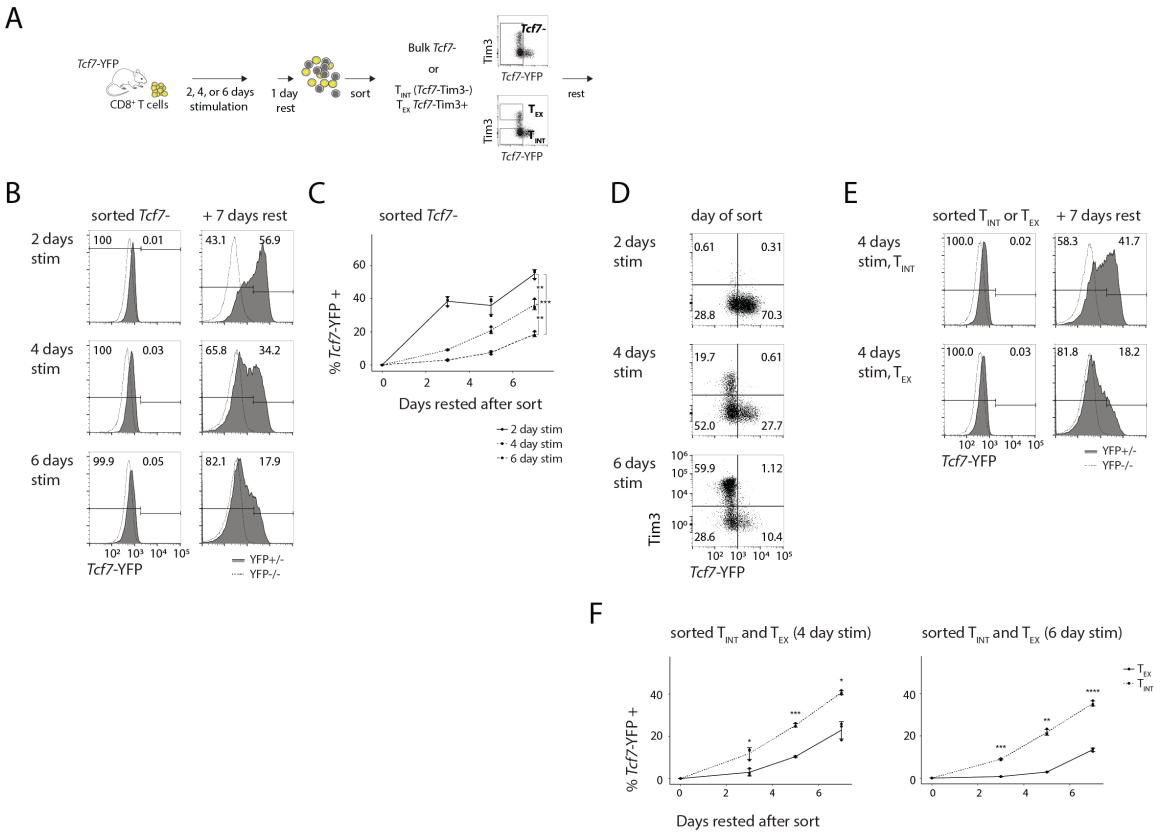


Figure 3.2: T cells lose Tcf7 reactivation potential with prolonged stimulation and exhaustion differentiation. (A) Experiment schematic: cells are stimulated for variable durations, rested in IL-2 only for one day, then sorted based on *Tcf7*-YFP alone or *Tcf7*-YFP and Tim3 before resting in IL-2. **(B)** Representative flow cytometry plots and **(C)** quantitation of *Tcf7* reactivation *Tcf7*- sorted cells stimulated for variable durations. **(D)** T_{INT} and T_{EX} distribution in cells after variable stimulation durations. **(E)** Representative flow cytometry plots and **(F)** quantitation of *Tcf7* reactivation T_{INT} and T_{EX} sorted cells stimulated for variable durations.

Given their different *Tcf7* reactivation potential, we reasoned that these early plastic and late terminal *Tcf7*- states must represent distinct epigenetic states. A defining feature of an epigenetic state is the maintenance of a given transcriptional response in the absence of the originating stimulus (Bonasio, Tu, & Reinberg, 2010). Under this definition and supported by our additional characterization of the cis-acting *Tcf7* silencing mechanism in **Chapter 2**, both the early

plastic and late terminal *Tcf7*- states qualify as epigenetic. Both states maintain transcriptional repression for a period after the originating signals, TCR stimulation and inflammation, are removed, but they differ significantly in the duration of continued repression before reactivation occurs. While cells stimulated in the context of IL-12 for 2 days become 38 (±2.8, n=3) percent reactivated after 3 days of rest, cells stimulated for 4 and 6 days are only 9.1 (±0.3, n=3) and 3.0 (±0.3%, n=3) percent reactivated after the same amount of rest, respectively (**Fig. 3.2C**).

We thus set out to elucidate the regulatory mechanisms that mediate these distinct epigenetic states. Broadly, epigenetic states can be categorized as either trans-acting or cis-acting (Bonasio et al., 2010). Trans-acting epigenetic states are achieved by self-propagating transcription factor feedback loops, and they are inherited across cell divisions through passing on of the cytosol from mother to daughter cells. Cis-acting epigenetic states, in contrast, are mediated by molecular modifications occurring on the chromosome at gene promoters or regulatory regions, such as DNA and histone methylation, and are inherited on the DNA passed from mother to daughter cells. Mechanisms in both of these categories have been demonstrated to play a role in regulation of T cell state during both acute and chronic stimulation and thus would be candidates for regulating the distinct Tcf7 off states (John T Chang, Wherry, & Goldrath, 2014; Henning, Roychoudhuri, & Restifo, 2018).

To search for epigenetic regulators that could influence *Tcf7* reactivation, we stimulated cells for variable durations as in **Fig. 3.2**, adding small molecule inhibitors of chromatin modifying enzymes upon initiation of rest (**Fig. 3.3A**). We hypothesized that inhibiting factors that support the terminally silent epigenetic state would increase *Tcf7* reactivation, while inhibiting factors that antagonize it would decrease *Tcf7* reactivation. We screened inhibitors targeting polycomb-mediated H3K27 methyltransferases (tazemetostat; UNC1999), H3K27 demethylases (GSKJ4),

DNA methyltransferases (azacytidine; decitabine, not shown), the bromodomain and extraterminal domain family member BRD4 ((+)-JQ1), H3K9 methyltransferases (BIX01294, not shown), and H3K9 demethylases (GSK-LSD1, not shown). These initial drug screens yielded a striking increase in *Tcf7* reactivation from both the T_{INT} and T_{EX} populations upon addition of tazemetostat, which inhibits Ezh2 of the Polycomb Repressive Complex 2 (PRC2). We replicated this result with the additional Ezh2 inhibitor, GSK126 (**Fig. 3.3B**). Notably, when we added tazemetostat during stimulation rather than during rest, it had the opposite effect, causing a striking reduction in *Tcf7* expression (**Fig. 3.3C**). GSK126 had a more modest effect when added during stimulation compared to tazemetostat, and further experiments using a wider range of GSK126 concentrations would be required to determine if its effect during stimulation is similar to that of tazemetostat.

These Ezh2 inhibition results are consistent with a model where PRC2-mediated H3K27 methylation contributes to the formation of the terminally silent *Tcf7* off state, such that its inhibition upon rest would slow or reverse terminal silencing, leading to increased *Tcf7* reactivation. This is in line with previous studies showing that PRC2-mediated H3K27me3 deposition on memory-associated genes locks in the terminal effector T cell state following early transcriptional state divergence (Gray et al., 2017), as well as studies demonstrating increase in the TCF1+ stem-like CD8 T cell state in tumors upon tazemetostat treatment (Y. Hou et al., 2023). It will be useful moving forward to test this model by measuring histone modification levels at the *Tcf7* locus in T_{INT} and T_{EX} cells with and without Ezh2 inhibition.

The opposite effects of Ezh2 inhibition during rest and stimulation highlight the context dependence of key regulators of T cell state, and it will be interesting to investigate the mechanisms underlying this and potentially other paradoxical results.

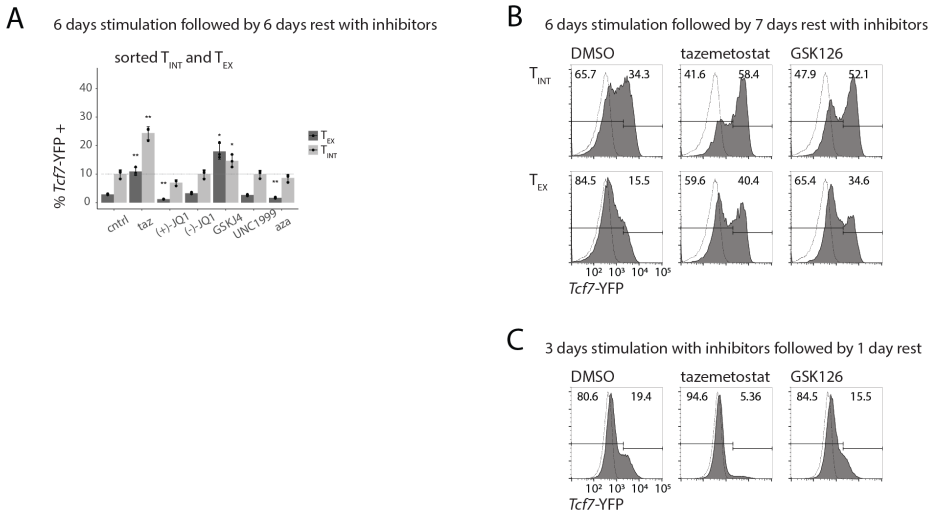


Figure 3.3: Ezh2 inhibition increases *Tcf7* reactivation upon addition during rest. (A) *Tcf7* reactivation in T_{INT} and T_{EX} cells sorted after 6 days of stimulation in the context of IL-12 and rested for an additional 6 days in IL-2 with or without inhibitors. Dashed lines indicated the mean *Tcf7*+ percent in T_{INT} and T_{EX} cells rested without inhibitors. (**B)** Representative flow cytometry plots of T_{INT} and T_{EX} cells stimulated for 6 days in the context of IL-12 and rested for 7 days with DMSO or inhibitors. (**C)** Representative flow cytometry plots of cells stimulated for three days and rested for one day prior to analysis. The indicated inhibitors and IL-12 were added 5 hours into the activation. [A] Inhibitor concentrations: tazemetostat, 1 μM; (+)-JQ1 and (-)-JQ1, 0.15 μM; GSKJ4, 1 μM; UNC1999, 1 μM; 5-Azacytidine, 1 μM. [B-C] inhibitor concentrations: tazemetostat, 3 μM; GSK126, 5 μM.

3.2.3 A SCREENING APPROACH TO IDENTIFICATION OF *TCF7* REGULATORS AT DISTINCT PHASES OF THE T CELL RESPONSE TO CHRONIC STIMULATION

T cell differentiation during any challenge, chronic or acute, spans multiple drastically different signaling and regulatory contexts, from the initial transition from a quiescent to a transcriptionally active, proliferative state, to a cytotoxic state, and ultimately to an inhibited dysfunctional or back

to a quiescent memory state. A limited number of regulators orchestrate the major epigenetic and transcriptional changes associated with each transition. Therefore, it is plausible that individual regulators such as histone and DNA methyltransferases and demethylases could have contrasting effects on cell state depending on the stage of differentiation. The opposite effects of Ezh2 inhibition that we observed on *Tcf7* regulation during stimulation and rest motivated us to test for such context-dependent effects at higher scale.

We designed a targeted CRISPR gene knockout screen to expand our interrogation of the mechanisms underlying plastic and terminal epigenetic *Tcf7* off states, and applied it to different phases of T cell differentiation during chronic stimulation. To address both cis- and trans- acting epigenetic mechanisms, we targeted 118 genes split between chromatin modifying enzymes and T cell differentiation associated transcription factors. We performed seven distinct screens using this library to test perturbation effects at different phases during chronic stimulation and in different starting cell populations and cytokine contexts, with DNA-sequencing of sorted *Tcf7*-YFP bins as the readout of each screen (**Fig. 3.4A**). To perturb early *Tcf7* silencing, we transduced cells from *Tcf7*-YFP$^{+/-}$ / Cas9$^{+/-}$ / P14^{+} mice with retrovirus delivering the sgRNA library one day after activation and sorted cells for sequencing after two or three additional days of stimulation. To perturb reactivation from the early plastic *Tcf7* off state, we sorted and transduced T_{INT} cells after 3 days of stimulation and sorted for sequencing after four days of rest. Finally, to perturb reactivation from later *Tcf7* off states, we separately sorted and transduced T_{INT} and T_{EX} cells after 6 days of stimulation and sorted for sequencing after six or seven days of rest. We performed each of these screens in duplicate and in the context of IL-12 or IFNα, with the exception of the day 3 T_{INT} screen, which we only performed in the IL-12 context. In combination, these screens revealed factors that affect *Tcf7* expression concordantly or discordantly across different phases of

differentiation, with reproducibility between replicates (**Fig. 3.4B**). Log2 fold changes (LFC) between *Tcf7* high and low sorted bins were calculated for each sgRNA, and z scores were calculated using the non-targeting control LFC distribution as a reference. Gene level z scores were then calculated as the mean z score of the top two scoring sgRNA for each gene.

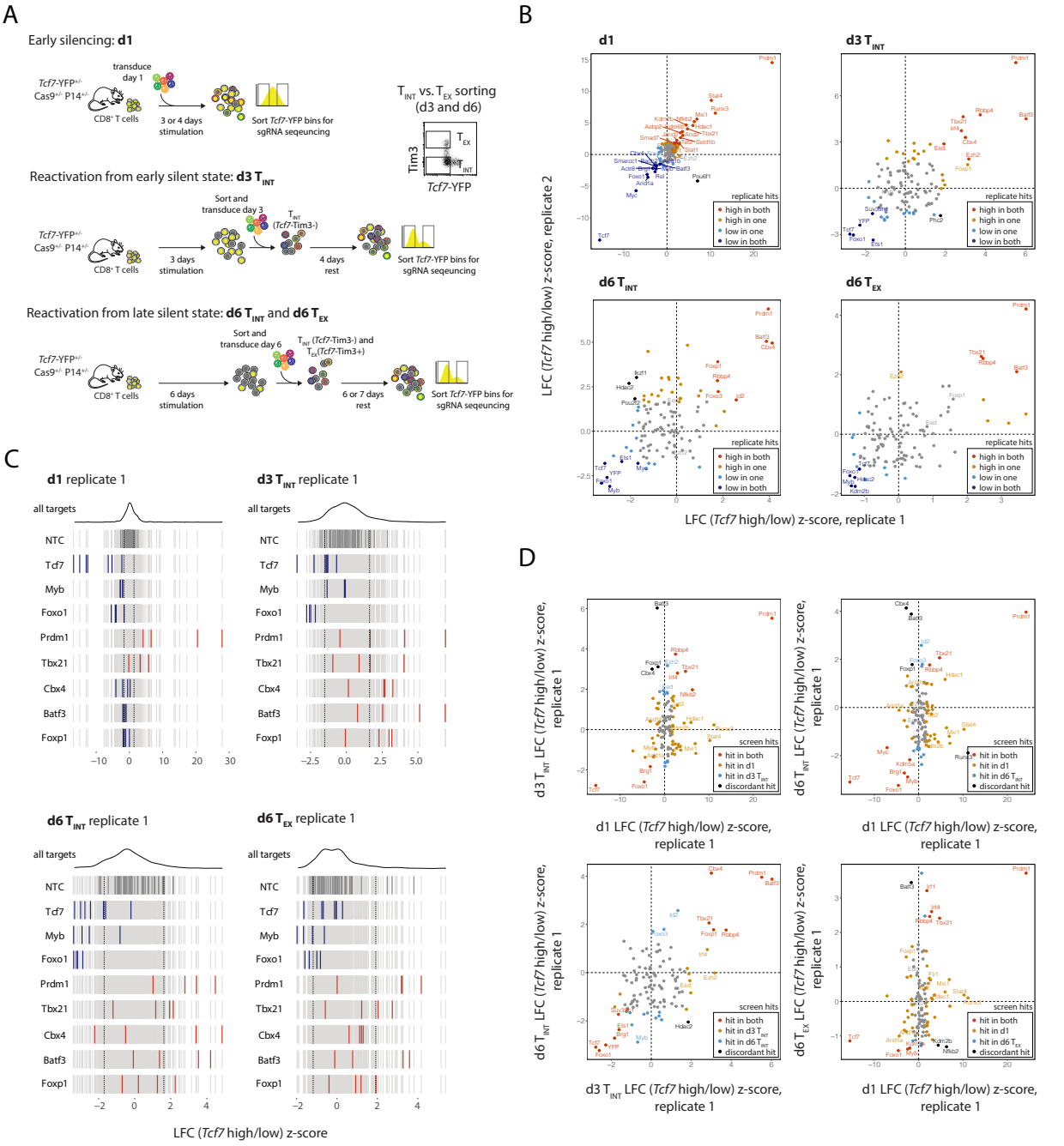

A
Early silencing: d1
transduce day 1
3 or 4 days stimulation
Sort Tcf7-YFP bins for sgRNA seqeuncing
T_INT vs. T_EX sorting (d3 and d6)
Tim3
Tcf7-YFP
Reactivation from early silent state: d3 T_INT
Sort and transduce day 3
3 days stimulation
4 days rest
Sort Tcf7-YFP bins for sgRNA seqeuncing
CD8+ T cells
Reactivation from late silent state: d6 T_INT and d6 T_EX
Sort and transduce day 6
6 days stimulation
6 or 7 days rest
Sort Tcf7-YFP bins for sgRNA seqeuncing
B
d1
d3 T_INT
d6 T_INT
d6 T_EX
replicate hits
high in both
high in one
low in one
low in both
LFC (Tcf7 high/low) z-score, replicate 2
LFC (Tcf7 high/low) z-score, replicate 1
C
d1 replicate 1
d3 T_INT replicate 1
d6 T_INT replicate 1
d6 T_EX replicate 1
all targets
NTC
Tcf7
Myb
Foxo1
Prdm1
Tbx21
Cbx4
Batf3
Foxp1
LFC (Tcf7 high/low) z-score
D
screen hits
hit in both
hit in d1
hit in d3 T_INT
hit in d6 T_INT
hit in d6 T_EX
discordant hit
d3 T_INT LFC (Tcf7 high/low) z-score, replicate 1
d6 T_INT LFC (Tcf7 high/low) z-score, replicate 1
d6 T_EX LFC (Tcf7 high/low) z-score, replicate 1
d1 LFC (Tcf7 high/low) z-score, replicate 1

Figure. 3.4: Interrogating epigenetic regulation of *Tcf7* at distinct phases of differentiation in response to chronic stimulation. (A) Diagram of targeted CRISPR screens for separate perturbation of early silencing (top, d1), reactivation from the early silent state (middle, d3 T_{INT}), and reactivation for late silent states (bottom, d6 T_{INT} and T_{EX}). Each screen except d3 T_{INT} was performed separately in IL-12 and IFNα contexts, but only the IL-12 screen results are shown in this figure. (**B)** Scatter plots of LFC z-scores (*Tcf7* high/low) for replicates of each screen, with the gene z-score score calculated as the mean of the highest two scoring sgRNA for each gene. (**C)** Individual sgRNA LFC z-scores for NTCs and hits of interest from (B). Vertical dashed lines represent the 5th and 95th percentiles of the NTC distribution, which was used for calling sgRNA- and gene-level hits (**D)** Comparison of gene LFC z-scores between different screens.

In each screen, the positive control sgRNAs targeting *Tcf7* were consistently enriched in the *Tcf7* low population (**Fig. 3.4B**). SgRNA targeting the transcription factors *Foxo1* and *Myb* were also consistently enriched in the *Tcf7* low population regardless of context, in line with their known roles in promoting the memory and self-renewal gene programs (Gray et al., 2017; Tsui et al., 2022). Genes with sgRNAs enriched in the *Tcf7* high population across all screens included *Prdm1* and *Tbx21*, which encode the transcription factors Blimp1 and Tbet, respectively, and are both widely demonstrated to promote effector differentiation in CD8 T cells (Dominguez et al., 2015; Rutishauser et al., 2009).

Comparing sgRNA scores across different screens also revealed genes that were discordantly enriched in different contexts (**Fig. 3.4C**). Most strikingly, sgRNAs targeting the polycomb repressive complex 1 subunit *Cbx4* and the transcription factors *Batf3* and *Foxp1* were enriched in the *Tcf7* high population when applied during rest but enriched in the *Tcf7* low population when applied during initial stimulation. Consistent with our observation of increased *Tcf7* expression upon *Cbx4* knockout during rest, *Cbx4* was recently demonstrated to induce effector T cell commitment in an LCMV model, with shRNA knock down of *Cbx4* increasing the memory population 8 days after infection (Melo et al., 2022). As *Cbx4* plays a role not only in PRC1 activity but also PRC2 activity through Ezh2, these results are also consistent with our

observation of increased *Tcf7* reactivation upon Ezh2 inhibition with rest (**Fig. 3.3**) (Wu et al., 2022). The contrasting enrichment of *Cbx4* sgRNA in the *Tcf7* low population when applied during stimulation could also reflect the increased *Tcf7* silencing we observed with Ezh2 inhibition during stimulation. Furthermore, others have reported a role for *Cbx4* in promoting Wnt/β-catenin activity, the pathway upstream of *Tcf7*, in intestinal stem cells and lung adenocarcinoma cells, suggesting that *Cbx4* knockout under some contexts could antagonize *Tcf7* expression (Chiacchiera et al., 2016; Wang et al., 2021). Taken together, these studies and the context-dependent influence of *Cbx4* on *Tcf7* revealed by our screens, suggest a nuanced role of *Cbx4* and polycomb activity across different stages of T cell differentiation.

In our screens, *Batf3* knockout during early stimulation decreased *Tcf7* expression, while knockout during rest increased *Tcf7* expression. The decrease in *Tcf7* expression with knockout during stimulation is consistent with the reported role of *Batf3* in establishing memory differentiation (Ataide et al., 2020). Critically, these findings were attributed to transient *Batf3* expression in the first 72 hours of activation, which spans the time frame of our initial *Tcf7* silencing screen. The role of *Batf3* at later phases of differentiation and in particular under chronic stimulation are less well-understood, but recent studies have demonstrated its upregulation during exhaustion. In a CAR T cell exhaustion model, overexpression of c-Jun decreased *Batf3* expression and chromatin binding of JunB-Batf3 complexes promoting exhaustion, which improved maintenance of a self-renewing stem-like population (Lynn et al., 2019). Thus, our screening approach may resolve the paradoxical role of *Batf3* in promoting self-renewal and memory differentiation early during stimulation but antagonizing the self-renewing stem-like population during chronic stimulation.

The discordant effects of *Foxp1* targeting sgRNAs on *Tcf7* during initial stimulation and during *Tcf7* reactivation after extended stimulation are also consistent with studies defining its role at different phases of CD8 T cell differentiation (Kaminskiy, Kuznetsova, Kudriaeva, Zmievskaya, & Bulatov, 2022). *Foxp1* actively maintains naive T cell quiescence by repressing IL-7Rα (Feng et al., 2011). Its deletion in naive cells causes spontaneous effector differentiation in response to IL-7, consistent with our observed enrichment of *Foxp1* targeting sgRNA in the *Tcf7* low population during early stimulation. However, *Foxp1* has also been shown to contribute to T cell exhaustion in tumor infiltrating lymphocytes, where its deletion is beneficial for tumor control, potentially consistent with the increase in *Tcf7* we observe upon deletion with rest after 6 days of chronic stimulation (Stephen et al., 2014; Zhang et al., 2019).

3.3 DISCUSSION

Manipulating epigenetic regulators of T cell state holds great promise for immunotherapy against chronic infection and cancer. The dysfunctional, or exhausted, T cell state that forms under chronic stimulation characteristic of these diseases is a distinct epigenetic state as opposed to a transient transcriptional state, and therefore, preventing or reversing this state must require epigenetic perturbation (Belk, Daniel, & Satpathy, 2022; Sen et al., 2016). However, a limited set of epigenetic regulators, including DNA and histone methyltransferases and transcription factors, orchestrates multiple cell state transitions over the course of activation of naive cells to establishment of the exhausted state. Therefore, to realize the potential of epigenetic perturbation to facilitate or prevent establishment of a cell state of interest without unintended perturbation of other state transitions, it is critical to understand how candidate epigenetic perturbations affect all relevant cells state transitions in the T cell response to chronic stimulation.

We set out in this study to ask whether chronically stimulated T cells that silence *Tcf7* can reactivate it and thus provide a means for replenishing the source of TCF1+ T_{PEX} cells in a response. We found that activated T cells that initially silence *Tcf7* are quite plastic in their ability to reactivate it, but this plasticity is lost over time with continued stimulation, in a manner associated with the acquisition of an exhausted T_{EX} phenotype. To understand how this loss of plasticity is regulated, we isolated cells at distinct phases with respect to *Tcf7* expression state during differentiation under chronic stimulation – naïve cells prior to *Tcf7* silencing, plastic cells that had recently silenced *Tcf7*, and more terminally *Tcf7* silent cells – and subjected them to epigenetic perturbation with small molecule inhibitors and genetic editing using CRISPR screening. By measuring the effect of each perturbation on *Tcf7* expression, we determined how epigenetic manipulation affected *Tcf7* expression across multiple cellular contexts and signaling environments.

This approach identified epigenetic regulators that had consistent effects on *Tcf7* expression across all cellular and signaling contexts, as well as regulators that had discordant effects in different contexts. In line with many studies characterizing transcription factor control of T cell state, we identified Blimp1 and Tbet as negative regulators and Foxo1 and Myb as positive regulators of *Tcf7* expression across all contexts. Hits that were discordant across contexts, such as Cbx4, Batf3, and Foxp1, may reconcile conflicting literature regarding their role in T cell state regulation. Between these two categories of hits were also many regulators that influenced *Tcf7* expression in some contexts but had no effect in others. Validation and further characterization of the mechanisms of these concordant and discordant effects in different contexts should provide deeper insight into the nuances of epigenetic control of the *Tcf7* locus and T cell state. Furthermore, while our study by design focused narrowly on *Tcf7* to understand its regulation and mechanisms

for maintenance of stemness over the course of chronic stimulation, it will also be useful to perform similar context-aware screens with more general readouts of T cell state.

Our approach may provide critical insight for design of immunotherapies for epigenetic control of T cell state. For example, the ability of Ezh2 inhibition and Cbx4, Batf3, and Foxp1 knockout to increase *Tcf7* reactivation after chronic stimulation identifies these factors as immunotherapy targets for promoting the TCF1+ stem-like T cell state critical for persistent responses. Indeed, the therapeutic relevance of Ezh2 (also related to Cbx4) and Batf3 repression is already being explored to improve T cell responses in cancer (Y. Hou et al., 2023; Lynn et al., 2019). However, the opposite effects of these factors on *Tcf7* expression during initial T cell stimulation could counter the intended therapeutic effect, if the perturbation is performed indiscriminately across different states. In order to effectively manipulate epigenetic regulators to promote desirable T cell states in chronic infection and cancer, the field will need to both gain context-aware understanding of key regulators and develop cell engineering strategies that can implement epigenetic manipulations at targeted phases during the T cells response. Careful design of genetic screening assays, in combination with signal and lineage recording methods that can connect transcriptomic measurements of T cell state with environmental cues, will aid in the first task. Synthetic biology approaches that couple CRISPR-based genetic activation, inhibition, and knockout with genetic circuits, will realize the second.

3.4 METHODS

Mice

Tcf7-YFP mice have been described (Harly et al., 2019). All mice used in experiments were heterozygous for the *Tcf7*-YFP reporter except where specified. WT C57BL/6 mice (Jackson Laboratory) were utilized as reporter negative controls, where applicable. Both male and female mice were used for *ex vivo* experiments, aged 8 to 12 weeks. Female CD45.1 mice, 8-12 weeks of age, were purchased from the Jackson Laboratory for use as recipients for adoptive transfer experiments. For donors for adoptive transfer experiments, homozygous *Tcf7*-YFP mice were crossed with an LCMV specific TCR transgenic strain (Pircher et al., 1989b) (P14) (Jackson Laboratory) and heterozygous offspring were used. P14 homozygous mice without *Tcf7*-YFP were utilized as non-fluorescent controls for sort gate setting. For CRISPR screening experiments mice homozygous for *Tcf7*-YFP and Cas9 (Jackson Laboratory) were crossed with homozygous P14 mice to generate offspring heterozygous for all transgenes. All mice were used in accordance with Institutional Animal Care and Use Committee guidelines for the University of Washington.

CD8+ T cell extraction

Spleens were harvested from mice, massaged between rough glass slides to generate a single-cell suspension, and filtered through 40 μm nylon mesh into HBH (HBSS, 10 mM HEPES, 0.5% BSA, pH 7.4). Cells were spun down for 5 min at 300g, resuspended in 3 ml red blood cell (RBC) lysis buffer (150 mM NH_4Cl, 10 mM $NaHCO_3$, 1 mM EDTA) for 3-5 min, and quenched with HBH. Cells were spun down for 5 min at 300g and resuspended in HBH with 2.4G2 blocking solution and incubated for 30 min on ice. Cells were counted, spun down again, and then enriched for CD8 T cells using a $CD8a^+$ T Cell Isolation Kit, mouse (Miltenyi, #130-104-075), with the volume and

amount of antibodies and microbeads used scaled down to 70% of that specified by the manufacturer. One LS column was used per spleen (Miltenyi, # 130-042-401).

***Ex vivo* T cell differentiation**

One day prior to T cell harvest and activation (day -1), plates were prepared by coating with anti-CD3e (Tonbo, #40-0031-U100) and anti-CD28 (Tonbo, #40-0281-U100). Unless otherwise specified, each well of a 96-well plate received 0.2 µg anti-CD3, 0.1 µg anti-CD28 in 50 µL of PBS. For differentiation in larger wells, these amounts were scaled up by well surface area. Plates were sealed with parafilm and incubated at 4°C overnight. On day 0, plates were allowed to come to room temperature for at least 30 min and washed 2x with PBS. Purified cells were added to wells in T cell media [85% RPMI 1640 with L-glutamine, 10% Fetal Bovine Serum, Pen-Strep-Glutamine, 20 mM HEPES, 1 mM Sodium Pyruvate, 0.1 mM NEAA, 50 µM BME] with indicated cytokine concentrations, mixed, and spun down for 1 min at 150g to ensure initial contact for all cells with the coated plate surface. Cytokines added to the media were 100 U/mL IL-2 (PeproTech, # 200-02), and 1 ng/mL IL-12 (PeproTech, #210-12) or 10 ng/mL IFNα (Biolegend, #752804) unless otherwise specified. The cell seeding concentration was 0.1 - 2.5 million cells / ml unless otherwise indicated. Cells were incubated at 37°C in 5% CO_2 and split every two days by mixing, removing half of the well volume, and topping off the volume with TCM and respective cytokines. Where applicable, prior to seeding, cells were stained with 5 µM CellTrace Violet (CTV) (Invitrogen, #C34557) following the manufacturer's instructions. Cells were rested in TCM with 100 U/mL IL-2 and small molecule inhibitors at indicated concentrations ((+)-JQ1, (-)-JQ1, GSKJ4, UNC1999, 5-Azacytidine, BIX01294, Decitabine, GSK-LD1, and GSK126 were from Cayman Chemical Company; tazemetostat was from Selleckchem).

sgRNA library design and cloning

The 500 sgRNAs were ordered as a Twist oligo pool with the following design: the 20 nucleotide spacer sequence was flanked on the 5' and 3' ends by gibson overhang sequences for cloning into lentiGuide-puro (Addgene #52963). Additional PCR handles were included for subcloning two separate libraries from the pool, one handle set for all of the sgRNA targeting transcription factors, and one for all of the sgRNA targeting chromatin modifiers. Each sublibrary was PCR amplified using 10 ng input and 12 cycles. Each amplified sublibrary was then further amplified for 6 cycles and cloned into lentiGuide-puro digested with BsmBI using NEBuilder HiFi DNA Assembly (NEB, Cat. 657 No. E2621S). For each sublibrary, a 433 base pair segment from the U6 promoter to the gRNA scaffold was then amplified from the lentiGuide-puro backbone and inserted to a MSCV backbone with an mRuby3 constitutive reporter. For all CRISPR screen experiments, equimolar amounts of the transcription factor and chromatin modifier sublibraries were mixed prior to HEK293T transfections for viral particle generation. All amplifications for cloning were performed using KAPA HiFi polymerase (Roche, #KK2601).

Retrovirus production and transduction

Retroviral particles were produced in 293T cells with the MSCV library and pCL-Eco plasmids using Lipofectamine 3000, with transfection of 293Ts at roughly 90% confluence in 10 cm dishes. Viral supernatant was harvested at 48 hours post-transfection, filtered through a 45 μM syringe filter, flash frozen in 1 mL aliquots, and stored at -80°C until use. Titration experiments were performed for transduction of T cells at different stages of differentiation, targeting 10-30% transduction efficiency at each stage. Immediately prior to transduction, supernatant was thawed

at 37°C and mixed with TCM for the appropriately titrated supernatant concentration, as well as polybrene and IL-2 for 6 ug/mL and 100 U/mL respectively in the final mixture of cells and viral particles. The viral supernatant was then added to wells already containing T cells and spun at 2000 g for 60 minutes at 32°C. The media was switched back to TCM between 4 and 24 hours after transduction.

sgRNA library gDNA extraction, amplification, and sequencing

Genomic DNA was extracted from sorted cell pellets by lysis in 50 µL per 100,000 cells of lysis buffer (20 mM Tris, 10 mM EDTA, 0.4 mg/mL Proteinase K), and then cleaned with a 1x ratio of KAPA Pure Beads (Roche, #KK8000). Extracted gDNA was the PCR amplified in two steps using KAPA HiFi polymerase (Roche, #KK2601) to prepare for sequencing. The first PCR added parts of Nextera Read 1 and Truseq Read 2, with one 50 µL PCR reaction performed for DNA from every 100,000 sorted cells. The first PCR reaction was cleaned with a 1.5x bead ratio, and reactions from the same starting sample were pooled before the second PCR, unless intentionally kept separate to assess PCR technical replication. The second PCR completed Nextera Read 1 and Truseq Read 2 and added i5 and i7 indices and P5 and P7 adapters. All PCRs were stopped at low cycle number in the logarithmic phase of amplification to prevent overamplification, which was roughly 21 cycles for the first PCR and 5 cycles for the second PCR. The plasmid library was amplified in the same manner, with the first PCR requiring roughly 12 cycles. After a 1x bead clean following the second PCR, amplicons from all samples were pooled and sequenced on a NextSeq 2000 P1 100 cycle kit with at least 45 cycles on Read 1 and 10 cycles on each index.

sgRNA screening data analysis

Reads for each sgRNA in each sample were counted using a Python script adapted from Joung et al (Joung et al., 2017). Only perfect matches were included in downstream analysis. All subsequent processing was performed in R. For each separate screen, the log2 fold change z-scores between sorted *Tcf7*-YFP high and low bins were calculated as follows. For each sgRNA, the number of reads for that sgRNA was divided by the total sample read count, then multiplied by 1,000,000, and a pseudocount of 1 was added. The log2 was taken to generate a log2 fold change (LFC). To generate z-scores, the mean LFC of the non-targeting control (NTC) sgRNA was subtracted from the sgRNA LFC, and that was then divided by the standard deviation of the NTC LFC. sgRNA with z-scores below the 5th percentile or above the 95th percentile of the NTC distribution were called hits. Gene level z-score were calculated as the average of the two highest scoring sgRNA for each gene.

Chapter 4. LOOKING AHEAD: IMMUNE CELL DIFFERENTIATION ACROSS TIME AND SPACE

4.1 INTRODUCTION

The progress of the past few decades in characterizing the immune cell populations that form in response to diverse challenges is stunning. Mouse models for different types of infection, cancer, and autoimmunity paired with ever-improving high resolution methods to measure cell proteomes, transcriptomes, and epigenomes, have generated detailed pictures of the immune populations present in countless healthy and diseased tissues. Application of these methods to human tissues

has furthered our understanding of these immune populations in a more therapeutically relevant setting. The missing piece in these rich pictures, however, is the connection of cell state to axes of time, signal, and space. Application of rapidly developing signal and lineage recording methods to complex in vivo models for immune cell differentiation will enable us not only to ask what immune cell states are present in a given tissue in a given disease but importantly, to ask how they got there and what signals drove them to their current state. With this information, we will be far better equipped to intervene therapeutically in the formation of dysfunctional immune states and enhance the states that best resolve disease. In this chapter, I will briefly review genomic technologies that will open up this line of questioning in immunology research.

4.2 TEMPORALLY-RESOLVED SINGLE-CELL RNA SEQUENCING

Parts of Chapter 4.2 have been adapted from the following manuscript:

Kirschenbaum, D.[*], Xie, K.[*], Ingelfinger, F.[*], Katzenelenbogen, Y.[*], Abadie, K.[*], Look, T., Sheban, F., Truong, S.P., Li, B., Yoffe, I., David, E., Mazuz, K., Hou, J., Chen, Y., Colonna, M., Yosef, N., Weiss, T., Weiner, A., Amit, I. (2023). Temporally resolved single-cell transcriptomics defines immune dysfunction trajectories in glioblastoma. *under revision and resubmission, Cell.*

All immunological responses function through dynamic and coordinated action of cell circuits that change their activity in response to evolving environmental signals. Single-cell RNA sequencing (scRNA-seq) has dramatically improved our ability to measure cell states at high resolution and scale. However, due to the destructive nature of the methodology, scRNA-seq can only capture static snapshots of gene expression at experimental endpoints, thus fundamentally omitting the temporal dimension. Despite this limitation, scRNA-seq at multiple time points in dynamic processes has provided insight into transcriptional drivers of development and differentiation in

myriad immunological systems (Ding, Sharon, & Bar-Joseph, 2022; Griffiths, Scialdone, & Marioni, 2018). Computational approaches, such as pseudotime and RNA velocity estimations, infer a temporal dimension in scRNA-seq experiments either by ordering cells along a trajectory based on gene expression similarity or estimating the time derivative of gene expression state by distinguishing spliced and unspliced mRNAs, respectively (La Manno et al., 2018; Trapnell et al., 2014). While useful, these algorithmic tools are sensitive to user-defined parameters and lack an empirically measured ground truth, making it difficult to test imputed conclusions, particularly in complex *in vivo* settings where multiple trajectories may coexist. To better connect the vast array of immune cell states that have been characterized by scRNA-seq to the history of signaling and differentiation that formed them, we require empirical *in vivo* genomic technologies that measure both cell state and time in single cells.

One promising approach to adding a time dimension to single-cell transcriptomics is metabolic labeling of new mRNA molecules (Erhard et al., 2022), used in **Chapter 2** of this dissertation (**Fig. 2.3-2.4**). This class of technologies improves resolution of rapid cellular responses to signals by detecting changes in new mRNA levels separate from the background of 'old' mRNA in a cell. Tiling metabolic labeling windows over longer differentiation responses can also temporally resolve cell state transitions that unfold over a few hours to close to a day. However, this approach is difficult to apply *in vivo* or to processes that take place over multi-day time scales.

An approach with potential to clarify multi-day differentiation processes *in vivo* is to physically label cells at a relevant time point or in a relevant *in vivo* location with a marker that can remain on the cell for multiple days and then be recovered along with the single-cell transcriptome upon harvest. Staining immune cells in an animal's vasculature with CD45 at a

reference time point and recovering labeled cells in tissues hours to days later can connect the transcriptomic state of the cell in the tissue to a measurement of its previous location at a known time (Potter et al., 2021). Similarly, photolabeling cells in one *in vivo* compartment and harvesting them at a later time in a separate compartment can connect cell state to prior *in vivo* location at known time points. This has proved useful for monitoring T cell trafficking between tumor and lymphoid tissues (Li et al., 2022). Continued development of these methods to empirically connect cell state to time and place in tissues will improve dynamic resolution of immune differentiation processes.

4.3 GENOMIC RECORDING OF SIGNALS AND LINEAGE

If immune cells kept a diary throughout their lifespan in which they diligently wrote where in the body they have been, what antigens they encountered, and what other cells they communicated with, then identifying connections between the signals cells perceive and the actions they take would be easy. Since cells do not do this, we as researchers have recently started forcing cells to write their own memoirs through DNA-based genomic recording. These methods, pioneered initially in cell lines and recently deployed in multiple complex *in vivo* systems, have great potential for resolving cell-state trajectories in immune differentiation and revealing causal relationships among signals, lineage, and state (Sheth & Wang, 2018; Wagner & Klein, 2020). A challenge with immunology research in mouse models is that immune cells migrate throughout the body and experience different signaling contexts over the course of their lifespan and the duration of an immune challenge. In general, even if researchers insert cells into a mouse at a defined time point through adoptive transfer and harvest them at a later time point in distinct tissues, the locations that cells traveled to and signals they experienced in the interim remain a

black box. By equipping cells with the machinery to make DNA edits in a lineage-specific manner or under the control of specified signals, we can begin to open this box.

Most genomic recording methods to date operate via CRISPR-based nucleases, integrases, or base editors. In these systems, the Cas-based effector molecule is recruited to a DNA 'scratchpad,' where it makes an edit via random nucleotide deletion or insertion or a defined base insertion or base edit, that can then be read out upon harvest by sequencing. In lineage recorders where the Cas9 machinery is constitutively active, continual scratchpad editing over multiple generations enables computational reconstruction of clonal phylogenies from a single endpoint harvest (McKenna et al., 2016; Yang et al., 2022). Construction of the DNA scratchpad to be expressed as mRNA furthermore enables joint recovery of the scratchpad with single-cell transcriptomes, enabling cell state placement along clonal phylogenetic trees. Using such methods, we could ask by sequencing rather than the *ex vivo* clonal imaging I used in **Chapter 2** how memory and effector differentiation unfolds within clonal lineages responding to challenges *in vivo*.

To record signals, editing machinery can be expressed selectively under defined signaling pathways (Kempton, Love, Guo, & Qi, 2022). A recent method from Chen et al. generates programmed insertions in a scratchpad by prime editing coupled to activity of hundreds of enhancers and to Wnt and NF-κB activity, as a proof of concept (Chen et al., 2021). Deploying such methods in immune cells to record intracellular and extracellular signaling events hypothesized to drive cell state changes could address many outstanding questions for immune cell differentiation.

A major challenge to use of these recording methods in immune cells is the task of getting all of the necessary components into primary immune cells. Transgenic mice that express the

necessary nuclease are helpful for mouse studies, as the nuclease itself is one of the largest and most difficult components to insert into cells. Availability of Cas9 mice has already enabled widespread CRISPR knockout screens, suggesting that development of mice expressing other Cas enzymes such as base editors and dCas9 fused to different effectors holds similar potential. Progress in human T cell engineering puts use of these recording methods in reach for human T cell studies (Roth et al., 2018). However, as these studies cannot be conducted in humans, advancement of organoid systems that recapitulate human immune differentiation trajectories will be necessary for this avenue (Wagar et al., 2021). In both mouse and human cells, continual progress in viral and non-viral DNA and protein delivery methods will also reduce the barriers to genomic recording.

Overall, methods to capture time and record signals in immune cells have the potential to build upon decades of knowledge and model design from immunologists to shed new light on immune differentiation in health and disease.

REFERENCES

Abadie, K., Pease, N. A., Wither, M. J., & Kueh, H. Y. (2019). Order by chance: Origins and benefits of stochasticity in immune cell fate control. *Current Opinion in Systems Biology*, *18*, 95–103.

Alves, R., & Savageau, M. A. (2000). Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics (Oxford, England)*, *16*, 786–798.

Arsenio, J., Kakaradov, B., Metz, P. J., Kim, S. H., Yeo, G. W., & Chang, J. T. (2014). Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nature Immunology*, *15*, 365–372.

Ataide, M. A., Komander, K., Knöpper, K., Peters, A. E., Wu, H., Eickhoff, S., … Kastenmüller, W. (2020). BATF3 programs CD8+ T cell memory. *Nature Immunology*, *21*, 1397–1407.

Bannard, O., Kraman, M., & Fearon, D. T. (2009). Secondary Replicative Function of CD8+ T Cells That Had Developed an Effector Phenotype. *Science*, *323*, 505–509.

Becker, A. J., McCulloch, E. A., & Till, J. E. (1963). *Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells*. Retrieved from https://tspace.library.utoronto.ca/handle/1807/2779

Belk, J. A., Daniel, B., & Satpathy, A. T. (2022). Epigenetic regulation of T cell exhaustion. *Nature Immunology*, *23*, 848–860.

Benjaafar, S., Morin, T. L., & Talavage, J. J. (1995). The strategic value of flexibility in sequential decision making. *European Journal of Operational Research*, *82*, 438–457.

Best, J. A., Blair, D. A., Knell, J., Yang, E., Mayya, V., Doedens, A., … Goldrath, A. W. (2013). Transcriptional insights into the CD8+ T cell response to infection and memory T cell formation. *Nature Immunology*, *14*, 404–412.

Bintu, L., Yong, J., Antebi, Y. E., McCue, K., Kazuki, Y., Uno, N., … Elowitz, M. B. (2016). Dynamics of epigenetic regulation at the single-cell level. *Science*, *351*, 720–724.

Blattman, J. N., Antia, R., Sourdive, D. J. D., Wang, X., Kaech, S. M., Murali-Krishna, K., … Ahmed, R. (2002). Estimating the Precursor Frequency of Naive Antigen-specific CD8 T Cells. *Journal of Experimental Medicine*, *195*, 657–664.

Bonasio, R., Tu, S., & Reinberg, D. (2010). Molecular Signals of Epigenetic States. *Science*, *330*, 612–616.

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*, 525–527.

Buchholz, V. R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Graf, P., … Busch,

D. H. (2013). Disparate Individual Fates Compose Robust CD8+ T Cell Immunity. *Science*, *340*, 630–635.

Buchholz, Veit R., Schumacher, T. N. M., & Busch, D. H. (2016). T Cell Fate at the Single-Cell Level. *Annual Review of Immunology*, *34*, 65–92.

Busch, D. H., Pilip, I. M., Vijh, S., & Pamer, E. G. (1998). Coordinate Regulation of Complex T Cell Populations Responding to Bacterial Infection. *Immunity*, *8*, 353–362.

Cao, J., Cusanovich, D. A., Ramani, V., Aghamirzaie, D., Pliner, H. A., Hill, A. J., … Shendure, J. (2018). Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science*, *361*, 1380–1385.

Cao, J., Packer, J. S., Ramani, V., Cusanovich, D. A., Huynh, C., Daza, R., … Shendure, J. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, *357*, 661–667.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., … Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, *566*, 496–502.

Cao, J., Zhou, W., Steemers, F., Trapnell, C., & Shendure, J. (2020). Sci-fate characterizes the dynamics of gene expression in single cells. *Nature Biotechnology*, *38*, 980–988.

Chang, J. T., Palanivel, V. R., Kinjyo, I., Schambach, F., Intlekofer, A. M., Banerjee, A., … Reiner, S. L. (2007). Asymmetric T Lymphocyte Division in the Initiation of Adaptive Immune Responses. *Science*, *315*, 1687–1691.

Chang, John T, Wherry, E. J., & Goldrath, A. W. (2014). Molecular regulation of effector and memory T cell differentiation. *Nature Immunology*, *15*, 1104–1115.

Chen, W., Choi, J., Nathans, J. F., Agarwal, V., Martin, B., Nichols, E., … Shendure, J. (2021, November 5). *Multiplex genomic recording of enhancer and signal transduction activity in mammalian cells* (p. 2021.11.05.467434). p. 2021.11.05.467434. bioRxiv.

Chiacchiera, F., Rossi, A., Jammula, S., Piunti, A., Scelfo, A., Ordóñez-Morán, P., … Pasini, D. (2016). Polycomb Complex PRC1 Preserves Intestinal Stem Cell Identity by Sustaining Wnt/β-Catenin Transcriptional Activity. *Cell Stem Cell*, *18*, 91–103.

Cho, Y.-L., Flossdorf, M., Kretschmer, L., Höfer, T., Busch, D. H., & Buchholz, V. R. (2017). TCR Signal Quality Modulates Fate Decisions of Single CD4 + T Cells in a Probabilistic Manner. *Cell Reports*, *20*, 806–818.

Chung, H. K., McDonald, B., & Kaech, S. M. (2021). The architectural design of CD8+ T cell responses in acute and chronic infection: Parallel structures with divergent fates. *The Journal of Experimental Medicine*, *218*, e20201730.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*, 37–46.

Connolly, K. A., Kuchroo, M., Venkat, A., Khatun, A., Wang, J., William, I., … Joshi, N. S. (2021). A reservoir of stem-like CD8 $^{+}$ T cells in the tumor-draining lymph node preserves the ongoing anti-tumor immune response. *Science Immunology*. https://doi.org/10.1126/sciimmunol.abg7836

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., … Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods*, *14*, 959–962.

Cusanovich, D. A., Hill, A. J., Aghamirzaie, D., Daza, R. M., Pliner, H. A., Berletch, J. B., … Shendure, J. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell*, *174*, 1309-1324.e18.

Danilo, M., Chennupati, V., Silva, J. G., Siegert, S., & Held, W. (2018). Suppression of Tcf1 by Inflammatory Cytokines Facilitates Effector CD8 T Cell Differentiation. *Cell Reports*, *22*, 2107–2117.

Ding, J., Sharon, N., & Bar-Joseph, Z. (2022). Temporal modelling using single-cell transcriptomics. *Nature Reviews Genetics*, *23*, 355–368.

Dominguez, C. X., Amezquita, R. A., Guan, T., Marshall, H. D., Joshi, N. S., Kleinstein, S. H., & Kaech, S. M. (2015). The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *The Journal of Experimental Medicine*, *212*, 2041–2056.

Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H. S., Holmberg, R., Hawkins, E. D., … Hodgkin, P. D. (2012). Activation-Induced B Cell Fates Are Selected by Intracellular Stochastic Competition. *Science*, *335*, 338–341.

Erhard, F., Saliba, A.-E., Lusser, A., Toussaint, C., Hennig, T., Prusty, B. K., … Dölken, L. (2022). Time-resolved single-cell RNA-seq using metabolic RNA labelling. *Nature Reviews Methods Primers*, *2*, 1–18.

Fearon, D. T., Manders, P., & Wagner, S. D. (2001). Arrested Differentiation, the Self-Renewing Memory Lymphocyte, and Vaccination. *Science*, *293*, 248–250.

Feng, X., Wang, H., Takata, H., Day, T. J., Willen, J., & Hu, H. (2011). Transcription factor Foxp1 exerts essential cell-intrinsic regulation of the quiescence of naive T cells. *Nature Immunology*, *12*, 544–550.

Fowler, A. C. (2021). Atto-Foxes and Other Minutiae. *Bulletin of Mathematical Biology*, *83*, 104.

Gerlach, C., Rohr, J. C., Perie, L., van Rooij, N., van Heijst, J. W. J., Velds, A., … Schumacher, T. N. M. (2013). Heterogeneous Differentiation Patterns of Individual CD8+ T Cells. *Science*, *340*, 635–639.

Graef, P., Buchholz, V. R., Stemberger, C., Flossdorf, M., Henkel, L., Schiemann, M., … Busch, D. H. (2014). Serial Transfer of Single-Cell-Derived Immunocompetence Reveals

Stemness of CD8+ Central Memory T Cells. *Immunity*, *41*, 116–126.

Gray, S. M., Amezquita, R. A., Guan, T., Kleinstein, S. H., & Kaech, S. M. (2017). Polycomb Repressive Complex 2-Mediated Chromatin Repression Guides Effector CD8 + T Cell Terminal Differentiation and Loss of Multipotency. *Immunity*, *46*, 596–608.

Gregor, T., Bialek, W., Steveninck, R. R. de R. van, Tank, D. W., & Wieschaus, E. F. (2005). Diffusion and scaling during early embryonic pattern formation. *Proceedings of the National Academy of Sciences*, *102*, 18403–18407.

Gregor, T., Wieschaus, E. F., McGregor, A. P., Bialek, W., & Tank, D. W. (2007). Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. *Cell*, *130*, 141–152.

Griffiths, J. A., Scialdone, A., & Marioni, J. C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. *Molecular Systems Biology*, *14*, e8046.

Guo, Y., Xie, Y.-Q., Gao, M., Zhao, Y., Franco, F., Wenes, M., … Tang, L. (2021). Metabolic reprogramming of terminally exhausted CD8+ T cells by IL-10 enhances anti-tumor immunity. *Nature Immunology*, *22*, 746–756.

Haghverdi, L., Lun, A. T. L., Morgan, M. D., & Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, *36*, 421–427.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*, 3573–3587.

Harly, C., Kenney, D., Ren, G., Lai, B., Raabe, T., Yang, Q., … Bhandoola, A. (2019). The transcription factor TCF-1 enforces commitment to the innate lymphoid cell lineage. *Nature Immunology*, *20*, 1150–1160.

Hastie, T., & Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association*, *84*, 502–516.

Hawkins, E. D., Markham, J. F., McGuinness, L. P., & Hodgkin, P. D. (2009). A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proceedings of the National Academy of Sciences*, *106*, 13457–13462.

He, B., Xing, S., Chen, C., Gao, P., Teng, L., Shan, Q., … Xue, H.-H. (2016). CD8 + T Cells Utilize Highly Dynamic Enhancer Repertoires and Regulatory Circuitry in Response to Infections. *Immunity*, *45*, 1341–1354.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., … Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, *38*, 576–589.

Henning, A. N., Roychoudhuri, R., & Restifo, N. P. (2018). Epigenetic control of CD8+ T cell differentiation. *Nature Reviews Immunology*, *18*, 340–356.

Herndler-Brandstetter, D., Ishigame, H., Shinnakasu, R., Plajer, V., Stecher, C., Zhao, J., … Flavell, R. A. (2018). KLRG1+ Effector CD8+ T Cells Lose KLRG1, Differentiate into All Memory T Cell Lineages, and Convey Enhanced Protective Immunity. *Immunity*, *48*, 716–729.

Hodgkin, P. D. (2018). Modifying clonal selection theory with a probabilistic cell. *Immunological Reviews*, *285*, 249–262.

Hou, S., Hyland, L., Ryan, K. W., Portner, A., & Doherty, P. C. (1994). Virus-specific CD8+ T-cell memory determined by clonal burst size. *Nature*, *369*, 652–654.

Hou, Y., Zak, J., Shi, Y., Pratumchai, I., Dinner, B., Wang, W., … Wu, P. (2023). *Transient EZH2 suppression by Tazemetostat during* in vitro *expansion maintains T cell stemness and improves adoptive T cell therapy* [Preprint]. Immunology.

Hudson, W. H., Gensheimer, J., Hashimoto, M., Wieland, A., Valanparambil, R. M., Li, P., … Ahmed, R. (2019). Proliferating Transitory T Cells with an Effector-like Transcriptional Signature Emerge from PD-1+ Stem-like CD8+ T Cells during Chronic Infection. *Immunity*, *51*, 1043-1058.e4.

Im, S. J., Hashimoto, M., Gerner, M. Y., Lee, J., Kissick, H. T., Burger, M. C., … Ahmed, R. (2016). Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature*, *537*, 417–421.

Jackson, C. H. (2011). Multi-State Models for Panel Data: The **msm** Package for *R*. *Journal of Statistical Software*, *38*. https://doi.org/10.18637/jss.v038.i08

Jacob, J., & Baltimore, D. (1999). Modelling T-cell memory by genetic marking of memory T cells in vivo. *Nature*, *399*, 593–597.

Jeannet, G., Boudousquie, C., Gardiol, N., Kang, J., Huelsken, J., & Held, W. (2010). Essential role of the Wnt pathway effector Tcf-1 for the establishment of functional CD8 T cell memory. *Proceedings of the National Academy of Sciences*, *107*, 9777–9782.

Joshi, N. S., Cui, W., Chandele, A., Lee, H. K., Urso, D. R., Hagman, J., … Kaech, S. M. (2007). Inflammation Directs Memory Precursor and Short-Lived Effector CD8+ T Cell Fates via the Graded Expression of T-bet Transcription Factor. *Immunity*, *27*, 281–295.

Joung, J., Konermann, S., Gootenberg, J. S., Abudayyeh, O. O., Platt, R. J., Brigham, M. D., … Zhang, F. (2017). Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols*, *12*, 828–863.

Kaech, S. M., & Ahmed, R. (2001). Memory CD8+ T cell differentiation: Initial antigen encounter triggers a developmental program in naïve cells. *Nature Immunology*, *2*, 415–422.

Kaech, S. M., & Cui, W. (2012). Transcriptional control of effector and memory CD8+ T cell differentiation. *Nature Reviews Immunology*, *12*, 749–761.

Kaech, S. M., Tan, J. T., Wherry, E. J., Konieczny, B. T., Surh, C. D., & Ahmed, R. (2003).

Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nature Immunology*, *4*, 1191–1198.

Kaech, S. M., Wherry, E. J., & Ahmed, R. (2002). Effector and memory T-cell differentiation: Implications for vaccine development. *Nature Reviews Immunology*, *2*, 251–262.

Kakaradov, B., Arsenio, J., Widjaja, C. E., He, Z., Aigner, S., Metz, P. J., … Yeo, G. W. (2017). Early transcriptional and epigenetic regulation of CD8+ T cell differentiation revealed by single-cell RNA sequencing. *Nature Immunology*, *18*, 422–432.

Kaminskiy, Y., Kuznetsova, V., Kudriaeva, A., Zmievskaya, E., & Bulatov, E. (2022). Neglected, yet significant role of FOXP1 in T-cell quiescence, differentiation and exhaustion. *Frontiers in Immunology*, *13*. Retrieved from https://www.frontiersin.org/articles/10.3389/fimmu.2022.971045

Kempton, H. R., Love, K. S., Guo, L. Y., & Qi, L. S. (2022). Scalable biological signal recording in mammalian cells using Cas12a base editors. *Nature Chemical Biology*, *18*, 742–750.

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021, February 1). *Fast gene set enrichment analysis* (p. 060012). p. 060012. bioRxiv.

Kueh, H. Y., Champhekar, A., Nutt, S. L., Elowitz, M. B., & Rothenberg, E. V. (2013). Positive Feedback Between PU.1 and the Cell Cycle Controls Myeloid Differentiation. *Science*, *341*, 670–673.

Kueh, Hao Yuan, Yui, M. A., Ng, K. K. H., Pease, S. S., Zhang, J. A., Damle, S. S., … Rothenberg, E. V. (2016). Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. *Nature Immunology*, *17*, 956–965.

Kurd, N. S., He, Z., Louis, T. L., Milner, J. J., Omilusik, K. D., Jin, W., … Chang, J. T. (2020). Early precursors and molecular determinants of tissue-resident memory CD8 $^{+}$ T lymphocytes revealed by single-cell RNA sequencing. *Science Immunology*, *5*, eaaz6894.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., … Kharchenko, P. V. (2018). RNA velocity of single cells. *Nature*, *560*, 494–498.

Ladle, B. H., Li, K.-P., Phillips, M. J., Pucsek, A. B., Haile, A., Powell, J. D., … Gamper, C. J. (2016). De novo DNA methylation by DNA methyltransferase 3a controls early effector CD8 $^{+}$ T-cell fate decisions following activation. *Proceedings of the National Academy of Sciences*, *113*, 10631–10636.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Li, Z., Tuong, Z. K., Dean, I., Willis, C., Gaspal, F., Fiancette, R., … Withers, D. R. (2022). In vivo labeling reveals continuous trafficking of TCF-1+ T cells between tumor and lymphoid tissue. *Journal of Experimental Medicine*, *219*, e20210749.

Lian, H., Thompson, W. A., Thurman, R., Stamatoyannopoulos, J. A., Noble, W. S., & Lawrence, C. E. (2008). Automated mapping of large-scale chromatin structure in

ENCODE. *Bioinformatics*, *24*, 1911–1916.

Lin, W.-H. W., Nish, S. A., Yen, B., Chen, Y.-H., Adams, W. C., Kratchmarov, R., … Reiner, S. L. (2016). CD8 + T Lymphocyte Self-Renewal during Effector Cell Determination. *Cell Reports*, *17*, 1773–1782.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550.

Lövkvist, C., Mikulski, P., Reeck, S., Hartley, M., Dean, C., & Howard, M. (2021). Hybrid protein assembly-histone modification mechanism for PRC2-based epigenetic switching and memory. *ELife*, *10*, e66454.

Lynn, R. C., Weber, E. W., Sotillo, E., Gennert, D., Xu, P., Good, Z., … Mackall, C. L. (2019). C-Jun overexpression in CAR T cells induces exhaustion resistance. *Nature*, *576*, 293–300.

Manz, M. G., & Boettcher, S. (2014). Emergency granulopoiesis. *Nature Reviews Immunology*, *14*, 302–314.

McKenna, A., Findlay, G. M., Gagnon, J. A., Horwitz, M. S., Schier, A. F., & Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, *353*, aaf7907.

Melo, G. A., Xu, T., Calôba, C., Schutte, A., Brum, G., Passos, T. O., … Pereira, R. M. (2022, October 7). *Polycomb Repressive Complex 1 subunit Cbx4 positively regulates effector responses in CD8 T cells* (p. 2022.10.03.510675). p. 2022.10.03.510675. bioRxiv.

Mescher, M. F., Curtsinger, J. M., Agarwal, P., Casey, K. A., Gerner, M., Hammerbeck, C. D., … Xiao, Z. (2006). Signals required for programming effector and memory development by CD8+ T cells. *Immunological Reviews*, *211*, 81–92.

Murali-Krishna, K., Altman, J. D., Suresh, M., Sourdive, D. J., Zajac, A. J., Miller, J. D., … Ahmed, R. (1998). Counting antigen-specific CD8 T cells: A reevaluation of bystander activation during viral infection. *Immunity*, *8*, 177–187.

Ng, K. K., Yui, M. A., Mehta, A., Siu, S., Irwin, B., Pease, S., … Kueh, H. Y. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. *ELife*, *7*, e37851.

Nguyen, P., Chien, S., Dai, J., Jr, R. J. M., Becker, P. S., & Kueh, H. Y. (2021). Unsupervised discovery of dynamic cell phenotypic states from transmitted light movies. *PLOS Computational Biology*, *17*, e1009626.

Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F., & Johnson, G. R. (2018). Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nature Methods*, *15*, 917–920.

Pai, J. A., Hellmann, M. D., Sauter, J. L., Mattar, M., Rizvi, H., Woo, H. J., … Satpathy, A. T. (2023). Lineage tracing reveals clonal progenitors and long-term persistence of tumor-

specific T cells during immune checkpoint blockade. *Cancer Cell*, *41*, 776-790.e7.

Pais Ferreira, D., Silva, J. G., Wyss, T., Fuertes Marraco, S. A., Scarpellino, L., Charmoy, M., … Held, W. (2020). Central memory CD8+ T cells derive from stem-like Tcf7hi effector cells in the absence of cytotoxic differentiation. *Immunity*, *53*, 1–16.

Pease, N. A., Nguyen, P. H. B., Woodworth, M. A., Ng, K. K. H., Irwin, B., Vaughan, J. C., & Kueh, H. Y. (2021). Tunable, division-independent control of gene activation timing by a polycomb switch. *Cell Reports*, *34*, 108888.

Pircher, H., Bürki, K., Lang, R., Hengartner, H., & Zinkernagel, R. M. (1989a). Tolerance induction in double specific T-cell receptor transgenic mice varies with antigen. *Nature*, *342*, 559–561.

Pircher, H., Bürki, K., Lang, R., Hengartner, H., & Zinkernagel, R. M. (1989b). Tolerance induction in double specific T-cell receptor transgenic mice varies with antigen. *Nature*, *342*, 559–561.

Plambeck, M., Kazeroonian, A., Löffler, D., Schroeder, T., Busch, D. H., Flossdorf, M., & Buchholz, V. R. (2021). *Heritable changes in division speed accompany the diversification of single T cell fate* [Preprint]. Immunology.

Plumlee, C. R., Sheridan, B. S., Cicek, B. B., & Lefrançois, L. (2013). Environmental Cues Dictate the Fate of Individual CD8+ T Cells Responding to Infection. *Immunity*, *39*, 347–356.

Polonsky, M., Rimer, J., Kern-Perets, A., Zaretsky, I., Miller, S., Bornstein, C., … Friedman, N. (2018). Induction of CD4 T cell memory by local cellular collectivity. *Science*, *360*, eaaj1853.

Potter, E. L., Gideon, H. P., Tkachev, V., Fabozzi, G., Chassiakos, A., Petrovas, C., … Roederer, M. (2021). Measurement of leukocyte trafficking kinetics in macaques by serial intravascular staining. *Science Translational Medicine*, *13*, eabb4582.

Pritykin, Y., van der Veeken, J., Pine, A. R., Zhong, Y., Sahin, M., Mazutis, L., … Leslie, C. S. (2021). A unified atlas of CD8 T cell dysfunctional states in cancer and infection. *Molecular Cell*, *81*, 2477-2493.e10.

Qiu, Q., Hu, P., Qiu, X., Govek, K. W., Cámara, P. G., & Wu, H. (2020). Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nature Methods*. https://doi.org/10.1038/s41592-020-0935-4

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, *344*, 1492–1496.

Roth, T. L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P. J., … Marson, A. (2018). Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature*, *559*, 405–409.

Rubinstein, M. P., Lind, N. A., Purton, J. F., Filippou, P., Best, J. A., McGhee, P. A., …

Goldrath, A. W. (2008). IL-7 and IL-15 differentially regulate CD8+ T-cell subsets during contraction of the immune response. *Blood*, *112*, 3704–3712.

Rutishauser, R. L., Martins, G. A., Kalachikov, S., Chandele, A., Parish, I. A., Meffre, E., … Kaech, S. M. (2009). Transcriptional Repressor Blimp-1 Promotes CD8+ T Cell Terminal Differentiation and Represses the Acquisition of Central Memory T Cell Properties. *Immunity*, *31*, 296–308.

Sade-Feldman, M., Yizhak, K., Bjorgaard, S. L., Ray, J. P., de Boer, C. G., Jenkins, R. W., … Hacohen, N. (2018). Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell*, *175*, 998-1013.e20.

Saunders, A., Core, L. J., Sutcliffe, C., Lis, J. T., & Ashe, H. L. (2013). Extensive polymerase pausing during Drosophila axis patterning enables high-level and pliable transcription. *Genes & Development*, *27*, 1146–1158.

Sen, D. R., Kaminski, J., Barnitz, R. A., Kurachi, M., Gerdemann, U., Yates, K. B., … Haining, W. N. (2016). The epigenetic landscape of T cell exhaustion. *Science (New York, N.Y.)*, *354*, 1165–1169.

Sheth, R. U., & Wang, H. H. (2018). DNA-based memory devices for recording cellular events. *Nature Reviews Genetics*, *19*, 718–732.

Smith, J. P., Corces, M. R., Xu, J., Reuter, V. P., Chang, H. Y., & Sheffield, N. C. (2021). PEPATAC: An optimized pipeline for ATAC-seq data analysis with serial alignments. *NAR Genomics and Bioinformatics*, *3*, lqab101.

Soerens, A. G., Künzli, M., Quarnstrom, C. F., Scott, M. C., Swanson, L., Locquiao, J. J., … Masopust, D. (2023). Functional T cells are capable of supernumerary cell division and longevity. *Nature*, *614*, 762–766.

Stemberger, C., Huster, K. M., Koffler, M., Anderl, F., Schiemann, M., Wagner, H., & Busch, D. H. (2007). A Single Naive CD8+ T Cell Precursor Can Develop into Diverse Effector and Memory Subsets. *Immunity*, *27*, 985–997.

Stephen, T. L., Rutkowski, M. R., Allegrezza, M. J., Perales-Puchalt, A., Tesone, A. J., Svoronos, N., … Conejo-Garcia, J. R. (2014). Transforming Growth Factor β-Mediated Suppression of Antitumor T Cells Requires FoxP1 Transcription Factor Expression. *Immunity*, *41*, 427–439.

Suda, T., Suda, J., & Ogawa, M. (1984). Disparate differentiation in mouse hemopoietic colonies derived from paired progenitors. *Proceedings of the National Academy of Sciences of the United States of America*, *81*, 2520–2524.

Taylor, J. J., Pape, K. A., Steach, H. R., & Jenkins, M. K. (2015). Apoptosis and antigen affinity limit effector cell differentiation of a single naive B cell. *Science*, *347*, 784–787.

Tello-Ramos, M. C., Branch, C. L., Kozlovsky, D. Y., Pitera, A. M., & Pravosudov, V. V. (2019). Spatial memory and cognitive flexibility trade-offs: To be or not to be flexible,

that is the question. *Animal Behaviour*, *147*, 129–136.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., … Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*, 381–386.

Tsui, C., Kretschmer, L., Rapelius, S., Gabriel, S. S., Chisanga, D., Knöpper, K., … Kallies, A. (2022). MYB orchestrates T cell exhaustion and response to checkpoint inhibition. *Nature*, *609*, 354–360.

Utzschneider, D. T., Charmoy, M., Chennupati, V., Pousse, L., Ferreira, D. P., Calderon-Copete, S., … Held, W. (2016). T Cell Factor 1-Expressing Memory-like CD8+ T Cells Sustain the Immune Response to Chronic Viral Infections. *Immunity*, *45*, 415–427.

Utzschneider, D. T., Gabriel, S. S., Chisanga, D., Gloury, R., Gubser, P. M., Vasanthakumar, A., … Kallies, A. (2020). Early precursor T cells establish and propagate T cell exhaustion in chronic infection. *Nature Immunology*, *21*, 1256–1266.

Verbist, K. C., Guy, C. S., Milasta, S., Liedmann, S., Kamiński, M. M., Wang, R., & Green, D. R. (2016). Metabolic maintenance of cell asymmetry following division in activated T lymphocytes. *Nature*, *532*, 389–393.

Vijh, S., & Pamer, E. G. (1997). Immunodominant and subdominant CTL responses to Listeria monocytogenes infection. *The Journal of Immunology*, *158*, 3366–3371.

Wagar, L. E., Salahudeen, A., Constantz, C. M., Wendel, B. S., Lyons, M. M., Mallajosyula, V., … Davis, M. M. (2021). Modeling human adaptive immune responses with tonsil organoids. *Nature Medicine*, *27*, 125–135.

Wagner, D. E., & Klein, A. M. (2020). Lineage tracing meets single-cell omics: Opportunities and challenges. *Nature Reviews Genetics*, *21*, 410–427.

Wang, Z., Fang, Z., Chen, G., Liu, B., Xu, J., Li, F., … Ji, H. (2021). Chromobox 4 facilitates tumorigenesis of lung adenocarcinoma through the Wnt/β-catenin pathway. *Neoplasia*, *23*, 222–233.

Weaver, C. T., Saparov, A., Kraus, L. A., Rogers, W. O., Hockett, R. D., & Bucy, R. P. (1998). Heterogeneity in the clonal T cell response: Implications for models of T Cell activation and cytokine phenotype development. *Immunologic Research*, *17*, 279–302.

Weinreb, C., & Klein, A. M. (2020). Lineage reconstruction from clonal correlations. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 17041–17048.

Wherry, E. J., Ha, S.-J., Kaech, S. M., Haining, W. N., Sarkar, S., Kalia, V., … Ahmed, R. (2007). Molecular Signature of CD8+ T Cell Exhaustion during Chronic Viral Infection. *Immunity*, *27*, 670–684.

Wu, L., Pan, T., Zhou, M., Chen, T., Wu, S., Lv, X., … Ma, X. (2022). CBX4 contributes to HIV-1 latency by forming phase-separated nuclear bodies and SUMOylating EZH2.

EMBO Reports, *23*, e53855.

Xiao, Z., Casey, K. A., Jameson, S. C., Curtsinger, J. M., & Mescher, M. F. (2009). Programming for CD8 T Cell Memory Development Requires IL-12 or Type I IFN. *The Journal of Immunology*, *182*, 2786–2794.

Xin, T., Gonzalez, D., Rompolas, P., & Greco, V. (2018). Flexible fate determination ensures robust differentiation in the hair follicle. *Nature Cell Biology*, *20*, 1361–1369.

Yang, D., Jones, M. G., Naranjo, S., Rideout, W. M., Min, K. H. (Joseph), Ho, R., … Weissman, J. S. (2022). Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell*, *185*, 1905-1923.e25.

Youngblood, B., Hale, J. S., Kissick, H. T., Ahn, E., Xu, X., Wieland, A., … Ahmed, R. (2017). Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature*, *552*, 404–409.

Zhang, M., Gao, D., Shi, Y., Wang, Y., Joshi, R., Yu, Q., … Min, W. (2019). MiR-149-3p reverses CD8+ T-cell exhaustion by reducing inhibitory receptors and promoting cytokine secretion in breast cancer cells. *Open Biology*, *9*, 190061.

Zhao, X., Shan, Q., & Xue, H.-H. (2021). TCF1 in T cell immunity: A broadened frontier. *Nature Reviews Immunology*. https://doi.org/10.1038/s41577-021-00563-6

Zhu, J., Jankovic, D., Oler, A. J., Wei, G., Sharma, S., Hu, G., … Paul, W. E. (2012). The Transcription Factor T-bet Is Induced by Multiple Pathways and Prevents an Endogenous Th2 Cell Program during Th1 Cell Responses. *Immunity*, *37*, 660–673.

Mathematical Appendix

Contents

1 Introduction

The immune system remembers prior pathogen encounters through the production of memory B- and T cells during an acute infection. It stores information about the nature and severity of the encountered pathogen, through the number, antigen specificity and functional states of the memory cells that form. The quantity of T cells generated, in particular, encodes information of the severity of a prior infection; across a range of pathogenic challenges, it has been observed that the number of memory cells that arise is a fixed fraction of the total number of T cells at the peak of the infection (Murali-Krishna et al. 1998; Vijh and Pamer 1997; Busch et al. 1998). This linear scaling in memory production allows the body to generate a proportionally larger quantity of memory when faced with a more severe pathogenic challenge.

This scalable control of memory cell generation may reflect a consequence of the fate decisions made by responding T cells during the course of infection. In the face of stimulatory signals, T cells decide whether to differentiate into a short-lived effector cell, or to maintain self-renewal potential and become a memory cell once the pathogen is cleared. Through our single-cell tracking studies, we have identified a strategy, whereby T cells initially retain memory potential shortly after activation, but stochastically transition into an effector state with a probability that increases with the strength of antigen stimulation. Importantly, the decision to become an effector cell is flexible, and this effector dedifferentiation occurs when antigenic signals are no longer present.

Here, we analyze the population dynamics of CD8 T cells to characterize memory formation using this flexible decision making strategy. We specifically evaluate whether this strategy permits quantitative encoding of infection severity through such scalable modulation of memory population sizes. To do so, we develop a mathematical framework to model the immune response to acute infection, that incorporates pathogen replication, control by effector T cells, as well as the flexible decision making strategy discussed above ("Flexible decision model"). Additionally, our model explicitly accounts for activity of the innate immune

system, which acts both as a first-line responder and also as a executor of T cell-directed effector activity. We further compare the memory encoding capabilities of the flexible strategy with two other decision making strategies proposed in the literature: one where all memory cells arise early, as a result of irreversibility in the decision to become short-lived effector cells ("Early decision model") (Lin et al. 2016; Chang et al. 2007; Kakaradov et al. 2017); and another, where all memory cells arise from effector cells, as a result of a direct conversion of naive cells to effector cells following antigen encounter ("Late decision model") (Youngblood et al. 2017; Bannard, Kraman, and Fearon 2009; Jacob and Baltimore 1999). By comparing the performance of these different models using modeling, we aim to gain insight as to why certain decision-making strategies may have been functionally beneficial for pathogen defense by the immune system and hence, selected for during evolution.

2 Models of memory formation

Flexible decision model

Here, we describe the reversible model for T cell memory decision making (Fig. 2.12). In this model, 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 increasing pathogen load. Both memory precursors and effector cells proliferate with a similar rate that increases with antigen level, as previously observed (Plambeck et al. 2022); 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 consider the innate immune system in both its inactive and active states, and collectively model the innate immune response in these two states using two variables. T cells and innate cells then mediate pathogen killing, both independently from each other and also in a cooperative manner. This model is described by the following system of equations:

$$
\begin{aligned}
\text{naive T cell:} \quad & \frac{\mathrm{d}T_n}{\mathrm{d}t} = -\alpha_n v T_n \\
\text{memory T cell:} \quad & \frac{\mathrm{d}T_m}{\mathrm{d}t} = \alpha_n v T_n + \left(\frac{v}{v+K_m}\right)\gamma_m T_m - \left(\frac{v}{v+K_{m,e}}\right)\alpha_{m,e} T_m + \left(\frac{K_{e,m}}{v+K_{e,m}}\right)\beta_{e,m} T_e \\
\text{effector T cell:} \quad & \frac{\mathrm{d}T_e}{\mathrm{d}t} = \left(\frac{v}{v+K_e}\right)\gamma_e T_e + \left(\frac{v}{v+K_{m,e}}\right)\alpha_{m,e} T_m - \left(\frac{K_{e,m}}{v+K_{e,m}}\right)\beta_{e,m} T_e - \delta_e T_e \\
\text{pathogen:} \quad & \frac{\mathrm{d}v}{\mathrm{d}t} = \left(\frac{v}{\epsilon+v}\right)^N \gamma_v v - \left(\delta_{v_1} T_e + \delta_{v_2} z^* + \delta_{v_3} z^* T_e\right) v \\
\text{inactive innate cell:} \quad & \frac{\mathrm{d}z}{\mathrm{d}t} = -\alpha_z v\, z \\
\text{active innate cell:} \quad & \frac{\mathrm{d}z^*}{\mathrm{d}t} = \alpha_z v\, z + \left(\frac{v}{v+K_{z^*}}\right)\gamma_{z^*}\, z^* - \delta_{z^*}\, z^*
\end{aligned}
\tag{1}
$$

Here, the subscripts n, m, and e denote naive, memory, and effector types, respectively, v denotes the pathogen population, and z and z^* denote the inactive and active innate immune cells, respectively. The rates $\alpha_{x,y}$ denote differentiation rates from the x to the y cell type (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 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 for half maximal rate of a process indicated in the subscript.

A detailed description of the model variables, parameters and initial conditions are given in Tables 1, 2. Parameters have been chosen based on the immune compartment sizes as measured in mice, as well as T cell biological parameters that we and others have measured. The initial conditions have been chosen to reflect the initial onset of an infection by a pathogen for which no prior immunological memory has been developed (Table 1); 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.

	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 1: **Variables and initial conditions for the reversible epigenetic switch model**

We point out that this system is fundamentally a Lotka-Volterra model where immune cells are predators and pathogens are prey. However, we have modified this framework to 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 further 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 (Fowler 2021). However, we note that the dynamics of simulated response does 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. 2.12B, Fig. 2.13A-B). Upon introduction to the system, pathogens increase exponentially in number, 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. During the course of the immune response, the numbers of activated innate immune cells increases rapidly and decreases steadily for the reminder of simulation (Fig. 2.13A-B). This heightened innate immune activity is critical for ensuring that pathogens clear after T cell contraction and do not rebound in number.

How does the size of the generated memory population depend on the severity of infection

	Parameter description	Value
α_n	naive cell activation rate constant	5/(units/µL)/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	8×10^{-4}/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	4.5×10^{-3}/(cells/µL)/hr
δ_{v_2}	rate constant for innate cell pathogen killing	1.5×10^{-3}/(cells/µL)/hr
δ_{v_3}	rate constant for T-cell assisted innate cell pathogen killing	1.5×10^{-3}/(cells/µL)$^{-2}$/hr
α_z	innate cell activation rate constant	2.5/(units/µL)/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	$2\times10^{-4}/hr$

Table 2: **Parameters for the reversible epigenetic switch model**

in the reversible switching model? In particular, 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 (Murali-Krishna et al. 1998). To ask this question, we performed simulations of the system 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. 2.12B).

However, the fraction f_{T_m} of memory cells to the total number of T cells present at the expansion peak becomes a fixed number (Fig. 2.12B, 6C: 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. 2.12C, top, shaded area). On the other hand, for slowly growing pathogens (small γ_v), the memory fraction f_{T_m} increases, with a non-linear inverse dependence on the virulence and the average viral load accumulated during the infection (Fig. 2.12C); in this regime, the number of memory cells depends strongly on the initial number of naive cells present. In summary, these results show that a flexible switching strategy for T cell memory generation allows for the amount of the generated T cell memory to scale with the size of the T cell response, in a way that depends on the severity of the infection. Our analytical results in the following section well recapitulates the behavior of this memory fraction, as indicated in Fig. 2.12.

Analytical results for the flexible decision model

Consider the dynamics of memory T_m and effector T_e populations, given by eq. 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. 1), necessary for different processes. Below the terms that are relevant for the dynamics of memory and effetor 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:

$$\text{memory T cell:} \quad \frac{\mathrm{d}T_m}{\mathrm{d}t} = \underbrace{\alpha_n v T_n + \left(\frac{v}{v+K_m}\right)\cdot\gamma_m T_m - \left(\frac{v}{v+K_{m,e}}\right)\cdot\alpha_{m,e}T_m}_{v\gg K_m, K_{m,e}} + \underbrace{\left(\frac{K_{e,m}}{v+K_{e,m}}\right)\cdot\beta_{e,m}T_e}_{v\ll K_{e,m}} \tag{2}$$

$$\text{effector T cell:} \quad \frac{\mathrm{d}T_e}{\mathrm{d}t} = \underbrace{\left(\frac{v}{v+K_e}\right)\cdot\gamma_e T_e + \left(\frac{v}{v+K_{m,e}}\right)\cdot\alpha_{m,e}T_m}_{v\gg K_e, K_{m,e}} - \underbrace{\left(\frac{K_{e,m}}{v+K_{e,m}}\right)\cdot\beta_{e,m}T_e}_{v\ll K_{e,m}} - \delta_e T_e. \tag{3}$$

We can formally integrate over the dynamical equation in eq. 2 to find a formal solution for the number of memory cells $T_m(t)$ at time t post infection,

$$\begin{aligned} T_m(t) &= T_m(0)\exp\left[\int_0^t \gamma_m - \alpha_{m,e} - \frac{\gamma_m K_m}{v(s)+K_m} + \frac{\alpha_m K_{m,e}}{v(s)+K_{m,e}}\mathrm{d}s\right] \\ &\quad + \left[\int_0^t \exp\left[\int_s^t \gamma_m - \alpha_{m,e} - \frac{\gamma_m K_m}{v(s)+K_m} + \frac{\alpha_m K_{m,e}}{v(s)+K_{m,e}}\mathrm{d}s\right]\right. \\ &\quad \left.\times\left(\alpha_n v(s)T_n(s) + \frac{K_{e,m}}{v(s)+K_{e,m}}\beta_{e,m}T_e(s)\right)\mathrm{d}s\right] \end{aligned} \tag{4}$$

$$\begin{aligned} &\approx T_m(0)\exp\left[(\gamma_m - \alpha_{m,e})\min(t,\tau_{v\approx 0})\right] \\ &\quad + \int_0^{\min(t,\tau_{v\approx 0})} \exp\left[(\gamma_m - \alpha_{m,e})(\min(t,\tau_{v\approx 0}) - s)\right]\alpha_n v(s)T_n(s)\mathrm{d}s \\ &\quad + H(t-\tau_{v\approx 0})\int_{\tau_{v\approx 0}}^t \beta_{e,m}T_e(s)\mathrm{d}s, \end{aligned} \tag{5}$$

where $\tau_{v\approx 0}$ is the time to effectively clear the infection, and $H(t-\tau_{v\approx 0})$ is a heaviside step function that takes value 1 for $t > \tau_{v\approx 0}$, and 0, otherwise. In arriving at eq. 5, we assumed that the typical viral load over the course of the infection is much higher than the differentiation thresholds $K_m, K_{m,e}, K_e$, and $K_{e,m}$, and thus, we approximated these processes by their maximal rates in eq. 4.

The following terms are important in determining the size of the memory pool:

1. $T_m(0)$: the initial memory size;

2. $\gamma_m - \alpha_{m,e}$: the effective growth rate of the memory pool;

3. $\tau_{v\approx 0}$: time to effectively clear the infection;

4. $\beta_{e,m}$: the transition rate from effector to memory.

Here, we are interested in an immune response to a primary infection, and therefore, we can assume that $T_m(0) = 0$. Moreover, given the rates reported in Table 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. 5 follows,

$$T_m(t) \approx \int_0^{\min(t,\tau_{v\approx 0})} \alpha_n v(s) T_n(s) ds + H(t - \tau_{v\approx 0}) \int_{\tau_{v\approx 0}}^{t} \beta_{e,m} T_e(s) ds.$$

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

$$T_m(t \to \infty) \approx \int_0^{\tau_{v\approx 0}} \alpha_n v(s) T_n(s) \mathrm{d}s + \int_{\tau_{v\approx 0}}^{\infty} \beta_{e,m} T_e(s) \mathrm{d}s,$$

then assuming $\tau_{v\approx 0} \approx \tau_{e^{\max}}$,

$$\begin{aligned} T_m(\infty) &\approx \int_0^{\tau_{e^{\max}}} \alpha_n v(s) T_n(s) \mathrm{d}s + \int_{\tau_{e^{\max}}}^{\infty} \beta_{e,m} T_e(s) \mathrm{d}s \\ &\approx \int_0^{\tau_{e^{\max}}} \alpha_n v(s) T_n(s) \mathrm{d}s + \int_{\tau_{e^{\max}}}^{\infty} \beta_{e,m} T_e^{\max} e^{-(\beta_{e,m}+\delta_e)s} \mathrm{d}s \\ &\approx \int_0^{\tau_{e^{\max}}} \alpha_n v(s) T_n(s) \mathrm{d}s + \beta_{e,m} \frac{T_e^{\max}}{\beta_{e,m} + \delta_e} \end{aligned} \tag{6}$$

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 ($v \ll K_{e,m}$) , from eq. 3.

From eq. 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. 6 follows,

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

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

$$T_m(\infty) \approx T_n(0) \left[1 - e^{-\alpha_n \tilde{V}}\right] + \beta_{e,m} \frac{T_e^{\max}}{\beta_{e,m} + \delta_e}, \tag{7}$$

resulting the following memory fraction,

$$f_{T_m} = \frac{T_m(\infty)}{T_e^{\max}} \approx \frac{x(0)}{T_e^{\max}} \left[1 - e^{-\alpha_n \tilde{V}}\right] + \frac{\beta_{e,m}}{\beta_{e,m} + \delta_e}. \tag{8}$$

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, the first term in our expression for f_{T_m} becomes,

$$\begin{aligned} C_0 &= \frac{1}{T_e^{\max}} \int_0^{\tau_{e^{\max}}} \exp\left[(\gamma_m - \alpha_{m,e})(\tau_{e^{\max}} - s)\right] \alpha_n v(s) T_n(s) \mathrm{d}s \\ &= \frac{\alpha_n T_n(0)}{T_e^{\max}} \int_0^{\tau_{e^{\max}}} \exp\left[(\gamma_m - \alpha_{m,e})(\tau_{e^{\max}} - s) - \alpha_n \int_0^s v(r)\mathrm{d}r\right] v(s)\mathrm{d}s \\ &\leq \frac{\alpha_n T_n(0)}{T_e^{\max}} \int_0^{\tau_{e^{\max}}} \exp\left[(\gamma_m - \alpha_{m,e})(\tau_{e^{\max}} - s)\right] v(s)\mathrm{d}s, \end{aligned} \tag{9}$$

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.

Early decision model

From experimental studies, it has been proposed that memory cells originate primarily from cells that have undergone little or no effector differentiation, and that memory precursors, upon silencing the memory regulator TCF1 and differentiating, are committed to becoming short-lived effectors (Lin et al. 2016; Pais Ferreira et al. 2020). Using the mathematical modeling framework developed above, we evaluate whether this irreversible effector decision strategy could also enable the asymptotic (long-term) memory T cell numbers $T_m(\infty)$ to scale linearly with the peak T cell number (primary effector cells) $T_e^{\max}$. To do so, we performed simulations of the above model, rendering effector differentiation irreversible by setting the rate of effector de-differentiation $\beta_{e,m}$ to zero.

From our simulations, we found that the number of memory T cells emerging depends on the balance between effector differentiation and memory precursor proliferation, as determined by the rate constants $\alpha_{m,e}$ and γ_m respectively. When these two rate constants are equal, the number of generated memory cells cannot exceed the initial number of naive cells (Fig. 2.13C-F). This is because the memory cell population, upon emerging from the naive cell pool, cannot further change in number as proliferation is balanced exactly by differentiation. We note that this regime captures the dynamics of obligate asymmetric division, where the division of each memory precursor necessarily gives rise to a precursor and a differentiated progeny. When the rate of effector differentiation $\alpha_{m,e}$ is smaller than that of memory precursor proliferation γ_m, the number of memory cells can exceed the initial naive cell number, due to a net proliferation of this population; in this regime, the size of the memory pool grows with increasing γ_v.

To perform a mathematically comparable comparison of the flexible and irreversible switching models (Alves and Savageau 2000), 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 irreversible switching model is unable to generate a constant fraction of memory cells amid changes in pathogen replication rates (Fig. 2.12C middle). The memory fraction was upheld at a similar value $f_{T_m} = 0.05$ at moderate pathogen replication rates $\gamma_v = 0.04/hr$; however, this fraction decreased steadily with increasing γ_v, eventually approaching less than 0.01 at high pathogen replication rates (Fig. 2.12C, middle; Fig. 2.13C-F). Indeed, from an approximate analytical solution of this system, we found that the memory cell fraction has an inherently inverse dependence on the peak T cell population size, as follows:

$$f_{T_m} \approx \frac{T_n(0)}{T_e^{\max}} \left[1 - e^{-\alpha_n \tilde{V}}\right] \tag{10}$$

This dependence cannot be offset when the fraction is much smaller than unity, which typically holds in the regime where the majority of cells generated at the height of an acute infection are those with effector function. Thus, from these results, we conclude that the T cell decision making strategy, where the memory precursors switch irreversibly into becoming short-lived effectors, cannot produce memory T cell numbers that scale proportionally with the peak T cell population sizes.

Late decision model (obligate reversible decision model)

An alternate strategy for memory differentiation is that, upon activation, all naive cells must first pass through an effector stage prior to the decision as whether to differentiate into memory (Fig. 2.12, bottom). This view is supported by the evidence that cells with a history of effector gene expression can become memory cells (Bannard, Kraman, and Fearon 2009), and that cells on the road 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 (Youngblood et al. 2017); this model can be thought as an obligate reversible decision model.

To evaluate such a decision making strategy, we alter our model above, such that naive cells, upon activation, directly transition to an effector state instead of a memory precursor state (Fig. 2.12, bottom, Fig. 2.13G-H). Following the model in eq. 1, the ordinary differential equations describing the T cell populations in the late decision model are modified as follows:

$$
\begin{aligned}
\text{naive T cell:} \quad \frac{\mathrm{d}T_n}{\mathrm{d}t} &= -\alpha_n v T_n \\
\text{memory T cell:} \quad \frac{\mathrm{d}T_m}{\mathrm{d}t} &= \big(\frac{v}{v+K_m}\big)\cdot\gamma_m T_m - \big(\frac{v}{v+K_{m,e}}\big)\cdot\alpha_{m,e}T_m + \big(\frac{K_e}{v+K_e}\big)\cdot\beta_{e,m}T_e \\
\text{effector T cell:} \quad \frac{\mathrm{d}T_e}{\mathrm{d}t} &= \alpha_n v T_n + \big(\frac{v}{v+K_e}\big)\cdot\gamma_e T_e + \big(\frac{v}{v+K_{m,e}}\big)\cdot\alpha_{m,e}T_m \\
&\quad - \big(\frac{K_{e,m}}{v+K_{e,m}}\big)\cdot\beta_{e,m}T_e - \delta_e T_e
\end{aligned}
$$

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. When pathogens replicate rapidly and give rise to a substantial T cell expansion, memory cells form robustly at defined fraction and number, similar to the flexible switching model (Fig. 2.12C, bottom, Fig. 2.13G-H, center, right); 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 the immune system to respond to a secondary challenge is no longer heightened, and is likely compromised.

Analysis of late decision model

Following the analytical analyses for the flexible switching model, we can again identify the dominant terms in different regimes of viral load. By assuming that the viral load triggered by the infection is much larger than the cellular differentiation thresholds $K_m, K_{m,e}, K_e$, and $K_{e,m}$, we arrive at the following approximate expression for the size of the memory pool $T_m(t)$ at time t in the late decision model.

$$
T_m(t) \approx T_m(0)\exp\left[(\gamma_m - \alpha_{m,e})\min(t, \tau_{v\approx 0})\right] + H(t-\tau_{v\approx 0})\int_{\tau_{v\approx 0}}^{t}\beta_{e,m}T_e(s)\mathrm{d}s.
$$

Then for primary immune response ($T_m(0) = 0$) we have,

$$
f_{T_m} = \frac{T_m(\infty)}{T_e^{\max}} \approx \frac{\beta_{e,m}}{\beta_{e,m}+\delta_e}. \tag{11}
$$

We note that this expression does not hold in the regime of slow/inefficient virus dynamics, consistent with results from simulations (Fig. 2.12).

VITA

Kathleen Victoria Abadie was born and raised in Montara, California. She did a B.S. in Chemical Engineering at Rice University and then worked in process engineering at Genentech for three years prior to starting her graduate degree in Bioengineering at the University of Washington.