THE UNIVERSITY OF CHICAGO

FURTHERING UNDERSTANDING OF T AND CAR-T CELL DIFFERENTIATION USING CUTTING-EDGE SEQUENCING AND IMAGING TECHNOLOGIES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE PRITZKER SCHOOL OF MOLECULAR ENGINEERING

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

GUOSHUAI CAO

CHICAGO, ILLINOIS

JUNE 2021

This dissertation is dedicated to

my parents, Xianfeng Cao and Congli Zhang,

and my brother, Shuaiquan Cao

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

There are truly too many people to thank, and without them, I would not have come this far, and this thesis work would not be possible.

First and foremost, I would like to express my deepest appreciation to my advisor, Dr. Jun Huang. Jun is an incredibly passionate, open-minded, and adventurous scientist and mentor, and I feel genuinely privileged that he took me on as a student even when I came with a Physics background and knew little about immunology. Jun provided me generous support and encouragement during the early stage of my Ph.D., when I was slowly picking up knowledge and skills in bioinformatics and immunoengineering. Later, after I passed my candidacy exam and finished my first project, he gave me all the space and freedom to continue to explore, learn and conduct analyses at my own pace in any topic of my own choosing. Jun holds high standards on our academic work, always encouraging us to push and achieve the best of ourselves. With his mentorship, I have grown tremendously as a scientist and bioinformatician.

I would also like to thank my committee members, Prof. Melody Swartz, Prof. Jeffrey Hubbell, and Dr. Mengjie Chen for their guidance, support, and encouragement. Particularly during my candidacy exam, their acknowledge of my early work and quick consensus on my passing greatly boosted my confidence and encouraged me to stay in this field and arrive where I am today.

I am incredibly grateful also to my fellow labmates in the Huang lab: Jill Rosenberg, Yifei Hu, Yu Zhao, Nick Asby, Devin Harrison, Dr. Nick Ankenbruck, Andrés Moya-Rodriguez, Tony Pan, Dr. Xiaolei Cai, Dr. Min Chen, Thao Cao, Erting Tang, and Laura Pulido. We were all very close and incredibly supportive to each other. I want to give my special thanks to Jill and Yifei,

for their continuous patience, support, and countless intellectual discussions in working with me in the collaborative projects in this thesis work.

I am also greatly indebted to Prof. Lilin Ye and Xiangyu Chen, for their generous offering of many collaborative projects, high-quality experimental work that complements and supports my analyses, and rigorous discussions and critical feedback in our work. Xiangyu has also been a great company during my PhD career, as we have been learning enormously from each other and grew to better scientists together. Prof. Ye has been one of my role model scientists, and he is extremely knowledgeable in immunology and rigorous in approaching science. I have truly learnt a lot from him.

Lastly, nothing in my life, this thesis work included, would have been possible without my family and friends. I am very grateful for my lifelong friends, who bring so much joy, company and meaning to my life: Yu Zhao, Yijun Quan, Feiyang He, Zhenyu Jin, Jingping Li, Bonan Zhu, and Xin Ai. I would like to give special thanks to Yu, who is not only a great lab mate, but also my best friend who helped me enormously in my later Ph.D days. Finally, I want to thank my parents and my brother Quan, who are forever with me when I feel down or lost and always give me everything to support my growth and wellbeing. My parents have always taught me to appreciate everything I have and live a happy and meaningful life by contributing back to the society and leave behind some legacy that can improve others' lives. They have always supported me in the pursuit of science, even though no one in my family had ever had an undergraduate degree before. Their faith in me allows me to clear all doubts and persist through all challenges whenever I wanted to give up. Without them, none of this would be possible.

ABSTRACT

Fundamental questions in T cell differentiation remain unanswered: 1) EZH2's role in Tfh fate commitment; 2) differences of T_{EX} differentiation in Tumor and in Chronic infection; 3) CAR-T cells post-infusion differentiation. Besides, there is no high-dimensional quantitative analysis and machine learning pipeline for TCR signaling studies. This thesis work addressed these issues using advanced sequencing and imaging technologies. Firstly, using ATAC-seq, we found that EZH2-mediated H3K27me3 modifications regulate early Tfh fate commitment, but not late Tfh differentiation or memory maintenance. Secondly, through single-cell RNA- and ATAC-sequencing, we revealed that terminal effector-associated genes and loci accessibility were preferentially enriched in Tex cells in chronic viral infection, while stem/memory-related genes and their loci accessibility more enriched in Tex subsets in tumor, leading to differential responsiveness to PD-L1 immune checkpoint blockade (ICB). Thirdly, through longitudinal, single-cell multi-omic sequencing of YESCARTA CAR-T cells in patients with DLBCL, we characterized the heterogeneity of both CAR-T and endogenous T cells. We also found post infusion enrichment of effector memory/effector $CD8^+$ CAR-T cells and regulatory CAR-T cells in responders and non-responders respectively. We also discovered an IRF7-centered regulatory module that can successfully predict clinical response. Furthermore, we developed an end-to-end pipeline that combines high resolution 4D LLSM data with machine learning and dimensionality reduction techniques to analyze TCR microcluster dynamics and predict T-cell signaling states. Taken together, this work demonstrates the power of single-cell sequencing and imaging technologies in enabling complex, dynamic, and deep-learning-based research on the molecular regulatory mechanisms and signaling patterns involved in T cell differentiation.

CHAPTER 1: INTRODUCTION

Overview of adaptive immunity and T cells

The mammalian immune system has evolved to a powerful system with two interconnected subcomponents: the innate and the adaptive immune systems. The innate immune system relies on recognition of a conserved set of pathogen-associated molecular patterns (PAMPs) or host-derived damage-associated molecular patterns (DAMPs) through the similarly evolutionary conserved pattern recognition receptors (PRRs) to mount an immediate response to contain pathogens and help activate the adaptive immune system through antigen presentation[1,2]. The adaptive immune system, in contrast, exhibits remarkable antigen recognition diversity and specificity through the combinatorial rearrangement of antigen receptors on T and B lymphocytes[3]. In addition, another cardinal feature of the adaptive immune system is immunological memory, where antigen-specific memory T and B cells are long-lived and maintained post antigen-clearance and can be quickly reactivated often in an enhanced manner for host defense upon antigen rechallenge[4,5].

T cells play particularly central roles in adaptive immune system and they can be classified into two major classes based on the expression of the co-receptors CD8 and CD4: $CD8^+$ T cells, often referred to as cytotoxic T lymphocytes (CTLs), recognize pathogen-derived or tumor-derived antigen peptides presented by major histocompatibility complex (MHC) class I (MHC-I) molecules on the surface of infected or malignant cells and directly kill them[6,7]; $CD4^+$ T cells, also known as helper T cells, provide help to B cells for antibody production and CTLs for optimal population expansion. In addition, they produce various cytokines and chemokines to mediate protection against a wide range of pathogenic microorganisms by enhancing macrophage microbicidal activity and recruiting neutrophils, eosinophils and basophils to cites of infection and

inflammation or in certain cases, by directly eliminating infected cells through cytotoxic killing[8,9]. Furthermore, certain $CD4^+$ T cells, the regulatory T cells (Treg), can also help maintain peripheral tolerance by suppressing the immune response towards self-antigens to prevent autoimmunity[10].

Overview of $CD4^+$ T cells heterogeneity and differentiation

The multi-faceted functions of $CD4^+$ T helper (Th) cells are performed by different subtypes of Th cells, with unique phenotypes, distinct effector functions and varying trafficking properties11. Each Th subtype is a unique lineage that naïve $CD4^+$ T cells can differentiate toward when they are activated by T cell receptor (TCR) binding to pathogenic peptides presented by MHC class II (MHC-II), and the definition of each Th lineage can be simplified to its production of one or more signature cytokines and the expression of a lineage-defining and lineage-driving master transcription factor. More specifically, Th1 cells produce IFN-γ and express T-bet; Th2 cells produce IL-4, IL-5, and IL-13 and express GATA-3; Th9 cells produce IL-9 and IL-10 and express PU.1; Th17 cells produce IL-17A, IL17F and IL-22 and express RORγt; Th22 cells produce IL-22 and express AHR; T follicular helper (Tfh) cells produce IL-21 and express BCL6; Treg in the periphery from naïve precursors express TGF-β, IL-10 and express FOXP3[11].

Despite the above simplification, it has been evident that $CD4^+$ T cell differentiation is a complex process coordinated by multiple transcription factors, the cytokine environment during TCR-mediated activation, and extensive chromatin remodeling events[12,13]. One indispensable class of transcription factors for Th cell fate determination, besides the mater transcription factors, is the signaling transducer and activator of transcription (STAT) proteins. Different members of the STAT proteins are regulated by posttranslational modifications mediated by different cytokine,

and activated STAT proteins then help induce the master transcription factors of different Th fates and together they regulate the production the key Th cytokines. For instance, pSTAT4 is required for Th1 differentiation and pSTAT5 is required for Th2 differentiation. Other transcription factors, induced through either TCR or cytokine-mediated signaling, form an intricate regulatory network to reinforce and fine-tune the Th fate choice and cell function. In addition, it is established that while one master transcription factor is driving a Th fate, the other master transcription factors are not completely shut off but are expressed at low level and can be upregulated to reprogram the Th cell to a different lineage[12]. This demonstrates that $CD4^+$ T cells are remarkably plastic despite their phenotypical diversity.

T follicular helper cell differentiation

Although some of these Th lineages, including the Th1, Th2, and Th17 cells, and their master regulators were established by the early 2000s[14-17], Tfh cells were only recognized as an distinct Th cell lineage a little more than a decade ago when several papers[18-20] independently identified and demonstrated Bcl6 as the master regulator of Tfh. Despite this, the function of Tfh cells were already well studied. While $CD4^+$ T cells have long been known to assist in humoral response by helping B cells make antibodies[21], Tfh was determined to be the dedicated B follicular helper subset with its high expression of CXCR5[22-24], which is a key chemokine receptor that promote Tfh cells to migrate towards the B cell follicles and interact with B cells.

Given the essential helper role of Tfh to B cells, Tfh differentiation is required in virtually all immune response to different diseases. The differentiation of Tfh is thus an early fate choice between Tfh and non-Tfh e.g. Th1, Th2, or Th17, and substantial evidences demonstrate that the

fate decision is made as early as in the first two cell divisions[25-27]. The main factors affecting this fate choice include TCR-signaling strength[27-29], cytokine environment especially IL-6, and costimulatory receptor signaling especially that through ICOS[30].

The essential master transcription factor for Tfh is BCL6, which is a zinc-finger transcription repressor. It has been well established that BCL6 represses BLIMP1[18-20,31], which is maintained at high level in non-Tfh cells and directs the differentiation away from Tfh by inhibiting the expression of canonical Tfh markers, including CXCR5, ICOS, and PD-1. In addition, new evidences recently show that BCL6 act as a hub for the repression of multiple transcriptional networks that inhibit Tfh in a BLIMP1-independent manner[31,32], which reinforce its indispensable role in Tfh fate commitment.

Despite the well-established transcriptional programs that drive Tfh fate commitment, the epigenetic regulation mechanisms that establish and maintain Tfh fate choice are still unclear, and this thesis aims to contribute some understanding on this topic as detailed later.

Epigenetic regulation through H3K27me3 of T cells

The term epigenetics, in the broad sense, refers to the underlying mechanisms that preserve cellular memory and maintain distinctive transcriptional profiles without alterations in the DNA sequence[12,33]. In recent years, epigenetics has adopted a more biochemical meaning, referring to a few specific chemical modifications that control whether a genomic segment or chromatin region is accessible/open for transcription and for regulation by trans-activating factors. Some of these epigenetic modifications include: 1) histone tail modifications e.g. H3K27 trimethylation as a

repressive histone mark; 2) DNA methylation resulting in gene expression; 3) histone variants; 4) chromatin interaction/chromosome conformation[12].

Histone tail modifications, or in short histone modifications, have been extensively studied and demonstrated to play key roles in Th cell differentiations[12,34,35]. These posttranslational modifications are remarkably diverse and include enzyme-mediated acetylation, methylation, phosphorylation, ubiquitination etc., and the equally diverse enzymes that write or erase those histone marks have also been identified. Although the combination of different histone marks at different histone tail positions generate a myriad of histone patterns, a few of them are particularly useful and informative to determine the local chromatin accessibility and gene activity.

One such representative histone mark is H3K27 trimethylation (H3K27me3), which is associated with repressive chromatin conformation that would silence the target genes. The H3K27me3 histone mark was found to be mediated by polycomb repressive complex 2 (PRC2) and the core PRC2 complex contains SUIZ12, EED, RpAp46/48, which are responsible for the resulting chromatin compaction, and the critical enzymatic functional component EZH2, which is mainly responsible for the trimethylation[36,37]. Substantial evidence in mice shows that EZH2-containing PRC2 is critical for T cell differentiation: it inhibits the differentiation and plasticity of naive T cells[38], inhibits expression of apoptotic molecules to promote survival of effector Th cells[38], , and it is required for differentiation of terminal effector $CD8^+$ T cells through repression of pro-memory genes[39].

Given this critical role of EZH2 containing PRC2 mediated H327me3 in T cell differentiation, it is reasonable to hypothesize that it also plays a critical role in Tfh differentiation. As alluded to in the previous section, such epigenetic event in Tfh development remains to be

studied, and this thesis aims to investigate the role of EZH2 in Tfh differentiation and attempts to elucidate the link to the known transcriptional regulations in Tfh.

Overview of $CD8^+$ T cells heterogeneity and Differentiation

During an acute viral infection, antigen-specific $CD8^+$ T cells are activated and undergo extensive proliferation and differentiation that can be characterized by three distinct phases: 1) clonal expansion; 2) contraction of T cell population; 3) Memory formation. During clonal expansion, naïve $CD8^+$ T cells robustly proliferate, undergo significant metabolic and transcriptional remodeling to acquire effector functions, and migrate to the site of infection to eliminate the pathogen. In the contraction state after antigen clearance, majority of the $CD8^+$ T cells (>90%) die and around 5-10% of the effector T cells survive to further differentiate into memory $CD8^+$ T cells, which is capable of TCR-independent[40], antigen-independent[41], but cytokine-dependent[42,43] (especially IL-7 and IL-15) homeostatic proliferation and thus persist for a long term. These memory $CD8^+$ T cells can rapidly respond to antigen rechallenge and quickly give rise to massive number of antigen-specific effector cells to clear the infection.

While terminally differentiated effector $CD8^+$ T cells are relatively clearly defined by their expression of cytokines (e.g. IFN-γ) and cytotoxic effector molecules (e.g. Granzyme B, perforin) and by their phenotype ($CD45RA^+CD62L^-CCR7^-IL7R^{low}KLRG1^{hi}CD69^+CD43^+$)[44,45], memory $CD8^+$ T cells are highly heterogeneous. Early studies around 2000 demonstrated that memory $CD8^+$ T cells consists of two subpopulations, central memory T cells (T_{CM}) and effector memory T cells (T_{EM}), based on differential expression of CCR7 and CD62L[46,47]. T_{CM} ($CCR7^{hi}CD62L^{hi}$) cells mainly localize and circulate among secondary lymphoid organs and have greater

proliferative potential, while T_{EM} ($CCR7^{low}CD62L^{low}$) cells often migrate to peripheral tissue and have constitutive effector functions (e.g. express Granzyme B) that allow for a more rapid recall response[48]. With emerging evidences in the following decade, two more memory $CD8^+$ T cell subpopulations were established, the memory stem T cells (T_{SCM}) and the tissue-resident memory T cells (T_{RM}). The T_{SCM} cells have a naïve-like $CD45RO^-CD45RA^+CD44^{low}CD62L^{high}$ phenotype but are distinguishable from naïve T cells by their high expression of CD95, CXCR3 and IL2Rβ[49]. They were also demonstrated to have even higher proliferative capacity than T_{CM}. The T_{RM} cells are memory T cells that reside long-term in the brain and mucosal tissues (e.g. lungs, gut and skin) and show limited recirculation thus the term "tissue-resident"[50-54]. They can be characterized by a $CD103^{hi}CD69^{hi}CD27^{low}$ phenotype, and like T_{EM}, they also express high levels of granzyme B. Recently, new evidence suggests that the T_{CM} and T_{SCM} cells can be reclassified into stem-like memory T cells (T_{STEM}), which are $CCR7^+PD\text{-}1^-TIGIT^-$, and progenitor exhausted T cells (T_{PEX})[55], which are $CCR7^+PD\text{-}1^+TIGIT^+$ and allude to the concept of $CD8^+$ T cell exhaustion that will detailed in the next section.

The differentiation of memory $CD8^+$ T cells have been extensively studied, and multiple models have been continuously proposed[48,56-59]. Most of these models can be summarized into two general ideas: 1) memory $CD8^+$ T cells arise from a subset of $CD8^+$ effector T cells that acquire better memory potential and survive the contraction phase; 2) memory $CD8^+$ T cells arise directly from a subset of naïve T cells that are likely activated by relative weaker TCR-dependent signaling. This debate was largely resolved in favor of the first model by two studies[60,61] in 2017 that tracked $CD8^+$ T cells differentiation in mice during a viral infection, and both studies examined the epigenetic signatures of naïve, effector and memory $CD8^+$ T cells through DNA-methylation

profiling. Both studies found that when naïve T cells differentiate into effector T cells, effector-function related genes get unmethylated and thus activated whereas stemness -associated genes get methylated and repressed. However, as effector cells then differentiate into memory T cells, the stemness-associated genes can be demethylated to re-express those genes, and the effector-related genes remain in an open state of low methylation. This is consistent with the ability of memory T cells to express some effector molecules and mount recall response rapidly. It has been established that this subset of effector $CD8^+$ T cells that differentiate into memory T cells have the $IL7R^{hi}KLRG1^{low}$ phenotype, and they are often referred to as memory precursor cells to differentiate from the terminal effector cells.

Differentiation of exhausted/dysfunctional T cells

In contrast to the above described differentiation of memory $CD8^+$ T cells in acute infection after antigen clearance, during chronic infection or in cancer where the antigen persists and T cells experience constant stimulation, they differentiate into a different dysfunctional state termed as exhausted T cells (T_{EX}). Accumulative evidences have demonstrated in recent years that T_{EX} is an independent $CD8^+$ T cell lineage with distinct metabolic, transcriptional, and epigenetic signatures from effector or memory $CD8^+$ T cells[62-64]. Specifically, T_{EX} cells are characterized by the loss of effector functions, increased and sustained expression of multiple inhibitory receptors (IRs), and inability to further differentiate into a long-lived, self-renewal state that is characteristic of memory T cells. Since the early studies of T_{EX} in chronic LCMV infection in mice[65,66], T_{EX} has been extensively observed and studied in many other chronic viral infections such as HIV[67-70], hepatitis

C virus (HCV)[71-73], hepatitis B virus (HBV)[74] and etc., as well as in multiple cancers such as melanoma[75,76], chronic myeloid leukemia[77,78], ovarian cancer[79] and etc.

Shortly after T_{EX} was established as a distinct lineage, it was quickly realized that T_{EX} cells are heterogeneous though the specific subsets are still being studied and defined. The current knowledge suggests that while T_{EX} cells can arise from the same subset of $IL7R^{hi}KILG1^{low}$ memory-precursor effector $CD8^+$ T cells[80] (T_{EFF}), they also arise from the aforementioned memory-like precursor exhausted $CD8^+$ T cells (T_{PEX}) that derive directly from naïve T cells, and these two paths seem to be parallel[55]. These memory precursor T_{EFF} or T_{PEX} then converge through the continuous upregulation of TOX, which was demonstrated to be necessary and sufficient to induce T_{EX}[81-84], and differentiate into $CX3CR1^+T\text{-}bet^{hi}$ effector-like transitory exhausted T cells that retain proliferative capacity and contribute to viral control[85]. These cells eventually differentiate into terminal exhausted T cells with $CD101^+T\text{-}bet^{low}CX3CR1^-$ phenotype[85].

In 2006, Barber et al.[86] demonstrated in chronic LCMV infection that blocking PD-1:PD-L1 interactions in vivo can reinvigorate T_{EX} cells and improve their proliferation and function, and this leads to extensive studies in other chronic infections including HIV, HCV, and HBV as well as in tumor models[62]. This promotes the development of checkpoint blockade therapies that have demonstrated remarkable efficacy in many cancers[87]. However, the mechanisms underlying the checkpoint-blockade mediated T_{EX} reinvigoration remains elusive. Further understanding of the T_{EX} differentiation regulations are required to fully explain the reinvigoration.

Despite the similarity of phenotype and function of the dysfunctional T cells in chronic infection and those in tumor, it has been suggested there is disease-specific exhaustion features and mechanisms, driven by different TCR signaling strength, inflammatory environment, and

costimulatory signals in tumor as compared with during a chronic infection[88]. This may in turn lead to differences in the dysfunctional $CD8^+$ T cells response to checkpoint blockade immunotherapies (CBI). Thus, this thesis aims to contribute to the understanding of such differences in exhaustion mechanisms and between the response to CBI.

CAR-T cells and their differentiation

Chimeric antigen receptor-T (CAR-T) cells are endogenous T cells taken from cancer patients and transduced with chimeric antigen receptors (CARs), which when adoptively transferred back into patients can direct CAR-T cells to bind to tumor-specific antigens and kill cancer cells. CAR-T cell therapy has proved to be remarkably successful in treatment of hematological malignancies, and the US Food and Drug Administration (FDA) has up to now approved four CD19-directed CAR-T cell therapies[89].

The molecular structures of the CAR greatly impact CAR-T cells differentiation and thus affect their effector function, persistence, memory formation and exhaustion. The first generation of CARs consisted of an extracellular antigen-specific single-chain variable fragment (scFv) derived form an antibody fused to a transmembrane hinge region and an intracellular signaling domain derived from the CD3ξ chain of the endogenous T cell receptor (TCR)[90-94]. Despite the novelty, the first generation CAR-T cells had little efficacy in clinical trials and fail to persist[95]. The second-generation CARs improved upon this structure by including an additional co-stimulatory endodomain, derived from CD28 or 4-1BB, fused to the CD3ξ signaling domain[90,96,97]. This greatly improved the efficacy of the CAR-T cells and conferred profound and

durable response in patients with B cell leukemia[98-100] and quickly led to the FDA approval of second-generation CD19-targeting CAR-T therapies.

Despite CAR-T cell therapy's remarkable 40% complete response rate, it appears that there are severe class-level side effects for CD19 CAR-T cells: cytokine release syndrome (CRS)[101,102] and neurotoxicity[103]. In addition, the mechanisms for CAR-T cell therapy non-response and the *in vivo* differentiation of CAR-T cells in patients remain to be studied. Pre-clinical CAR studies have shown that immunosuppressive tumor microenvironments induce CAR-T cell exhaustion[104], and clinical CAR studies have shown that non-response correlate with molecular signatures of exhaustion before infusion[105,106]. Furthermore, emerging evidences suggest that CD28 costimulatory domain promotes rapid CAR-T cells expansion but limits its long-term persistence and memory formation, whereas the 4-1BB costimulatory domain confers slower expansion but longer persistence and better memory formation[107-111]. Despite these early findings, a comprehensive CAR-T differentiation model is still lacking, and the mechanisms that regulate memory formation, exhaustion, and the side effects in CAR-T therapy are still elusive. Thus, this thesis aims to address some of these by elucidating the transcriptional, clonal, and proteomic dynamics of anti-CD19 CAR-T cell differentiation *in vivo*.

T cell differentiation research fueled by single-cell sequencing technologies

Since the Human Genome Project, the cost of genome sequencing has dropped more than 50,000-fold[112] due to the rapid development of massively parallel sequencing technologies in the past 10-15 years, which is often referred to as the next-generation sequencing (NGS) to differentiate from the legacy Sanger sequencing[113]. Although NGS was first used to determine

DNA sequences, it has since been expanded and now routinely used to map and quantify a wide range of other biological molecules such as RNA[114] (RNA-seq), DNA-binding proteins[115] (CHIP-seq), histone modifications[116] (CHIP-seq), DNA methylation[117] (DNAse-seq), and several specific steps in the cascade of information from transcription to translation[118].

The advent of single-cell sequencing technologies in the past decade has further transformed many areas of biology and medicine, allowing researchers to interrogate the immune system at the level of individual cells, and it was named method of the year in 2013 by *Nature Methods*[119]. Virtually all bulk sequencing methods have been adapted and extended into single cell resolution, providing access to an almost full range of molecular information in a cell including but not limited to DNA (scDNA-seq, scWG-seq), RNA (scRNA-seq), transcription factors and histone modifications (scCHIP-seq), DNA modifications (scBS-seq, scABa-seq), chromosome organization (scHIC-seq), DNA accessibility (scATAC-seq), surface protein expression (CITE-seq, REAP-seq), and BCR/TCR clonotype (scBCR/scTCR-seq)[120,121]. In the past few years, single-cell methods have been further extended to map spatial information in tissue[122-124], opening up even more possibilities of multiomics temporospatial integrative analysis.

These sequencing technologies, especially the single-cell technologies, have truly revolutionized T cell research and have led to many breakthroughs in understandings of the mechanisms of T cell differentiation. To start with, scRNA-seq has demonstrated remarkable power in resolving cellular heterogeneity and led to the discovery of novel T cell subsets such as the stem-like precursor exhausted T cells T_{PEX}[55] and heterogeneous Treg cells along the spectrum of tissue adaptation from lymph node to non-lymphoid tissue[125]. Besides this, bulk/single-cell RNA-seq combined with bulk/single-cell ATAC-seq, CHIP-seq or other sequencing methods

enable advanced regulatory network analyses to determine key regulators or regulatory modules during T cell differentiation. For instance, TOX, the key transcription factor that drives T cell exhaustion, was simultaneously discovered by Chen *et al.*[81] through scRNA-seq analyses, Khan[84] *et al.* and Scott[82] *et al.* through bulk ATAC-seq and RNA-seq analyses. Furthermore, scRNA-seq, along or combined with scATAC-seq, have been used to reconstruct dynamic cellular trajectories of differentiation from multiple static snapshot data by taking advantage of the asynchronous nature of differentiation process and advanced computational approaches. For example, both endogenous and transplanted human hematopoietic stem cell differentiation landscapes have been mapped using scRNA-seq and/or scATAC-seq[126-128], and the differentiation bifurcation trajectories of Naïve $CD4^+$ T cells into Th1 or Tfh cells during blood-stage Plasmodium infection in mice was also traced and reconstructed using scRNA-seq[129]. Given such power of sequencing technologies, this thesis aims to address some of the standing questions in Tfh, T_{EX}, and CAR-T differentiation by combining multimodal bulk or single-cell sequencing technologies as detailed later.

TCR signaling and clustering research fueled by single-molecule microscopy

Besides sequencing technologies, single-molecule fluorescence microcopy has also seen tremendous advancement in the past decade, especially with the advent of new live-cell 4D imaging techniques such as Lattice-light-sheet microscopy[130] and deep learning-based image segmentation, restoration, enhancement and tracking methods[131]. This enabled and accelerated single-molecule research across fields in biosciences[132], including TCR clustering and signaling.

TCR clustering and signaling plays a critical role in regulating T cell activation, proliferation, survival, and differentiation into effector T cells[133,134], and single-molecule microscopy has continuously yielded valuable insights into the dynamics of TCR and structure of TCR clusters[135]. For instance, in the early 2000s, pioneers in TCR signaling used Interference reflection microscopy or Total Internal Reflection Fluorescence (TIRF) microscopy and determined the putative signaling assemblies on T cell surface to be the TCR microclusters[136-138], which form seconds after T cell contacts an antigen-presenting surface and before calcium fluxing. With the advancement of super-resolution microcopy techniques such as the light sheet direct stochastic optical reconstruction microscopy (dSTORM) and structured illumination microscopy (SIM), the idea of microclusters was revised with new observations of nanoscale TCR clusters[139,140], and the exact size of the TCR clusters are still under activate debate. Besides the structure of the TCR clusters, their distribution, associated signaling molecules, mechanotransduction mechanisms have all been elucidated by single-molecule microscopies[132,141].

Further research on TCR clustering and signaling can benefit enormously from cutting-edge single-molecule microscopy such as lattice light-sheet microscopy (LLSM). Compared to other major single-molecule microscopy techniques such as TIRF microscopy, confocal microscopy, and single-molecule localization microscopy (SMLM), LLSM achieves a more optimal combination of four-dimensional (4D) (x, y, z, and time) imaging, high spatiotemporal resolution (~0.1μm along x-y, ~0.4 μm along z; ~100 frames/s, ~1 cell volume/s), and minimal photobleaching. It achieves this by first using a spatial light modulator and Fraunhofer diffraction through an annulus to create an interference pattern that forms a light sheet of optical lattice. It then uses this lattice light sheet to dither along the x-y direction and scan through the z direction

to collect a z-stack of x-y images per cell volume. Since only samples at the lattice points are illuminated at any time, photobleaching is minimized. Given such high spatiotemporal resolution and low photobleaching, it opens new possibilities to study TCR clustering and signaling dynamics on live T cell surface with the field of view of a whole cell and real time for a relatively long time. For example, LLSM has been used to directly examine the microvillar dynamics on live T cell surface and the stabilization mechanism was revealed[142].

Despite the possibility to directly visualize TCR clustering dynamics using LLSM, the LLSM images contain much more information than what can be directly perceived. Advanced deep learning and high-dimensional analyses methods can be applied to extract the hidden/invisible quantitative information about each TCR cluster and give insights into the underlying signaling states of T cells. Thus, this thesis aims to explore the general pipeline that can take advantage of advanced computational methods to fully realize the potential of the data from LLSM and hope to inspire a paradigm-shift in the quantitative analyses of single-molecule microscopy data.

Summary: Aims of this thesis

As introduced above, fundamental questions in T cell differentiation remain unanswered: 1) while EZH2-mediated H3K27me3 has been implicated in the differentiation of naïve $CD4^+$ T cells and certain lineage of Th cells such as Th1 and Th2 cells[38], its role in Tfh fate commitment is unknown; 2) the heterogeneity and response to checkpoint blockade immunotherapy of T_{EX} cells, while extensively explored in either tumor or during chronic infection, are yet to be contrasted between within tumor and during chronic infection; 3) the *in vivo* differentiation of

CAR-T cells, the therapy failure/relapse mechanisms, and the clinical response biomarkers all remain to be elucidated. One additional issue is the lack of adaptation and application of advanced high-dimensional and machine learning computational methods in the analysis of LLSM data in general, but particularly in the quantitative analysis of TCR clusters.

Hence, this thesis aims to address the above gaps by taking advantage of advanced sequencing imaging technologies through the completion of the following specific aims:

1. Determine the role of EZH2 in Tfh cells fate commitment and elucidate on the underlying regulatory mechanism
2. Determine the differences in phenotypic heterogeneity, transcriptional and epigenetic signatures of each subset, and response to checkpoint blockade immunotherapy of T_{EX} cells in tumor and those during chronic infection.
3. Map the in vivo differentiation trajectories of CAR-T cells in both responders and non-responders to CAR-T therapy and investigate the mechanisms of failure/relapse as well as identify biomarkers of response.
4. Design a pipeline for high-dimensional quantitative analysis and machine learning of LLSM images of TCR clusters on live T cell surface and demonstrate the utility of such analyses in discovering the invisible underlying states of T cell signaling.

CHAPTER 2: INVESTIGATE THE ROLE OF EZH2 IN TFH DIFFERENTIATION

Note:

The following section (Chapter 2) is reproduced with slight modifications from my co-first authored publication "The histone methyltransferase EZH2 primes the early differentiation of follicular helper T cells during acute viral infection" (Chen et al. 2019), which was published in Nature Cellular & Molecular Immunology on March 06, 2019 under the Creative Commons license (http://creativecommons.org/licenses/by/4.0/).

Authors:

Xiangyu Chen*, **Guoshuai Cao***, Jialin Wu*, Xinxin Wang*, Zhiwei Pan, Jianbao Gao, Qin Tian1, Lifan Xu, Zhirong Li, Yaxing Hao, Qizhao Huang, Pengcheng Wang, Minglu Xiao, Luoyingzi Xie, Shupei Tang, Zhenyu Liu, Li Hu, Jianfang Tang, Ran He, Li Wang, Xinyuan Zhou, Yuzhang Wu, Mengjie Chen, Beicheng Sun, Bo Zhu, Jun Huang and Lilin Ye.

* These authors contributed equally to this work.

Author Contributions:

X.C., J.W., X.W., J.G., Z.P., Q.T., L. Xu., Z. Li., Y.H., Q.H., P.W., M.X., L. Xi., S.T., Z. Liu., L.H, J.T, R. H., S.T. and Z.L. performed the experiments. **G.C. analyzed the ATAC-Seq data and microarray data** with supervision from M.C.; L.Y. designed the study, analyzed the data and wrote the paper with assistance from J.H., X.C., G.C., L.W., X.Z., Y.W. and B.S.; and L.Y., B.Z. and J.H. supervised the study.

Abstract

Epigenetic modifications to histones dictate the differentiation of naïve $CD4^+$ T cells into different subsets of effector T helper (T_H) cells. The histone methyltransferase enhancer of zeste homolog 2 (EZH2) has been implicated in the mechanism regulating the differentiation of T_H1, T_H2 and regulatory T (T_{reg}) cells. However, whether and how EZH2 regulates follicular helper T (T_{FH}) cell differentiation remain unknown. Using a mouse model of acute lymphocytic choriomeningitis virus (LCMV) infection, we observed abundant EZH2 expression and associated H3K27me3 modifications preferentially in the early committed virus-specific T_{FH} cells compared to those in T_H1 cells. Ablation of EZH2 in LCMV-specific $CD4^+$ T cells leads to a selective impairment of early T_{FH} cell fate commitment, but not late T_{FH} differentiation or memory T_{FH} maintenance. Mechanistically, EZH2 specifically stabilizes the chromatin accessibility of a cluster of genes that are important for T_{FH} fate commitment, particularly B cell lymphoma 6 (Bcl6), and thus directs T_{FH} cell commitment. Therefore, we identified the chromatin-modifying enzyme EZH2 as a novel regulator of early T_{FH} differentiation during acute viral infection.

Full Text

Introduction

During pathogenic infections, follicular T helper (T_{FH}) cells provide essential assistance to cognate pathogen-specific B cells that enable them to initiate and sustain germinal center (GC) reactions in B-cell follicles within secondary lymphoid tissues[143]. GC reactions lead to both the rapid production of high-affinity antibodies that protect against the immediate infection and the subsequent generation of persistent humoral immune memory for long-term protection[143,144]. In the early stage of GC reactions, intimate interactions between early differentiated T_{FH} cells and

newly activated cognate B cells in the interfollicular regions of secondary lymphoid tissues direct the migration of both cell types to B-cell follicles, in turn cooperating to prime early GC responses. During ongoing GC reactions, T_{FH} cells promote the survival, proliferation, class switching and hypermutation of cognate B cells and eventually drive them to differentiate into long-lived memory B cells and antibody-secreting plasma cells by secreting important cytokines such as IL-21[145], IL-4[146] and IL-9[147] and engaging certain cell surface-bound receptors and their matching ligands, including CD40-CD40L[148], ICOS-ICOSL[148] and PD-1-PD-L1[149]. Unlike other lineages of $CD4^+$ helper T (T_H) cells, T_{FH} cells are programmed to express the chemokine receptor CXCR5[22-24]. In response to the chemoattractant CXCL3[150,151], $CXCR5^+$ T_{FH} cells migrate to B-cell follicles, where they physically interact with cognate B cells[143].

Given the critical role of T_{FH} cells in B-cell-mediated humoral immunity, investigations of the early fate commitment of T_{FH} cells are very important, as the differentiation of T_{FH} cells occurs within 48 h after acute viral infections[152,153]. The transcriptional repressor B cell lymphoma-6 (Bcl-6) functions as a "master regulator" to govern early T_{FH} cell differentiation[154-156]. A rapid induction of Bcl-6 expression in de novo activated virus-specific $CD4^+$ T cells represents a key step toward the T_{FH} fate commitment. A wide variety of transcription factors (TFs) have recently been shown to regulate Bcl-6 expression during T_{FH} differentiation[30]. STAT family TFs, including STAT1[157], STAT3[158] and STAT4[159], promote Bcl-6 expression in response to stimulation with the corresponding cytokines, such as IL-6, IL-12 and IL-21. Batf[160], IRF4[161], Notch1 and Notch2[162] have also been reported to induce Bcl-6 expression and promote T_{FH} differentiation. In contrast, the TFs Blimp-1[156] and Foxo1[163] inhibit Bcl-6 expression and subsequently repress T_{FH} differentiation. Despite the known positive and negative effects of these TFs on Bcl-6 expression,

researchers have not conclusively determined whether these regulators exert their effects during the early stage of T_{FH} commitment. Transcription factor-1 (TCF-1) initiates the T_{FH} fate commitment by directly inducing Bcl-6 expression and suppressing Blimp-1 expression during an acute viral infection[152,164,165]. Although TCF-1 is expressed at high levels in naïve $CD4^+$ T cells[166], activated virus-specific $CD4^+$ T cells must further increase TCF-1 expression levels for subsequent T_{FH} conversion following an acute viral infection[152]. Currently, additional factors that are responsible for the early induction of the fate commitment of T_{FH} cells remain to be determined.

Epigenetic modifications of histones have been extensively implicated in the mechanisms regulating T cell differentiation[38,167]. The epigenetic regulator enhancer of zeste homolog 2 (EZH2), which is the catalytic subunit of polycomb repressive complex 2 (PRC2), functions as a methyltransferase to induce the trimethylation of histone H3 at lysine 27[168,169]. The resulting H3K27me3 protein recruits chromatin-compressing protein complexes to certain loci to silence the expression of corresponding genes[168,169]. EZH2-mediated H3K27me3 plays a critical role in the differentiation and lineage stability of various types of $CD4^+$ T_H cells, including T_H1, T_H2 and regulatory T (T_{reg}) cells[170-175]. In the T_H1 and T_H2 lineages, several groups reported that EZH2 and the associated H3K27me3 directly bind to Tbx21[172,174] (which encodes the TF T-bet, a master regulator of T_H1 differentiation) and Gata3[172] (encoding the TF Gata3, which specifies T_H2 differentiation) to inhibit the transcription of both TFs and eventually suppress the differentiation of naïve $CD4^+$ T cells into both T_H1 and T_H2 cells. Additionally, EZH2 and H3K27me3 inhibit the differentiation of T_H1 and T_H2 cells through the direct epigenetic marking and silencing of genes that encode lineage-specific cytokines, such as Ifng[172,174] and Il13[172,174]. Paradoxically, EZH2 also induces T_H1 cell differentiation by increasing the stability of T-bet and inducing the production of

T_H1 cytokines[171,175]. In T_{reg} cells, CD28-induced EZH2 expression promotes H3K27me3 deposition at gene loci that are normally repressed in T_{reg} cells and thus plays an important role in stabilizing the lineage specification of activated T_{reg} cells[170]. Despite the profound effects of EZH2 on T_H1, T_H2 and T_{reg} differentiation, whether and how EZH2 regulates the T_{FH} fate commitment remains to be investigated.

Here, we first defined a strict lineage-specific mode of chromatin accessibility in virus-specific T_{FH} cells compared to virus-specific T_H1 cells in response to an acute infection. Bona fide differentiated virus-specific T_{FH} cells exhibited increased EZH2 expression and the associated H3K27me3 modification compared to naïve T cells and T_H1 cells in the early days after an acute lymphocytic choriomeningitis virus (LCMV) infection. Furthermore, EZH2 was required for governing the chromatin accessibility of a cluster of T_{FH}-lineage-associated genes, particularly Bcl6, which are essential for T_{FH} fate commitment. Accordingly, inactivation of EZH2 in virus-specific $CD4^+$ T cells led to a pronounced reduction in the early commitment but not late differentiation or maintenance of T_{FH} cells.

Results

Chromatin states of virus-specific T_{FH} and T_H1 cells in response to acute viral infection

In response to an acute viral infection, activated virus-specific $CD4^+$ T cells differentiate into either T_H1 or T_{FH} cells[152,176,177]. We first adoptively transferred LCMV-specific naïve SMARTA cells (expressing a transgenic T cell receptor specific for the LCMV glycoprotein epitope I-AbGP66–77) into WT C57BL/6J recipients and subsequently infected the chimeric recipients with the LCMV Armstrong strain to investigate the potential regulatory regions involved

in this bifurcated differentiation at the genome level. Then, we sorted virus-specific SMARTA T_{FH} cells and T_H1 cells from the SMARTA chimera mice on days 2, 5 and 8 after the infection (Supplementary Figure S2.1a–c) and subsequently performed an ATAC-Seq assay to measure the transposase-accessible chromatin[178]. We also sorted naïve SMARTA cells ($CD4^+CD25^-CD62L^+CD44^-$) as a control (Supplementary Figure S2.1d). QC of the ATAC-Seq libraries revealed the characteristic DNA fragment length distribution and the expected peak distribution across genomic features (Supplementary Figure S2.2a, b). With these high-quality libraries, we assessed the differences in chromatin-accessible regions (ChARs) and found that compared to naïve cells, dramatic changes in the numbers of ChARs emerged as early as day 2 postinfection in both virus-specific T_{FH} and T_H1 cells (Supplementary Figure S2.2c). Furthermore, chromatin- accessible patterns were discerned in T_{FH} and T_H1 cells, respectively, at different time points postinfection (Fig. 2.1a). The greatest differences in ChAR patterns were observed between T_{FH} and T_H1 cells at day 8 postinfection (Fig. 2.1a), reflecting the multistage differentiation of fully functional T_{FH} and T_H1 cells[143]. A ChAR-based PCA further revealed that T_{FH} and T_H1 cells started to enter two distinct $CD4^+$ T cell differentiation states on day 2 postinfection, and these differentiation trajectories continuously developed through the entire process (Fig. 2.1b).

K-means clustering of the differential peaks revealed four distinct areas, and the third and fourth clusters revealed opened ChARs that were specifically detected in T_H1 and T_{FH} cells, respectively (Fig. 2.1c), compared to naïve cells. Notably, when focusing on cluster 4, we found an array of T_{FH} lineage-associated gene loci that were enriched in this region, including Bcl6, Tcf7, Id3[179], Ascl2[180], Cxcr5 and Il21 (Fig. 2.1d). The T_H1-associated genes Tbx21, Prdm1, Havcr2, Ccl3, Ccl4 and Ccl5 were observed in cluster3 (Fig. 2.1e).

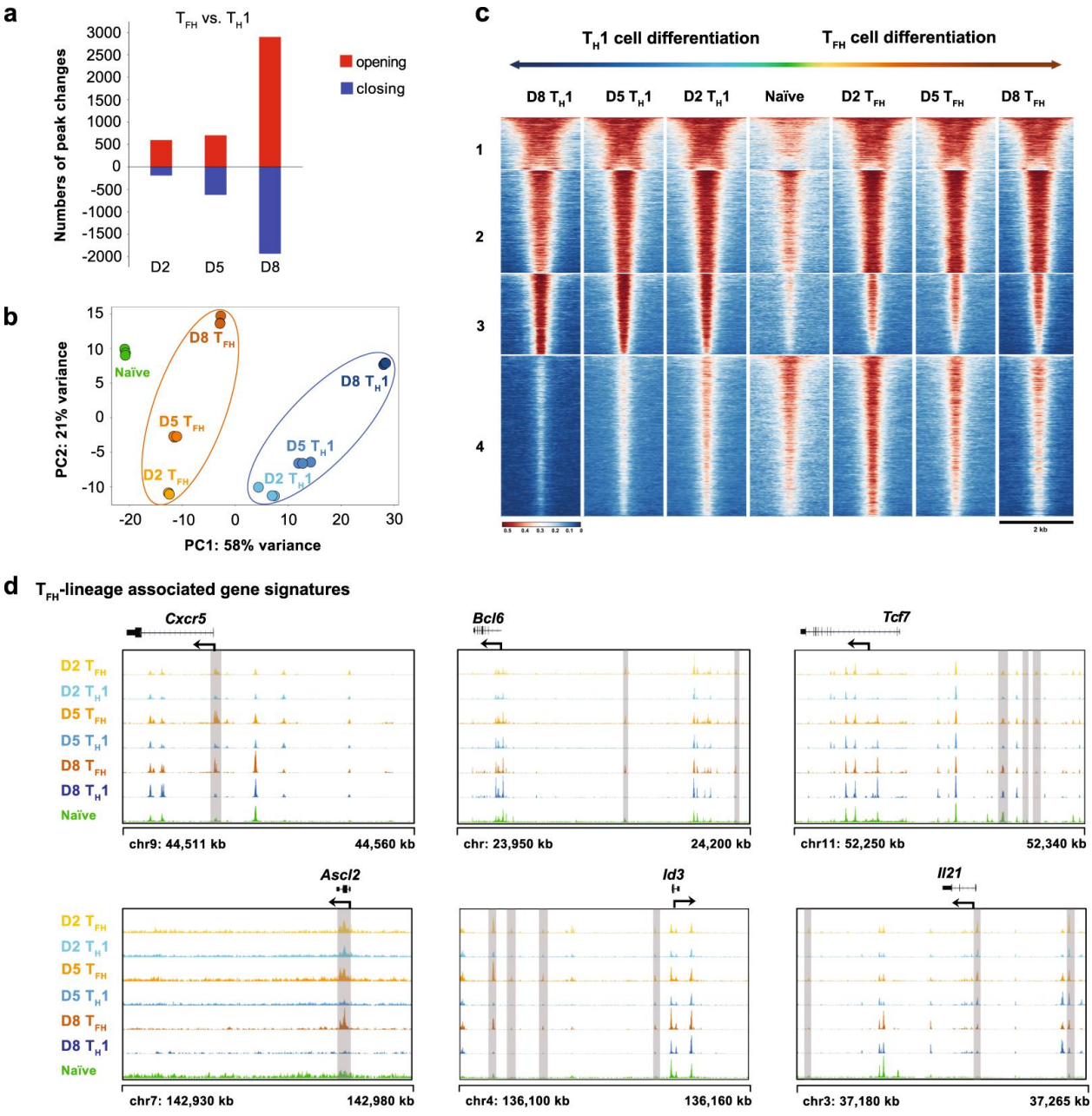


Figure 2. 1 Chromatin states of the virus-specific T_{FH} and T_H1 cells in response to an acute viral infection.

a Numbers of chromatin peaks with differential accessibility (FDR < 0.05; FC > 4) between SMARTA T_{FH} cells and SMARTA T_H1 cells at the indicated time points after LCMV Armstrong infection. **b** PCA plot of the peak accessibilities in naïve SMARTA $CD4^+$ T cells, SMARTA T_{FH} cells (days 2, 5 and 8 postinfection) and SMARTA T_H1 cells (days 2, 5 and 8 postinfection). Each dot represents a replicate of the indicated group.

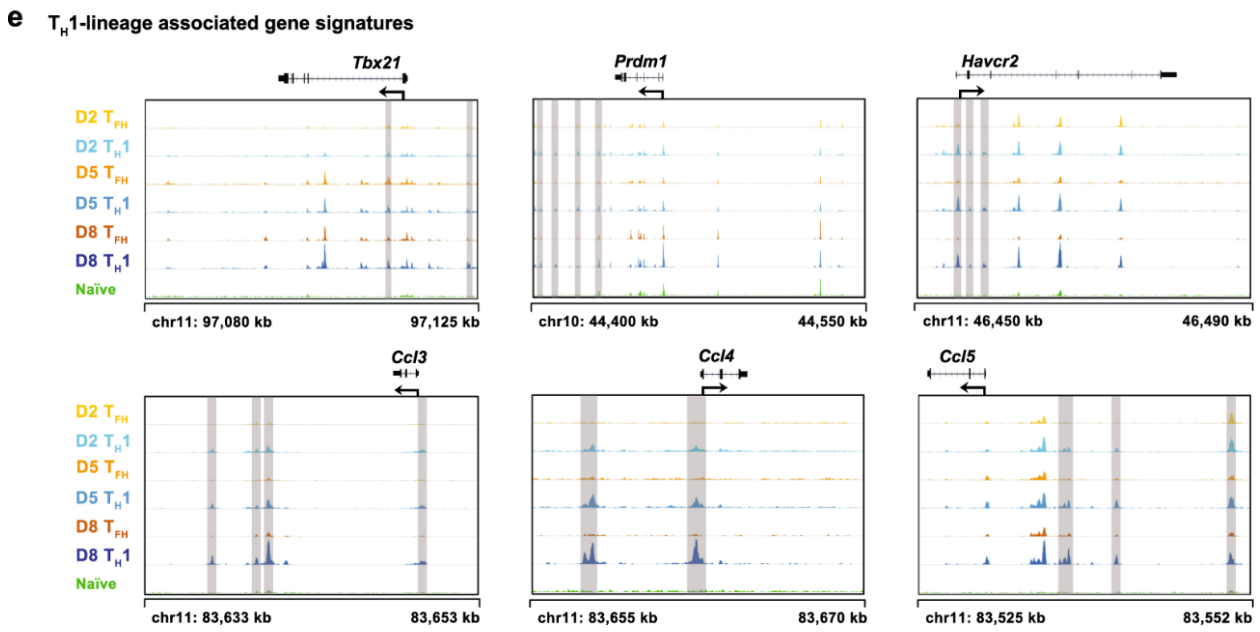


Figure 2.1, Continued

c Chromatin accessibility heat map of differential peaks from **a**. Each row represents one of the 15,600 differential peaks that was center-aligned and extended upstream and downstream by 1 kb from the center. The peaks are K-means clustered. **d** ATAC-Seq signal profiles of T_{FH} lineage-associated gene loci. **e** ATAC-Seq signal profiles of T_H1 lineage-associated gene loci. Differential peaks are highlighted in gray (**d** and **e**). The data were obtained from one experiment with three biological replicates (pooled from at least five mice per group) (**a**–**e**)

Further analysis of the ChARs for each individual gene locus revealed the stringent lineage-specific mode of chromatin accessibility; i.e., the chromatin accessibility of T_{FH}-associated genes was more prominent in T_{FH} cells than in T_H1 cells, and vice versa (Fig. 2.1d and e). Based on these results, chromatin remodeling is tightly associated with the T_{FH} but not T_H1 lineage commitment and differentiation in response to an acute viral infection.

Dynamic EZH2 expression and H3K27me3 modification in virus- specific T_{FH} cells

The EZH2-mediated H3K27me3 modification plays a critical role in chromatin remodeling[181]. Next, we sought to investigate the relationship between EZH2 expression and H3K27me3 modification during T_{FH} cell differentiation in response to an acute viral infection. We first adoptively transferred naïve SMARTA cells into WT C57BL/6J recipients, followed by an infection with the LCMV Armstrong strain. On day 2 after infection, we compared EZH2 expression and H3K27me3 modification between naïve ($CD44^{lo}CD25^{-}$), T_H1 ($CD25^{hi}CXCR5^{-}$) and T_{FH} ($CD25^{lo}CXCR5^{+}$) SMARTA cells. We observed the highest levels of EZH2 expression in T_{FH} cells compared to those in T_H1 and naïve cells, and T_H1 cells expressed a relatively higher level of EZH2 than the naïve cells (Fig. 2.2a). This phenotypic pattern was consistent with the levels of the EZH2-mediated H3K27me3 modification in T_{FH}, T_H1 and naïve SMARTA cells (Fig. 2.2b). Next, we analyzed the kinetics of EZH2 expression in T_{FH} SMARTA cells at different time points after infection and found that virus-specific T_{FH} cells rapidly increased EZH2 expression upon infection and reached a peak on day 2 postinfection (Fig. 2.2c). EZH2 expression then decreased to a comparable level to naïve cells on day 8 postinfection (Fig. 2.2c).

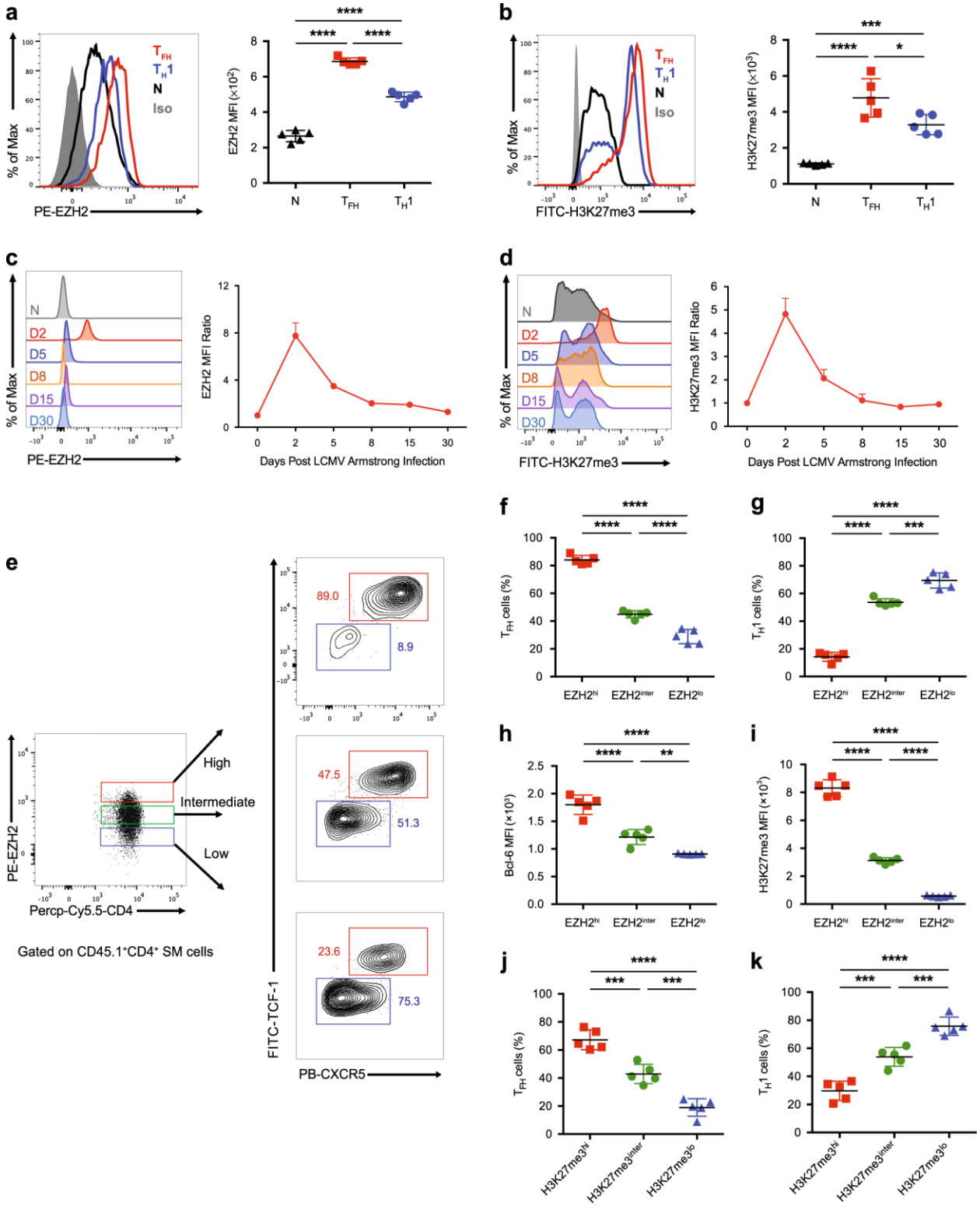


Figure 2. 2 Dynamic changes in EZH2 expression and the H3K27me3 modification in virus-specific T_{FH} cells.

Figure 2. 2, Continued

a, b Comparison of EZH2 (a) and H3K27me3 (**b**) levels between SMARTA T_{FH} cells ($CD25^{lo}CXCR5^{+}$) and SMARTA T_H1 cells ($CD25^{hi}CXCR5^{-}$) from the spleens of $CD45.2^{+}$ wild-type mice that underwent adoptive transfer of $CD45.1^{+}$ SMARTA cells and analyzed on day 2 after LCMV Armstrong infection and in naïve ($CD44^{lo}CD62L^{hi}$) SMARTA cells (N). **c, d** Flow cytometry analysis of EZH2 (**c**) and H3K27me3 (**d**) levels in T_{FH} cells derived from the mice listed in a on days 2, 5, 8, 15 and 30 after the LCMV Armstrong infection and in naïve SMARTA cells (N). The EZH2 or H3K27me3 mean fluorescence intensity (MFI) ratio at each time point was calculated as the MFI of T_{FH} cells / MFI of $CD4^{+}CD44^{lo}$ T cells in an identical mouse. **e** Flow cytometry analysis of the $EZH2^{hi}$, $EZH2^{inter}$ and $EZH2^{lo}$ subsets of T_{FH} cells and T_H1 cells from the mice shown in a on day 2 after the LCMV Armstrong infection. The proportions of T_{FH} cells and T_H1 cells and the MFI of Bcl-6 in each population are summarized in **f, g** and **h**, respectively. **i** The MFI of H3K27me3 in the $EZH2^{hi}$, $EZH2^{inter}$ and $EZH2^{lo}$ subsets of T_{FH} cells described in **e**. The proportions of T_{FH} cells (**j**) and T_H1 cells (**k**) among the $H3K27me3^{hi}$, $H3K27me3^{inter}$ and $H3K27me3^{lo}$ subsets of SMARTA cells from mice shown in a on day 2 after the LCMV Armstrong infection. $^{*}P < 0.05$, $^{***}P < 0.001$ and $^{****}P < 0.0001$ (unpaired two-tailed t-test). The data are representative of two independent experiments with at least four mice per group (**a–d** and **f–k**; error bars in **a–d** and **f–k** indicate the s.d.)

Consistent with the EZH2 expression kinetics, the direct target of EZH2, H3K27me3, exhibited similar kinetics during T_{FH} cell differentiation (Fig. 2.2d).

The fate commitment of virus-specific T_{FH} cells occurs within the first 48 h after an acute viral infection. Therefore, the high expression of EZH2 in early differentiated T_{FH} cells on day 2 postinfection prompted us to examine whether the level of EZH2 expression correlated with T_{FH} cell commitment. For this purpose, on day 2 after infection, we divided SMARTA cells into three subsets, $EZH2^{hi}$, $EZH2^{inter}$ and $EZH2^{lo}$, according to their EZH2 expression levels and observed that the $EZH2^{hi}$ subset was much more poised to adopt the T_{FH} fate than the $EZH2^{inter}$ and $EZH2^{lo}$ subsets (Fig. 2e, f), while the opposite phenotype was observed during T_H1 differentiation (Fig. 2.2e–g). Compared to the $EZH2^{inter}$ and $EZH2^{lo}$ subsets, the $EZH2^{hi}$ subset expressed a significantly higher level of Bcl-6 (Fig. 2.2h), suggesting the enhanced propensity of this subset to become T_{FH} cells. Consistent with these findings, the extent of the H3K27me3 modification was also positively correlated with EZH2 expression and an early T_{FH} fate choice in SMARTA cells on day 2 after an acute viral infection (Fig. 2.2i–k). Collectively, high EZH2 expression and the associated H3K27me3 modification in activated virus-specific $CD4^+$ T cells preferentially drove them to differentiate into T_{FH} cells during the early response to acute viral infection, while the cells displaying low EZH2 and H3K27me3 levels were more prone to differentiate into T_H1 cells.

Role of EZH2 in early T_{FH} commitment during acute viral infection Considering the positive correlation between the EZH2 expression level and T_{FH} commitment, we next investigated whether EZH2 is indeed essential for early T_{FH} differentiation upon acute viral infection. We crossed mice harboring *loxP*-flanked *Ezh2* alleles ($Ezh2^{fl/fl}$) with TCR-transgenic SMARTA mice that expressed the congenic marker CD45.1 to generate $Ezh2^{fl/fl}$ SMARTA mice (hereafter

designated *Ezh2*$^{fl/fl}$ SM). Subsequently, we used a retroviral transduction system to overexpress codon-improved Cre recombinase (iCre) for the conditional deletion of EZH2 in virus-specific *Ezh2*$^{fl/fl}$ SM cells. Then, we transferred retrovirus-transduced and nontransduced *Ezh2*$^{fl/fl}$ SM cells or WT SM cells (*Ezh2*$^{+/+}$ SM) into naïve WT recipients that expressed the congenic marker CD45.2 and subsequently infected animals with the LCMV Armstrong strain (Fig. 2.3a). On day 2 after infection, we visualized the efficient deletion of EZH2 expression in *Ezh2*$^{fl/fl}$ SM cells overexpressing iCre (Supplementary Figure S2.3a). Notably, only 4.7% of iCre-expressing *Ezh2*$^{fl/fl}$ SM cells differentiated into CD25^{lo}CXCR5^{+} T_{FH} cells, whereas ~20% of virus-specific *Ezh2*$^{+/+}$ SM cells (with and without iCre recombinase) and *Ezh2* SM cells (without iCre recombinase) were committed to the T_{FH} cell fate (Fig. 2.3b, c), highlighting the importance of EZH2 in T_{FH} fate determination. *Ezh2*$^{+/+}$

We bred *Ezh2*$^{fl/fl}$ mice with SMARTA mice that transgenically expressed Cre recombinase downstream of the *Cd4* enhancer, promoter and silencer sequences (hereafter called *Ezh2*$^{fl/fl}$ SM-Cre) to further confirm the role of EZH2 in the early differentiation of virus-specific T_{FH} cells. We next mixed equal numbers of *Ezh2*$^{+/+}$ SM-Cre cells with *Ezh2*$^{fl/fl}$ SM-Cre cells and cotransferred these cells into WT recipients, which were subsequently infected with the LCMV Armstrong strain (Fig. 2.3d). On day 2 after infection, we validated the efficiency of *Ezh2* deletion in *Ezh2*$^{fl/fl}$ SM-Cre cells (Supplementary Figure S2.3b) and observed a remarkable reduction in the T_{FH} cell commitment of the *Ezh2*$^{fl/fl}$ SM-Cre cells compared to *Ezh2*$^{+/+}$ SM-Cre cells (Fig. 2.3e, f). Furthermore, we compared the expression levels of molecules that are closely associated with T_{FH} fate commitment, including TCF-1, Bcl-6, ICOS and CXCR5, between T_{FH} cells that differentiated from *Ezh2*$^{fl/fl}$ SM-Cre cells and *Ezh2*$^{+/+}$ SM-Cre cells (cells within the red circle in Fig. 2.3e). The

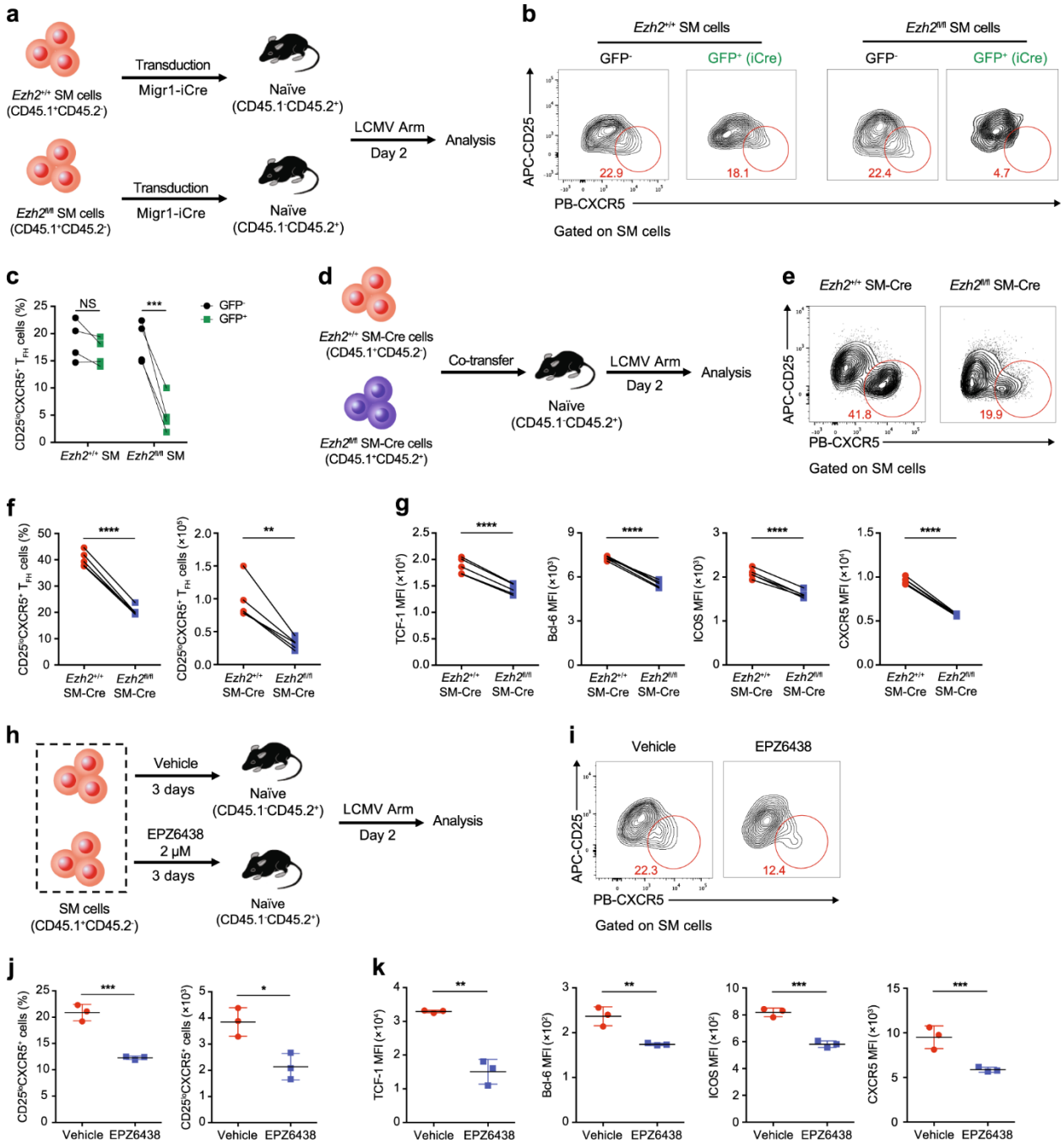


Figure 2. 3 Role of EZH2 in early T_{FH} commitment during an acute viral infection.

a Experimental setup. A retrovirus overexpressing iCre was introduced into CD45.1^{+}Ezh2$^{+/+}$ SMARTA (SM) cells and CD45.1^{+}Ezh2$^{fl/fl}$ SM cells, which were transferred into CD45.2^{+} WT recipients. Then, the recipients were infected with the LCMV Armstrong strain and analyzed on day 2 postinfection.

Figure 2. 3, Continued

b Flow cytometry analysis of $Ezh2^{+/+}$ and $Ezh2^{fl/fl}$ SM cells transduced with a retrovirus expressing iCre. The numbers adjacent to the outlined areas indicate the percentages of $CD25^{lo}CXCR5^{+}$ T_{FH} cells, which are summarized in **c**. **d** Experimental setup. $Ezh2^{+/+}Cd4^{-}Cre$ SMARTA cells ($Ezh2^{+/+}$ $SM^{-}Cre$; $CD45.1^{+}CD45.2^{-}$) and $Ezh2^{fl/fl}$Cd4-Cre SMARTA cells ($Ezh2^{fl/fl}$ SM-Cre; $CD45.1^{+}CD45.2^{+}$) were cotransferred into WT recipients ($CD45.1^{-}CD45.2^{+}$), which were infected with the LCMV Armstrong strain and assessed on day 2 postinfection. **e** Flow cytometry analysis of $Ezh2^{+/+}$ SM-Cre and $Ezh2^{fl/fl}$ SM-Cre SMARTA cells. The numbers adjacent to the outlined areas indicate the percentages of $CD25^{lo}CXCR5^{+}$ T_{FH} cells, which are summarized in f (left panel). Total numbers of $CD25^{lo}CXCR5^{+}$ T_{FH} cells in **e** are presented (f, right panel). **g** Quantification of TCF-1, Bcl-6, ICOS and CXCR5 levels in the $CD25^{lo}CXCR5^{+}$ T_{FH} cells shown in **e**. **h** Experimental setup. After treatment with either EPZ6438 or vehicle for 3 days, $CD45.1^{+}$SMARTA cells were transferred into WT $CD45.2^{+}$ recipients that were subsequently infected with the LCMV Armstrong strain. The adoptively transferred SMARTA cells were analyzed on day 2 postinfection. **i** Flow cytometry analysis of EPZ6438-treated and vehicle-treated SMARTA cells. The numbers adjacent to the outlined areas indicate the proportions of $CD25^{lo}CXCR5^{+}$ T_{FH} cells, which are summarized in **j** (left panel). Total numbers of $CD25^{lo}CXCR5^{+}$ T_{FH} cells analyzed in **i** are presented in **j** (right panel). **k** Quantification of TCF-1, Bcl-6, ICOS and CXCR5 levels in the $CD25^{lo}CXCR5^{+}$ T_{FH} cells shown in **j**. NS not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$ (paired two-tailed t-test (**c, f** and **g**) or unpaired two-tailed t-test (**j** and **k**)). The data are representative of two independent experiments with at least three mice (**c, f, g, j** and **k**) per group (error bars in j and k indicate the s.d.)

levels of all examined molecules were substantially reduced in *Ezh2*$^{fl/fl}$ SM-Cre cell-derived T_{FH} cells compared to those in T_{FH} cells that differentiated from *Ezh2*$^{+/+}$ SM-Cre cells (Fig. 2.3g). Additionally, these phenotypes characterizing the early T_{FH} differentiation of *Ezh2*$^{fl/fl}$ SM-Cre cells in response to LCMV infection were largely confirmed to be the same cells upon an acute intracellular bacterial infection with a recombinant L. monocytogenes strain that expressed the LCMV glycoprotein epitope I-AbGP61–80 (Supplementary Figure S2.3e–i).

Next, we investigated whether the inhibition of the EZH2-mediated H3K27me3 modification also led to defects in the T_{FH} cell commitment of virus-specific $CD4^+$ T cells in response to a viral infection. For this purpose, we treated SMARTA cells with vehicle or the small molecule EPZ6438, which is a specific inhibitor of the H3K27me3 modification[182], in vitro for 3 days and subsequently transferred these cells into WT recipients that were then infected with the LCMV Armstrong strain (Fig. 2.3h). Treatment with EPZ6438 efficiently reduced the levels of the H3K27me3 modification in SMARTA cells without affecting EZH2 expression (Supplementary Figure S2.3c and d). Notably, the inhibition of the H3K27me3 modification by the EPZ6438 treatment resulted in a substantial reduction in T_{FH} cell differentiation on day 2 postinfection compared to the T_{FH} commitment of vehicle-treated SMARTA cells (Fig. 2.3i, j). Furthermore, similar to results obtained using *Ezh2*$^{fl/fl}$ SM-Cre cell-derived T_{FH} cells (Fig. 2.3g), T_{FH} cells that differentiated from EPZ6438-treated SMARTA cells also exhibited lower expression of T_{FH} lineage-associated molecules, including TCF-1, Bcl-6, ICOS and CXCR5, compared to that of T_{FH} cells derived from vehicle-treated SMARTA cells (Fig. 2.3k). Taken together, these data validated the hypothesis that EZH2 expression and the subsequent H3K27me3 modification in virus-specific

$CD4^+$ T cells were important for early commitment to the T_{FH} cell fate in response to an acute viral infection.

Requirement for EZH2 expression in endogenous virus-specific T_{FH} cell differentiation

We bred *Ezh2*$^{fl/fl}$ mice with *Cd4*-Cre transgenic mice to further analyze the role of EZH2 in non-TCR transgenic, endogenous virus-specific T_{FH} cell differentiation. The resulting *Ezh2*$^{fl/fl}$*Cd4*-Cre mice conditionally lost EZH2 expression in their $CD4^+$ T cell population (Supplementary Figure S2.4a). On day 8 after infection with the LCMV Armstrong strain, *Ezh2*$^{fl/fl}$*Cd4*-Cre mice exhibited a remarkable reduction in the virus-specific GP66–77 tetramer-positive $CXCR5^+ICOS^+$ T_{FH} cell population in the spleen compared to that in *Ezh2*$^{fl/fl}$ control mice (Fig. 2.4a, b). Furthermore, the tetramer-positive T_{FH} cells from *Ezh2*$^{fl/fl}$*Cd4*-Cre mice exhibited decreased levels of T_{FH} differentiation-associated molecules, such as Bcl-6, TCF-1, PD-1 and CXCR5, compared to T_{FH} cells from *Ezh2*$^{fl/fl}$ control mice (Fig. 2.4c). These data revealed the effect of EZH2 deficiency on the T_{FH} differentiation of endogenous virus-specific $CD4^+$ T cells, which were highly consistent with our observations in the *Ezh2*-null TCR-transgenic SMARTA cells shown in Fig. 2.3. Additionally, we compared the T_{FH} differentiation of bulk virus-activated $Foxp3^-CD44^{hi}$ $CD4^+$ T cells from *Ezh2*$^{fl/fl}$*Cd4*-Cre mice and *Ezh2*$^{fl/fl}$ control mice after LCMV infection. Similar to the impaired T_{FH} differentiation of GP66–77 tetramer-positive $CD4^+$ T cells in *Ezh2*$^{fl/fl}$*Cd4*-Cre mice upon infection, the bulk-activated $Foxp3^-CD44^{hi}CD4^+$ T cells from these mice also exhibited compromised $CD44^{hi}CXCR5^+T_{FH}$ differentiation, as evidenced by the decreases in the proportion of T_{FH} cells, the T_{FH} to T_H1 ratio and the absolute T_{FH} number compared to that in *Ezh2*$^{fl/fl}$ control mice (Supplementary Figure S2.4b–d). As expected, T_{FH} cells lacking

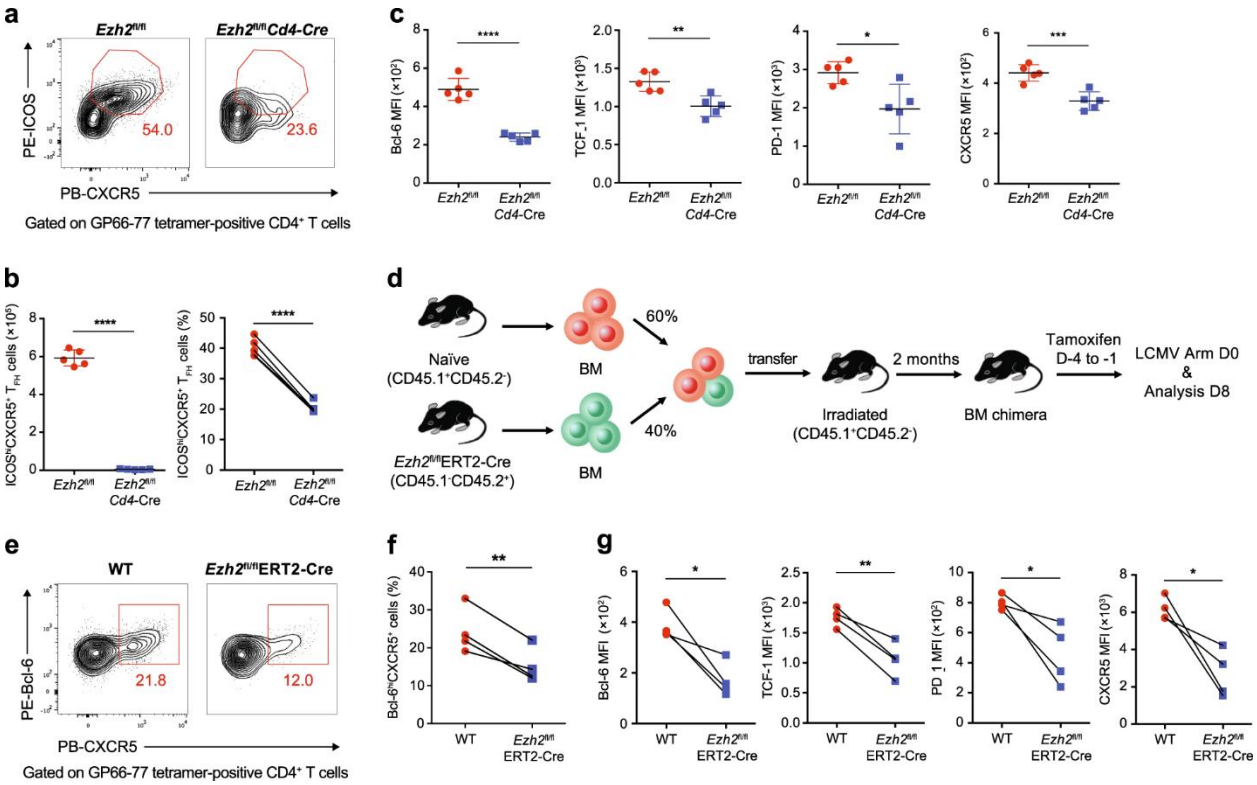


Figure 2. 4 Requirement for EZH2 expression in endogenous virus-specific T_{FH} cell differentiation.

a Flow cytometry analysis of GP66–77 tetramer-positive $CD4^+$ T cells in the spleens of $Ezh2^{fl/fl}$Cd4-Cre mice and $Ezh2^{fl/fl}$ mice on day 8 after LCMV Armstrong infection. The numbers adjacent to the outlined areas indicate the proportions of $ICOS^{hi}CXCR5^+$ T_{FH} cells. **b** The percentage (left panel) and number (right panel) of $ICOS^{hi}CXCR5^+$ T_{FH} cells in **a**. **c** Quantification of Bcl-6, TCF-1, PD-1 and CXCR5 levels in the $ICOS^{hi}CXCR5^+$ T_{FH} cells shown in **a**. **d** Setup of the BM chimera experiment. Irradiated $CD45.1^+$ WT recipients underwent the adoptive transfer of $CD45.1^+$ WT BM cells (60%) and $CD45.2^+$ $Ezh2^{fl/fl}$ ERT2-Cre BM cells (40%). Two months after reconstitution, the recipients were treated with tamoxifen and then infected with the LCMV Armstrong strain. **e** Flow cytometry analysis of GP66–77 tetramer-positive $CD4^+$ T cells in the spleens of recipient mice shown in d on day 8 after the LCMV Armstrong infection. The numbers adjacent to the outlined areas indicate the proportions of $Bcl\text{-}6^{hi}CXCR5^+$ T_{FH} cells, which are summarized in **f**. **g** Quantification of Bcl-6, TCF-1, PD-1 and CXCR5 levels in the $Bcl\text{-}6^{hi}CXCR5^+$ T_{FH} cells shown in **e**. NS not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$ (unpaired two-tailed t-test (**b** and **c**) or paired two-tailed t-test (**f** and **g**)). The data are representative of two independent experiments with at least four mice (**b**, **c**, **f** and **g**) per group (error bars in **b** and c indicate the s.d.)

EZH2 expression displayed reduced levels of an array of T_{FH} lineage-associated proteins, including TCF-1, Bcl-6, PD-1 and ICOS (Supplementary Figure S2.4e).

Consistent with the defective T_{FH} differentiation caused by *Ezh2* deficiency in $CD4^+$ T cells, *Ezh2*$^{fl/fl}$*Cd4*-Cre mice exhibited remarkably reduced frequencies and absolute numbers of PNA^+FAS^+GC B cells in the spleen on day 8 postinfection (Supplementary Figure S2.4f, g). The typical GC structure was also rarely observed in tissue sections of the spleens of *Ezh2*$^{fl/fl}$*Cd4*-Cre mice compared to sections from *Ezh2*$^{fl/fl}$ control mice, as revealed by a confocal microscopy analysis (Supplementary Figure S2.4h). Consistently, we noted less differentiation of B220loCD138hi plasma cells in the spleens of *Ezh2*$^{fl/fl}$*Cd4*-Cre mice than that in the spleens of *Ezh2*$^{fl/fl}$ control mice (Supplementary Figure S2.4i and j). Given the defective T_{FH} differentiation and consequent scarcity of the GC B cell population in response to LCMV infection, *Ezh2*$^{fl/fl}$*Cd4*-Cre mice exhibited substantially reduced LCMV-specific IgG titers on days 8 and 90 postinfection compared with those of their *Ezh2*$^{fl/fl}$ counterparts (Supplementary Figure S2.4k).

We generated BM chimeras by mixing congenitally marked BM cells from *Ezh2*$^{fl/fl}$ERT2-Cre mice (expressing a fusion protein with a tamoxifen-sensitive estrogen receptor variant and Cre; CD45.2, 40%) with BM cells from WT mice (CD45.1, 60%) and subsequently injecting these mixtures into irradiated WT recipients to more precisely evaluate the role of cell-autonomous EZH2 in the mechanism regulating endogenous T_{FH} differentiation (Fig. 2.4d). After reconstitution, we first administered tamoxifen to efficiently delete EZH2 (Supplementary Figure S2.5a) and then infected these chimeric mice with the LCMV Armstrong strain. On day 8 after infection, a decreased percentage of Bcl-$6^{hi}CXCR5^+$ T_{FH} cells was observed among GP66–77 tetramer-positive $CD4^+$ T cells originating from *Ezh2*$^{fl/fl}$ ERT2-Cre mice compared with that of

cells of WT origin (Fig. 2.4e, f). Significantly lower levels of TCF-1, Bcl6, PD-1 and CXCR5 were also detected in EZH2-deficient T_{FH} cells (*Ezh2*$^{fl/fl}$ ERT2-Cre) than those in control T_{FH} cells (WT) (Fig. 2.4g). We observed similar phenotypes in bulk T_{FH} cells from *Ezh2*$^{fl/fl}$ERT2-Cre and WT mice (Supplementary Figure S2.5b–e). Altogether, these data highlighted the importance of cell-intrinsic EZH2 in the mechanism regulating endogenous T_{FH} differentiation in response to an acute viral infection.

EZH2 is not required for the late differentiation and maintenance of virus-specific T_{FH} cells during an acute viral infection

We next determined the role of EZH2 in the late differentiation of T_{FH} cells during an acute infection. BM chimeras were generated by mixing BM cells from *Ezh2*$^{fl/fl}$ERT2-Cre mice (CD45.2, 40%) with BM cells from WT mice (CD45.1, 60%) and subsequently injecting the mixtures into irradiated WT recipients (Fig. 2.5a). After reconstitution, these chimeric mice were initially infected with the LCMV Armstrong strain and then administered tamoxifen from days 4 to 7 after infection to induce the deletion of EZH2 (Fig. 2.5b). Notably, we observed comparable virus-specific Bcl-6^{hi}CXCR5^{+} T_{FH} cell proportions between cells from *Ezh2*$^{fl/fl}$ ERT2-Cre and WT mice (Fig. 2.5c, d). Consistent with the intact late T_{FH} cell differentiation, the levels of molecules related to T_{FH} cell differentiation, including Bcl-6, PD-1 and ICOS, were also similar between cells from *Ezh2*$^{fl/fl}$ ERT2-Cre and WT mice (with the exception of a reduction in TCF-1 expression) (Fig. 2.5e). BM chimeric mice were infected with LCMV Armstrong and then administered tamoxifen from days 9 to 12 to ablate EZH2 expression and further determine whether EZH2 is

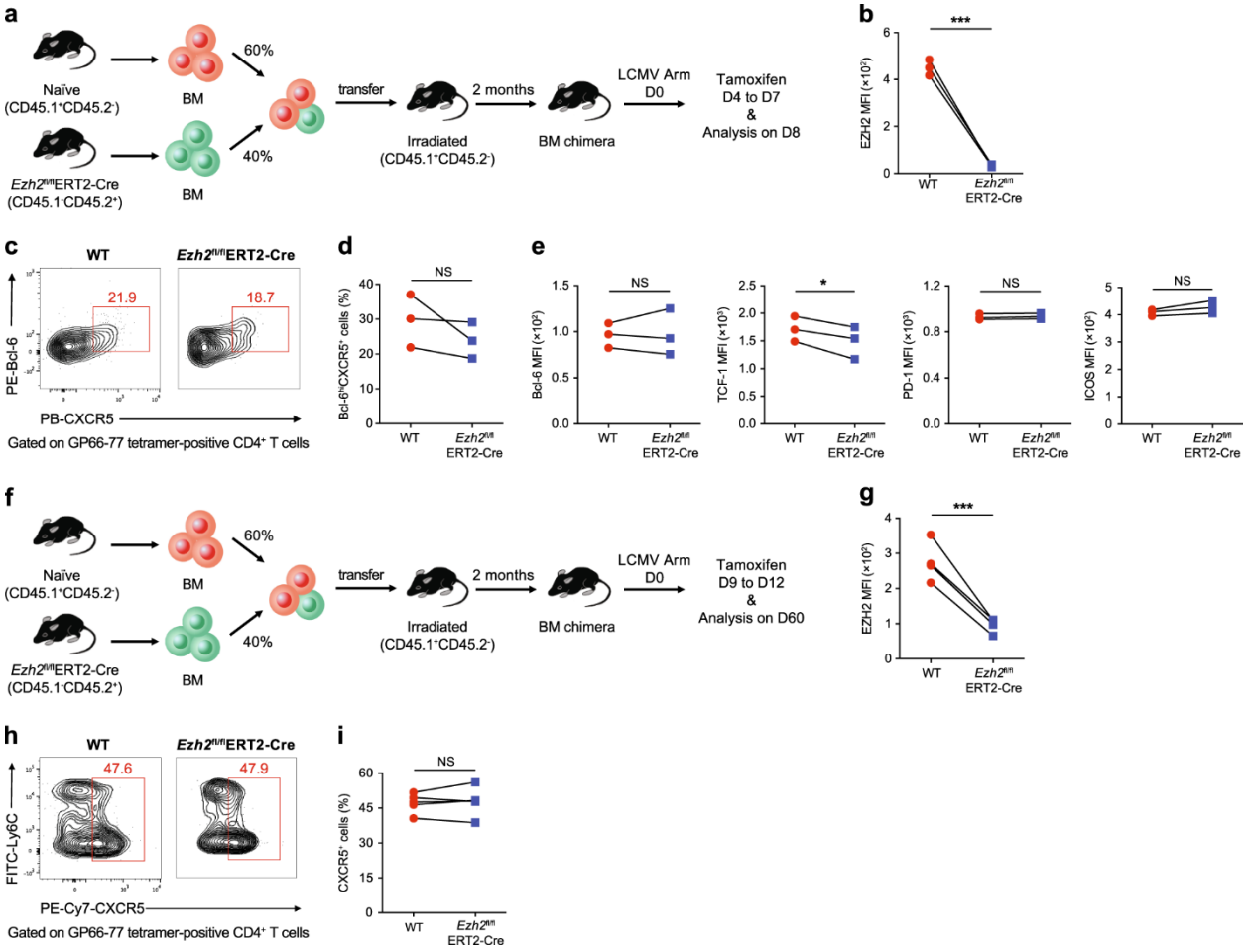


Figure 2. 5 EZH2 is not essential for the late differentiation and maintenance of virus-specific T_{FH} cells during an acute viral infection.

a Setup of the BM chimera experiment. Irradiated CD45.1$^+$ WT recipients underwent adoptive transfer of CD45.1$^+$ WT BM cells (60%) and CD45.2$^+$ Ezh2$^{fl/fl}$ ERT2-Cre BM cells (40%). Two months after reconstitution, the recipients were infected with LCMV Armstrong, followed by the administration of tamoxifen from days 4 to 7 after infection and an analysis on day 8. **b** Expression of EZH2 in virus-specific Bcl6^{hi}CXCR5$^+$ T_{FH} cells from the BM chimeric WT and Ezh2$^{fl/fl}$ ERT2-Cre mice shown in **a**. **c** Flow cytometry analysis of GP66–77 tetramer-positive CD4$^+$ T cells in the spleens of the chimeras described in a on day 8 after the LCMV Armstrong infection. Numbers adjacent to outlined areas indicate the percentage of Bcl-6^{hi}CXCR5$^+$ T_{FH} cells, which were summarized in **d**. **e** Levels of Bcl-6, TCF-1, PD-1 and ICOS in the Bcl-6^{hi}CXCR5$^+$ T_{FH} cells shown in **c**. **f** The BM chimeras described in a were infected with LCMV Armstrong, treated with tamoxifen on days 9 to 12 after infection and analyzed on day 60. **g** Quantification of EZH2 expression in virus-specific CXCR5$^+$ T_{FH} cells originating from BM chimeric WT and Ezh2$^{fl/fl}$ ERT2-Cre mice shown in **f**.

Figure 2. 5, Continuted

h Flow cytometry of GP66–77 tetramer-positive $CD4^+$ T cells in the spleens of the chimeras described in **f** on day 60 after the LCMV Armstrong infection. Numbers adjacent to outlined areas indicate the percentage of $CXCR5^+$ T_{FH} cells, which were summarized in **i**. NS not significant; $*P < 0.05$ and $***P < 0.001$ (paired two-tailed t-test (**b**, **d**, **e**, **g** and **i**)). The data are representative of two independent experiments with at least three mice (**b**, **d**, **e**, **g**, **h** and **i**) per group.

required for the maintenance of virus-specific T_{FH} cells (Fig. 2.5f). On day 60 postinfection, normal virus-specific T_{FH} cell proportions were observed in the absence of EZH2 (Fig. 2.5g–i). Hence, EZH2 is not required for the late differentiation and the maintenance of virus-specific T_{FH} cells in response to an acute viral infection, suggesting a specific effect on early T_{FH} fate commitment.

EZH2 remodels T_{FH} lineage-associated chromatin accessibility during viral infection

We investigated the impact of the EZH2 deficiency on chromatin state changes in virus-specific T_{FH} cells to obtain an understanding of the epigenetic modifications mediated by EZH2 to regulate T_{FH} differentiation. Therefore, we analyzed ATAC-Seq libraries (Supplementary Figure S2.6a, b) generated from WT T_{FH} cells, EZH2-KO T_{FH} cells and naïve $CD4^+$ T cells. Compared to naïve $CD4^+$ cells, both WT T_{FH} and KO T_{FH} cells showed significant changes in opening and closing peaks in the genome, suggesting the presence of different chromatin states (Fig. 2.6a). This finding was clearly revealed in the PCA plot (Fig. 2.6b): the KO T_{FH} cells showed a chromatin state that differed from WT T_{FH} cells and naïve T cells. The EZH2 deficiency resulted in the remodeling of a cluster of specific gene loci that are closely associated with T_{FH} differentiation, which were almost identical to the loci identified in Fig. 2.1d, including Bcl6, Tcf7, Id3, Ascl2, Cxcr5, Icos, Il21 and Sh2d1a[143] (Fig. 2.6c). Notably, these loci were less open in Ezh2 KO T_{FH} cells than those in WT T_{FH} cells (Fig. 2.6c). Because the EZH2-directed H3K27me3 modification is generally associated with a closed state of chromatin accessibility, we further probed whether these loci were marked by the H2K27me3 modification. Therefore, we sorted virus-specific T_{FH}

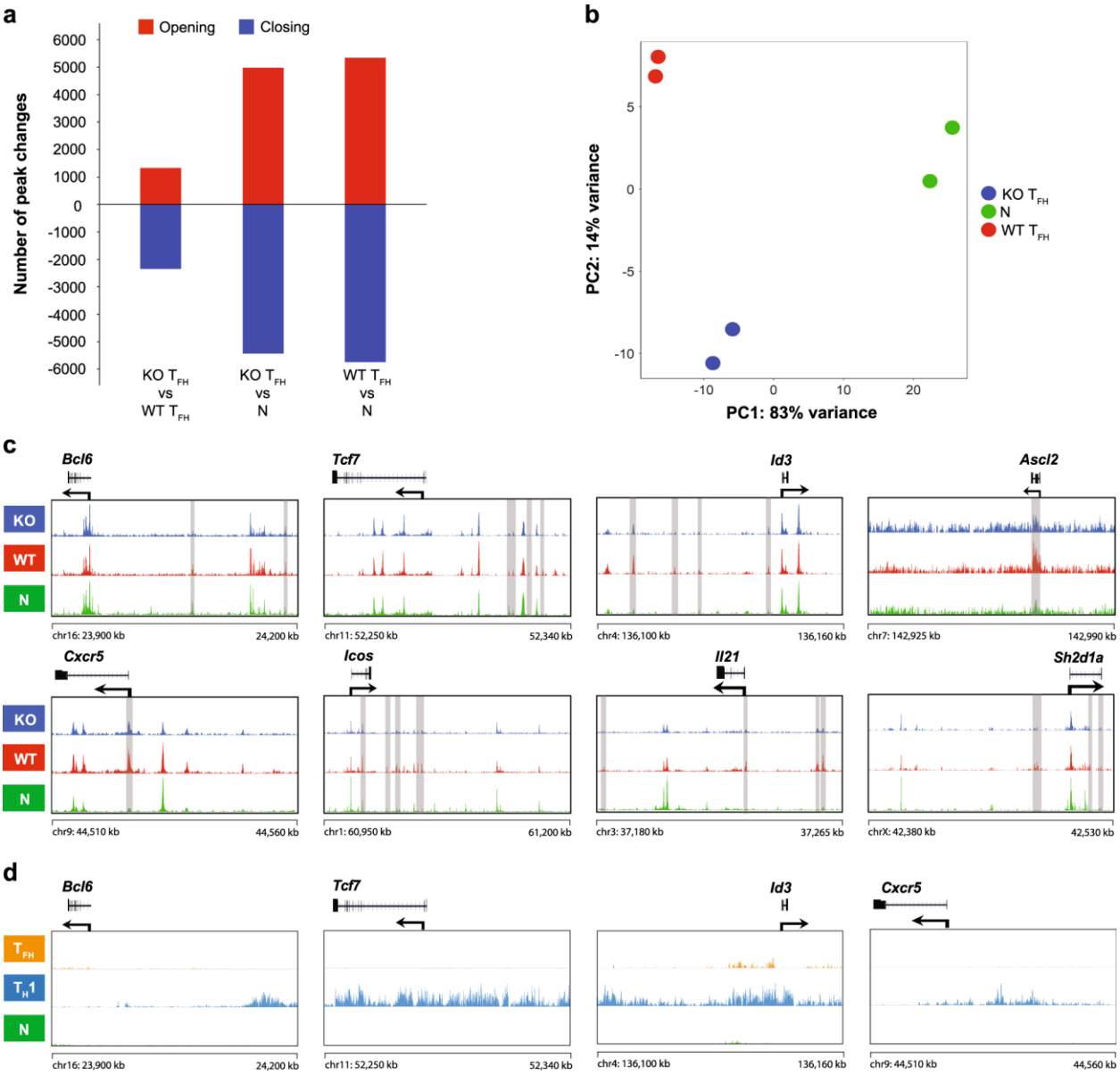


Figure 2. 6 Role of EZH2 in the remodeling of T_{FH} lineage-associated chromatin accessibility during viral infection.

a Numbers of chromatin peaks with differential accessibility (FDR < 0.05; FC > 1.5) in naïve $CD4^+$ T cells (N), WT T_{FH} cells and KO T_{FH} cells. **b** PCA plot of the peak accessibilities in naïve $CD4^+$ T cells (green), WT T_{FH} cells (red) and KO T_{FH} cells (blue). Each dot represents a replicate of the group.

Figure 2. 6, Continued

c ATAC-Seq signal profiles of T_{FH} lineage-associated gene loci in naïve $CD4^+$ T cells (N), WT T_{FH} cells and KO T_{FH} cells. **d** Tracks of H3K27me3 ChIP-Seq data from T_{FH} lineage-associated gene loci in naïve $CD4^+$ T cells (N), T_{FH} cells and T_H1 cells. The data were obtained from two independent experiments with one biological replicate (pooled from at least five mice per group) in each experiment (**a–c**) and from one experiment with one biological replicate (**d**)

cells and T_H1 cells from the SMARTA chimeric mice on day 2 after the LCMV Armstrong infection and naïve $CD4^+$ T cells ($CD4^+CD25^-CD62L^+CD44^-$) from naïve mice and subsequently performed H3K27me3 ChIP-Seq experiments. Strikingly, T_{FH}-associated genes (e.g., *Bcl6, Tcf7, Id3* and *Cxcr5*) were selectively marked by the EZH2-associated H3K27me3 modification in T_H1 cells, but not in T_{FH} cells, during the bifurcated $CD4^+$ T cell differentiation (Fig. 2.6d). Thus, although EZH2 functions as a chromatin repressor, it appeared to rewind the chromatin states of certain T_{FH}-associated genes to favor T_{FH} differentiation.

Bcl-6 overexpression rescues compromised H3K27me3 modification-induced T_{FH} cell differentiation

Consistent with the reduced accessibility of important ChARs in T_{FH} cells caused by the EZH2 deficiency (Fig. 2.6c), the expression of the corresponding genes was substantially reduced, as revealed by microarray analyses (Fig. 2.7a). In addition to the genes analyzed in Fig. 2.6c (including *Bcl6, Tcf7, Id3, Ascl2, Cxcr5, Icos, Il21 and Sh2d1a*), we also observed the downregulation of the mRNA levels of several other T_{FH} differentiation-associated genes, such as *Tcf3*[179], *Cd40lg*[148], *Pdcd1*, *Egr3*[183], *Lef1*, *Batf* and *Cd28*[184] (Fig. 2.7a). We further confirmed our discoveries by performing reverse-transcription quantitative PCR (RT-PCR) assays (Fig. 2.7b). These results raised the possibility that EZH2-mediated H3K27 trimethylation increases the chromatin accessibility, primarily at T_{FH}-differentiation-associated genes, which then induce a favorable transcription program directing T_{FH} lineage commitment.

Bcl-6 is the master regulator of T_{FH} cell differentiation[154-156]. Because virus-specific $CD4^+$ T cells with abundant Bcl-6 and EZH2 expression tend to differentiate into the T_{FH} cell fate (Fig.

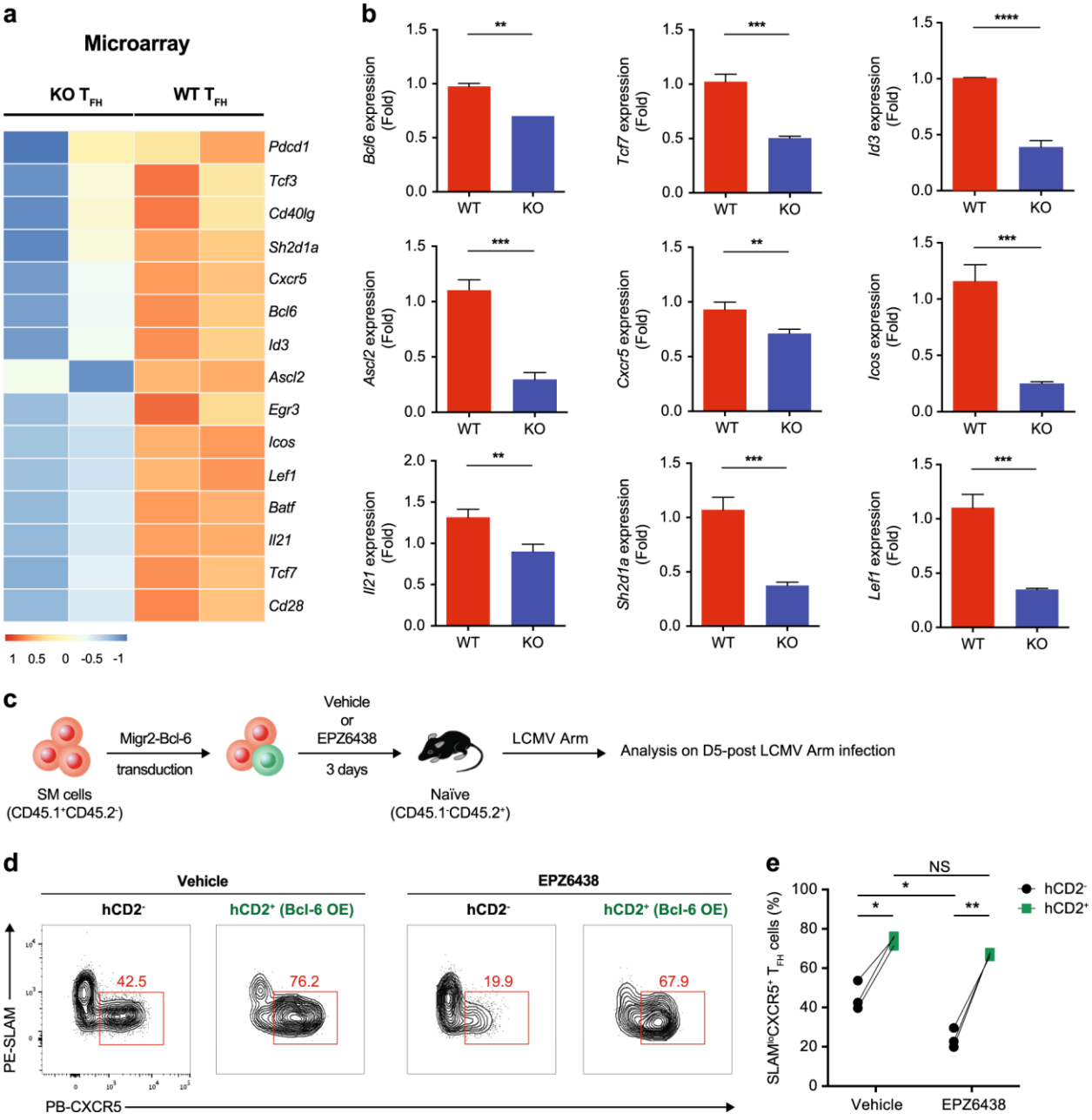


Figure 2. 7 Bcl-6 overexpression rescues compromised H3K27me3 modification-induced T_{FH} cell differentiation.

a Heat map of the canonical T_{FH} lineage-associated genes in WT T_{FH} cells and EZH2 KO T_{FH} cells based on data obtained from the microarray analysis. **b** Quantitative RT-PCR of selected genes in a (normalized to their expression in WT T_{FH} cells). **c** Experimental setup. After treatment with either EPZ6438 or vehicle for 3 days, $CD45.1^{+}$SMARTA (SM) cells transduced with an empty retrovirus or retroviral vector overexpressing Bcl6 were transferred into WT $CD45.2^{+}$ recipients that were subsequently infected with LCMV Armstrong.

Figure 2. 7, Continued

The adoptively transferred SM cells were analyzed on day 5 postinfection. **d** Flow cytometry analysis of vehicle/EPZ6438-treated SM cells that were transduced with ($hCD2^+$) or without ($hCD2^-$) a retrovirus expressing Bcl-6. The numbers adjacent to the outlined areas indicate the percentages of SLAMloCXCR5+ T_{FH} cells, which are summarized in **e**. NS not significant; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ and $****P < 0.0001$ (unpaired two-tailed t-test (**b**); paired two-tailed t-test (**e**)). The data were obtained from one experiment with two biological replicates pooled from at least ten mice per group (**a**), are representative of two independent experiments with three technical replicates pooled from at least ten mice per group (**b**) or are representative of two independent experiments with at least three mice (**e**) per group (**b**; error bars indicate the s.d.)

2.2f–h), and lower levels of the Bcl-6 protein (Figs. 2.3g, k, 4c and g) and mRNA (Fig. 2.7a, b) and reduced chromatin accessibility (Fig. 2.6c) are observed in EZH2-deficient T_{FH} cells, we hypothesized that the impaired T_{FH} cell commitment observed in the absence of EZH2 is potentially driven by compromised Bcl-6 expression and that Bcl6 overexpression might rescue the defective T_{FH} differentiation of activated virus-specific $CD4^+$ T cells with decreased EZH2 activities. Naïve SMARTA cells were transduced with a retrovirus encoding Bcl-6 and then treated with vehicle or the H3K27me3 inhibitor EPZ6438 in vitro for 3 days to evaluate this hypothesis. After treatment, these cells were adoptively transferred into WT recipients, which were immediately infected with the LCMV Armstrong strain (Fig. 2.7c). Consistent with published studies[154-156], forced Bcl-6 expression biased vehicle-treated SMARTA cells to T_{FH} cell differentiation on day 5 postinfection (Fig. 2.7d, e; 42.5% vs. 76.2%). Meanwhile, a substantial reduction in the T_{FH} cell differentiation of nontransduced SMARTA cells was observed in the EPZ6438-treated group compared to that in the vehicle-treated group (Fig. 2.7d, e; 42.5% vs. 19.9%), consistent with our aforementioned data (Fig. 2.3j). Strikingly, Bcl-6 overexpression in EPZ6438-treated SMARTA cells restored the T_{FH} cell commitment to a similar level as the forced Bcl-6 expression-amplified T_{FH} cell commitment in the vehicle group (Fig. 2.7d, e; 76.2% vs. 67.9%). Taken together, the EZH2-associated H3K27me3 modification is a crucial positive determinant of Bcl-6 expression, which further directs T_{FH} cell differentiation.

Discussion

During the initial phase of an acute viral infection, activated virus-specific $CD4^+$ T cells differentiate into either T_H1 or T_{FH} cells[152,176,177]. Bcl-6 is required to specify the T_{FH} fate choice[154-

[156], and Blimp-1 antagonizes this effect by suppressing Bcl-6 expression and Bcl-6-mediated transcriptional activity[156]. Thus, the Bcl-6–Blimp-1 balance coordinates the early T_{FH} commitment. The preserved and elevated TCF-1 expression in activated virus-specific $CD4^+$ T cells induces Bcl-6 expression but inhibits Blimp-1 expression, and thus, it plays a critical role in guiding the T_{FH} commitment pathway[152]. Currently, little is known about the epigenetic mechanism that stabilizes these TF networks during early T_{FH} fate determination. In this study, we defined a distinct chromatin state in virus-specific T_{FH} cells compared to virus-specific T_H1 cells in response to an acute infection. Importantly, early fate-committed T_{FH} cells exhibited higher expression of the histone methyltransferase EZH2 in response to an acute viral infection and accordingly increased levels of the H3K27me3 modification compared with that of their T_H1 counterparts. The ablation of EZH2 by genetic deletion in virus-specific $CD4^+$ T cells substantially reduced the early T_{FH} commitment by disrupting the T_{FH}-specific chromatin state. Specifically, EZH2 is required for the stabilization of chromatin accessibility and transcription pattern of Bcl6, which is essential for the T_{FH} fate commitment. Thus, our results highlighted the importance of EZH2, a widely recognized chromatin repressor, in the priming of T_{FH} cell fate commitment by reinforcing Bcl-6 expression in activated T_{FH} precursors.

T_{FH} differentiation is characterized as a stepwise process[143]: early fate commitment (2 days after an acute infection, occurring at the T–B border), maturation (3–4 days after an acute infection, occurring in B-cell follicle and pre-GC) and memory formation (~1–2 months after an acute infection). In our study, the chromatin identity of virus-specific T_{FH} cells was discerned from virus-specific T_H1 cells within the first 2 days after viral infection. In particular, chromatin remodeling of a cluster of T_{FH} lineage differentiation-associated genes, including *Bcl6, Tcf7, Id3, Ascl2, Cxcr5,*

Icos, Il21 and *Sh2d1a*, was observed in virus-specific T_{FH} cells. This distinct chromatin pattern is maintained throughout the differentiation process and maximized on day 8 postinfection, indicating a progressive mechanism regulating the chromatin states of T_{FH} cells in response to an acute infection.

TCR stimulation is crucial for the effective induction of EZH2 expression in $CD4^+$ T cells[170,185]. The enhanced strength and duration of the interaction between TCR and pMHC favors T_{FH} differentiation[29], which may partially explain the elevated EZH2 expression levels observed in T_{FH} cells compared to those of their T_H1 counterparts following an acute viral infection. During T_{FH} differentiation, EZH2 expression spikes within 2 days after infection and then decreases to the baseline level that is comparable to naive $CD4^+$ T cells. This dynamic pattern of EZH2 expression is further evidenced by a crucial role for EZH2 in the early commitment but not late differentiation or memory maintenance of T_{FH} cells. Strikingly, the ablation of EZH2 in early committed T_{FH} cells leads to the expression of a cluster of T_{FH} lineage differentiation-associated genes with less chromatin accessibility, including *Bcl6, Tcf7, Id3, Ascl2, Cxcr5, Icos, Il21* and *Sh2d1a*. In this regard, TCR-trigged EZH2 seems to imprint a T_{FH}-associated chromatin pattern in T_{FH} precursors at an early stage of the bifurcated T_{FH} and T_H1 differentiation. The EZH2-mediated H3K27me3 modification generally leads to a closed state of chromatin accessibility, but our results identified an essential role for this epigenetic modification in promoting the chromatin accessibility of a cluster of T_{FH}-associated genes in T_{FH} cells. These effects appeared to be indirect, since we did not observe the deposition of H3K27me3 marks at these T_{FH}-associated loci in T_{FH} cells. Alternatively, EZH2 and its associated H3K27me3 modification may orchestrate other chromatin modifications that directly target specific loci of these genes to promote T_{FH} differentiation. Further studies will

be required to identify potential candidate chromatin modifiers downstream of EZH2 and to dissect the underlying regulatory mechanisms.

EZH2 has been reported to negatively regulate the differentiation of an array of effector T_H cell subsets, including T_H1, T_H2 and T_H17 cells[172-174]. These suppressive effects are likely mediated by the EZH2-driven H3K27me3 modifications that directly target and silence a variety of genes required for the differentiation of these cell subsets, including lineage-specifying TFs and/or cytokines, such as *Tbx21*, *Gata3* and *Ifng*, *Il13* and *Il17*. In sharp contrast to its negative role in modulating the differentiation of other T_H lineages, our study revealed that EZH2 instructed the early commitment to T_{FH} cell differentiation by stabilizing a cluster of T_{FH} cell lineage-associated genes, including *Bcl6*. Consistent with our results, a newly published study[186] revealed a requirement for EZH2 in governing T_{FH} cell differentiation by integrating phosphorylation-dependent *Bcl6* activation and H3K27me3-dependent repression of p19Arf. Therefore, EZH2 regulates T_H differentiation in a cell-type-specific manner.

In conclusion, the present study is the first to define a distinct chromatin state in early committed T_{FH} cells that is maintained throughout the acute viral infection. The histone methyltransferase EZH2 is involved in this process and plays a critical role in priming the early T_{FH} cell fate commitment by promoting the stabilization of T_{FH} lineage-specific chromatin pattern. These findings provide valuable insights into strategies targeting EZH2 to improve vaccine efficacy and may aid in the development of novel therapeutic strategies for diseases associated with aberrant T_{FH} differentiation.

Material and Methods

Mice, virus, bacteria and tamoxifen treatment

CD45.1$^+$ SMARTA mice were a gift from Dr. R. Ahmed (Atlanta, USA). The *Ezh2*$^{fl/fl}$, *Cd4*-Cre transgenic, ERT2-Cre transgenic and C57BL/6J (CD45.1 and CD45.2) mice were obtained from Jackson Laboratories. The LCMV Armstrong strain was provided by Dr. R. Ahmed (Atlanta, USA), and 2×10^5 plaque-forming units (PFUs) of this strain were intraperitoneally injected into mice to establish an acute viral infection. *Listeria monocytogenes* expressing the LCMV glycoprotein-specific I-A^b-restricted CD4$^+$ T cell epitope GP61–80 (LM-GP61) was created from a vector strain[187], and 1×10^7 colony-forming units (CFUs) of the recombinant bacteria were intravenously injected to establish a bacterial infection in mice. Six- to ten-week old mice of both sexes were infected without randomization or "blinding". Bone marrow (BM) chimera mice were infected 2 months after reconstitution. Tamoxifen (T5648; Sigma-Aldrich; 10 mg/ml) in sunflower oil (S5007; Sigma-Aldrich) was intraperitoneally injected into mice at a daily dose of 1 mg/mouse for 4 days. Infected mice were housed in accordance with the institutional biosafety regulations of the Third Military Medical University. All mouse experiments were performed according to the guidelines of the Institutional Animal Care and Use Committees of the Third Military Medical University.

ATAC-Seq library preparation

The ATAC-Seq libraries were prepared as previously described[178]. Briefly, 50,000 target cells were washed with PBS and then treated with lysis buffer, followed by labeling with the Nextera enzyme (15027865; Illumina). The labeled samples were immediately amplified by 9–10

cycles of polymerase chain reaction (PCR) with barcoded primers and sequenced with a HiSeq4000 instrument in a 150 bp/150 bp paired-end run or a NextSeq500 instrument in a 76 bp/76 bp paired-end run.

ATAC-Seq data preprocessing

Raw sequencing reads were first trimmed of adapters to improve the quality using Trim Galore! v0.4.4 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), which is a wrapper based on CutAdapt v1.14[188] and FastQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired-end reads that passed quality control (QC) were then aligned to mm10 using Bowtie2 v2.2.9[189]. The resulting BAM files were then filtered again to remove unmapped reads, mate-unmapped reads, nonprimary aligned reads, reads that failed platform quality checks and PCR duplicate reads using SAMtools v1.4.1[190] (-F 1804). In addition, reads mapped to ChrM were also removed and PCR duplicate reads were further identified and removed using Picard v2.16.0 MarkDuplicate (https://broadinstitute.github.io/picard/).

The insert size distributions were then calculated using Picard v2.16.0 CollectInsertSizeMetrics. Since Tn5 transposase binds as a dimer and inserts two adaptors separated by 9 bp[191], all aligned reads were shifted by + 4 bp on the positive strand and −5 bp on the negative strand using deepTools v2.5.2 alignmentSieve[192]. Afterward, peak calling was performed using MACS2 v2.1.1[193], with a q-value threshold of 0.01. Peaks that overlapped with the Encode black list regions were then removed using bedtools intersect v2.26.0[193]. We then merged peaks from all replicates and filtered peaks that are not reproducible, based on the

Irreproducible Discovery Rate (IDR)[194] (IDR < 0.005), across at least one pair of replicates in each sample group. Subsequently, depending on the comparisons required for different purposes, filtered peaks from multiple sample groups were merged using the bedtools v2.26.0 merge algorithm to create genome-wide atlas of accessible chromatin regions for further analysis.

Assignment of ATAC-Seq peaks to genes

Each ATAC-Seq peak was uniquely assigned to a gene, with the corresponding feature annotations (e.g., exon, intron and intergenic) in R v3.4.1 using ChipPeakAnno v3.10.2[195], with TxDb.Mmusculus.UCSC.mm10.knownGene v3.4.0 as the reference database. The following rules listed in order of priority were used for assignment: peaks that overlap with the transcribed region of a gene or within 2 kb upstream or downstream of the transcribed region are assigned to that gene; any remaining peaks are assigned to the nearest gene based on the distance from the center of the peak to the transcription start site (TSS) or 3′ end of the transcribed region. Regions within 5 kb of the TSS are defined as the promoter regions.

Differential peak analysis and principal component analysis

For each analysis that compares data from multiple sample groups, all required shifted BAM files from all replicates of those samples were used to generate an accessibility matrix by counting the normalized reads (normalized by DESeq2-calculated size factors) within each peak region of the corresponding atlas peak file using deepTools v.2.5.4 multiBamSummary in BED-mode. The resulting matrix was input to DESeq2 v1.16.1[196] to calculate the differential

accessibility of the peaks of the relevant pairs. Principal component analysis (PCA) plots were then generated using DESeq2 v1.16.1.

Coverage plots and chromatin accessibility heat map

For each sample group, shifted and RPM normalized bam files were first converted to bedgraph files by counting the normalized reads in 10 bp bins and removing reads in the Encode blacklisted regions. Then, reads obtained from replicates of the same sample group were pooled using bedtools v2.26.0 unionbedg, and the resulting merged bedgraph files were used to generate a single coverage plot for each sample group, which was visualized in IGV v2.4.4[197] and used to generate the chromatin accessibility heat map. Using deepTools v2.5.4 computeMatrix, each atlas peak was extended to a ±1 kb region from the peak center, and the reads from the final coverage plot were then separated into 10 bp bins to be represented as a row in the heat map. Selected differential peaks were then stacked together to generate the overall heat map using deepTools v2.5.4 plotHeatmap. Bins with read counts greater than a threshold, which was defined as the 75th percentile +1.5 × inter quartile range (IQR), were capped at that threshold value to increase the visibility of low-signal regions. Capping was only performed to plot the heat map, and all other analyses were performed with uncapped values.

Flow cytometry and antibodies

Flow cytometry data were obtained with a FACSCanto II instrument (BD Biosciences) and analyzed using FlowJo software (Tree Star). The major histocompatibility complex (MHC) class II (I-A^b) tetramer of LCMV epitope of GP66–77 was provided by Dr. Rafi Ahmed (Emory

University). CXCR5 staining has been described previously[152]. Surface staining was performed in PBS containing 2% fetal bovine serum (wt/vol) on ice. Intracellular staining of Foxp3, EZH2, H3K27me3, TCF-1 and Bcl-6 was performed using the Foxp3/Transcription Factor Staining Buffer Set (00–5523; eBioscience). MHC II GP66–77 tetramer staining was performed by incubating cells with the tetramer for 1 h at 37 °C.

Immunofluorescence staining

Immunofluorescence staining was performed using previously described methods[198]. Briefly, frozen sections of the spleen were fixed with cold acetone for 10 min at 4 °C and blocked with 5% normal goat serum for 30 min. Sections were then stained with biotin-conjugated PNA (RL-1072; Vector), Alexa Fluor 647-conjugated IgD (11–26c.2a; eBioscience) and PE-conjugated CD4 (RM4–5; BD Biosciences) antibodies, followed by Alexa Fluor 488-conjugated streptavidin (25–4317–82; Invitrogen). Cover slips on which both types of cells were cultured were mounted on slides with the ProLong Antifade Kit (P-7481; Life Technologies) and examined under a Zeiss LSM 510 confocal fluorescence microscope. The images were processed with ImageJ software.

Retroviral constructs and transduction

The sequences encoding the codon-improved Cre (iCre) gene or Bcl6 were amplified and cloned into the vector MIGR1 (MSCV-IRES-GFP) or MIGR2 (MSCV-IRES-hCD2), respectively. Retroviruses were packaged by transfecting 293T cells with the retroviral vectors along with the pCL^{eco} plasmid. SMARTA cells were activated in vivo by injecting 200 μg of the GP61–77 peptide into SMARTA mice. Eighteen hours later, activated SMARTA cells were purified and “spin-

infected" by centrifugation (800 g) with retrovirus supernatants, 20 ng/ml IL-2 (130–098–221; Miltenyi Biotec) and 8 μg/ml polybrene (H9268; Sigma-Aldrich) at 37 °C for 90 min. SMARTA cells were then transferred into recipient mice, followed by the infection of the hosts with LCMV Armstrong.

Adoptive transfer

A total of 5×10^5 (for analysis on days 2, 3 or 5) or 1×10^4 (for analysis on day 8 or later) CD45.1$^+$ SMARTA cells (naïve or retrovirus-transduced) were adoptively transferred into CD45.2$^+$ recipients. On the following day, the recipients were intraperitoneally injected with 1×10^6 PFUs of LCMV Armstrong (day 2 or 5) or 1×10^7 CFUs of LM-GP66 (day 3) or were intraperitoneally injected with 2×10^5 PFUs of LCMV Armstrong (day 8 or later). For the EPZ6438-treated SMARTA cell transfer experiment, naïve CD45.1$^+$ SMARTA cells were treated with EPZ6438 (2 μM; E-7438, Active Biochem) or vehicle at 37 °C for 3 days, and then transferred into CD45.2$^+$ recipient mice, followed by infection with LCMV Armstrong.

BM chimeras

A total of 2×10^6 BM cells harvested from *Ezh2*$^{fl/fl}$ERT2-Cre (CD45.2$^+$) mice and C57BL/6J wild-type (WT) (CD45.1$^+$) mice were mixed at a ratio of 4:6. Mixed BM cells were then intravenously injected into lethally irradiated (two doses of 550 rads each) CD45.1$^+$ WT recipients. After a 2-month reconstitution, the recipient mice were infected with LCMV Armstrong.

ELISA

The LCMV-specific serum IgG titers were measured using an ELISA, as previously described[199,200]. Briefly, plates were coated with LCMV-infected cell lysates, and LCMV-specific antibodies were detected with HRP-conjugated goat anti-mouse IgG secondary antibodies (Southern Biotech).

Microarray

T_{FH} cells were isolated using a previously described method[152]. Briefly, total splenocytes obtained from WT and *Ezh2*$^{fl/fl}$*Cd4*-Cre mice on day 8 after LCMV Armstrong infection were subjected to the depletion of lineage marker-positive cells (Lin^+ cells) using biotin-conjugated antibodies (anti-B220 (RA3–6B2), anti-CD8 (53.6.7), anti-CD11c (N418), anti-TER-119 (TER-119) and anti-NK1.1 (PK136); all from BioLegend), followed by coupling to BeaverBeads Mag500 Streptavidin (22302; Beaver). The enriched Lin^- cells were then stained with anti-CD4, anti-CD44, anti-GITR, anti-CD25 and anti-CXCR5 antibodies. The $CD4^+CD25^-GITR^-CD44^+CXCR5^+$ T_{FH} cells were sorted with a FACS Aria II cell sorter (BD Biosciences) and then immediately lysed with TRIzol LS reagent (10296; Life Technologies). Then, total RNA was extracted and submitted to CapitalBio for a microarray analysis.

Quantitative reverse-transcription PCR

Sorted T_{FH} cells were sorted directly into TRIzol LS reagent (10296; Life Technologies). Total RNA was extracted and reverse transcribed with the RevertAid Minus First Strand cDNA Synthesis Kit (K1632; Thermo Scientific). The relative expression of various genes was

determined using AceQ qPCR SYBR Green Master Mix (Q111; Vazyme) with a CFX96 Touch Real-Time System (Bio-Rad).

ChIP

ChIP assays of sorted T_{FH} cells, T_H1 cells and naïve $CD4^+$ T cells were performed using a Simple Enzymatic Chromatin IP Kit (Magnetic Beads) (9003; Cell Signaling Technology). The resulting chromatin fragments were immunoprecipitated with an anti-H3K27me3 antibody (9733; Cell Signaling Technology), followed by binding to ChIP Grade Protein G Magnetic Beads (9006; Cell Signaling Technology) and purification with a PCR purification kit (28104; Qiagen). Sequence-indexed libraries were prepared from immunoprecipitated chromatin fragments using the Illumina TruSeq indexed pair-ended DNA library preparation protocol and ultimately sequenced using the NextSeq500 platform.

ChIP-Seq analysis and coverage plot

The raw sequencing reads obtained from the Chip-Seq analysis were first trimmed and filtered with Trim Galore! v0.4.4, and then aligned in single-end mode and searched against mm10 using Bowtie2 v2.2.9. The reads were then filtered using samtools to remove low-quality reads and unmapped reads. Duplicate reads were also filtered using Picard v2.16.0 MarkDuplicates. Peak calling was performed using MACS2 v2.1.1 with a q-value threshold of 0.05 and the --SPMR flag. The resulting bedgraph files were used to build a coverage plot using MACS2 v2.1.1 bdgcmp with the logLR method and a P-value of $1e^{-5}$.

Statistical analysis

Statistical analyses were conducted with Prism 6.0 software (GraphPad). An unpaired two-tailed t-test with a 95% confidence interval was used to calculate P-values. For retroviral transduction, SMARTA cell cotransfer, spleen chimera and BM chimera experiments, and a paired two-tailed t-test with the 95% confidence interval were used to calculate P-values.

Supplementary Figures

Supplementary figures for this chapter are included in Appendix A: Supplementary Figures.

CHAPTER 3: CONTRAST DYSFUNCTIONAL CD8 T CELL DIFFERENTIATION IN CANCER AND CHRONIC INFECTION

Note:

The following section (Chapter 3) is largely adapted from a manuscript in preparation where I am a co-first author.

Authors:

Xiangyu Chen*, Guoshuai Cao*, Yao Lin, Nick Asby, Yifei Hu, Zhiwei Pan, Shuai Yue, Qifeng He, Yang Yang, Zhirong Li, Lifan Xu, Qizhao Huang, Yaxing Hao, Jianfang Tang, Li Hu, Beicheng Sun, Jun Huang, Lilin Ye.

* These authors contributed equally to this work.

Author Contributions:

X.C., Y.L., N.A., Y.H., Z.P., S.Y., Q.H., Y.Y., Z.L., L.X., Q.H., Y.H., J.T. and L.H. performed the experiments. **G.C. performed the bioinformatic analyses.** L.Y. designed the study, analyzed the data and wrote the paper with X.C., G.C., and J.H.; and B.S., J.H. and L.Y. supervised the study.

Abstract

Persistent antigen exposure during chronic viral infection or tumorigenesis causes functional exhaustion of antigen-specific $CD8^+$ T cells. Blocking PD-1/PD-L1 pathway of exhausted $CD8^+$ T (Tex) cells can partially restore their effector functions, resulting in viral load decrease or tumor remission. Despite intensive studies in characterizing the commonalities of T cell exhaustion between chronic viral infection and cancer, little is known about their differences. To decipher disease-specific T cell exhaustion program, we engineered a system that generates Tex cells bearing a same transgenic TCR stimulated by the same epitope presented in both chronic viral infection and tumor. The transcriptional module analysis of differential gene expression of bulk Tex cells revealed distinct effector- and memory-like T cell exhaustion programs in chronic viral infection and tumor, respectively. Consistently, single-cell transcriptional analysis also demonstrated that canonical effector-associated gene signatures selectively enriched in the functional subsets of Tex cells in chronic viral infection, while memory-related genes more enriched in Tex subsets in tumor. Such dichotomous T cell exhaustion programs further dictated the differential responsiveness of Tex cells to PD-L1 immune checkpoint blockade (ICB). Our study uncovered disease-dependent T cell exhaustion programs and highlighted the importance of developing disease-specific ICB to respectively treat chronic viral infection and cancer.

Full Text

Introduction

In contrast to effector to memory $CD8^+$ T cell differentiation followed by antigen clearance in acute viral infection, persistent antigen stimulation in chronic viral infection and tumor leads to the generation of exhausted $CD8^+$ T (Tex) cells[201,202]. Previous studies have found that Tex cells

in both diseases share many similar exhaustion-associated phenotypes, including impaired effector functions (cytokine secretion and cytotoxic activities), reduced proliferative capacities, lost memory potential, as well as elevated and sustained expression of an array of inhibitory immune checkpoint receptors that dampen TCR signaling, such as PD-1, CTLA-4 and Tim-3, as compared to effector and memory $CD8^+$ T cells[201,202]. In addition, previous studies also have uncovered common epigenetic and transcriptional profiles of T cell exhaustion in chronic viral infection and cancer[62].

However, chronic viral infection and cancer are two different types of diseases with different physiopathology[203]. Despite intensive studies in T cell exhaustion, very little is known about the transcriptional and epigenetic differences of T cell exhaustion between these two distinct diseases[88,204]. The immune checkpoint blockade (ICB) therapies have been approved for treating a broad range of cancers with great clinical successes[205,206], but none of these have been approved to treat chronic viral infection[207,208]. Understanding disease-specific T cell exhaustion programs will provide important insights into the development of specific, effective immunotherapies to better treat each disease, particularly chronic viral infection.

Results

Distinct transcriptional modules identify disease-specific T cell exhaustion

To probe the possible disease-specific transcriptional profiles of Tex cells, we sought to compare Tex cells that originated from the same antigen-specific naïve $CD8^+$ T cells but developed in either chronic viral infection or tumor microenvironment. To this end, we first generated a mouse melanoma cell line B16F10 that expresses lymphocytic choriomeningitis virus Clone 13

strain (LCMV Cl13) glycoprotein (GP) using CRISPR-Cas9 gene knock-in technology (hereinafter referred to as B16F10-GP) (Supplementary Figure S3.1 a, b). As expected, TCR-transgenic $CD8^+$ T cells, P14, specifically recognizing LCMV glycoprotein epitope H-$2D^b$GP33-41, proliferated extensively in B16F10-GP but not in control B16F10 tumor (Supplementary Figure S3.1 c, d). Next, congenic naïve ($CD45.1^+CD44^{lo}CD62L^{hi}$) P14 cells were adoptively transferred into LCMV Cl13 chronically infected or B16F10-GP cell engrafted C57/BL6 recipients ($CD45.2^+$). In this scenario, we were able to compare the characteristics of Tex cells of the same origin but developed in distinct diseases. As a control, we also transferred the same naïve P14 cells into recipients that were subsequently challenged with LCMV Armstrong strain to establish an acute viral infection. P14 cells from the late stage (day 21 post cell transfer) of both chronic viral infection and tumor exhibited canonical phenotypic and functional exhaustion, such as high inhibitory receptor expressions (PD-1, Tim-3 and 2B4) and impaired effector function (CD107, IFN-γ, TNF-α and IL-2) compared to functional effector P14 cells at day 8 post-acute infection (Supplementary Figure S3.1 e, f). Next, P14 cells from the spleens of recipients in acute viral infection (referred as A), the spleens of chronic viral infection (referred as C), or tumor tissues (referred as T) were sorted at indicated time points and RNA from these cells were extracted for subsequent bulk RNA sequencing (RNA-seq) (Fig. 3.1a).

Principal component analysis (PCA) of the bulk RNA-seq data revealed primary exhaustion-specific transcriptional profiles (PC1, 58% variance) coupled with secondary disease-specific transcriptional signatures (PC2, 19% variance). Overall, the transcriptional profiles of P14 Tex cells (C4, C8 and C21; T4, T8 and T21) in chronic viral infection and tumor were more similar compared to those of P14 effector (A4 and A8) and memory (A90 and A300) cells in acute

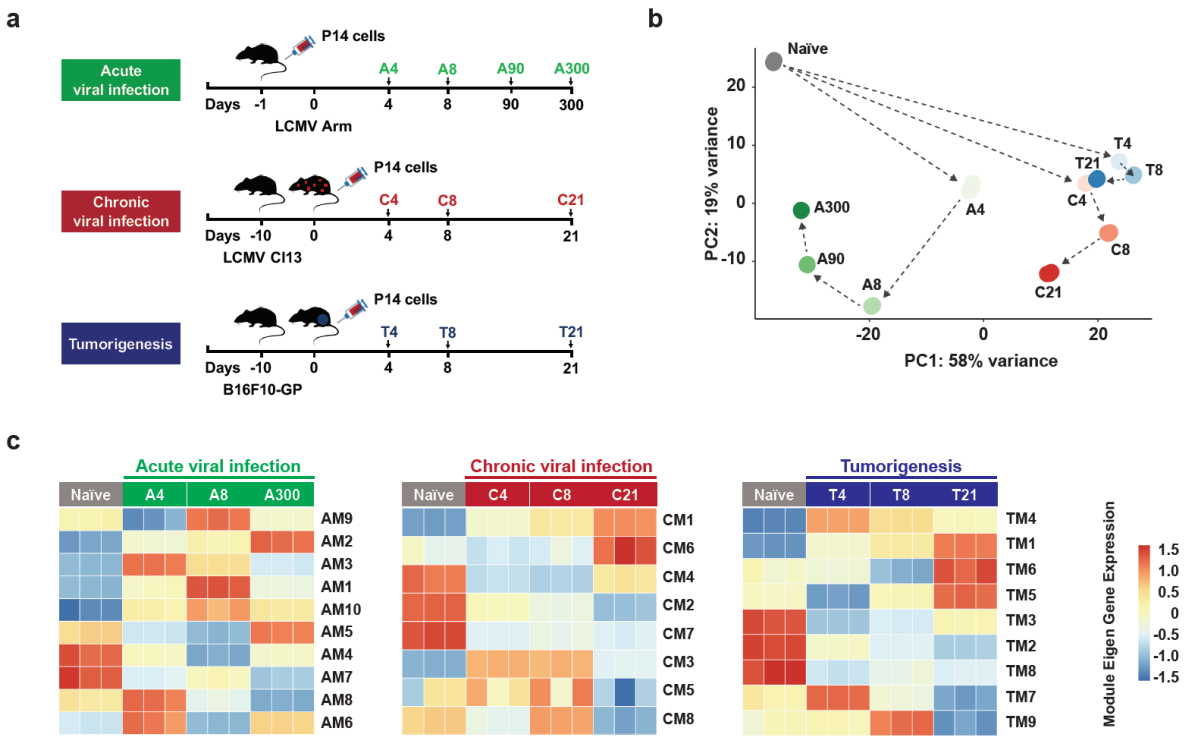


Figure 3. 1 Distinct transcriptional modules identify disease-specific T cell exhaustion.

a, Experimental scheme of bulk RNA-seq. **b**, Principal component analysis (PCA) of transcriptional profiles of P14 $CD8^+$ T cells during acute viral infection (green), chronic viral infection (red) and tumorigenesis (blue). **c**, Weighted gene coexpression network analysis (WGCNA) was performed on the transcriptome of P14 $CD8^+$ T cells in each disease to identify gene modules and the eigengenes of the modules were visualized as heatmaps. AM, acute viral infection module; CM, chronic viral infection module; TM, tumor module.

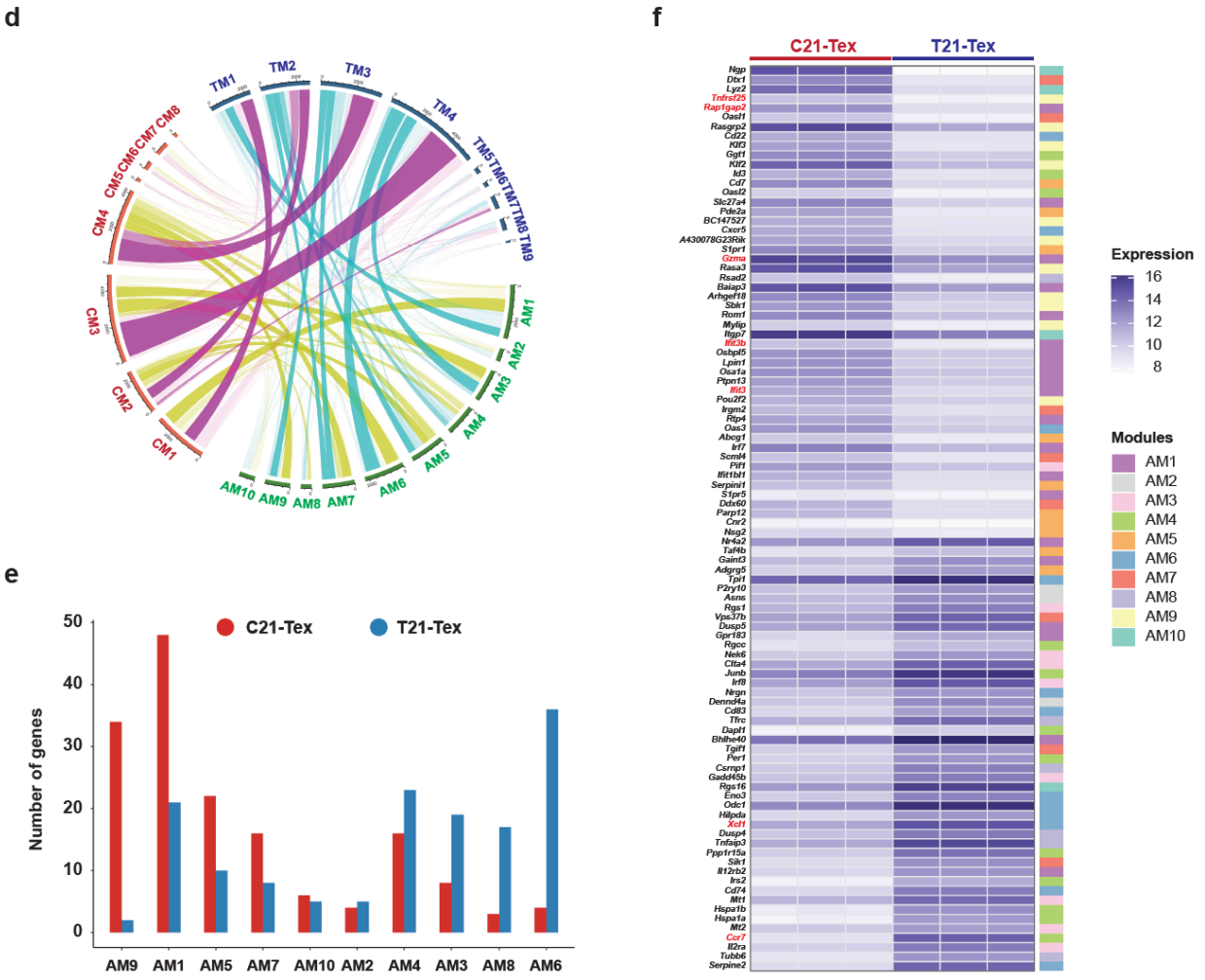


Figure 3.1, Continued

d, Chord diagram of gene overlap among AM, CM and TM modules. The width of each ribbon shows the number of genes overlapped, and the color indicates the pair of two disease modules: AM/CM (yellow), AM/TM (blue) and CM/TM (purple). Only statistically significant overlaps with FDR < 0.05 are shown (Fisher's test). The opacity indicates the statistical significance of the overlap. Specific statistics can be found in Supplementary Figure S3.2a. **e**, Distribution of differential expression genes (DEGs) between C21-Tex cells and T21-Tex cells across modules of acute viral infection. **f**, Heatmap showing the normalized expression of top 100 DEGs between C21-Tex cells and T21-Tex cells.

infection (Fig. 3.1b). However, the two exhaustion trajectories (C versus T) also became increasingly divergent along disease progression (Fig. 3.1b), indicating accumulative and distinct transcriptional remodeling in Tex cells as a result of each unique disease microenvironment.

We next globally examined the gene profiles and performed disease-specific weighted gene co-expression network analysis (WGCNA)[63] using the top 10,000 variable genes along T-cell differentiation path and identified 10 acute viral infection (AM), 8 chronic viral infection (CM) and 9 tumorigenesis (TM) modules in each disease setting (Fig. 3.1c). In the acute viral infection, different sets of functional modules were observed at proliferation (AM3, AM6 and AM8) (day 4, A4), effector (AM1 and AM9) (day 8, A8) and memory (AM2, AM4 and AM5) (day 300, A300) phases. Similarly, in both the chronic viral infection and tumor, specific modules were also found to function at corresponding exhaustion stages (e.g. CM3/5 and TM4/7 at day 4, C4/T4; CM3/5/8 and TM9 at day 8, C8/T8; CM1/6 and TM1/6 at day 21, C21/T21), consistent with current understanding that different transcriptional programs govern different stages of Tex differentiation[209]. To compare the transcriptional modules across diseases, we conducted pair-wise comparisons of modules and identified shared and distinct genes (Fig. 3.1d, Supplementary Figure S3.2a, b) and functional gene sets (Supplementary Figure S3.2c) between each pair. While each AM had small and diverse overlaps with multiple CMs and TMs activated at different time points, the substantial gene overlaps were more observed between certain pairs of CMs and TMs (e.g. CM3 and TM4) (Fig. 3.1d and Supplementary Figure S3.2a), suggesting significant transcriptional rewiring between effector/memory T cells and Tex cells. Furthermore, even among the CMs and TMs, one-to-one significant gene overlaps exist among only a third of total modules (Fig. 3.1d and Supplementary Figure S3.2a-c), highlighting disease-specific elements of their transcriptional

networks. Coincident with this point, we identified ~1,389 differential expression genes (DEGs) between chronic viral infection- and tumor-derived Tex cells at day 21 post-naïve P14 cell transfer (Supplementary Figure S3.2d).

Tex cells are generally characterized as less proliferative, dysfunctional and memory potential losing, as compared to canonical effector and memory $CD8^+$ T cells[62,201,202]. To examine whether Tex cells from chronic viral infection or tumor microenvironment preserve certain traits of effector function and memory properties, we grouped these identified DEGs into aforementioned transcriptional modules enriched in functionally effector and memory P14 cells differentiated in acute viral infection. Notably, DEGs specifically upregulated in C21-Tex cells were more biased towards effector-associated modules (AM1 and AM9), whereas DEGs of T21-Tex cells were most enriched in two modules that were re-activated in memory phase from effector phase (AM4 and AM6) (Fig. 3.1e and Supplementary Figure S3.2e). Accordingly, signature genes of effector $CD8^+$ T cells, such as *Tnfrsf25*, *Rap1gap2*, *Gzma*, *Ifit3*, and *Ifit3b*, were among the top 50 DEGs enriched in C21-Tex cells. By contrast, genes associated with memory characteristics, including *Xcl1* and *Ccr7*, were highlighted in the top 50 DEGs in T21-Tex cells (Fig. 3.1f). Taken together, these results revealed the differential transcriptomic profiles of Tex cells developed in chronic viral infection and tumor, which specified the nature of their exhaustion programs reminiscent of effector- or memory-characteristics, respectively.

Single-cell deconvolution of disease-specific exhausted T cells.

Tex cells are heterogeneous, encompassing various cellular subsets at distinct differentiation or functional states[85,199,209-212]. Given the transcriptomic differences between

chronic viral infection- and tumor-derived Tex cells at the bulk level, we further sought to compare the transcriptional differences at the single-cell level. To this end, single-cell RNA-seq (scRNA-seq) was performed on C21-Tex and T21-Tex cells isolated from spleens of chronically infected mice and subcutaneous tumor tissues of tumor engrafted mice, respectively (Fig. 3.2a, I and II, herein as C21_spl and T21_sub). To avoid potential confounding tissue-specific transcriptomic signatures, we further transferred congenic naïve P14 T cells ($CD45.1^+$) into recipients ($CD45.2^+$) that were subsequently infected with LCMV Cl13 or engrafted with B16F10-GP cells intravenously to establish chronic viral infection or melanoma metastasis in the lung tissues. Lung-resident P14 $CD8^+$ T cells were isolated from recipients with chronic viral infection or lung metastasis on day 21 and analyzed by scRNA-seq (Fig. 3.2a, III and IV, hereinafter as C21_lung and T21_lung; Supplementary Figure S3.3a, b).

By integrating the two sets of datasets (Fig. 3.2a, I-IV) and performing uniform manifold approximation and projection (UMAP) analysis, we found that C21- and T21-Tex cells clearly partitioned into distinct regions (Fig. 3.2b, left panel). Notably, Tex cells of the same disease but from different tissues largely overlapped (Fig. 3.2b, middle and right panels), indicating that the disease type primarily discriminates transcriptional profiles of Tex cells. To validate the differential effector- or memory-like characteristics of C21-or T21-Tex cells based on the bulk RNA-seq, we calculated module scores of AM1 (effector associated), AM4 and AM6 (memory associated) modules in each individual cell. Consistent with the distinct transcriptional profiles of C21- and T21-Tex cells at bulk level, UMAP showed that AM1 module was more enriched in C21-Tex cell regions, while higher scores of AM4 module and AM6 module were found in

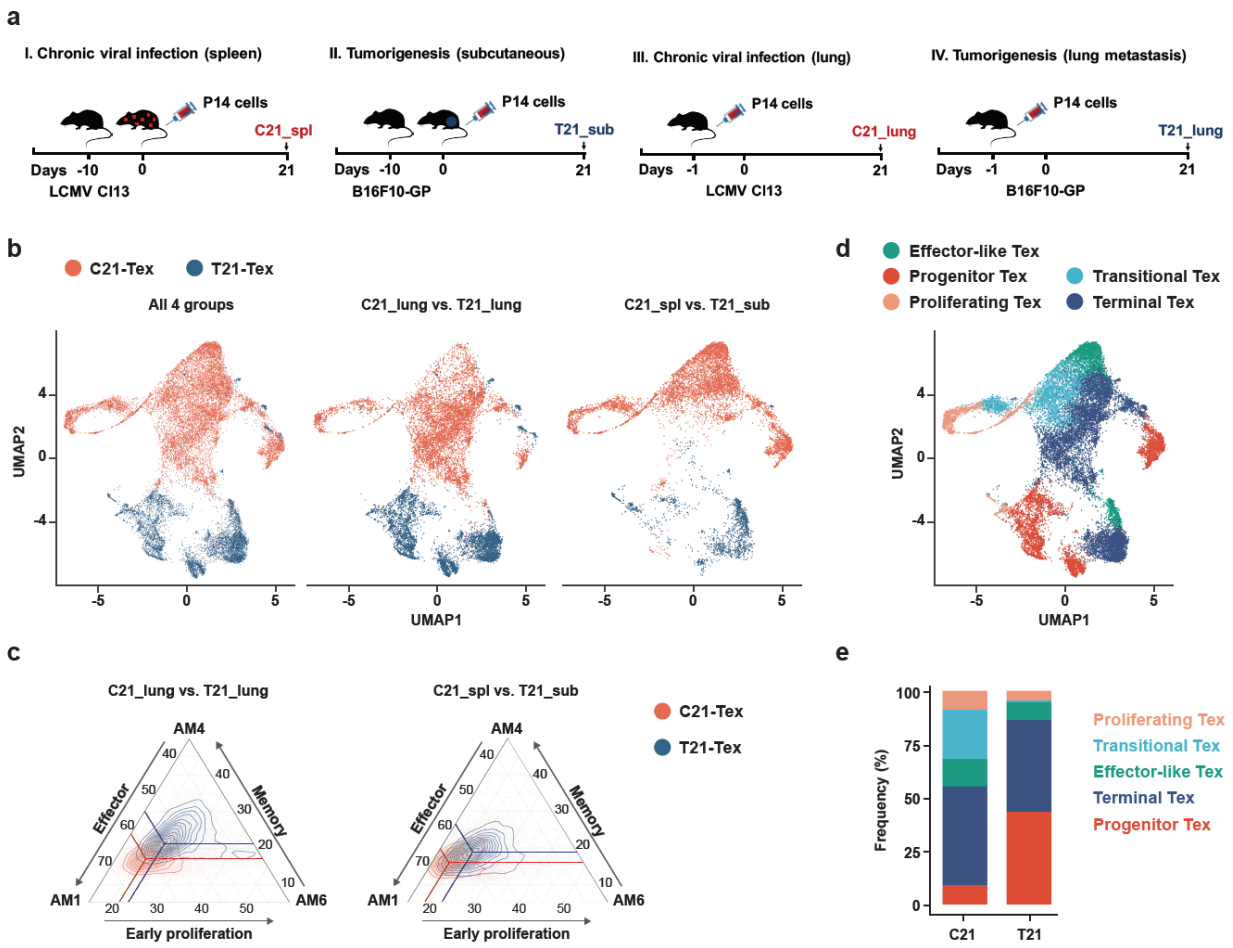


Figure 3. 2 Single-cell transcriptomes of disease-specific exhausted T cells

a, Experimental scheme of scRNA-seq. **b**, UMAP of C21_spl, C21_lung, T21_sub and T21_lung Tex cells. Cells are colored by disease types from which they were collected. (Red represents C21-Tex, and blue represents T21-Tex). Left: all 4 groups. Middle: C21_lung Tex versus T21_lung Tex. Right: C21_spl Tex versus T21_sub Tex. **c**, Ternary maps based on AM1 (effector), AM4 (memory) and AM6 (early proliferation) modules derived from bulk RNA-seq in Fig. 3.1c. Shown are the module score distributions of C21_lung Tex versus T21_lung Tex (left) and C21_spl Tex versus T21_sub Tex (right). The red contour line indicates C21-Tex and the blue contour line indicates T21-Tex. **d**, UMAP of all the Tex cells color-labeled with inferred Tex subset type, including effector-like (*Klrg1*^{hi}*Tcf7*^{lo}), progenitor (*Tcf7*^{hi}*Havcr2*^{lo}), proliferating (*Mki67*^{hi}), transitional (*Tcf7*^{lo}*Havcr2*^{hi}*Cx3cr1*^{hi}) and terminal (*Tcf7*^{lo}*Havcr2*^{hi}*Cd101*^{hi}) Tex subsets, frequencies of which were shown in **e**.

Figure 3. 2, Continued

f, Heatmaps showing normalized and scaled expression of the DEGs between C21-Tex (C21_spl and C21_lung) and T21-Tex (T21_sub and T21_lung) in each Tex subset, including effector-like, progenitor, proliferating and terminal Tex subsets.

T21-Tex cell clusters (Supplementary Figure S3.3c). We further summarized such functional predispositions of all cells onto a ternary map (Fig. 3.2c). The results again supported the notion that C21-Tex cells (red in Fig. 3.2c) were relatively effector poised (AM1) while T21-Tex cells (blue in Fig. 3.2c) were more memory like (AM4) with higher proliferative potential (AM6), though both were dysfunctional relative to canonical effector and memory $CD8^+$ T cells.

Next, based on differential gene expression patterns[85,199,210-212], all Tex cells were further classified into five phenotypically distinct subsets, including effector-like (*Klrg1*^{hi}*Tcf7*^{lo}), progenitor (*Tcf7*^{hi}*Havcr2*^{lo}), proliferating (*Mki67*^{hi}), transitional (*Havcr2*^{hi}*Cx3cr1*^{hi}) and terminal (*Havcr2*^{hi}*Cx3cr1*^{lo}*Cd101*^{hi}) subsets (Fig. 3.2d and Supplementary Figure S3.3d). While the transitional and proliferating subsets were highly enriched in C21-Tex cells, the progenitor subset was most significantly enriched in T21-Tex cells (Fig. 3.2e). Next, we conducted DEG analysis for each cell subset and identified 788, 733, 630 and 454 DEGs in effector-like, progenitor, proliferating and terminal Tex cells, respectively (Fig. 3.2f). These DEGs are involved in regulating a wide range of biological processes (Supplementary Figure S3.4). More notably, the top ranking DEGs upregulated in each C21-Tex cell subset include a substantial number of effector-associated genes, such as *Fos, Fosb, Cx3cr1, Jun, Ccl5, Zeb2, Tbx21* in effector-like subset, *Ccl5, Batf, Fos, Fosb, Gzmk, Tbx21* in progenitor subset, *Fos, Fosb, Prf1, Cx3cr1, Ccl5,* in proliferating subset and *Ccl5, Cx3cr1, Batf, Fos, Fosb, Gzmb, Prf1, Tbx21* in terminal subset; conversely, memory-associated signature genes were more enriched in DEGs upregulated in each individual subset of T21-Tex cells, e.g., *Vps37b, Cd28* and *Foxo1* in effector-like subset, *Vps37b, Foxp1, Il7r* and *Ccr7* in progenitor subset, *Tcf7, Ccr7* and *Xcl1* in proliferating subset and *Tcf7, Il7r* and *Foxp1* in terminal subset. Furthermore, Tox, recently identified as a transcription factor

suppressing the transcription of memory-associated gene signatures[81-84], such as *Ccr7, Tcf7* and *Il7r*, had higher expression in each Tex-cell subset in chronic viral infection than in tumor (Fig. 3.2f). These results together suggested that while the Tex cells had different compositions of each functional subset in each disease, the disease-specific effector- or memory-like transcriptional features were intrinsic to chronic viral infection and tumor, respectively.

To further elucidate whether disease-specific effector- or memory-poised transcriptional signatures of Tex cells were epigenetically imprinted, we assessed chromatin accessible states in C21- and T21-Tex cells by single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq). As shown by the UMAP (Fig. 3.3a), each disease-specific Tex population occupied a different region without apparent overlaps, revealing disease-dependent epigenetic regulation of T cell exhaustion programs. Annotation of the scATAC-seq data using label transfer from the corresponding scRNA-seq data in Fig. 3.2d revealed disease-specific epigenetic profiles of Tex cells (Fig. 3.3b and Supplementary Figure S3.5). Further analysis of chromatin-accessible regions identified 489 differentially accessible regions (DARs), in which many of the corresponding genes coincided with aforementioned DEGs in different cell subsets by scRNA-seq analysis (Fig. 3.2f). For example, the loci of effector-associated genes *Tox, Rap1gap2, Tnfaip3, Nfatc1, Gzmk and Irf2* were more accessible in C21-Tex cells while the gene loci of memory-related signatures *Tcf7, Ccr7* and *Foxp1* were more accessible in T21-Tex cells. Specifically, the *Tox* gene loci was also more accessible in C21-Tex cells than in T21-Tex cells (Fig. 3.3c). *Tox* loci downstream of transcription start site (TSS) (e.g., +100 to +114 kb, +119 to +140 kb and +147 to +166 kb) were more accessible in C21 Tex cells, while *Ccr7* loci upstream (e.g., -14 to -11 kb and -9 to -6 kb) and downstream (e.g. +6 to +14 kb, +17 to +20 kb, +21 to

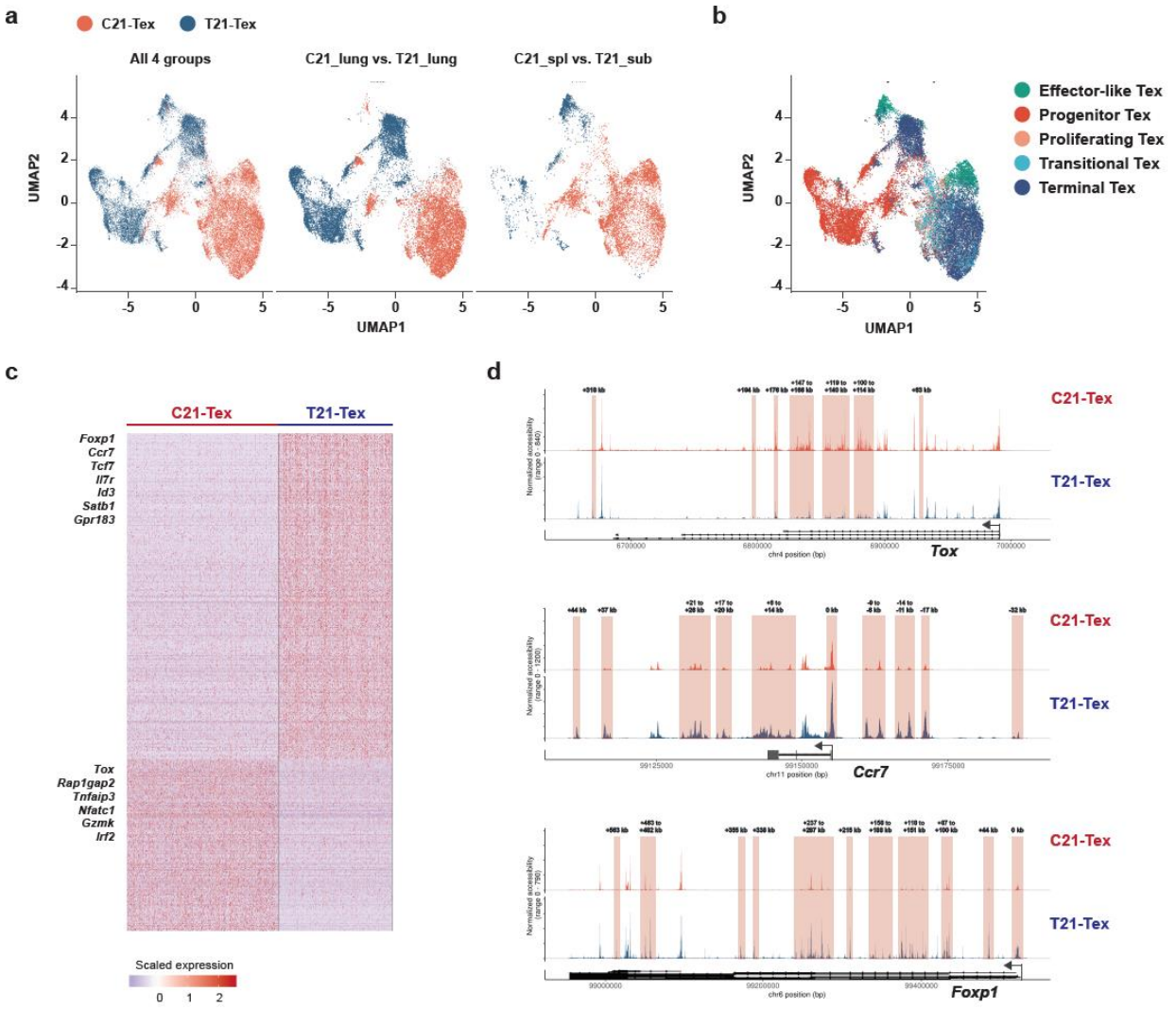


Figure 3. 3 Single-cell chromatin accessibility of disease-specific exhausted T cells.

a, UMAP of scATAC-seq profiles of C21_spl, C21_lung, T21_sub and T21_lung Tex cells. Clusters are color-coded with disease type (Red represents C21-Tex, and blue represents T21-Tex). Left: all 4 groups. Middle: C21_lung Tex versus T21_lung Tex. Right: C21_spl Tex versus T21_sub Tex. **b**, UMAP of scATAC-seq profiles of all the Tex cells denoted by color labeled with inferred Tex subset type, including effector-like (*Klrg1*^{hi}*Tcf7*^{lo}), progenitor (*Tcf7*^{hi}*Havcr2*^{lo}), proliferating (*Mki67*^{hi}), transitional (*Tcf7*^{lo}*Havcr2*^{hi}*Cx3cr1*^{hi}) and terminal (*Tcf7*^{lo}*Havcr2*^{hi}*Cd101*^{hi}) Tex subsets. **c**, Heatmap showing the 489 differentially accessible regions (DARs) between C21 (spl + lung) and T21 (sub + lung) Tex cells. **d**, scATAC-seq signal profiles of *Tox*, *Ccr7* and *Foxp1* loci. Differential accessible peaks between C21-Tex and T21-Tex were highlighted in red.

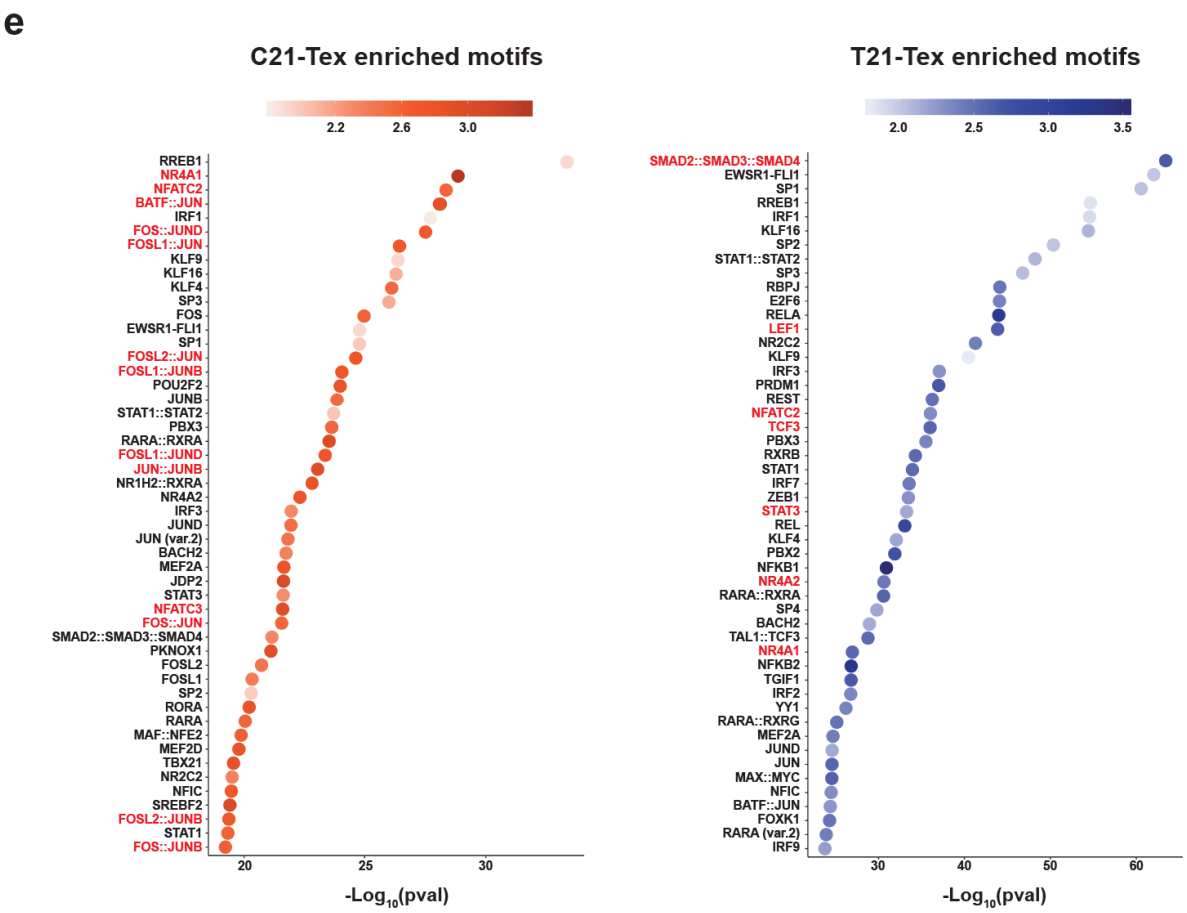


Figure 3.3, Continued

e, Top 50 enriched transcriptional factor (TF) motifs in more accessible DARs in C21-Tex or T21-Tex.

+26 kb, +37 kb and +44 kb) of TSS and several *Foxp1* loci downstream of TSS (e.g., +44 kb, +87 to +100 kb, +118 to 151 kb) were more accessible in T21-Tex cells (Fig. 3.3d). Furthermore, according to DAR profiles, we conducted transcription factor (TF) motif enrichment analysis and identified the top 50 TF motifs in each disease condition (Fig. 3.3e). The motifs for key exhaustion regulatory TFs, like NFAT[213], NR4A[214,215], were comparably enriched in both C21- and T21-Tex cells. Notably, a wide variety of AP-1 family heterodimers (e.g., BATF::JUN, FOS::JUND, FOSL1::JUN, FOSL2::JUN, FOSL1::JUNB, JUN:: JUNB, FOS:: JUN, FOSL2: :JUNB and FOS:: JUNB) associated with effector T-cell differentiation were most significant in DARs of C21-Tex cells. In contrast, T21-Tex DARs more enriched TFs regulating memory-cell differentiation, including SMAD2::SMAD3::SMAD4, LEF1, STAT3 and TCF3 (Fig. 3.3e), which corroborated the distinct transcriptional characteristics between C21-Tex cells and T21-Tex cells. Collectively, these results indicated the involvement of epigenetic layers in the regulation of disease-specific transcriptional features of Tex cells developed in chronic viral infection and tumor.

Differential Tex cell responsiveness to PD-L1 ICB between chronic viral infection and tumor.

PD-1/PD-L1 ICB can functionally reinvigorate exhausted $CD8^+$ T cells[79,86,216]; however, it is unclear whether Tex cell responses to PD-1 ICB are disease-specific. To this end, we performed scRNA-seq and scATAC-seq after the anti-PD-L1 antibody treatment and then integrated transcriptional and epigenetic profiles (Supplementary Figure S3.6a, e and Supplementary Figure S3.7a). As expected, PD-L1 ICB was effective in both chronic viral infection and tumor conditions, evidenced by decreased viral loads and blunted tumor progression (Supplementary Figure S3.6). Consistent with published studies[217], both transcriptional and epigenetic landscapes

of C21-Tex cells were only marginally altered following PD-L1 ICB at the single-cell level, compared to its isotype control (Fig. 3.4a, Supplementary Figure S3.7a, b and Supplementary Figure S3.8a); however, compared to its isotype control, T21-Tex cells showed moderately changes in both epigenetic and transcriptional profiles upon ICB (Fig. 3.4a, Supplementary Figure S3.7a, b and Supplementary Figure S3.8a).

To identify PD-L1 ICB-responding Tex cell subsets in both disease settings, we applied Louvain clustering to assign Tex profiles into 24 distinct clusters (Fig. 3.4b, c and Supplementary Figure S3.7c, d). As anticipated, compared to the isotype control, little changes in relative cell frequency post-PD-L1 blockade were observed among C21-Tex dominant clusters, except small increases in the frequencies in cluster 1 and cluster 19. Notably, cluster 4 and cluster 20 of T21-Tex cells were significantly expanded upon PD-L1 blockade (Fig. 3.4d and Supplementary Figure S3.7e). To gain further insights into these PD-L1 blockade-responding clusters, we probed genes that are differentially expressed in those clusters. C21-Tex derived cluster 1 and cluster 19 exhibited higher expression of *Pdcd1*, *Lag3*, *Ctla4* and *Tox*, reflecting a minimal reprogramming of exhaustion by PD-L1 ICB; furthermore, effector genes (e.g., *Cx3cr1*, *Gzmb*, *Prf1* and *Ccl5*) were most abundant in cluster 1, and several genes associated with cell proliferation (e.g., *Mki67*, *Stmn1* and *Ube2c*) were more pronounced in cluster 19 (Fig. 3.4e). Given that the relatively smaller size of cluster 19 cluster, these results demonstrated that effector-like cluster 1 primarily responded to PD-L1 ICB by reengaging limited effector-associated molecules in chronic viral infection. In contrast, T21-Tex-dominant cluster 20 were characterized by a mixture of genes downstream of the T cell activation (e.g., *Hspa1a, Hspa1b, Jun, Junb, Jund, Fos, Fosb* and *Cd69*) and memory-associated genes (*Id3, Il7r, Tcf7, Ccr7, Sell, Lef1* and *Xcl1*) after PD-L1 ICB (Fig. 3.4e). Notably,

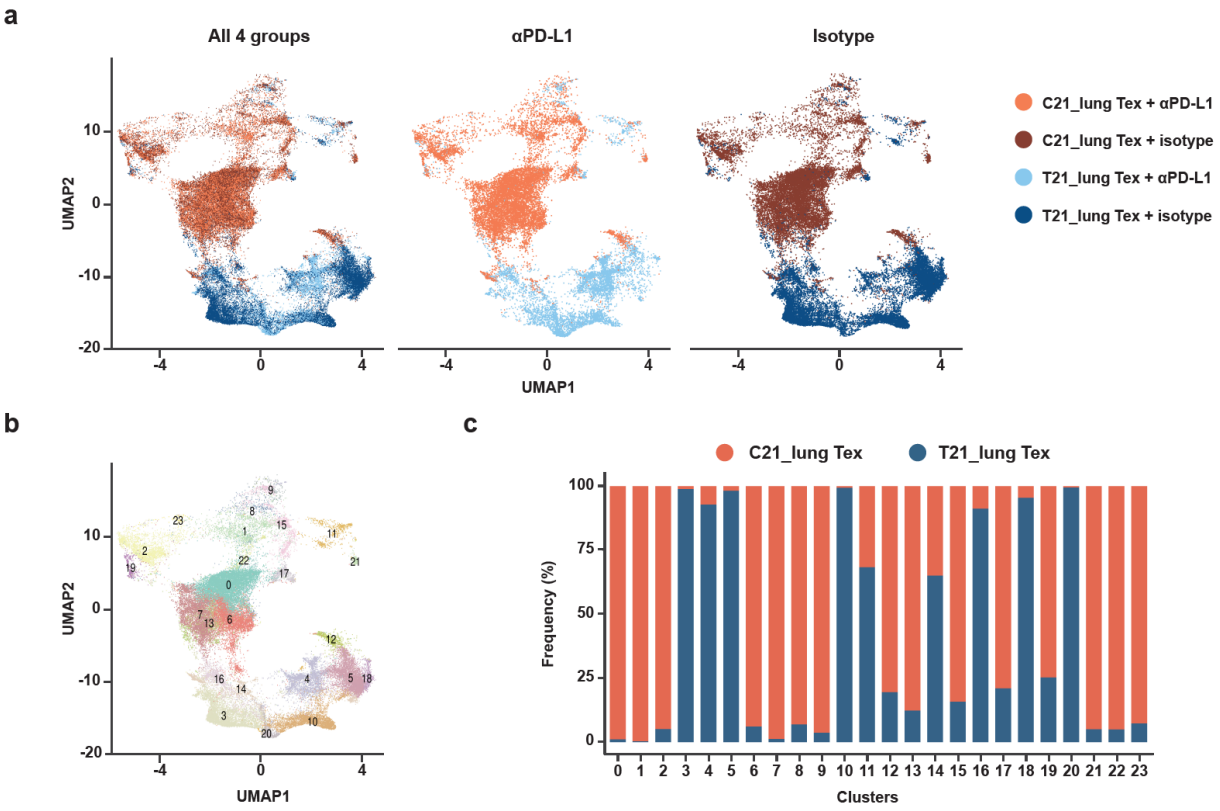


Figure 3. 4 Differential Tex cell responsiveness to PD-L1 ICB between chronic viral infection and tumor.

a, UMAP of lung-resident C21-Tex and T21-Tex with αPD-L1 treatment or isotype treatment. αPD-L1- or isotype- treated C21-Tex cells are denoted as light red and dark red, respectively. αPD-L1- or isotype-treated T21-Tex cells are denoted as light blue and dark blue, respectively. b, UMAP of lung-resident C21-Tex and T21-Tex with αPD-L1 treatment or isotype treatment. Shown are color-coded 24 clusters determined by Louvain clustering. c, Frequencies of C21-Tex and T21-Tex (both αPD-L1- and isotype-treated) in each cluster from Fig. 3.4b.

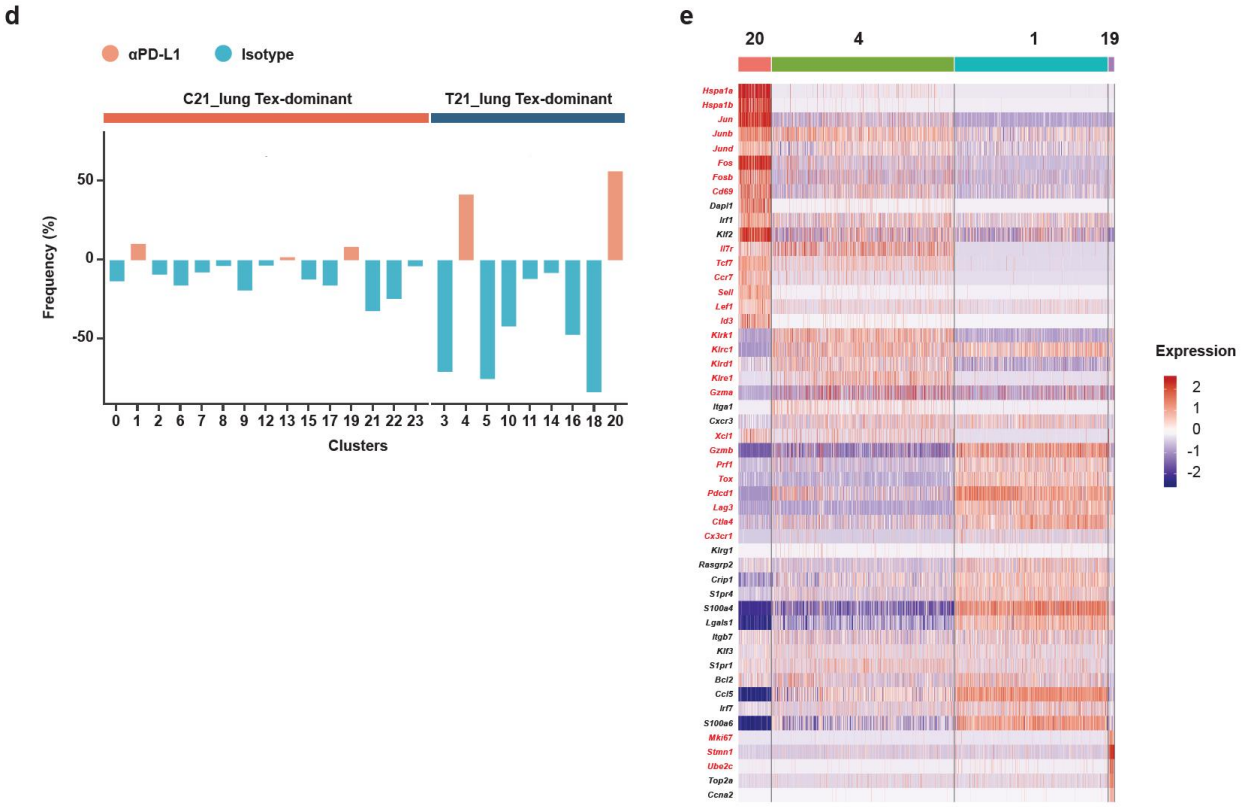


Figure 3. 4, Continued

d, Frequency change of each cluster in Fig. 3.4b upon PD-L1 blockade treatment. The frequency change value of each cluster is calculated as: % of αPD-L1 treatment in one cluster 9 % of isotype treatment in one cluster. **e**, Heatmap showing normalized and scaled expression of representative functional genes of clusters 1, 4, 19 and 20.

cluster 20 had low expression of killing genes *Klrk1, Klrc1, Klrd1* and cytolytic molecules *Gzma, Gzmb and Prf1*. These data suggested that PD-L1 ICB rewired the early activation program in memory-like cluster 20 of Tex cells in tumor. Given that PD-L1 blockade may promote the replenishment of tumor-infiltrating Tex cells from the peripheral less/non-exhausted T cells[218,219], we speculated that cells of cluster 20 with memory characteristics were newly recruited tumor-specific $CD8^+$ T cells from the peripheral. Indeed, several top genes enriched in cluster 20, including *Hspa1a, Hspa1b, Jun, Junb, Fos, Cd69* and *Il7r,* were also upregulated in the replenished human tumor-specific cells after PD-1 ICB therapy in patients with basal cell carcinoma. Similarly, cluster 4 also showed mixed gene signatures related to activation, naïve/memory, effector and exhaustion, but with downregulated memory-associated molecules (*Id3, Il7r, Tcf7, Ccr7, Sell, Lef1* and *Xcl1*) and upregulated killing and cytolytic gene signatures (*Klrk1, Klrc1, Klrd1, Gzma, Gzmb and Prf1*) as well as exhaustion signatures (*Pdcd1* and *Ctla4*). In this regard, cluster 4 seemed to be an intermediate state between newly replenished memory-reminiscent Tex cells and dysfunctional Tex cells, indicating that PD-L1 ICB drove the further differentiation of memory-like but early-activated cluster 4 into cluster 20 under tumor condition. Together, these data illustrated the effector- or memory-like characteristics of clusters of Tex cells that dictated their differential responsiveness to PD-L1 ICB between chronic viral infection and tumor.

Disease-specific regulons of exhausted T cells following PD-L1 ICB.

Upon PD-L1 ICB, Tex cells in both chronic viral infection and tumor showed 219 DEGs in total, indicating limited transcriptomic changes; furthermore, only 27 out of these DEGs were

shared (Supplementary Figure S3.8 a, b), suggesting distinct molecular mechanisms underlying differential Tex cell responses to PD-L1 ICB between these two diseases. Moreover, among top ranking PD-L1 ICB induced DEGs in each disease, many genes showed similar expression patterns as C21-Tex PD-L1 ICB responding clusters 1/19 (e.g., *Hspa1a, Hspa1b, Cd69, Jun, Fos, Fosb, Klrc1* and *Id3*) and T21-Tex PD-L1 ICB responding clusters 4/20 (e.g., *Gzma, Gzmb* and *Ctla4*), respectively (Supplementary Figure S3.8c, d). To elucidate the potential molecular mechanisms, we conducted in-depth integrative analysis of single-cell transcriptomic and epigenetic profiles by constructing regulons specific to lung-resident isotype antibody-treated and αPD-L1-treated Tex cells in each disease (Supplementary Figure S3.9a). Each regulon consists of a sample-specific differentially upregulated central TF and its putative target genes whose expressions correlate with that of the TF and whose loci contain DARs with that TF's consensus DNA binding motif. Regulons derived from each condition of each disease are then connected via co-targeted genes to generate a regulon network for that sample. Hence, the resultant regulon network from each isotype sample highlights the transcriptional network that was active before the PD-L1 ICB but was suppressed post-treatment in each disease context, whereas the regulon networks derived from the αPD-L1 samples represent the transcriptional networks that are repressed before PD-L1 ICB but are re-engaged/activated post-treatment.

By comparing the pre-treatment regulon networks in T21-Tex cells and C21-Tex cells, we found that PD-L1 ICB suppressed a shared set of TFs, including Fosl2, Ets1, Nfkb1, Ikzf1, and Irf2, in both diseases. These TFs, together with some shared targets in both contexts (e.g., *Pdcd1, Nr4a2, Ccnd2*) (Fig. 3.5a), likely represent the conserved exhaustion subnetwork that is reversed or disengaged by ICB, embodying the aspect of "de-exhaustion". In terms of disease-specific PD-

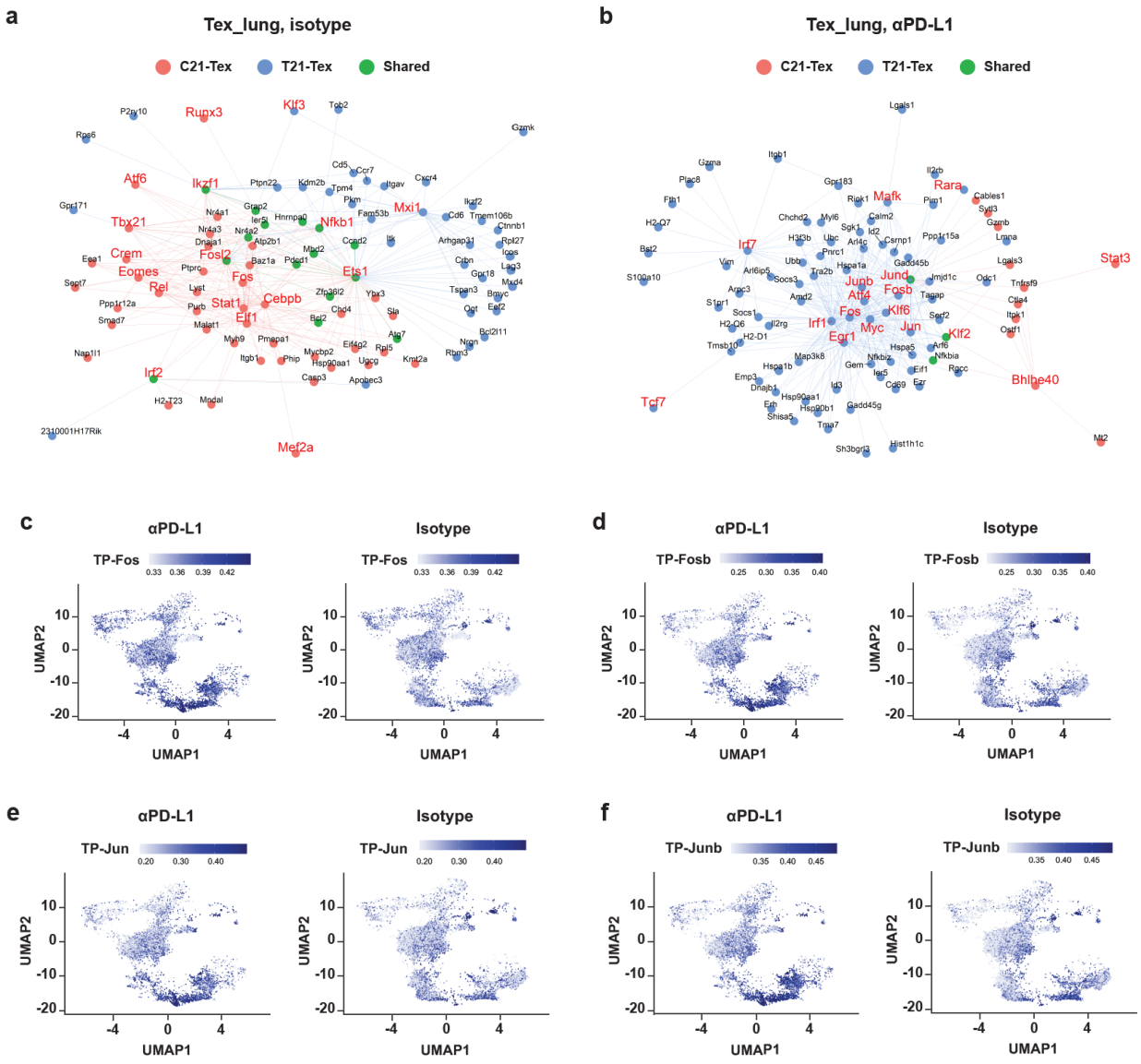


Figure 3. 5 Disease-specific regulons of exhausted T cells following PD-L1 ICB

a, b, the regulon networks of isotype-treated Tex (**a**) and αPD-L1-treated Tex (**b**). Each node represents a central transcription factor (TF, red font) or a target gene (black font), and each edge represents a potential regulation derived from correlation analysis. Node color indicates whether that gene is differentially expressed in C21-Tex (red) or T21-Tex (blue) or both (green). **c-f**, UMAP showing the scores of PD-L1 ICB T21-Tex-derived Fos- (**c**), Fosb- (**d**), Jun- (**e**) and Junb-centred regulons (**f**) in lung-resident Tex receiving αPD-L1 treatment (left) or isotype treatment (right).

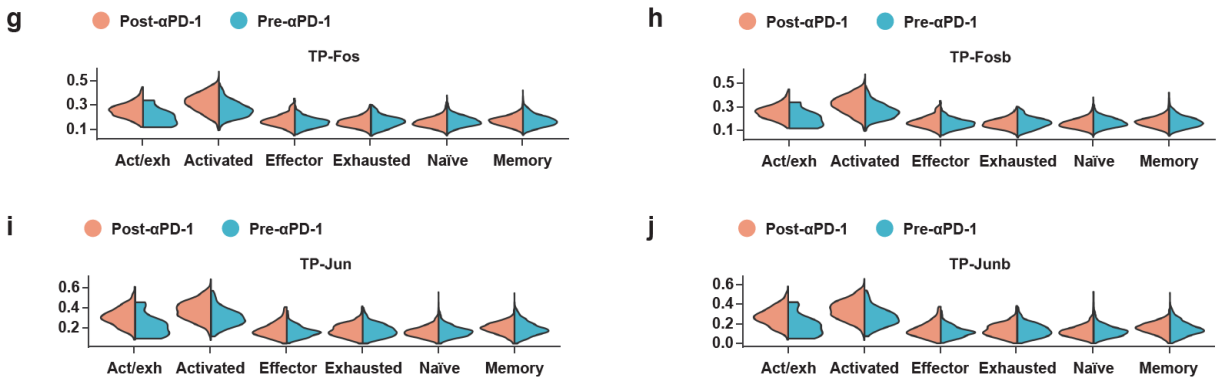


Figure 3. 5, Continued

g-j, Scores of the regulons of αPD- L1-treated lung-resident T21-Tex, including Fos- (**g**), Fosb- (**h**), Jun- (**i**) and Junb-centred regulons (**j**), in human BCC TIL subsets (including activated/exhausted (Act/exh), activated, effector, exhausted, naïve and memory) pre- and post-PD-1 ICB treatment.

L1 ICB suppressed TFs, downregulation of C21-Tex specific TFs (e.g., Tbx21, Eomes, Runx3 and Fos) led to downregulation of their target genes (e.g., *Nr4a1, Nr4a3, Ptprc*), whereas in tumor, the disease-specific suppressed TFs include Klf3 and Mxi1 and the suppressed target genes include *Ccr7, Icos, Lag3* and *Bcl2l11* (comparing Fig. 3.5a with Fig. 3.5b).

Comparison of the post-treatment regulon networks revealed that PD-L1 ICB re-activated a large number of TFs in T21-Tex but only few TFs in C21-Tex (Fig. 3.5b). As with the shared suppressed TFs, there were also two shared TFs among the activated ones, including Klf2 and Jund, and one shared target gene, *Nfkbia*. These shared upregulated TFs further explained how PD-L1 ICB works in general on Tex cells, embodying the "re-invigoration" aspect that is complementary to the "de-exhaustion" aspect. While in T21-Tex cells, additional disease-specific activated TFs included a few AP-1 components (Jun, Junb, Jund, Fosb and Fos), Myc, Mafk, Irf7, Klf6, Tcf7, etc., while the additional TFs in C21-Tex cells only included Bhlhe40 and Stat3. Such differential TF activation, led to an array of differential target gene upregulation between T21-Tex cells (e.g. *Tcf7, Id3, Il2rg, Il2rb, Gpr183, Gzma, Hsp1a, Hspa1b,* and *CD69*) and C21-Tex cells (e.g. *Ctla4* and *Gzmb*), consistent with differential gene expression patterns in PD-L1 ICB responding Tex clusters 4/20 and 1/19 in tumor and chronic viral infection, respectively. In addition, many of the re-activated TFs are known to be antagonist-pair with the suppressed TFs. For instance, in response to PD-L1 ICB, Mxi1 was specifically repressed in T21-Tex cells (Fig. 3.5a) and it is known to be the repressor of c-Myc[220]. As a result, it was specifically activated in T21-Tex cells after PD-L1 ICB (Fig. 3.5b). Such antagonist-pair highlights the inter-connection between the "de-exhaustion" and the "re-invigoration". Taken together, the regulatory networks

demonstrated that there are largely unshared rewirings of the transcriptional sub-networks by the PD-L1 ICB between two diseases.

We next evaluated the specificity of the lung tumor metastasis-derived regulons across all mouse Tex-cell clusters. As expected, we found an exclusive enrichment in clusters 4 and 20 (Fig. 3.4d, Fig. 3.5c-f and Supplementary Figure S3.10a-d), again confirming that these two clusters were PD-L1 ICB responders in metastatic lung tissues. Notably, although the functional regulons of the clusters 4 and 20 were derived from mouse model of lung metastasis, they were also found in mouse model of subcutaneous melanoma, but not mouse model of chronic viral infection (Supplementary Figure S3.10e, f), highlighting the role of these regulons in regulating tumor-specific Tex cell responses to PD-L1 ICB.

Finally, we asked whether Tex cell regulons driven by PD-L1 ICB in murine tumor model can be applied to tumor-infiltrating lymphocytes (TILs) in human cancer upon PD-1 ICB. We test these regulons in TILs of basal cell carcinoma (BCC) patients before and after anti-PD-1 treatment using single-cell sequencing data published in a recent study. The regulon scores were calculated in each individual cell for each regulon that was activated in PD-L1 ICB responding T21-Tex cells. First, we observed higher baseline regulon scores in activated and activated/exhausted clusters (as defined in the original paper) when compared to other counterparts (Fig. 3.5g-j and Supplementary Figure S3.10g-j). The gene signatures of activated and activated/exhausted clusters in human BCC were largely reminiscent of those in cluster 4 and 20 in murine melanoma (Supplementary Figure S3.11a). Upon PD-1 ICB, regulon activities were further elevated and restricted to activated and activated/exhausted TILs (Fig. 3.5g-j and Supplementary Figure S3.10g-j), suggesting that murine-tumor derived Tex-cell regulons function similarly to reinvigorate human TILs in response

to PD-1 blockade. As a control, we further examined TILs of human BCC patients using the regulons enriched in ICB-treated Tex cells in chronic viral infection (Supplementary Figure S3.12a, b). We found that the signature scores of these regulons were very low (Supplementary Figure S3.12e-h), reinforcing disease-specific regulatory network of Tex cells in response to PD-1/PD-L1 ICB between chronic viral infection and tumor.

Discussion

Understanding T cell exhaustion is critical for developing effective immunotherapies to treat chronic viral infection and cancer. As Tex cells shared many common characteristics, many studies do not differentiate T cell exhaustion between these two diseases. However, chronic viral infection and cancer are very different diseases and each creates a unique immune microenvironment to drive T cell exhaustion. To define disease-specific transcriptomic and epigenetic characteristics of T cell exhaustion, we created a unique system to unbiasedly compare Tex cells derived from chronic viral infection and tumor. Different from most studies, these Tex cells have an identical TCR recognizing the same epitope presented in both diseases. Furthermore, we compared lung-resident Tex cells of both diseases. Our system eliminated potential confounding factors including TCR affinity, epitope diversity and tissue specificity, thus enabling the identification of disease-specific T cell exhaustion features before and after ICB treatment.

As expected, we found shared transcriptomic/epigenetic features between two Tex cells that have been described in previous studies by both bulk and single-cell analyses. However, we also identified disease-specific transcriptomic and epigenetic features unique to each disease. In the bulk RNA-seq analysis, we identified two diverged differentiation trajectories, major distinct

transcriptional modules and 1,389 DEGs between C21- and T21-Tex cells. These transcriptional differences marked C21-Tex cells as more effector-like while T21-Tex cells as more memory-characteristic, although both Tex cells were still dysfunctional. Consistent with bulk RNA-seq, scRNA-seq also showed that each individual subset of C21-Tex cells are more effector-like than the corresponding subset of Tex-cells, while T21-Tex cell subsets are overall more memory-like than subsets of C21-Tex cells, which were further confirmed by scATAC-seq. In addition to identifying DARs corresponding to DEGs found in scRNA-seq, scATAC-seq further identified different TF sets potentially involved in regulating disease-specific transcriptomic profiles. In particular, AP-1 heterodimers downstream of TCR activation that promote effector T-cell differentiation were enriched in C21-Tex cells while memory differentiation-related TFs were found in T21-Tex cells. The disease-specific T cell exhaustion programs may also dictate the differential responses of Tex cells to ICB therapy, given that memory-like cluster 20 and effector-like cluster 1 primarily responded to PD-L1 ICB in tumor and chronic viral infection, respectively. Disease-specific regulons further uncovered the "de-exhaustion" and "functional reinvigoration" mechanisms underlying differential responses of Tex cells to PD-L1 ICB in chronic viral infection and tumor.

In summary, we demonstrated disease-specific transcriptomic and epigenetic profiles of T cell exhaustion programs established during chronic viral infection and tumor development, which further dictate the differential responses of disease-specific Tex-cell subsets to PD-1/PD-L1 ICB. These findings provide fundamental insights into developing new immunotherapy by targeting disease-specific regulatory networks of T cell exhaustion for effectively treating chronic viral infection and tumors, respectively.

Material and Methods

Mice, viruses and cell lines.

CD45.1$^+$ P14 mice were a gift from Dr. Rafi Ahmed (Emory University). The C57BL/6J mice were obtained from Jackson Laboratories. The Lymphocytic choriomeningitis virus (LCMV) Armstrong (Arm) and clone 13 (Cl13) strains were from Dr. Rafi Ahmed (Emory University). Mice were infected intraperitoneally with LCMV Arm (2×10^5 PFU) or intravenously with LCMV Cl13 (2×10^6 PFU). The B16F10 cell line was acquired from ATCC. The B16F10 cell line expressing LCMV Cl13 GP was generated by CRISPR/Cas9-mediated insertion of LCMV Cl13 GP gene sequence into the genome of B16F10 cell line. Cells were selected in puromycin. The LCMV Cl13 GP gene knock-in B16F10 cell line (herein referred as B16F10-GP) was conducted by Beijing Biocytogen Co.Ltd, China. In subcutaneous tumor model, mice were subcutaneously implanted with 5×10^5 B16F10 or B16F10-GP cells. In lung metastasis model, mice were intravenously injected with 5×10^5 B16F10-GP cells. 6-10-week mice of both sexes were infected with virus or engrafted with tumor cells without randomization or "blinding". All mouse experiments were performed with the guidelines of the Institutional Animal Care and Use Committees of the Southern Medical University.

Quantitative reverse-transcription PCR.

Cultured B16F10 or B16F10-GP cells were added into TRIzol LS reagent (Life Technologies, 10296). Then, total RNA was extracted and reverse-transcribed using RevertAid Minus First Strand cDNA Synthesis Kit (Thermo Scientific, K1632). The relative expression of gene coding LCMV-Cl13 GP was determined using AceQ qPCR SYBR Green Master Mix

(Vazyme, Q111) on a CFX96 Touch Real-Time System (Bio-Rad). Forward primer (LCMV Cl13 GP gene): 5'-GCAACTGCTGTGTTCCCGAAAC-3'; reverse primer (LCMV Cl13 GP gene): 5'-CATTCACCTGGACTTTGTCAGACTC-3'.

Adoptive transfer.

In acute viral infection, 5×10^3 naïve CD45.1$^+$ P14 cells were adoptively transferred into CD45.2$^+$ recipients, which were infected with LCMV Arm on the next day. For the analysis of spleen P14 cells in chronic viral infection, CD45.2$^+$ recipients were firstly infected with LCMV Cl13. Ten days after infection, recipients were adoptively transferred with 5×10^5 (for day 4 analysis) or 5×10^3 (for day 8 and 21 analysis) naïve CD45.1$^+$ P14 cells. For lung-resident P14 cells in chronic viral infection, recipients were adoptively transferred with 5×10^3 naïve CD45.1$^+$ P14 cells and then infected with LCMV Cl13. In subcutaneous tumor model, recipients were firstly implanted with B16F10-GP, followed by administration of 4 mg CTX (Sigma, PHR1404) and transfer of 2×10^6 (for day 4 analysis) or 5×10^5 (for day 8 and 21 analysis) naïve CD45.1$^+$ P14 cells on day 9 and 10 post-tumor implantation, respectively. In lung metastasis model, recipients were administrated with 4 mg CTX, transferred with 5×10^5 naïve CD45.1$^+$ P14 and then intravenously injected with B16F10-GP cells on 3 consecutive days. For proliferation assay, enriched 5×10^5 naïve CD45.1$^+$ P14 cells were labeled with proliferation-indicating dye (CellTrace CFSE; Thermo Fisher, C34554) by following manufacture's guide and intravenously transferred to CD45.2$^+$ recipients. The recipients were engrafted with B16F10 or B16F10-GP cells 10 days before cell transfer.

Anti-PD-L1 treatment.

On day 11, 14, 17 and 20 after P14 cell transfer, recipients were administrated with 200 μg αPD-L1 (BioXCell, BE0101, clone 10F.9G2) or isotype-matched control antibody (BioXCell, rat IgG2b) via intraperitoneal injection. In subcutaneous tumor model, tumors were measured staring at day 6 after P14 cell transfer every 3 days. Tumor size was determined following the ellipsoid volume formula: length × $width^2$ / 2.

Virus titration.

The LCMV Cl13 viral loads in spleen and lung were determined by quantitative RT-PCR as previously described[221].

Tissue-infiltrating lymphocytes isolation.

To obtain tumor infiltrating lymphocytes (TILs), tumors were dissected from euthanized mice and mechanically minced. Then, TILs were enriched by using Percoll (GE Healthcare, 17-0891-09) density gradient. For lung-resident lymphocytes isolation, euthanized mice were firstly intravenously injected with 2 μg of PE anti-CD8 (Biolegend, 100708). Five minutes later, lungs were dissected, perfused with PBS, mechanically minced and treated with type 1 collagenase (Sigma, C0130) for 30 min at 37ºC. Next, lung-resident lymphocytes were harvested using Percoll (GE Healthcare, 17-0891-09) density gradient and stained with PerCP anti-CD8 (Biolegend, 100732). Lung-resident $CD8^+$ T cells were identified as PerCP anti-CD8-positive, but PE anti-CD8-negative.

Flow cytometry and antibodies.

Surface staining was performed in PBS containing 2% FBS (wt/vol) on ice for 30 min. For intracellular cytokine staining, harvested lymphocytes were stimulated by GP33 peptides (0.2 μg/ml), brefeldin A (BD Biosciences, 555029), anti-CD107a (BD Biosciences, 553793) and anti-CD107b antibodies (BD Biosciences, 558758) for 5 hours at 37^{o}C, followed by surface staining and intracellular cytokine staining with Cytofix/Cytoperm Fixation/permeabilization Kit (BD Biosciences, 554714). The following antibodies were used: anti-CD8α (Biolegend, clone 53-6.7), anti-CD44 (Invitrogen, clone IM7), anti-CD45.1 (Biolegend, clone A20), anti-CD45.2 (Biolegend, clone 104), anti-PD-1 (Biolegend, clone 29F.1A12), anti-Tim-3 (Biolegend, clone RMT3-23), anti-2B4 (Biolegend, clone C1.7), anti-IFN-γ (BD Biosciences, clone XMG1.2), anti-TNF-α (BD Biosciences, clone MP6-XT22), anti-IL-2 (Biolegend, clone JES6-5H4), anti-CD107a (BD Biosciences, clone 1D4B) and anti-CD107b (BD Biosciences, clone ABL-93). Flow cytometry data were collected with a FACSCanto II (BD Biosciences) and analyzed by using FlowJo software (TreeStar). Cell sorting was performed with a FACS Aria III (BD Biosciences).

Bulk RNA-seq library preparation and sequencing.

Transferred CD45.1^{+} P14 CD8^{+} T cells were sorted from spleens, lungs or tumors of recipient mice based on the following gating strategy: live CD8^{+}CD44^{+}CD45.1^{+}CD45.2^{-}Lin^{-}. Total RNA was extracted from the sorted cells with a Micro Total RNA Isolation Kit (Thermo Fisher, AM1931), reverse-transcribed using SMART-Seq v4 Ultra Low Input RNA Kit (Takara, 634891) and library prepared following the protocol of a Nextera XT DNA Library Preparation Kit (illumina, FC-131-1002). Then, the bulk RNA-seq libraries were sequenced on an illumina

NextSeq with 75bp/75bp paired-end runs, with a minimum sequencing depth per sample of 84M reads and a maximum of 120M reads.

Bulk RNA-seq analysis.

Raw sequencing reads were first trimmed of adapters to improve the quality using Trim Galore! (version 0.4.4)[222]. Paired-end reads that passed quality control (QC; Phred score >20) were then aligned to mm10 reference genome using HISAT2 (version 2.1.0)[223] with the "--dta" flag. The resulting SAM files were converted to BAM files using SAMtools (version 1.9)[224]. Following that, the aligned reads were then assembled to transcripts using StringTie (version 1.3.5)[225] and the gene expressions were quantified in R using the *featureCounts* function in Rsubread package (version 1.34.7.)[226].

Sources for reference sequencing data.

Reference bulk RNA-seq data of P14 T cells from LCMV Arm infected mice (Day 4.5) were obtained from Gene Expression Omnibus (GEO) (GSE119943)[227]. Reference single-cell RNA-seq data of $CD8^+$ T cells from basal cell carcinoma patients before and after αPD-1 treatment were obtained from GEO (GSE123814)[228].

Differential gene expression (DGE) analysis on bulk RNA-seq data.

Raw count matrix was built from the featureCounts results and was passed into DESeq2 (version 1.24.0)[229] to conduct differential gene expression analysis with the default Benjamini-

Hochberg correction to the pvalues. A default cut-off of abs (log2(Fold Change)) > 1.5 and padj < 0.05 is used to filter the differentially expressed genes unless otherwise stated.

Weighted gene coexpression network (WGCNA) analysis.

The WGCNA were conducted in R (version 3.6.1) using the WGCNA package (version 1.68)[230]. To ensure all groups have data at the corresponding time points, we first combined the RNA-seq data of P14 T cells from GSE119943 with our data. We then normalized the data using the variance-stabilized transformation in DESeq2 (version 1.24.0) during which we explicitly added the batch (i.e. reference data and our data as 2 batches) into the model design. Following that, we also removed the batch effect using the *removeBatchEffect* function in limma (version 3.40.6)[231]. Eventually, we only retained the D4.5 Armstrong RNA-seq data from the reference datasets and for convenience, we labeled it as D4 Armstrong, with the assumption that the transcriptional profile does not change significantly in 0.5 day.

The resultant expression matrix was split into 3 matrices, corresponding to T cells in LCMV Arm, LCMV Cl13 and B16F10-GP respectively. For each matrix, we first filtered the genes and retained only the top 10k genes ranking by mean absolute deviation. We then followed the workflow as described in the WGCNA package tutorials. Briefly, we first calculated the adjacency matrix using the "signed" network option and the "bicor" correlation function to be more robust, with the appropriate power beta determined interatively using the *pickSoftThreshold* function. The adjacency matrix was then converted first to a weight matrix by 1-adjacency and then to a distance matrix using *as.dist* function. Hierarchical clustering was then used to partition the network using the *hclust* function from stats (version 3.6.1) package with the "average"

method. The threshold used to cut the dendrogram and obtain the corresponding number of modules was determined iteratively by checking whether downstream annotation (e.g. using Gene Ontology Enrichment) of each module is sufficiently module-specific, mutually independent, and biologically relevant. Once the modules are determined, the eigengenes were calculated through the *moduleEigengenes* function.

Annotation of modules.

For each module in each of the 3 conditions (LCMV Arm; LCMV Cl13; B16F10-GP), we performed gene set enrichment analyses using the clusterProfiler package (version 3.12.0)[232] and the msigdbr package (version 7.0.1)[233,234]. The results were then summarized and plotted as a heatmap using the ComplexHeatmap package (version 2.0.0)[235]. All enriched terms were filtered by imposing FDR adjusted *p*-value to be less than 0.05.

Module overlap analysis.

For any pair of the 3 conditions (LCMV Arm, LCMV Cl13, B16F10-GP), every module from condition 1 was overlapped with every module from condition 2. The number of genes overlapped in each pair of modules was put as entries of the module overlap matrix. The significance of each overlap was determined using a Fisher's exact test with FDR adjustment of the p-values.

The 3 overlap matrices were then used to construct an undirected graph of modules. Each significant (padj<0.05) overlap is represented as an edge, and the corresponding two modules are represented as the connected nodes. The graph was then converted to the corresponding adjacency

matrix and partitioned using the leiden algorithm from the leiden package (version 0.3.1)[236] with the default resolution (=1).

Single-cell RNA-seq library preparation and sequencing.

The single-cell RNA-seq (scRNA-seq) libraries were prepared using Chromium Next GEM Single Cell 3' reagent Kits (10x Genomics) according to the manufacturer's instructions. Briefly, FACS-sorted $CD8^+$ T cells were suspended in DPBS containing 0.04% BSA and captured in droplets using a Chromium Single Cell Controller instrument (10x Genomics). After reverse transcription and cell barcoding in droplets, cDNA was purified and PCR amplified, followed by fragmentation, end-repair, double-size selection and PCR amplification with sample-indexing primers. Then, the scRNA-seq libraries were sequenced with paired-end sequencing of 150nt each end on one lane of NovaSeq 6000 per sample.

Single-cell ATAC-seq library preparation and sequencing.

The single-cell ATAC-seq (scATAC-seq) libraries were prepared following the guide of Chromium Next GEM Single Cell ATAC Reagent Kits (10X genomics), which was also described by Yost K.E. *et al.*[237]. In brief, optimal amounts of resuspended $CD8^+$ T cell nuclei were treated with ATAC Buffer and ATAC Enzyme. Then, the mixture was further combined with Reducing Agent B and Barcoding Enzyme and loaded on a Chromium Chip E. Resulting single-cell Gel Beads-in-emulsion (GEMs) were further linear amplified, purified and sample-indexing PCR amplified. Sequencing libraries were loaded on an Illumina NovaSeq using PE50 sequencing.

Single-cell RNA-seq data processing.

The scRNA-seq reads were aligned to the mm10 reference genome and quantified using the cellranger count (10X Genomics, version 3.0.0). The filtered gene-barcode matrices that contained only barcodes with unique molecular identifier (UMI) counts that passed the quality control were used for downstream analyses. We obtained a total of 48,937 cells from 8 samples with on average 84,288 reads per cell and a median of 2,417 genes per cell.

Single-cell ATAC-seq data processing.

The scATAC-seq reads were aligned to the mm10 reference genome and quantified using the cellranger-atac count (10x Genomics, version 1.1.0). The filtered feature-barcode matrices that contained only barcodes with unique molecular identifier (UMI) counts that passed the quality control were used for downstream analyses. All matrices from 8 samples were then aggregated without normalization using cellranger-atac aggr so that all samples share the same universe of peaks. We obtained a total of 57,172 cells from 8 samples with on average 10,394 fragments per cell, 79.4% fraction of fragments overlapping any targeted region, and 69.6% fraction of transposition events in peaks in cell barcodes.

UMAP analysis and clustering on the isotype single-cell RNA-seq data.

UMAP analysis and clustering were performed using the Seurat package (Version 3.1.2)[238]. Raw count matrices from 8 samples were first converted to Seurat object before further merged into one Seurat object. Following that, cells with less than 1000 genes detected or greater than 25% mitochondrial RNA content were excluded from further analysis, with 43,043 retained after

filtering. While the number of cells removed seems to be large, it is not unexpected given the harsh environment from which the exhausted T cells were isolated and the consequent poor states of a lot of them. Stringent filtering ensures the reliability of our downstream analysis.

For UMAP and clustering analysis, the isotype data were subset from the filtered master Seurat object. Then, the raw count was log-normalized using the *NormalizeData* function with default options. Top 10,000 variable features were then identified using the *FindVariableFeatures* function with the default "vst" method. The data were then centered and scaled using the *ScaleData* function, with additional regression against the percent of mitochondrial RNA content. Scaled data were then used as input for principal component analysis (PCA) on the basis of variable genes using the *RunPCA* function. To remove the batch effect between T cells taken from lung (batch1) and those from the spleen and subcutaneous tumor (batch2), we then used the harmony package (version 1.0) to integrate the data, grouped by the "batch" label and based on the top 100 principal components. UMAP was then constructed based on the first 50 harmony components, with "n.neighbors=20" and "min.dist=0.1". The same harmony components were then used to construct the shared nearest neighbor (SNN) graph using the *FindNeighbors* function, which was then partitioned to identify clusters using the *FindClusters* function with default Louvain algorithm and "resolution =2.5". These finely resolved clusters were then aggregated into larger functional clusters based on current understanding of markers of different exhausted T cell subtypes[221,239-241]. Note that a few "Tumor cell clusters" that lie far away from the T cell clusters, contain very few cells (normally <50), and do not express $CD8^+$ T cell markers (e.g. *Cd3*, *Cd8a*, *Cd8b*) were removed from the visualization and further analyses.

Differential gene expression (DEG) analysis on single-cell RNA-seq data.

DEG analyses were all performed using the *FindMarkers* function in Seurat (Version 3.1.2) package, with default parameters and the appropriate 'ident.1" and "ident.2" set as contrast. Unless otherwise stated, the results were then filtered with p_val_adj<0.05 and abs(avg_logFC)>0.25.

Gene expression signature scoring.

Individual cells were scored using the AUCell package (version 1.6.1)[242] for a particular gene set derived either from bulk RNA-seq module or from a single-cell-derived regulon as follows. The normalized gene expression was first used as input into the *AUCell_buildRankings* function to score each cell for gene set enrichment and build a ranking. The signature score was then calculated as an AUC score using the *AUCell_calcAUC* function with all default parameters. In the case when a gene set and the expression matrix are derived from different species, the gene set is converted to the corresponding homologous genes with NA values removed. Ternary maps based on 3 bulk modules were plotted using the ggtern package (version 3.1.0)[243] and the ggplot2 package (version 3.2.1)[244].

UMAP analysis and label transfer on isotype single-cell ATAC-seq data.

UMAP analysis on scATAC-seq data was performed using the Signac package (Version 0.1.4)[245]. The aggregated filtered feature matrix was first loaded using the *Read10X_h5* function. It was then used to construct the Seurat object (as Signac shares the same object class), with the appropriate metadata and "min.cells=1" option to remove peaks inaccessible in any cell. Following that, the filepath to the corresponding fragment file was added into the Seurat object using the

SetFragments function. The data were then filtered with the following cutoff: 1) peak_region_fragments > 1000 i.e. fragments in peak regions in each cell greater than 1000; 2) peak_region_fragments < 100,000; 3) pct_reads_in_peaks >0.5 i.e. more than half of the reads in that cell are in the peak regions; 4) nulecosomal_signal<10 i.e. approximate ratio of mononucleosomal to nucleosome-free fragments less 10.

For UMAP and clustering analysis, the isotype data were subset from the filtered master Seurat object. Then, the raw count was normalized using the *RunTFIDF* function with default options i.e. through frequency-inverse document frequency (TF-IDF) normalization. The variable features were then set to be all peaks using the *FindTopFeatures* function with "min.cutoff='q0' " option. Singular value decomposition (SVD) was then performed on the TD-IDF normalized data with "n=100". To remove the batch effect between T cells taken from lung (batch1) and those from the spleen and subcutaneous tumor (batch2), we then used the harmony package (version 1.0) to integrate the data, grouped by the "batch" label and based on the top 100 latent semantic indexing (LSI) components. UMAP was then constructed based on the first 50 harmony components, with "n.neighbors=15" and "min.dist=0.05". Note that a few "Tumor cell clusters" that lie far away from the T cell clusuters, contain very few cells (normally <50), and have few accessible regions in $CD8^+$ T cell marker genes (e.g. *Cd3*, *Cd8a*, *Cd8b*) were removed from the visualization and further analyses.

The functional cluster labels from isotype scRNA-seq data were then transferred to the scATAC-seq data as follows. A gene activity matrix was first created using the *FeatureMatirx* function, by summing the reads intersecting the gene body and promoter region (with addition of 2000 bp upstream and 500 bp downstream). The gene activity matrix was then added to the Seurat

object and log-normalized using the *NormalizeData* function with default options. To transfer labels, the transfer anchors were first calculated using the *FindTransferAnchors* function with "reduction='cca'" (canonical correlation analysis) and "dims=1:50" (calculate top 50 components). The functional labels were then transferred using the *TransferData* function with the previous LSI reduction as the "weight.reduction". Further details on this integration and label transfer method can be found in Signac documentation and the corresponding publication[238].

Differential accessibility analysis and signal tracks on single-cell ATAC-seq data.

Differentially accessible peaks were calculated using the FindMarkers function in Signac package (version 0.1.4), with the appropriate 'ident.1" and "ident.2" set as contrast and additional parameters of "min.pct=0.15, test.use='LR', latent.vars = 'peak_region_fragments'".Unless otherwise stated, the results were then filtered with p_val_adj<0.05 and abs(avg_logFC)>0.2. Signal tracks of differentially accessible regions were then plotted using the CoveragePlot function, with a few plot style modifications.

Motif enrichment analysis on single-cell ATAC-seq data.

To conduct motif enrichment in differentially accessible regions, we first extracted transcription factors' (TF's) position frequency matrices (PFM) from the JASPAR2020 package (version 0.99.8)[246] by specifying "species=9606". We note here that we decide to use the human collection of PFMs from the JASPAR data base for two reasons: 1) DNA binding profiles and sequence specificity of orthologous human and mouse TFs are almost identical, making most of the PFM interchangeable; 2) The corresponding mouse PFM list (if set "species=10090") is

significantly smaller and cover much fewer transcription factors. In the rare case where a human TF does not have a homologous gene in mouse, we remove it where applicable.

With the PFM list, we created a motif matrix using the *CreateMotifMatrix* function by scanning all peaks for the presence of each motif in the list. The matrix was then added into the scATAC-seq Seurat object as a matrix object using the *AddMotifObject* function. Following this, additional statistics such as the GC content in each peak were calculated using the *RegionStats* function. Finally, the motif enrichment was calculated in the selected set of differentially accessible peaks using the *FindMotifs* function with default parameters. The p-values were then FDR adjusted, and the results were filtered with a cutoff of padj<0.05.

Integration of single-cell RNA and single-cell ATAC-seq data.

To investigate the similarities and differences between αPD-L1 and the isotype data, we integrated the scRNA-seq and the corresponding scATAC-seq data as follows. The master scRNA-seq Seurat object containing all 8 samples and the master scATAC-seq Seurat object containing the corresponding 8 samples were first loaded. Transfer anchors between the scRNA-seq data and the scATAC-seq data were calculated using the FindTransferAnchors function in Seurat package (version 3.1.2) with "reduction='cca'" (canonical correlation analysis) and "dims=1:50" (calculate top 50 components). An scRNA expression matrix was then imputed from the scATAC-seq data using the transfer anchors and the TransferData function, with LSI reduction as the "weight.reduction". The imputed scRNA expression matrix was then added into the scATAC-seq Seurat object as the "RNA" assay. With the imputed RNA assay, the scATAC-seq object were then merged with the scRNA-seq object, resulting in a master integrated Seurat object.

Data in the integrated Seurat object was then re-scaled using the *ScaleData* function, and the scaled data was used to conduct PCA using the *RunPCA* function. The top 100 PCs then served as input into the RunHarmony function from the harmony package (version 1.0), to integrate across both the modality (scRNA + scATAC) and the batch (batch1-Lung Datasets + batch2-Spleen&Subcatenous Tumor Datasets). The top 50 harmony components were then used to construct UMAP, with "n.neighbors=20" and "min.dist=0.1". The same harmony components were then used to construct the shared nearest neighbor (SNN) graph using the *FindNeighbors* function, which was then partitioned to identify clusters using the *FindClusters* function with default Louvain algorithm and "resolution =0.8".

DGE analysis on integrated data.

Differential gene expression analyses were performed first by subsetting the scRNA data from the master integrated data. Then we used the *FindMarkers* function in Seurat (Version 3.1.2) package, with default parameters and the appropriate 'ident.1" and "ident.2" set as contrast. Unless otherwise stated, the results were then filtered with p_val_adj<0.05 and abs(avg_logFC)>0.25.

Sample-specific regulon construction on single-cell RNA-seq data.

To construct regulons, we used only our Lung datasets, consisting of $CD8^+$ T cells taken from the lung. For each of the 4 samples (CV-Tex-αPD-L1; T-Tex-αPD-L1; CV-Tex-isotype; T-Tex-isotype), we conduct an independent analysis to construct regulons functional and specific to that sample. We did that by supplying the sample raw count matrix as input into the SCENIC package and running through the workflow as detailed in its documentation and publication[242].

Each resultant regulon is a gene list with a central TF and all its putative target genes determined through correlation in expression.

To filter the regulons, we filtered the TF and the putative targets separately as follows. Taking isotype T-Tex sample as example, we first obtain the list of differentially upregulated genes specific in T-Tex-isotype. Then, we perform motif enrichment analysis in all peaks that are within the genomic regions of those genes (does not need to be differentially accessible). The significantly enriched motifs were then further filtered to select the specifically upregulated TFs. This final list serves as a putatively functional TF pool, against which we filtered our regulons.

To further filter the targets and refine each regulon, we impose the following two criteria: 1) the target genes have to be specifically upregulated in that sample; 2) the target genes have accessible peaks (does not need to be differentially accessible; but number of reads > 0.05*number of cells) in its loci that contain the regulon's central TF's motif.

All differential expression analyses performed during the regulon construction used a cutoff of log2(FC)>0.2. The final filtered and refined regulons in each sample were then combined into a graph, through shared target genes across regulons.

Cross-sample regulons comparative analysis.

For any pair of the 4 samples (CV-Tex-αPD-L1; T-Tex-αPD-L1; CV-Tex-isotype; T-Tex-isotype), every regulon from condition 1 was overlapped with every regulon from condition 2. The number of genes overlapped in each pair of regulons was put as entries of the regulon overlap matrix. The significance of each overlap was determined using a Fisher's exact test with FDR adjustment of the p-values.

The 4 overlap matrices were then used to construct an undirected graph of regulons. Each significant (padj<0.05) overlap is represented as an edge, and the corresponding two regulons are represented as the connected nodes. The graph was then converted to the corresponding adjacency matrix and partitioned using the leiden algorithm from the leiden package (version 0.3.1) with the default resolution (=1).

Plots.

Unless otherwise stated, all plots were constructed in R (version 3.6.1) using ggplot2 (version 3.2.1).

Data availability.

All bulk RNA-seq, scRNA-seq and scATAC-seq data have been deposited in the GEO and are available (GSE152628).

Supplementary Figures

Supplementary figures for this chapter are included in Appendix A: Supplementary Figures.

CHAPTER 4: MAP YESCARTA CAR-T CELL DIFFERENTIATION IN DIFFUSE LARGE B-CELL LYMPHOMA PATIENTS

Note:

The following section (Chapter 4) is a summary of current results from an ongoing project, in which I am expected to eventually be a co-first author. Most of the sequencing data are ready, and experimental validations and extensions are being planned and conducted by the other expected co-first author, Yifei Hu. This chapter reports preliminary results from 3 pairs of patients, while samples and data from more patients will continue to be collected.

Authors:

Yifei Hu*, Guoshuai Cao*, Justin P. Kline, Jun Huang.

* These authors contributed equally to this work.

Author Contributions:

Y.H performed the experiments. **G.C. performed the bioinformatic analyses.** J.H. and J.K. supervised the study.

Introduction

Anti-CD19 chimeric antigen receptor-T (CAR-T) cell therapy has revolutionized treatment for relapsed/refractory large B-cell lymphoma (DLBCL)[247,248]. During treatment, T cells taken from cancer patients (autologous), transduced with CARs, and then adoptively transferred back into patients to target and kill $CD19^+$ lymphoma cells, in an MHC-independent manner[90,249]. Despite CAR-T cell therapy's remarkable 40% complete response rate, the mechanisms for CAR-T cell therapy resistance/non-response remain largely unknown.

All current FDA-approved CARs are second-generation CARs that each consists of an extracellular CD19-targeting single-chain variable fragment (scFv) derived form an antibody, a transmembrane hinge region, and an intracellular CD28 or 4-1BB co-stimulatory domain fused to a signaling domain derived from the CD3ξ chain of the endogenous T cell receptor (TCR)[90,250]. Accumulative evidences demonstrate that the different co-stimulatory domains confer different CAR-T cell dynamics and phenotype, with CD28-based CAR-T having a more rapid expansion but less persistence and a more effector-like phenotype with enhanced glycolytic metabolism[107,109,247,248]. However, it has been well established that T cells are highly heterogeneous and polyfunctional[13,59,251], and it remains elusive how the CD28-based CAR-T cells differentiate in vivo and what mechanisms/subsets are driving the effector-like phenotype and limiting their persistence.

In addition, recent studies found that the effector-like phenotype of CD28-based CAR-T cells correlated with fast activation and greater signaling intensity partly due to the constitutive association of the tyrosine kinase LCK with the CD28 intracellular CAR domain[252]. T cell research

has demonstrated that strong T cell activation partially drives exhaustion[253,254], and mouse model studies have suggested that CD28-based CAR-T cells mostly become exhausted[255,256]. Clinical CAR studies have shown that non-response correlates with molecular signature of exhaustion before infusion, but the post-infusion exhaustion mechanisms and exhausted T cell heterogeneity in non-responding patients have not been studied.

In this study, we conducted comprehensive single-cell multi-omic profiling (transcriptome, proteome, and TCR clonotype) of the apheresis, infusion, and post-infusion early, peak, and late-stage PBMCs from six YESCARTA (CD28 co-stimulation) CAR-T patients with DLBCL, 3 complete responders (CRs) and 3 non-responders (NRs). We found the CAR-T cells were highly heterogenous at all timepoints, but there is a differential enrichment of CAR^+ Treg cells in the non-responders. In addition, through gene regulatory network analysis and deep-learning, we determined a type-1 interferon inducing IRF7-driving gene module that accurately differentiate the complete responder and non-responder CAR-T cells in the infusion product and thus can be used as an early predictor of clinical response. Furthermore, we revealed the possibility that gamma-delta T cells and cytotoxic CD4 T cells may be some of the key effector subpopulations that mediate the CAR-T therapy response.

Summary of Current Results

Single-cell transcriptional profiling of total peripheral T cells

To characterize the YESCARTA CAR-T cells *in vivo* and reveal their heterogeneity at various stages of differentiation, we performed droplet-based single-cell RNA sequencing (scRNA-seq; 10X Genomics) coupled with single-cell protein and TCR profiling (Total-seq and scTCR-seq; 10X Genomics) on PBMCs taken from 3 CRs and 3 NRs with DLBCL at different stages of the CAR-T therapy: the apheresis product, the infusion product, and post-infusion cells before, during and after peak expansion (Fig. 4.1a). For apheresis and infusion product, we sorted out and sequenced the $CD3^+$ T cells. For post-infusion cells, we applied our in-house developed CD19-CAR tetramer and sorted the cells into CAR-T cells, endogenous T cells, and non-T cells, which were then sequenced separately. We then focused our analyses on the sorted $CD3^+$ T cells (both the CAR-T and the endogenous T cells) from the infusion and post-infusion samples, while preserving the apheresis and non-T data (not shown) for later analyses and validation. After standardized preprocessing (see Methods) and filtering of doublets and high-mitochondrial-gene-content cells, we obtained scRNA-seq profiles from 100,358 T cells, with 49544 cells from CRs and 50814 cells from NRs.

Through integrative analyses and graph-based clustering, we first identified 10 major cell types based on canonical gene markers (Fig. 4.1a-e; Apoptotic T based on relatively high mitochondrial gene content), including five T cell clusters ($CD8^+$ T, $CD4^+$ T, Apoptotic T, Proliferating T, and Gamma-delta T) and five small contaminating non-T cell clusters (NK, B, Dendritic Cell, Monocytes, and Platelets). Within the T cell compartments, we observed homogenous mixing of cells from CRs and NRs, suggesting no response-specific clusters.

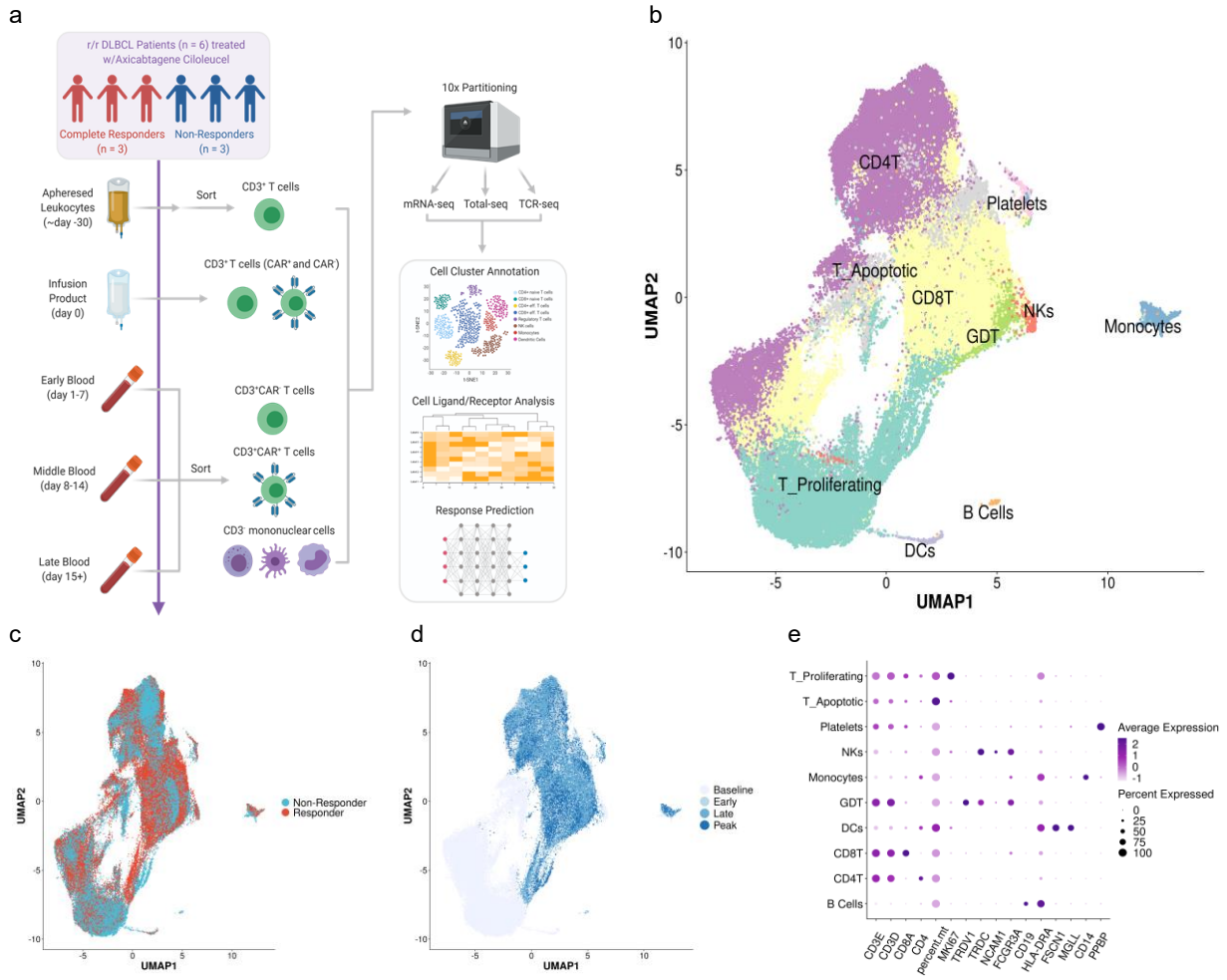


Figure 4. 1 Longitudinal, multi-omic single cell profiling of T cells in DLBCL patients receiving YESCARTA CAR-T therapy.

a, Flowchart describing the overall experimental design and analysis plan of this study. Not all cells and analyses are shown in this chapter, as project is still ongoing. **b**, UMAP representation of scRNA-seq data of 100,358 $CD3^+$ sorted T cells from infusion products and post-infusion samples as shown in **a**, color-coded for the major cell types. **c**, UMAP representation from **b** color-coded by clinical response. **d**, UMAP representation from **b** color-coded by different time points when samples were taken. **e**, Canonical cell gene markers were used to annotate clusters by major cell types as represented in the UMAP plot in **b**.

However, we observed clear distinction between T cells from infusion product (Baseline) and those from post-infusion samples taken before (Early), duing (Peak), and after (Late) peak expansion. Thus, we proceeded our analyses by first removing the non-T and Apoptotic T clusters, and then re-analyzed cells from pre-infusion and post-infusion samples independently.

Heterogeneity and differential compositions of post-infusion T cells

To characterize the T cell heterogeneity more clearly, we first re-clustered the post-infusion T cells, which consisted of the presorted CAR-T cells ($CD3^+CAR^+$) and endogenous T cells ($CD3^+CAR^-$) using our in-house designed CD19-CAR targeting tetramers. We then identified 20 T cell subtypes, including 9 subtypes of $CD4^+$ T cells, 8 subtypes of $CD8^+$ T cells, proliferating T cells, γδ T cells, and mucosal associated invariant T (MAIT) cells (Fig. 4.2 a-b), and most of them were observed with different proportions at all timepoints and both NRs and CRs (Fig. 4.2 c).

To understand what phenotypes correlate with CAR expression, we determined the CAR transgene sequence and mapped transcripts on it to estimate the CAR expression in every cell. Of the $CD4^+$ T cells, we identified 5 memory T cell subpopulations: the Th1 effector memory ($TCF7^+CCR7^-IL7R^+KLRB1^+CXCR3^+$) cells, the Th17 central memory ($CCR7^+TCF7^+IL7R^+CCR6^+$) cells, and 3 other central memory ($CCR7^+TCF7^+IL7R^+$) clusters, which were partitioned by their relative expression of KLRB1 and CISH (CISH-KLRB1-; $CISH^+KLRB1^-$; $CISH^-KLRB1^+$). All the central memory $CD4^+$ T cell subpopulations consisted mostly of endogenous T cells that had no CAR expression (Fig. 4.2 d-e) and thus were not activated by the tumor cells. The Th1 effector memory subpopulation, in contrast, consisted of a significant

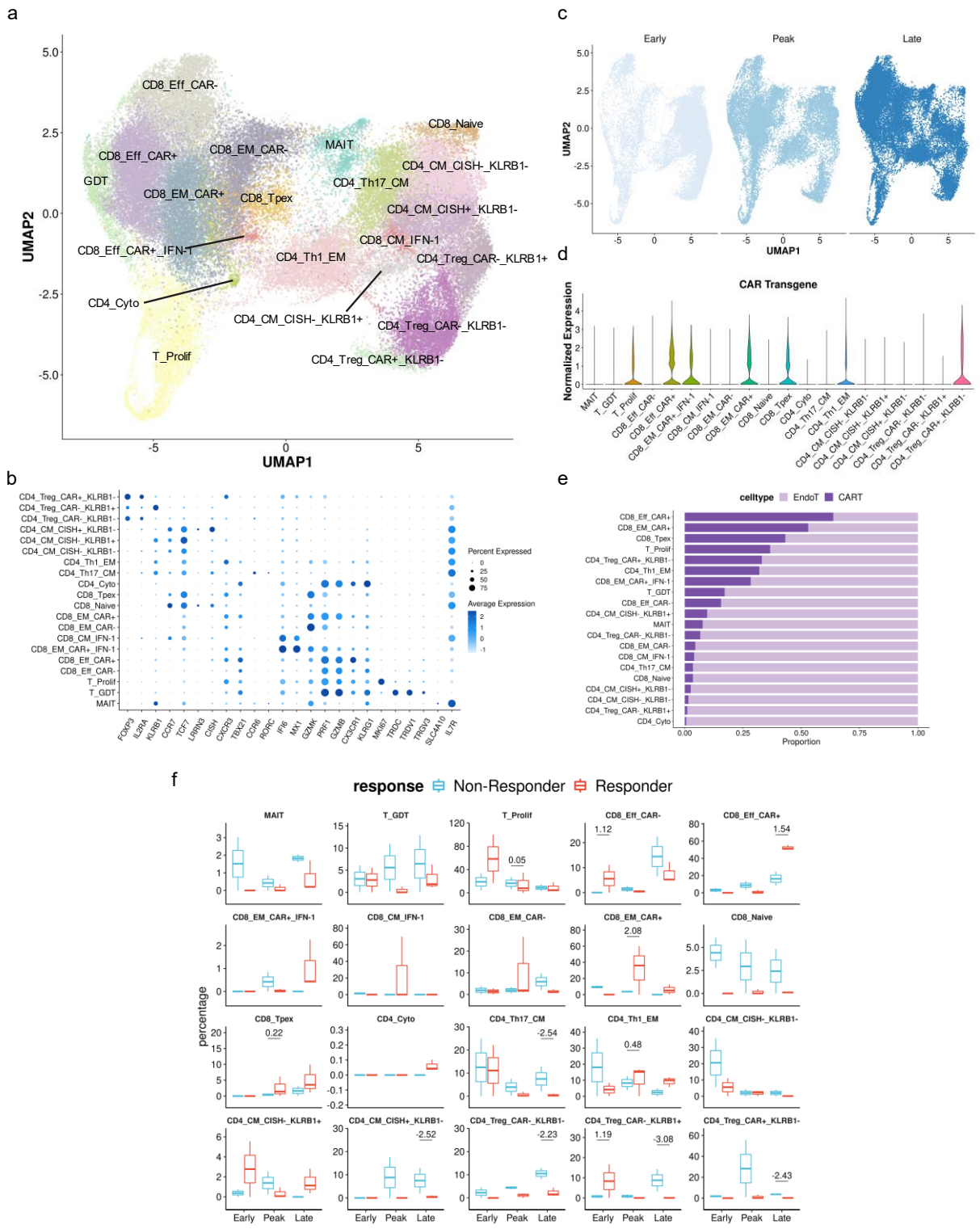


Figure 4. 2 Heterogeneity and differential compositions of post-infusion CAR-T and endogenous T cells.

Figure 4. 2, continued

a, UMAP representation of scRNA-seq data of 14,515 $CD3^+CAR^+$ CAR-T and 46,883 $CD3^+CAR^-$ endogenous T cells from post-infusion samples before, during or after peak expansions. Colors encode different T cell subsets. **b**, Canonical cell gene markers that were used to annotate T cell subsets in the UMAP plot in **a**. **c**, UMAP in **a** separated and color-coded by timepoints. **d**, Violin plots showing the CAR transgene expression distributions among different clusters. **e**, Bar plot depicting the proportions of CAR-T cells and endogenous T cells in each cell subset. **f**, Box plots of the percentage of each T cell subset among all CAR-T cells at each time point in responders or non-responders. Differential composition analysis between responders and non-responders was done using a Bayesian approach provided by scCODA. Significant differences were marked by horizontal lines, and the numbers on top indicate the magnitude of difference (responder vs non-responder) in terms of log2-fold change.

proportion of CAR-T cells with high CAR expression, which potentially led to their activation and the remodeling from a central memory to an effector memory phenotype. Similarly, little CAR T cells or CAR expression were observed in the Naïve T cells ($CCR7^{hi}LRRN3^{hi}TCF7^{+}IL7R^{+}$) or the central memory type-1 interferon response T cells that we identified in the $CD8^{+}$ T cell compartment. We note here that we took advantage of the expression of CD45RO/CD45RA (data not shown here) we obtained from total-seq as a reference to differentiate the memory population ($CD45RO^{+}$) from the naïve populations ($CD45RA^{+}$).

In addition to the memory $CD4^{+}$ T cells, we also identified three large regulatory T cell (Treg; $FOXP3^{+}IL2RA^{+}$) clusters in the $CD4^{+}$ T cell compartment: the $CAR^{+}KLRB1^{-}$ Treg, $CAR^{-/low}KLRB1^{-}$ Treg, and $CAR^{-}KLRB1^{+}$ Treg cells. There was a clear CAR expression gradient from the top to the bottom of the UMAP among the Treg cells, which was echoed in other parts of the UMAP. Among the $CD8^{+}$ T cells, both terminal effector ($CX3CR1^{+}KLRG1^{+}TCF7^{-}$) and effector memory cells ($CX3CR1^{low}KLRG1^{low}TCF7^{+}$) consist of CAR^{+} and CAR^{-} subpopulations, with the CAR^{-} subpopulation on top. It was also apparent that there is another effector-to-stem phenotype gradient from left to right of the UMAP. Among the effector T cells, we also identified 3 other unique clusters: 1) a cytotoxic $CD4^{+}$ T cell ($CD4^{+}TBX21^{+}KLRG1^{+}$) cluster with high expression of genes associated with cytotoxicity such as GZMB and PRF1; 2) a cluster with a transcriptional profile of precursor exhausted $CD8^{+}$ T cells, including both high expression of stem-like genes such as CCR7 and TCF7 and effector-associated genes such as GZMK, CXCR3, and KLRG1; 3) a $CD8^{+}$ effector T cell cluster that also express genes associated with type-1 interferon response such as IFI6 and MX1. The large Treg population, many of which are CAR^{+}, and the highly

heterogeneous effector T cell populations suggest complex cell compositions of the T cells that mediate resistance or response.

To further investigate how the differential composition of T cells associate with response/resistance, we evaluated the distribution of each cluster in both responders and non-responders across different time points (Fig. 4.2 f). Applying a Bayesian model from scCODA, we assessed the significance of the composition comparison and estimated the magnitude in terms of log2-fold change. Notably, during the peak expansion stage, there were significantly higher proportions of CAR^+ effector memory $CD8^+$ T cells and effector memory Th1 T cells (many CAR^+ as in Fig 4.2 d) in the responders. Towards the late stage, there were significantly higher proportions of both CAR^+ and CAR^- Treg cells in the non-responders, whereas the proportion of CAR^+ effector $CD8^+$ T cells was significantly increased in responders. In addition, while not statistically significantly, there appeared to be more naïve $CD8^+$ T cells and stem-like $CISH^+KLRB1^-$ central memory $CD4^+$ T cells in the non-responders across all timepoints. Overall, this indicates that the response to CAR-T therapy were largely mediated by the CAR^+ effector and effector memory $CD8^+$ T cells which successfully expanded, while the non-response was associated with a lack of activated effector or effector memory $CD8^+$ T cells but instead a larger Treg populations. We note here that some large change was determined insignificant by scCODA due to our limited patient number (n=3 each for NRs and CRs) but this will be reassessed once data from more patients are collected.

Heterogeneity and differential compositions of pre-infusion T cells

Next, to gain insight into T cells prior to the infusion, we re-clustered the pre-infusion T cells as well. We identified 11 T cell subtypes, including 6 subtypes of $CD4^+$ T cells, 4 subtypes of $CD8^+$ T cells, and proliferating T cells (Fig. 4.3 a-b), and most of them were confirmed to be CAR^+ as expected (Fig. 4.3 c). In addition, no response-specific clusters were found (Fig 4.3 d).

The dominant factors that dictate the heterogeneity of the T cells in the infusion product was different from that of post-infusion T cells. UMAP1 correlates most strongly with a resting-to-proliferation phenotype differences, and UMAP2 does not show clear correlation. Among the resting T cells, we identified both naïve ($CCR7^{hi}TCF7^+LRRN3^+$) $CD4^+$ and naïve $CD8^+$ T cells, as well as central memory ($CCR7^+TCF7^+LRRN3^{low}$) $CD4^+$ T cells. However, notably, all these resting T cells expressed at low level genes associated with activation or effector function such as CXCR3, PRF1 and GZMB, indicating activation and transcriptional remodeling by the ex vivo stimulation which is part of the CAR-T manufacturing process. We also identified Th1 and Th17 effector memory cells re-expressing CD45RA, again suggesting activation of effector memory $CD4^+$ T cells. In addition, we observed two distinct Treg ($FOXP3^{hi}IL2RA^{hi}$) clusters, with one of them displaying an activated phenotype and expressing genes associated with cytotoxicity, such as GZMB and PRF1, as well as CSF2 which encodes GM-CSF. In the effector (expressing high GZMB and PRF1) $CD8^+$ T cell compartment, we identified 3 clusters: CAR^- effector $CD8^+$ T cells, type2 CAR^+ effector ($GATA3^+$; expressing IL5 and IL13) T cells, and $LRRN3^+CAR^+$ effector $CD8^+$ T cells. Notably, only the CAR^- effector $CD8^+$ T cells express high level of CX3CR1 and KLRG1, which were key effector markers to identify *in vivo* $CD8^+$ T cells. This

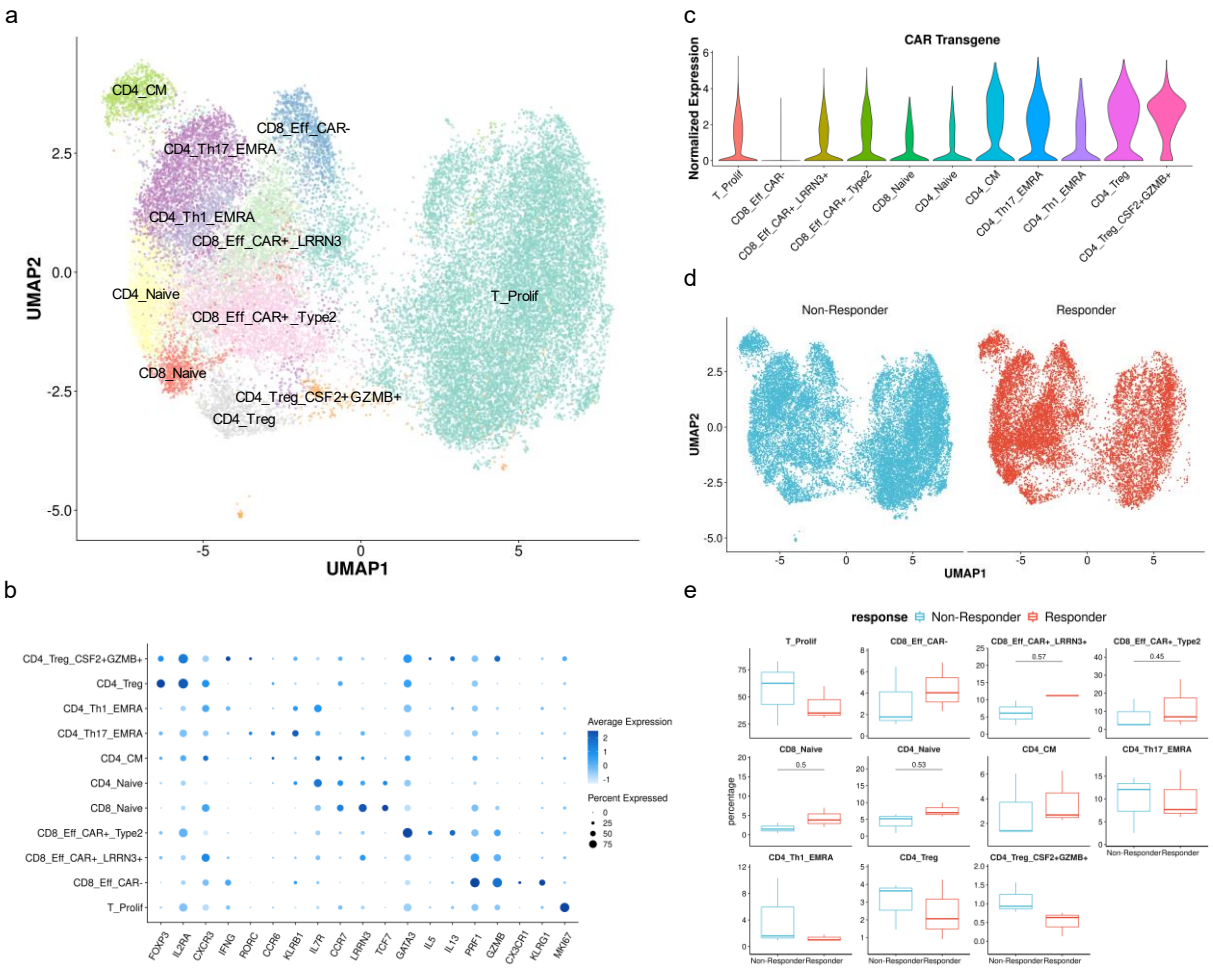


Figure 4. 3 Heterogeneity and differential compositions of pre-infusion CAR-T cells in the infusion products.

a, UMAP representation of scRNA-seq data of 31,586 $CD3^+$ CAR-T cells from the pre-infusion infusion products. Colors encode different T cell subsets. **b**, Canonical cell gene markers that were used to annotate T cell subsets in the UMAP plot in **a**. **c**, Violin plots showing the CAR transgene expression distributions among different CAR-T subsets. **d**, UMAP in **a** separated and color-coded by clinical response. **e**, Box plots of the percentage of each T cell subset among all CAR-T cells in responders or non-responders. Differential composition analysis between responders and non-responders was done using a Bayesian approach provided by scCODA. Significant differences were marked by horizontal lines, and the numbers on top indicate the magnitude of difference (responder vs non-responder) in terms of log2-fold change.

again suggests that the *ex vivo* stimulation and expansion of CAR-T cells led to certain unique transcriptional remodeling in different T cell subsets.

To investigate differences in the T cell compositions in the infusion product, we again evaluated the distribution of each cluster in both responders and non-responders applied the Bayesian model from scCODA. As expected, around 30-70% of the cells were actively proliferating in both the responder-derived and non-responder-derived CAR-T cells. However, we found significantly higher proportions of both types of effector $CD8^+$ T cells in the responder-derived CAR-T cells. In addition, both CD4 and CD8 naïve T cells, which showed signatures of activation as described above, were also more significantly expanded in the responder-derived CAR-T cells. Furthermore, though not significant, we could also observe slightly higher proportions of Treg subpopulations in the non-responder-derived CAR-T cells, which is consistent with the greater Treg populations post-infusion in the non-responders.

Predictions on post-infusion clinical outcome using infusion products

To confirm previous findings that exhaustion signatures in $CD8^+$ T cells in the infusion product correlate with poor clinical response[77], we calculated the exhaustion signature scores in all the $CD8^+$ T cell clusters in the infusion products and compared the score distributions of responders and non-responders. Indeed, except for naïve $CD8^+$ T cells, all proliferating and effector $CD8^+$ CAR-T cells appeared to be more exhausted for those derived from non-responders compared to those from responders (Fig. 4.4 a). In addition, the difference in exhaustion signature is more significant in proliferating and CAR^+ effector $CD8^+$ T cells than the CAR^- ones. This

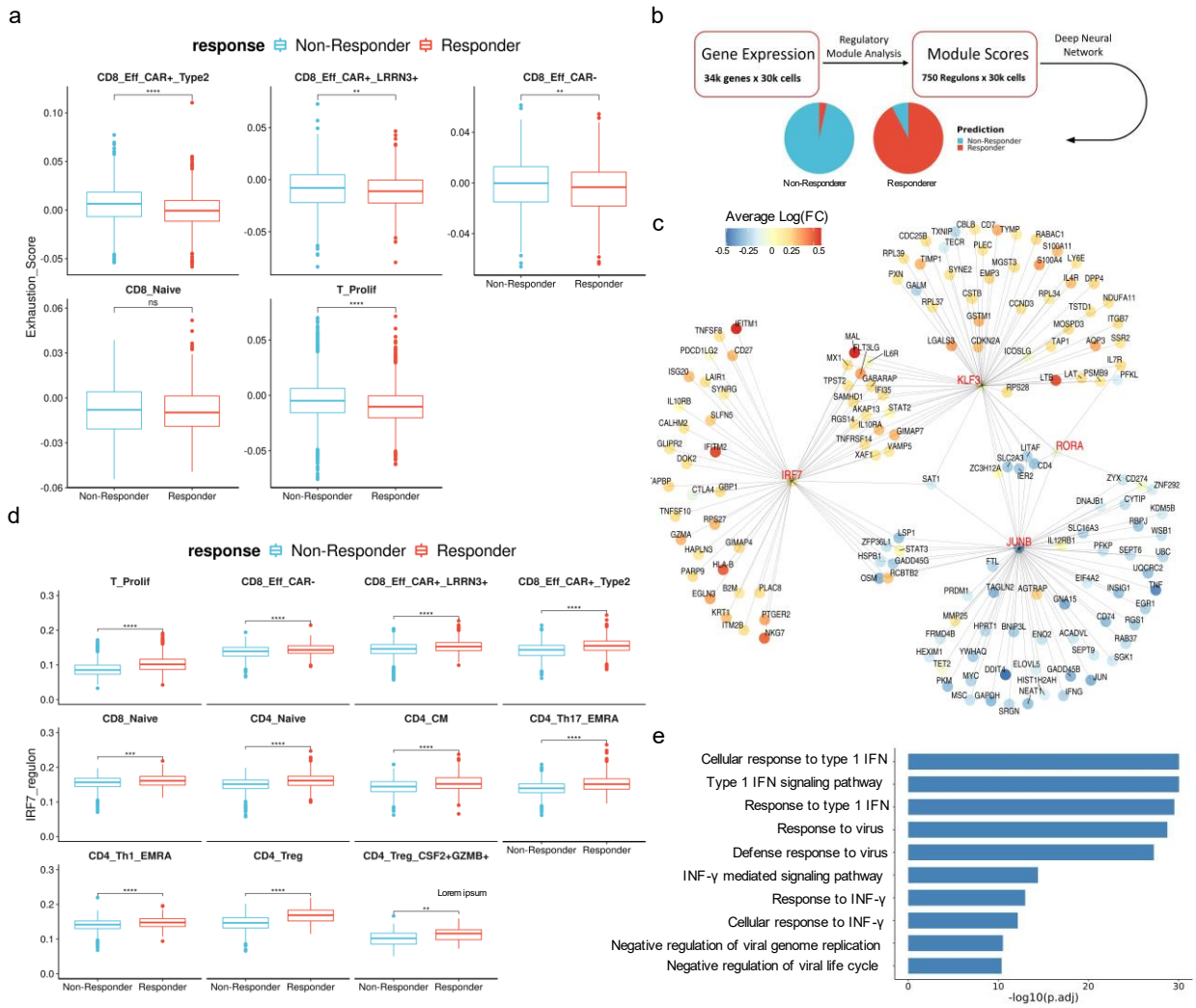


Figure 4. 4 Predictions on post-infusion clinical outcome using infusion products signatures

a, Box plots showing the distribution of $CD8^+$ T cell exhaustion scores across different CAR-T subsets in the infusion products in responders or non-responders. The differences were assessed using two-sided Wilcoxon rank-sum test with Benjamini-Hochberg p-value adjustment, and the significance levels were labeled: ns, p.adj>0.05; *, p.adj<0.05; **, p.adj < 0.01; *** , p.adj<0.001; ****, p.adj<0.0001. **b**, Workflow of deep learning to predict clinical outcome using regulatory module scores of CAR-T cells in the infusion products. **c**, Gene network plot showing top 4 informative regulatory modules with the central transcription factors highlighted in red. The putative targets in each module were filtered based on differential gene expression (DGE) of responders vs non-responders, and their names are shown in black. Each node is colored by the average log2(Fold Change) values of the DGE. **d**, Box plots showing the distribution of IRF7-regulon scores across different CAR-T subsets in the infusion products in responders or non-responders. Statistical tests were performed, and the results were labelled the same way as in **a**. **e**, Bar plot showing the top 10 pathways enriched among the IRF7-regulon genes.

validates $CD8^+$ T cell exhaustion signature as a predictor for non-response/resistance to CAR-T immunotherapy.

To discover other gene markers that can also be used as response predictors using an unbiased and more comprehensive approach, we took advantage of regulatory module analysis and deep neural networks (Fig. 4.4 b). Briefly, we first analyzed transcription factor and putative target genes correlations and built a set of regulatory modules (or simply known as regulons), each consisted of a single transcription factor and its highly correlated target genes. We then scored every cell using every regulatory module and trained a deep neural network based on the regulon scores to predict the clinical response associated with every cell. With this approach, we obtained a final overall prediction accuracy of around 93%. We then evaluated the relative importance of each regulon in informing the neural network to make a correct prediction, and we found 4 regulons to be most informative: IRF7-, KLF3-,RORA- regulons were upregulated in responders and JUNB-regulon was upregulated in non-responders. After filtering and examining the target genes differential expression between responders and non-responders in each regulon (Fig. 4.4 c), we decided to investigate further the IRF7-regulon, which was also the most informative regulon for predicting responders.

To determine if the IRF7-regulon differentially affect certain T cell subsets like how exhaustion signature is most pronounced in $CD8^+$ effector T cells, we evaluated the IRF7-regulon scores in responders and non-responders derived cells across different clusters. Surprisingly, the IRF7-regulon scores were significantly higher across all T cell clusters in those cells from responders than those from non-responders (Fig. 4.4 d). We then conducted pathway enrichment analysis and found Type-1 interferon response across the top terms enriched (Fig. 4.4 e). This

strongly suggests that IRF7 preferentially mediated type-1 interferon signaling in CAR-T cells derived from the responders, which may play an important role in conferring the efficacy of the therapy post infusion. This is currently being investigated in our lab through *in vitro* studies and xenograft mouse models.

Discussion and Future Work

In this project, we performed single-cell multi-omic profiling (transcriptome, proteome, and TCR clonotype) of the infusion product, and post-infusion early, peak, and late-stage PBMCs from six YESCARTA (CD28 co-stimulation) CAR-T patients with DLBCL, 3 complete responders (CRs) and 3 non-responders (NRs). We characterized the heterogeneity of all the T cells, including both the CAR^+ T cells and the CAR^-/endogenous T cells, and we assessed the differential proportions of each cell subtype between the cells from responders and those from non-responders. In particular, we found significant enrichment of effector memory and effector $CD8^+CAR^+$ CAR-T cells in the responders during and after the early expansion whereas in the non-responders, we found significant increase in the proportions of both CAR^+ and CAR^- regulatory T cells. Consistent with these post-infusion observations, we also found significantly higher proportions of $CD8^+CAR^+$ CAR-T cells among infusion product cells derived from the responders than from the non-responders. Although not statistically significant, we also found higher proportions of regulatory T cells among the non-responder derived infusion product CAR-T cells. Moreover, we validated the use of exhaustion signature as a predictor of clinical response, and we discovered using deep learning an additional IRF7-centered regulatory module that can successfully predict clinical response.

Unlike past studies where an enrichment of $CD8^+$ memory T cells were found to be enriched in infusion products of responders[77,257], we did not observe memory $CD8^+$ T cells in our infusion product cells. However, the naïve $CD4^+$ and $CD8^+$ CAR-T cells we identified in our samples may correspond to the memory cells detected by other studies, as naïve and central memory T cells share most of the transcriptional signatures. The reason why we labeled those cells

as naïve cells were largely due to the $CD45RO^{-}CD45RA^{+}$ surface protein expression profiles we obtained from the paired total-seq. Indeed, we also observed much higher proportions of what we labeled as naïve $CD8^{+}$ CAR-T cells. In addition, our naïve $CD8^{+}$ T cells resemble the $CD45RO^{-}$ $CD27^{+}$ $CD8^{+}$ T cells that Fraietta et al.[77] reported to be correlate to positive response. Furthermore, these naïve $CD8^{+}$ CAR-T cells had high expression of CCR7, which was essentially the main memory-signature gene that past studies have concluded to correlate with positive clinical response. Therefore, our results also support the enrichment of CCR7 expression among CAR-T cells could be used as a predictor of clinical response.

Our study also discovered a IRF7-driven regulatory module that induces type-1 interferon response as a novel positive clinical response predictor. Besides activation signal through TCR-pHMC binding and costimulatory signals, the differentiation of naïve T cells into effector T cells also require cytokine support, often referred to as the third signal. A typical example is type-I interferons, which are known to support $CD8^{+}$ T cell function[258]. While type-1 interferon is produced by many cell types such as plasmacytoid dendritic cells and monocytes[259,260], Zhao et al.[261] showed that human CAR-T cells can also produce significant levels of IFNβ through an IRF7/IFNβ signal pathway. In addition, they demonstrated that IRF7/IFNβ activation also enhances CAR-T cell-mediated tumor elimination. While their study showed the IRF7/IFNβ induction in a third generation 1928z-41BBL CAR-T cells, our study suggests that the IRF7-mediated type-1 interferon response may be directly relevant to the efficacy of YESCARTA CAR-T cell therapy as well.

There are several limitations to the current work, but as noted earlier, this is still an ongoing project and further work will continue to be added to this study. Firstly, there are currently only 3

pairs of responders and non-responders profiled, and more patient samples will be collected and profiled as the study continues. Secondly, the functional relevance of the IRF7 regulatory module discovered needs to be validated experimentally using *in vitro* cell lines and xenograft mouse models. Experiments are currently being planned by my lab collaborator, Yifei Hu, to specifically test if culturing CAR-T cells with and without type-1 interferons will modify the CAR-T cell effector functions. In addition, further analyses on the TCR data are still being done and the results will validate any direct link between the regulatory T cells in the infusion product and those ones enriched post infusion in the non-responders. We also planned to sort out these regulatory T cells at different time points for further *in vitro* functional studies.

Taken together, this longitudinal, multi-omic, integrated multicellular profiling of YESCARTA CAR-T cells in DLBCL patients fills in the gap of little *in vivo* single-cell differentiation studies of CD-19 targeting CAR-T cells, and it lays the foundation for future characterization of the complex, dynamic differentiation processes and underlying mechanisms of CAR-T cells. Furthermore, the IRF7-regulatory module, once validated experimentally, could also provide a novel biomarker for prognosis and even enable modification of the current manufacturing process to favor the generation of more efficacious CAR-T cells and improve the clinical outcome. Lastly, our multimodal single-cell data can also serve as a rich resource for deeper understanding of the pre-infusion and post-infusion CAR-T cell heterogeneity in future studies.

Material and Methods

Generation of CD19-tetramers

Multimers were constructed from AviTag-biotinylated human His-tagged CD19 (Acro Biosystems, CD9-H82E9). Reconstituted CD19 was stored at -80°C at 0.2 mg/mL with a biotin-to-protein ratio of 0.5 (manufacturer-tested) until experiments. To generate CD19-tetramers, biotinylated CD19 was added to Alexa Fluor 647-labeled tetrameric streptavidin (BioLegend, 405237) at a 4:1 molar ratio for 30 minutes at 4°C in the dark. This mixture was diluted with PBS to 50 nM for staining.

Infusion product cell preparation for single-cell omics assays

Infusion product cells (IPs) were collected from the residual cells from patient infusion bags by washing with PBS, and cryopreserved (CellBanker 1, Amsbio, 11888) in liquid nitrogen until analyses.

Cryopreserved IP biospecimens were thawed in warm RPMI+10% FBS. Cells were first washed with cold FACS buffer (PBS, 2% BSA, 0.05% sodium azide). Next, Fc receptors were blocked by incubation with Human TruStain FcX (BioLegend, 422301) at 1:50 dilution for 5 minutes at 4°C. Then, IP cells were incubated for 30 minutes at 4°C in the dark with a staining solution containing oligo-conjugated antibodies (Total-seq C, BioLegend): anti-CD3ε (clone UCHT1, 300479), anti-CD5 (clone UCHT2, 300637), anti-TCRα/β (clone IP26, 306743), anti-TCRγ/δ (clone B1, 331231), anti-CD4 (clone SK3, 344651), anti-CD8α (clone SK1, 344753), anti-CD45RA (clone HI100, 304163), anti-CD45RO (clone UCHL1, 304259), anti-CCR7 (clone G043H7, 353251), anti-CD95 (clone DX2, 305651), anti-CD57 (clone QA17A04, clone 393321),

anti-CD25 (clone BC96, 302649), anti-CD127 (clone A019D5, 351356), anti-CD103 (clone BerACT8, 350233), anti-CXCR3 (clone G025H7, 353747), anti-CCR4 (clone L291H4, 359425), anti-CCR6 (clone G034E3, 353440), anti-PD-1 (clone EH12.2H7, 329963), anti-TIM-3 (clone F38-2E2, 345049), anti-LAG-3 (clone 11C3C65, 369335), anti-CD39 (clone A1, 328237), anti-TIGIT (clone A15153G, 372729), anti-CD27 (clone O323, 302853), anti-CD40L (clone 24-31, 310849), anti-GITR (clone 108-17, 371227), anti-OX40 (clone Ber-ACT35, 350035), anti-4-1BB (clone 4B4-1, 309839), anti-CD28 (clone CD28.2, 302963), mouse IgG1, κ isotype control (clone MOPC-21, 400187), mouse IgG2a, κ isotype control (clone MOPC-173, 400293), and mouse IgG2b, κ isotype control (clone MPC-11, 400187). All antibodies were generally used at manufacturer recommendations.

Subsequently, IP cells were washed three times in cold cell media and counted, before single-cell partitioning.

Peripheral blood cell preparation for single-cell omics assays

Patient peripheral blood mononuclear cells (PBMCs) were collected from patient blood biospecimens by Ficoll-Paque PLUS (Cytiva, 95021-205), and cryopreserved (RPMI supplemented with 10% FBS and 10% DMSO) in liquid nitrogen until analyses.

Cryopreserved PBMC biospecimens were thawed in warm RPMI+10% FBS. Cells were first washed with cold FACS buffer (PBS, 2% BSA, 0.05% sodium azide). Next, Fc receptors were blocked by incubation with Human TruStain FcX (BioLegend, 422301) at 1:50 dilution for 5 minutes at 4°C. Then, blood cells were incubated for 30 minutes at 4°C in the dark with a staining solution containing BV421-anti-CD3ε (clone SK7, BioLegend, 344833), oligo-conjugated antibodies, and Alexa Fluor 647-labeled antigen-tetramers at 3 nM final concentration for CAR-

detection. Oligo-conjugated antibodies (Total-seq C, BioLegend) consisted of anti-CD5 (clone UCHT2, 300637), anti-TCRα/β (clone IP26, 306743), anti-TCRγ/δ (clone B1, 331231), anti-CD4 (clone SK3, 344651), anti-CD8α (clone SK1, 344753), anti-CD45RA (clone HI100, 304163), anti-CD45RO (clone UCHL1, 304259), anti-CCR7 (clone G043H7, 353251), anti-CD95 (clone DX2, 305651), anti-CD57 (clone QA17A04, clone 393321), anti-CD25 (clone BC96, 302649), anti-CD127 (clone A019D5, 351356), anti-CD103 (clone BerACT8, 350233), anti-CXCR3 (clone G025H7, 353747), anti-CCR4 (clone L291H4, 359425), anti-CCR6 (clone G034E3, 353440), anti-PD-1 (clone EH12.2H7, 329963), anti-TIM-3 (clone F38-2E2, 345049), anti-LAG-3 (clone 11C3C65, 369335), anti-CD39 (clone A1, 328237), anti-TIGIT (clone A15153G, 372729), anti-CD27 (clone O323, 302853), anti-CD40L (clone 24-31, 310849), anti-GITR (clone 108-17, 371227), anti-OX40 (clone Ber-ACT35, 350035), anti-4-1BB (clone 4B4-1, 309839), anti-CD28 (clone CD28.2, 302963), mouse IgG1, κ isotype control (clone MOPC-21, 400187), mouse IgG2a, κ isotype control (clone MOPC-173, 400293), and mouse IgG2b, κ isotype control (clone MPC-11, 400187). All antibodies were generally used at manufacturer recommendations.

Subsequently, blood cells were incubated with LIVE/DEAD Fixable Near-IR viability dye (Invitrogen, L34975) diluted 1:1000 in PBS for 5 minutes at room temperature. Finally, cells were washed three times in cold cell media before fluorescence-activated cell sorting. To draw sorting gates based on a biological control, PBMCs from a healthy donor were similarly stained and analyzed. Antibody compensation was performed with UltraComp eBeads Plus (Invitrogen, 01-3333-41). Dye compensation was performed with ArC Amine Reactive Compensation Beads (Invitrogen, A10628). Sorted cells were used for single-cell partitioning

Single-cell partitioning

Infusion product (20,000 cells) and peripheral blood cells (≤12,000 each of sorted non-T cells, endogenous T cells (Endo-T cells), and CAR-T cells) were partitioned into droplets for single-cell omics assays via Chromium Next GEM Single-Cell 5'Kit v2 (10x Genomics, 1000263). From the resulting barcoded nucleic acids, TCR-seq libraries were prepared via the Chromium Single-Cell Human TCR Amplification Kit (10x Genomics, 1000252). Total-seq libraries were prepared via the 5' Feature Barcode Kit (10x Genomics, 1000256). Samples indices originated from Dual Index Kit TN Set A (10x Genomics, 1000250) and Dual Index Kit TT Set A (10x Genomics, 1000215). Sequencing libraries were quantified via the Qubit dsDNA HS Assay Kit (Invitrogen, Q32851), quality-checked for fragment sizes via a high-sensitivity D5000 screentape (Agilent, 5067-5592), pooled, and sequenced (Illumina, NovaSeq-6000).

Single-cell RNA-seq/Total-seq data processing

The scRNA-seq reads were aligned to the GRCh38 reference genome and quantified using the *cellranger count* (10X Genomics, version 6.0.0). Total-seq reads were quantified simultaneously by passing the data to the same function. The filtered gene-barcode matrices that contained only barcodes with unique molecular identifier (UMI) counts that passed the quality control were used for downstream analyses.

CAR Transgene Mapping

The axicabtagene ciloleucel CAR design is documented[262]. Its sequence was confirmed by Sanger sequencing of genomic DNA extracted from axicabtagene ciloleucel infusion products. The CAR

sequence was then added to the GRCh38 FASTA and GTF files accordingly, and a custom reference for cellranger was built from these annotation files using cellranger mkref (10x Genomics, version 6.0.0). The resultant custom reference was used for the CAR transgene mapping through *cellranger count*.

Single-cell TCR-seq Data Processing and Analysis

The scTCR-seq reads were aligned to the 10x curated GRCh38 vdj reference genome and quantified using the *cellranger vdj* function(10x Genomics, version 6.0.0). The resultant clonotype and filtered contig annotation data were used for downstream analyses. The clonotype and contig data were then added to the scRNA data by matching the cell barcodes. The frequency of each clonotype in each sample was recalculated based on those successfully matched to the corresponding scRNA data. Each unique clonotype is defined by the nucleotide sequence of CDR3 regions for a productive TCRα chain paired with a productive TCRβ chain.

UMAP analysis and clustering on single-cell RNA-seq data

UMAP analysis and clustering were performed using the Seurat package[263] (Version 4.0.0). Raw count matrices from $CD3^+CAR^+$ CAR-T cells ("CAR-T") and $CD3^+CAR^-$ endogenous T cells ("EndoT") were first converted to Seurat objects before being further merged into one Seurat object. Following that, cells with less than 200 genes detected or greater than 25% mitochondrial RNA content were excluded from further analysis.

Then, the raw count was log-normalized using the *NormalizeData* function with default options. Top 5,000 variable features were then identified using the *FindVariableFeatures* function with the default "vst" method. The data were then centered and scaled using the *ScaleData*

function, with additional regression against the percent of mitochondrial RNA content. Scaled data were then used as input for principal component analysis (PCA) based on variable genes using the *RunPCA* function. Data harmonization was then performed with patients as the covariate on the principal components using the Harmony[264] package through the *RunHarmony* function. UMAP was then constructed based on the first 50 harmony components, with "n.neighbors=30" and "min.dist=0.3". The same harmony components were then used to construct the shared nearest neighbor (SNN) graph using the *FindNeighbors* function, which was then partitioned to identify clusters using the *FindClusters* function with default Louvain algorithm and "resolution =0.8". These clusters were then manually aggregated and classified as T cell subsets based on known markers.

Differential gene expression (DEG) analysis on single-cell RNA-seq data

DEG analyses were all performed using the *FindMarkers* function in Seurat (Version 4.0.0) package, with default parameters and the appropriate 'ident.1" and "ident.2" set as contrast. Unless otherwise stated, the results were then filtered with p_val_adj < 0.05 and abs(avg_logFC) > 0.25.

CAR transgene mapping

The axicabtagene ciloleucel CAR design is documented.3 Its sequence was confirmed by Sanger sequencing of genomic DNA extracted from axicabtagene ciloleucel infusion products. The CAR sequence was then added to the GRCh38 FASTA and GTF files accordingly, and a custom reference for cellranger was built from these annotation files using cellranger mkref (10x Genomics, version 6.0.0). The resultant custom reference was used for the CAR transgene mapping through *cellranger count*.

Regulon/Regulatory module analysis and scoring

We constructed the regulons by supplying the raw scRNA-seq count matrix as input into the SCENIC package and running through the workflow as detailed in its documentation and publication[265]. Each resultant regulon is a gene list with a central TF and all its putative target genes determined through correlation in expression. Individual cells were scored using the AUCell package (version 1.6.1) for a particular regulon as follows. The normalized gene expression was first used as input into the AUCell_buildRankings function to score each cell for gene set enrichment and build a ranking. The signature score was then calculated as an AUC score using the AUCell_calcAUC function with all default parameters.

Exhaustion score calculation

For exhaustion signature score calculation, we retrieved the "GSE9650_EFFECTOR_VS_EXHAUSTED_CD8_TCELL_DN" gene set from the msigdb database[266]. We then used the AddModuleScore function in the Seurat package (Version 4.0.0) with default parameters to calculate the signature scores.

Pathway enrichment analysis

Pathway enrichment analysis was performed using the clusterProfiler[267] package (Version 3.18.1) with default parameters.

Deep neural networks

Deep neural networks were constructed with TensorFlow2[268] and the Tensorflow package (Version 2.4.0) in R. The network itself essentially consists of 8 dense layers, with relu activation,

batch normalization (momentum=0.01, epsilon=0.001) and dropout (rate=0.1) in each dense layer, and a final prediction layer using sigmoid activation. The model was trained for 30 epochs with a validation split of 0.1 and batch size of 32. 5% of all data were also reserved as independent test set. After training and optimization, the feature importance i.e. the assessment of how informative each regulon is was done using the LIME[269] method.

Plotting

Unless otherwise stated, all plots for multi-omics data analysis were constructed in R (version 4.0.0) using ggplot2 (version 3.3.1).

CHAPTER 5: REVEAL AND PREDICT T CELL SIGNALING STATES USING LATTICE LIGHT SHEET MICROSCOPY MULTI-DIMENSIONAL ANALYSES

Note:

The following section (Chapter 5) is reproduced with slight modifications from my co-first authored publication "Lattice Light-Sheet Microscopy Multi-dimensional Analyses (LaMDA) of T-Cell Receptor Dynamics Predict T-Cell Signaling States" (Rosenberg et al. 2020), which was published in Cell Systems on May 20, 2020 under the Creative Commons CC-BY-NC-ND license (https://creativecommons.org/licenses/).

Authors:

Jillian Rosenberg*, **Guoshuai Cao***, Fernanda Borja-Prieto, Jun Huang

* These authors contributed equally to this work.

Author Contributions:

Conceptualization, J.R. and J.H. with input from G.C.; Methodology, J.R. and F.B.-P. with input from J.H. and G.C.; **Formal Analysis G.C.** with input from J.R. and J.H.; Writing, J.R., G.C., and J.H.; Supervision, J.H.

Abstract

Lattice light-sheet microscopy provides large amounts of high-dimensional, high-spatiotemporal resolution imaging data of cell surface receptors across the 3D surface of live cells, but user-friendly analysis pipelines are lacking. Here, we introduce lattice light-sheet microscopy multi-dimensional analyses (LaMDA), an end-to-end pipeline comprised of publicly available software packages that combines machine learning, dimensionality reduction, and diffusion maps to analyze surface receptor dynamics and classify cellular signaling states without the need for complex biochemical measurements or other prior information. We use LaMDA to analyze images of T-cell receptor (TCR) microclusters on the surface of live primary T cells under resting and stimulated conditions. We observe global spatial and temporal changes of TCRs across the 3D cell surface, accurately differentiate stimulated cells from unstimulated cells, precisely predict attenuated T-cell signaling after CD4 and CD28 receptor blockades, and reliably discriminate between structurally similar TCR ligands. All instructions needed to implement LaMDA are included in this paper.

Full Text

Introduction

Lattice light-sheet microscopy (LLSM) is a recently developed microscopy technique that allows for four-dimensional (4D) (x, y, z, and time) imaging with exceptionally high temporal resolution (~100 frames/s, ~1 cell volume/s) and minimal photobleaching[130]. LLSM provides high-dimensional, high spatiotemporal resolution imaging data of cell surface receptors or receptor microclusters (≥1,000) over a long duration of time (≥4 min) on the entire 3D surface of live primary cells. As a result of these capabilities, LLSM and other cutting-edge microscopy

techniques generate larger and more complex high-dimensional data. These data are often underutilized due to a lack of comprehensive and efficient high-dimensional analysis pipelines that are accessible to the general user. The lack of such an analysis pipeline represents a key limitation in the use of imaging systems, including LLSM, that generate big data and are crucial to answering biological questions at the single-molecule level.

Here, we introduce lattice light-sheet microscopy multi-dimensional analyses (LaMDA), a pipeline to apply big data analysis techniques to high-dimensional LLSM data. LaMDA provides a complete end-to-end pipeline from upstream data collection and feature engineering to downstream machine learning and dimensionality reduction analyses. We intentionally constructed LaMDA using publicly available packages to enable easy adaptation and incorporation by general users to facilitate broad application across all areas of cell biology. Moreover, it was designed to motivate a paradigm shift wherein, rather than focusing on the cell as the unit of study, we focus on a molecular unit. The LaMDA pipeline supersedes the need for high-throughput collection of single-cell time-lapse data, which is difficult to capture using LLSM, by enabling extraction of sufficient molecular content for the use of high-dimensional analyses.

We applied LaMDA to image and analyze the dynamics of T-cell receptors (TCR) microclusters on the primary T-cell surface with high dimensionality to understand T-cell signaling states. T cells play a central role in adaptive immunity by mediating immune responses against cancer and infection[203,270,271]. As the dominant receptor, TCRs govern the recognition, activation, differentiation, and function of T cells in health and disease[271,272]. Our LaMDA approach quantitatively revealed the global spatiotemporal dynamics of TCRs, reliably deciphered TCR microclusters from T cells at different signaling states, precisely identified the roles of co-

stimulatory receptors, and accurately differentiated between T-cell stimulations triggered by structurally similar peptide ligands of different affinities. In addition to uncovering new T-cell biology, LaMDA can also be used to guide the pre-clinical design, development, and improvement of immunotherapies and vaccines for cancer, infection, and autoimmunity. Furthermore, as a flexible and broadly applicable pipeline, LaMDA provides a framework for future studies of other surface receptors or intracellular molecules on different cell types by directly linking molecular dynamics to cell signaling and function.

Results

LaMDA

LaMDA was designed to enable big data analysis of high-dimensional LLSM data (Figure 5.1). To begin using LaMDA, LLSM is employed to capture single-cell images in four dimensions (x, y, z, and time) with high spatiotemporal resolution (~1 s/cell volume, x and y resolution ~200 nm, and z step 400 nm) (Figure 5.1A). Then, advanced imaging software, such as Imaris, is utilized to track thousands of surface receptor microclusters (or other structures) from individual cells (Figure 5.1B) and extract multiple (36 in this study) statistical parameters for each tracked object simultaneously (Figure 5.1C). Due to the large size of collected datasets, machine learning and dimensionality reduction methods can be incorporated for analysis (Figures 5.1D and 5.1E). For example, a machine learning-based classifier[273] could be built to learn the inherently subtle yet very complex differences between the tracked receptors of two or more known cellular conditions. This classifier could then be used to predict the state of the receptor microclusters in new or less-understood cellular conditions (Figure 5.1F). Furthermore, feature importance metrics from the

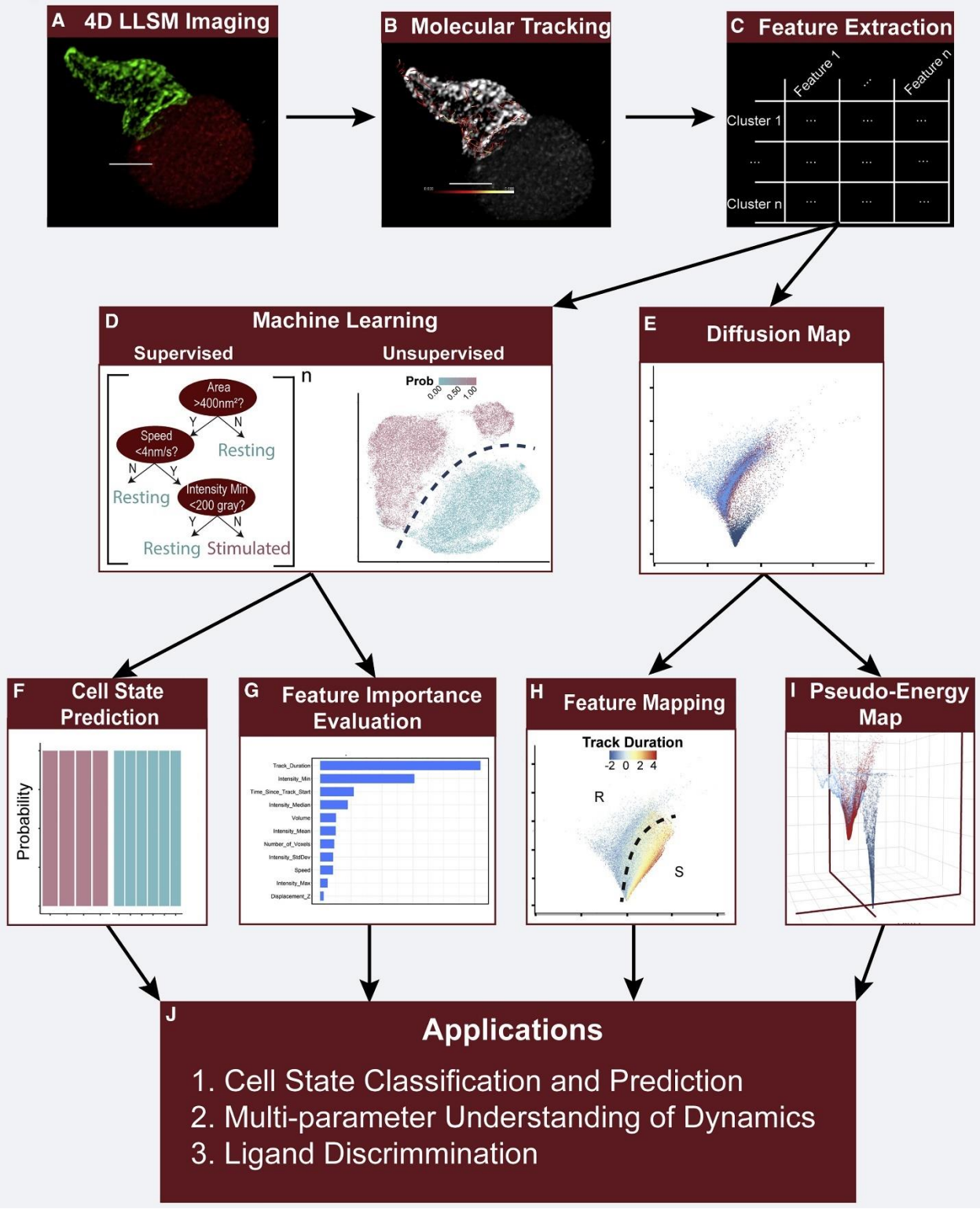


Figure 5. 1 LaMDA Pipeline.

Figure 5. 1, Continued

(A) 4D imaging is conducted with LLSM. Receptors (green) on a cell are fluorescently labeled and 4D images are collected. Scale bar represents 5 mm. (B) After deconvolution and debleaching, receptor microclusters are tracked with Imaris software. Scale bar represents 5 mm. (C) Microcluster features are extracted from tracked TCR microclusters. (D) Machine learning is conducted on extracted features for prediction. (Left) To classify between different cell states, supervised XGboost decision tree ensembles are trained as a binary classifier then tested and validated. (Right) Unsupervised UMAP clustering is used to independently validate the XGboost binary classifier. The XGboost binary classifier is used to predict the states of cells. (E) Diffusion maps are used as a dimensionality reduction technique to further quantify variations among cell states. The diffusion map is built from features selected based on their importance. (F) The XGboost classifier is utilized to provide cell state predictions based on TCR microcluster features. (G) The importance of each feature in the XGboost classifier is evaluated using SHAP, and the top feature is individually analyzed using appropriate statistical modeling. (H) Diffusion map is colored by normalized features (e.g., track duration shown here). Dotted line indicates approximate divide between resting (R) and stimulated (S) cells. (I) The diffusion map is transformed into a 3D pseudo-energy map by estimating local density of datapoints on the diffusion map. The 3D pseudo-energy map is used to analyze stability of tracked receptor microclusters across cell states. (J) LaMDA pipeline applications, further detailed throughout paper.

classifier can inform on the underlying biological differences in the system (Figure 5.1G). Once the classifier metrics are used to select important features, they can be studied individually by building statistical models. To understand if multiple parameters are working together to drive biological differences, dimensionality-reduction techniques, such as uniform manifold approximation and projection (UMAP)[274] and diffusion maps[275,276], could be applied in parallel (Figures 5.1D and 5.1E) to graphically emphasize three important points: (1) these techniques cluster objects into several groups representative of different physical states; (2) the relative locations of these object states inform us on their relationships, indicating directionality and identifying the properties that change with that direction (Figure 5.1H); and (3) the local density at each coordinate on these dimensionality-reduction maps can be further transformed into a pseudo-energy surface to show relative stability of each object (Figure 5.1I). Together, LaMDA uses high-dimensional imaging and big data analyses to discover differentiating properties between cellular phenotypes and to reveal and predict underlying cellular states (Figure 5.1J).

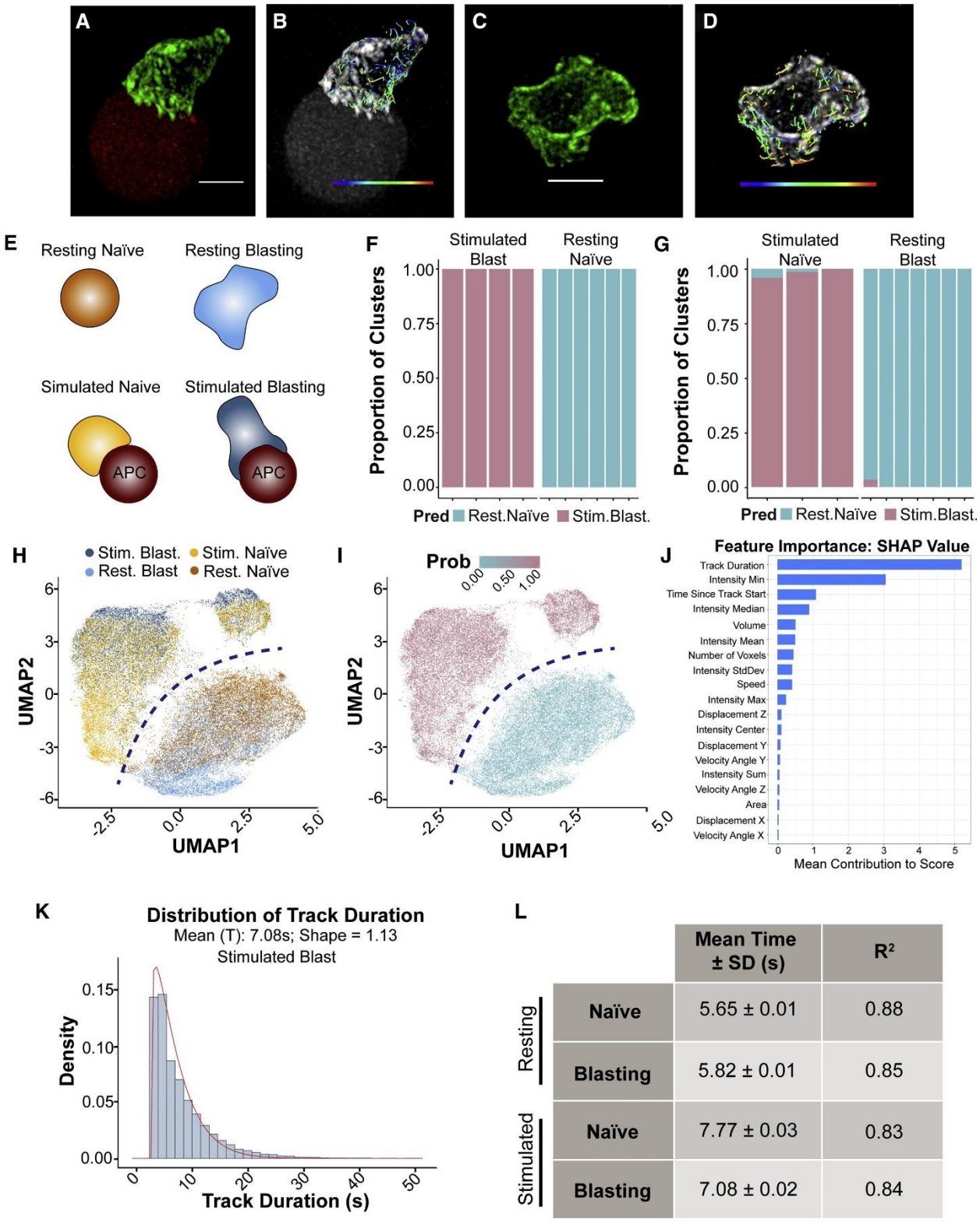
LaMDA Discovers that TCR Microclusters Reflect T-Cell Activation State (Resting versus Stimulated), Not the Functional Phenotype (Naïve versus Blasting)

As a proof of concept of the LaMDA pipeline, we investigated the dynamics of TCRs, one of the most critical molecules in adaptive immunity and their relationship to T-cell activation. TCRs specifically recognize rare agonist peptide-major histocompatibility complexes (pMHCs) among numerous self-pMHCs on the surface of the cell being surveyed to trigger adaptive immune responses and therefore are essential molecules to combat infection and cancer[277,278]. TCRs are

known to form microclusters on the surface of the T cell and represent a key mechanism toward understanding T-cell signaling and function[136,137,140,279-294].

To visualize the 4D TCR microcluster dynamics by LLSM, we either fused a green fluorescent protein (GFP) to the C-terminus of the CD3ζ chain of the TCRs or used an Alexa Fluor 488 (AF488)-conjugated anti-TCRβ Fab to fluorescently label the TCRs at the membrane of live primary 5C.C7 $CD4^+$ T cells. Cytosolic mCherry-transduced CH27 cells were used as the antigen-presenting cells. After fluorescent labeling of TCRs on T cells and loading agonist moth cytochrome C (MCC) peptide onto the antigen-presenting cells, both the T cells and antigen-presenting cells were added to the LLSM imaging chamber for 4D imaging using 488- and 561-nm lasers. We recorded videos (4–6 min) of TCR microclusters across the entire 3D cell surface (x and y resolution ~200 nm, z step 400 nm, ~1 s/cell volume; Figures 5.2A–5.2D and S5.1A–S5.1G) for four T-cell states: resting naïve, naïve stimulated by antigen-presenting cells, resting blasting, and blasting stimulated by antigen-presenting cells (Figure 5.2E).

We then tracked ~10,000 individual TCR microclusters on each T cell, measuring 36 parameters (including speed, direction, volume, intensity, area, location, and track duration; see Methods for details) for each microcluster across the videos (Figures 5.2B, 5.2D, and S5.1K). The data were pre-processed and an extreme gradient boosted (XGboost) decision tree ensemble[273] with logistic loss was built on 19 of these parameters (Figures 5.2F and S5.1H–S1K, referred to henceforth as the XGboost classifier) to classify TCR microclusters as microclusters from resting T cells or stimulated T cells. To avoid strong correlation within the same cells, and to prevent the confounding effect of "past stimulation" (i.e., naïve versus blast), the XGboost



L

		Mean Time ± SD (s)	R^2
Resting	Naïve	5.65 ± 0.01	0.88
	Blasting	5.82 ± 0.01	0.85
Stimulated	Naïve	7.77 ± 0.03	0.83
	Blasting	7.08 ± 0.02	0.84

Figure 5. 2 XGboost Binary Classifier Differentiates between T-Cell Signaling States.

Figure 5.2, Continued

(**A**) 3D rendering of a blasting $CD4^+$ T cell stably transduced with CD3z-GFP (green) as it encounters a CH27 cell stably transduced with cytosolic mCherry (red). (**B**) Dragon tails showing particle positions over the previous eight frames overlaid onto (**A**). Color bar represents velocity angle X from 0 (purple) to 180 (red). (**C**) 3D rendering of a blasting resting $CD4^+$ T cell stably transduced with CD3z-GFP (green). (**D**) Dragon tails showing particle positions over the previous eight frames overlaid onto (**C**). Color bar represents velocity angle X from 0 (purple) to 180 (red). (**E**) Diagram depicting the four cell states, resting naı¨ve, resting blasting, stimulated naïve, and stimulated blasting, graphically. (**F**) XGboost training results. Each bar represents an independent cell (resting naı¨ve cells n = 58,784 microclusters; stimulated blasting cells n = 97,237 microclusters). Pred, prediction. See also Figure S5.1K for number of microclusters in each cell. (**G**) XGboost testing results. Each bar represents an independent cell (stimulated naı¨ve cells, n = 38,809 microclusters; resting blasting cells, n = 60,116 microclusters). Pred, prediction. See also Figure S5.1K for number of microclusters in each cell. (**H**) UMAP of data from (**F–G**) with color indicating cell group. Boundary line represented in dark blue. (I) UMAP of data from (**F–G**) with color indicating probability of microcluster to be a stimulated blasting microcluster as predicted by XGboost binary classifier. Probability was indicated by color scale (0.00–1.00). Boundary line represented in dark blue. (**J**) SHAP values of each property of the microcluster in XGboost binary classifier. (**K**) Weibull distribution fitting of microcluster track duration from stimulated blasting cells to obtain mean track duration. See also Figures S5.2C–S5.2F. (**L**) Table of mean track duration for each cell group obtained by Weibull distribution fittings. Data are presented as mean ± standard deviation (SD) and the goodness of fit is indicated by R2 values. Scale bars, 5 mm.

classifier was first trained on microclusters from stimulated blast cells and resting naïve cells (Figure 5.2F). The classifier was later tested on stimulated naïve cells and resting blast cells (Figure 5.2G). Within the training set, a train-validate-test approach was utilized to avoid overfitting (see Methods; Figure S5.1I). The XGboost classifier consists of 150 decision trees, each of which selects up to three parameters to fit a logistic regression model. The weighted values from all 150 trees are used to classify each microcluster (Figure S5.1H).

This classifier allows us to identify which observable features of TCR dynamics can function as signatures of a cell's state and allows for classification. For example, we hypothesized that because TCR microclusters behave differently in response to a stimulation (e.g., TCR-pMHC interaction)[137,279,281,287], TCR microclusters switch from a "resting state" into a "stimulated state" and thus could be differentiated by machine learning. Indeed, in the internal test, the XGboost classifier could precisely differentiate between TCR microclusters on the resting naïve T cell from those on a stimulated blasting T cell with a validation accuracy of 100% and an internal test accuracy of 99.96% (Figure 5.2F). This suggests that our machine learning model can reliably predict the cell state, that is resting naïve versus stimulated blasting, by TCR microcluster dynamics alone. Accordingly, when a similar analysis was performed on two other T-cell states, stimulated naïve and resting blasting, we found large differences between TCR microcluster dynamics on stimulated and resting T cells, which allowed for accurate discrimination based on cell state by XGboost (Figure 5.2G). These results demonstrate that TCR microcluster dynamics are sufficient to predict cell state. Notably, however, the prediction of cell state is irrespective of prior antigen experience; the XGboost classifier could not distinguish between naïve and blasting T cells (Figures 5.2F and 5.2G). These results demonstrate that machine learning can precisely

predict T cell signaling states (resting state versus stimulated state) using surface TCR dynamics, but that from the perspective of the TCR dynamics, naïve and blasting T cells are indistinguishable. This observation is not unique to the XGboost classifier. TCR microclusters can also be segregated into two groups, resting-state TCR microclusters and stimulated-state TCR microclusters, using the dimensionality reduction method, UMAP[274] (Figures 5.2H and 5.2I, decision boundary in dark blue and Figure S5.2A).

Since the dynamic features of TCR microclusters represent a direct reflection of the T-cell signaling states, we next sought to use big data analysis techniques to investigate which biological features contributed to the observed differences. To further understand the features predicted by the XGboost classifier, we plotted the shapley additive explanations (SHAP) values[295,296] and found that track duration (i.e., the length of time a microcluster could be detected on the cell surface) was the most important feature (Figures 5.2J and S5.2B). While large SHAP values only directly indicate the features most informative to the decision trees, these features are often biologically relevant. Thus, we plotted the distribution of track duration for resting naïve, stimulated naïve, resting blasting, and stimulated blasting T cells and fitted three-parameter Weibull distribution models, a commonly used model for lifetime analyses[297] (Figures 5.2K and S5.2C–S5.2F). The mean duration "T" was derived from the Weibull distributions (see Methods) and describes the average lifetime of TCR microclusters on the surface of a T cell. We determined that antigen stimulation increased the mean duration of TCR microclusters significantly when compared with resting state, for both naïve and blasting T cells (Figures 5.2L and S5.2F). This implies a signaling pathway was initiated to alter global control of TCR microclusters on the surface.

Next, we characterized TCR microcluster dynamics using diffusion maps, a nonlinear dimensionality reduction technique that focuses on identifying the underlying manifold of the data and reveals diffusion-like behavior between different states[275,276]. We sampled 8,000 microclusters from each cell group and chose a subset of 7 selected features (mean intensity, minimum intensity, intensity sum over the surface, area, volume, speed, and track duration) based on the XGboost classifier feature importance (as previously shown in Figure 5.2J) and their mutual independence (see Methods). This diffusion map allowed us to graphically compare resting-state TCRs with stimulated-state TCRs (Figure 5.3A). We next estimated the local density of TCR microclusters on the diffusion map (Figure 5.3B) and derived the 3D pseudo-free energy surface (see Methods) by adapting a method commonly used in molecular simulation research[275] (Figure 5.3C). For easy visualization and comparison, these 3D energy wells were projected along dimension 2 to become 2D energy wells (Figure 5.3D). Assuming all other terms that contribute to energy are held constant, the depth of the energy well directly revealed the stability of the TCR microclusters on each cell (Figure 5.3D).

We found the stimulated-state TCR microclusters occupy a much deeper energy well than the resting-state TCR microclusters, on both blasting and naïve T-cell surfaces (Figure 5.3D), demonstrating that, assuming all other terms that contribute to energy are held constant, TCR stimulation stabilizes TCR microclusters on the cell surface. We also analyzed the diffusion maps to compare individual dynamic features between resting and stimulated TCR microclusters (Figures 5.3E–3H, middle panel). For track duration, we found clear separation between resting-state and stimulated-state TCR microclusters in the diffusion map (Figure 5.3E). Similar distinctions between resting-state and stimulated-state TCR microclusters were found for

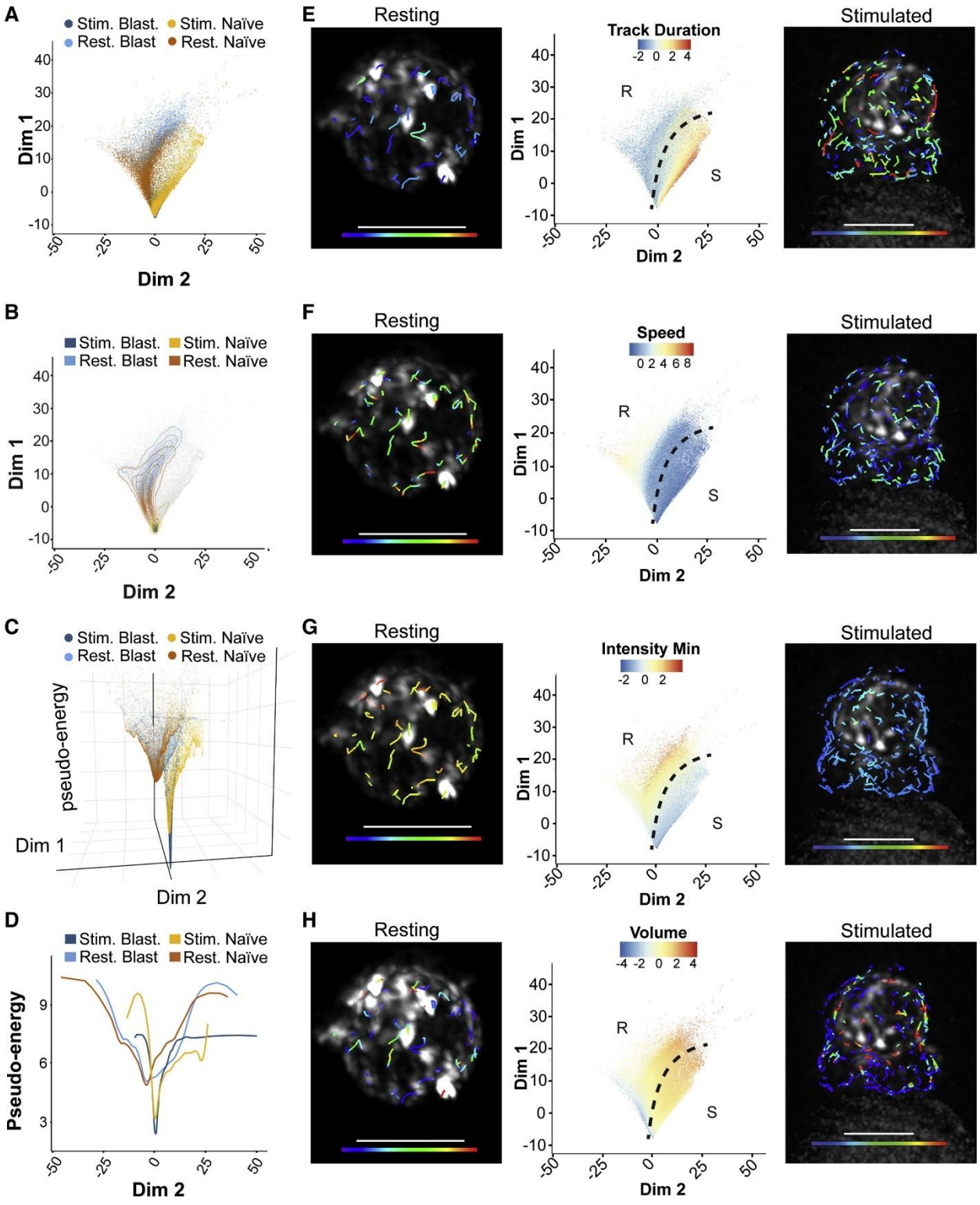


Figure 5. 3 Diffusion Maps Differentiate between T-Cell Signaling States

Figure 5. 3, Continued

(**A**) Diffusion map built from 7 selected features. Colors indicate cell group. (**B**) Density map created by estimating local density on the diffusion map (**A**). Colors indicate cell group. (**C**) Pseudo-energy map created from density map (**B**). Colors indicate cell group. (**D**) Projection of pseudo-energy map (**C**) along dimension 2. Colors indicate cell group. (**E**–**H**) Left: resting naı̈ve cell from Figure S5.1A overlaid with dragon tails showing particle positions over the previous eight frames are color coded to show particle track duration from 0 s (purple) to 25 s (red) (**E**), particle speed from 0.0 mm/s (purple) to 0.5 mm/s (red) (**F**), particle intensity minimum from 0 AU (purple) to 500 AU (red) (**G**), or particle volume from 0.0 mm2 (purple) to 0.2 mm2 (red) (**H**). Center: diffusion map from Figure 5.3A colored by normalized track duration (**E**), speed (**F**), intensity minimum (**G**), or volume (**H**). Approximate boundary line between resting (R) and stimulated (S) cells (see also Figure 5.3A) represented in black. Right: Stimulated blasting cell from Figure S5.1D overlaid with dragon tails showing particle positions over the previous eight frames are color coded to show particle track duration from 0 s (purple) to 25 s (red) (**E**), particle speed from 0.0 mm/s (purple) to 0.5 mm/s (red) (**F**), particle intensity minimum from 0 AU (purple) to 500 AU (red) (**G**), or particle volume from 0.0 mm2 (purple) to 0.2 mm2 (red) (**H**).

speed (Figure 5.3F) and minimum intensity (Figure 5.3G), whereas the differences in volume (Figure 5.3H) were less obvious. These analyses reinforced the reliability of our machine learning prediction. Spatially, we observed that the longest lasting microclusters (red) are not localized to the synapse as would be expected due to the TCR-pMHC bonds, but rather are dispersed throughout the whole cell (Figure 5.3E, right panel); similar global distributions were found for speed, minimum intensity, and volume (Figures 5.3F–5.3H, right panels), indicating that T cell activation is a global event.

Taken together, our data suggest that TCR microclusters directly reveal different signaling states of a T cell and support current biological understanding of TCR dynamics. It is known that TCR microcluster dynamics are closely linked with the actin network, which globally reorganizes with TCR-pMHC binding and signaling[298-305]. Accordingly, we observe that TCRs initially concentrated at the distal pole, quickly dispersed into small TCR microclusters and then continuously trafficked to the immunological synapse across the 3D spherical surface of the T cell (Figures S5.3A–S5.3D). Notably, after dividing the TCR microclusters between inside and outside synapse (Figures S5.3E–S5.3G), LaMDA was unable to identify differences between microclusters inside and outside the synapse (Figures S5.3H–S5.3J). This suggests that TCR triggering at the synapse propagates rapid, global re-organization of TCRs for effective antigen recognition and signaling. To our knowledge, such global, dynamic, and directional structural changes of TCR microclusters on the 3D T-cell surface had not yet been observed using conventional microscopy techniques or other biochemical or biophysical assays.

4D TCR Microcluster Dynamics Enable Ligand Discrimination

TCR ligand discrimination is essential for adaptive immunity. The dysfunction of TCR ligand discrimination can directly lead to cancer, infection, or autoimmunity. Moreover, TCR ligand discrimination displays two important characteristics: (1) high sensitivity—TCRs can recognize even a single agonist pMHC in the presence of abundant self-pMHCs[290,306]; and (2) high specificity—TCRs can discriminate between structurally similar peptides and elicit distinct immune responses[307-309]. Despite its critical importance, the molecular mechanism of TCR ligand discrimination remains controversial[277]. A common problem is that most existing mechanistic models were proposed based on in vitro studies that often cannot reliably predict physiological events in vivo. One example is that the high-affinity, slow off-rate, and high-potency (as defined by in vitro measurements) K5 peptide triggers attenuated in vivo T-cell responses[310], and such experimental data cannot be explained by the prevailing kinetic models that were built based on in vitro binding affinity and/or half-life[277,278,310]. In order to overcome the limitations imposed by in vitro studies, we applied our LaMDA approach to study TCR ligand discrimination to evaluate whether it can accurately predict physiological in vivo T-cell responses.

First, we confirmed that our experimental setup accurately reflects well-understood TCR biology by recapitulating previous observations of TCR's co-receptor CD4, a monomeric polypeptide that plays an important role in augmenting TCR signaling through associated tyrosine kinase Lck[311-316], and CD28, a co-stimulatory receptor required to fully activate the T cells without causing apoptosis[317-321]. As expected, we observed that CD4 blockade (and Lck inhibition) and CD28 blockade impairs entry into the full stimulated state and significantly reduces the stability of TCR microclusters (Figures S5.4A–S5.4N and S5.2F). These observations support the

conclusion that using LaMDA for observations of TCR dynamics, diverse signaling states can be identified without the need for complex biochemical or functional assays. Next, we measured the TCR microcluster dynamics stimulated by three structurally similar peptides K5, MCC, and 102S using LLSM (Figure 5.4B)[310]. We then applied the same XGboost classifier trained on MCC-stimulated versus unstimulated TCR microclusters to this dataset. The XGboost classifier predicted that ~89.3% and ~76.1% of the microclusters stimulated by variants K5 and 102S, respectively, are in the stimulated state (Figure 5.4A). This indicates that both K5 and 102S stimulations resulted in partial activation, consistent with previous in vivo studies, which showed that both K5 and 102S peptides resulted in attenuated immune responses compared with the MCC peptide[310]. Consistent with machine learning, we found that the Weibull-derived average lifetimes of 102S- and K5-stimulated microclusters were similar to those of MCC-stimulated microclusters but significantly different from those of resting microclusters (Figures 5.4C–5.4H and S5.2F).

To systematically study TCR ligand discrimination, we plotted a diffusion map of the TCR microclusters stimulated by the three structurally similar peptides K5, MCC, and 102S (Figures 5.4I and 5.4J). After converting the diffusion map to 3D (Figure 5.4K) and 2D (Figure 5.4L) energy wells, we were able to clearly visualize different activation states of T cells. Compared with the resting-state T cells without stimulation (blue), MCC-mediated stimulation (dark blue) induced the formation of the most stable TCR microclusters, while K5 (dark green) and 102S (green) resulted in less stable TCR microclusters. Our findings are well aligned with a previous in vivo study by Allison and colleagues showing that the MCC peptide, rather than the K5 and 102S, triggers the optimal intracellular signaling, cytokine production, and cell proliferation[310].

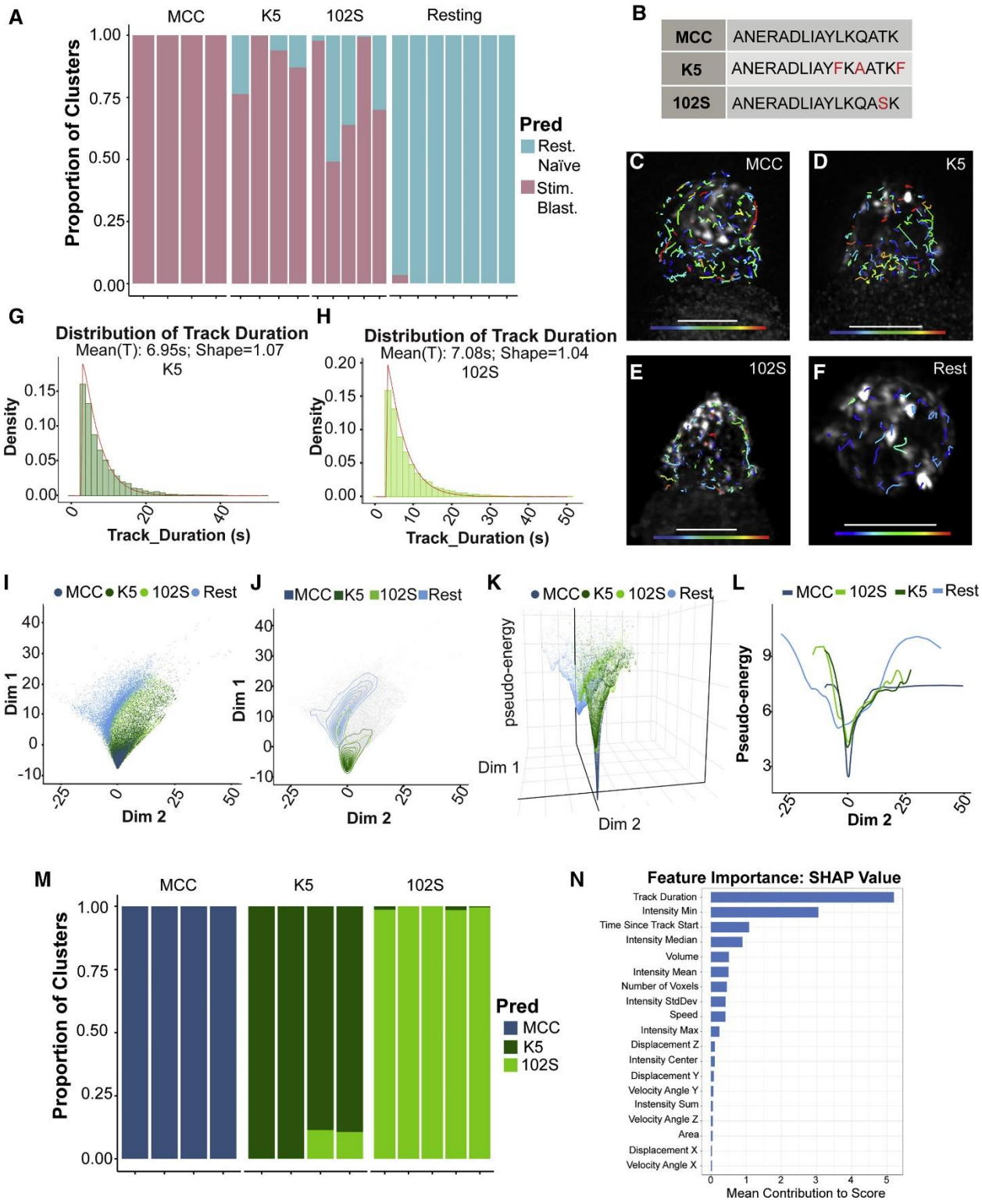


Figure 5. 4 TCR Ligand Discrimination by LaMDA

Figure 5. 4, Continued

(**A**) XGboost prediction results on microclusters from blasting cells stimulated with MCC (n = 97,237 microclusters), K5 (n = 24,786 microclusters), and 102S (n = 46,218 microclusters), and unstimulated resting cells (n = 58,784 microclusters). Each bar represents an independent cell. See also Figure S5.1K. (**B**) Table of peptide sequences for MCC, K5, and 102S. (**C**–**F**) 3D renderings of blasting cells stimulated with MCC (C), K5 (D), and 102S (**E**) and unstimulated resting blasting cells (**F**). All images are overlaid with dragon tails showing particle positions over the previous eight frames are color coded to show particle track duration (0–25 s). Scale bars, 5 mm. (**G** and **H**) Weibull distribution fitting of microcluster track duration from blasting cells stimulated with K5 (**G**) or 102S (**H**). See also Figure S5.2F. (**I**) Diffusion map built from the same selected features as Figure 5.3A, with colors indicating cell groups. (**J**) Density map created from (**I**). (**K**) 3D pseudo-energy map created from (**J**) of blasting cells stimulated with MCC, K5, and 102S peptides, and resting blasting cells, with colors indicating cell groups. (**L**) Projection of pseudo-energy map from (**K**) along dimension 2. (**M**) Separate XGboost classifier built to differentiate microclusters from cells stimulated with MCC, K5, or 102S. (**N**) SHAP values of each feature as used in XGboost binary classifier in (**M**).

Additionally, to fully test whether we can apply machine learning to TCR ligand discrimination, we trained another XGboost softmax classifier to distinguish between the TCR microclusters on T cells stimulated by MCC, K5, and 102S. We found that this alternative machine learning approach could be used to precisely differentiate between the TCR microcluster organization following stimulation by the three structurally similar peptides with ~96% accuracy (for training, validation, and test datasets) (Figures 5.4M and 5.4N). We expect our LaMDA pipeline can be extended to a wide range of ligand discrimination systems and detect subtle yet important dynamic differences between cell states, which provides a more physiologically relevant approach to study immune molecules and cells in cancer, infection, and autoimmunity.

Discussions

We developed a modular pipeline, LaMDA, that combines high resolution 4D LLSM data with machine learning and dimensionality reduction techniques to analyze TCR microcluster dynamics and predict T-cell signaling states. In contrast to many studies that emphasize differences between naïve and blasting T cells, our LaMDA pipeline identified that the dynamics and distribution of TCR microclusters were primarily determined by the T-cell activation state (resting versus stimulated), rather than by prior antigen exposure (naïve versus blasting) (Figures 5.2 and 5.3). Using LaMDA, we also demonstrated that, assuming all other terms that contribute to energy are held constant, TCR-pMHC ligation stabilizes TCR microclusters globally across the entire T-cell surface (Figure S5.3), which has largely been overlooked in the past. Finally, we demonstrated that the same pipeline can be extended to study ligand discrimination (Figure 5.4).

In this study, we found that both K5- and 102S-mediated stimulation of T cells resulted in partial activation when compared with MCC (Figure 5.4). Our observations complement in vivo

studies that have shown the K5 peptide results in attenuated immune responses compared with the MCC peptide[310] but stand in contrast to in vitro studies that have previously shown that K5 is a super-agonist with the highest TCR affinity and results in the strongest activation of T cells among three peptides[310,322-324]. This suggests that unlike common in vitro approaches, our LaMDA pipeline may be able to predict in vivo peptide potency in activating T cells and could be extended toward the development of peptide vaccines to treat infection, cancer, and autoimmunity[325,326] or be used to study thymic education and/or peripheral tolerance, two very important topics in T-cell biology.

LLSM has been widely used across many biological fields to address a variety of questions[130,142,302,305,327-335]. While LLSM has revealed important insights, current analysis approaches are not capable of considering all of the collected data to maximize the biological understanding or conclusions that could be drawn from the data. Most analysis approaches employed for investigating cell biology primarily focus on single features, such as intensity, colocalization, autocorrelation, diffusion rates, resolution, or directional analyses[316,336]. The LaMDA pipeline is able to take full advantage of the 4D (x, y, z, and time) data intrinsically provided by LLSM and utilizes it in a myriad of advanced high-dimensional analysis tools, such as dimensionality reduction techniques and machine learning. However, LLSM is a low-throughput method regarding cell number, and if each cell is treated as a single data point, sufficient numbers for high-dimensional analysis methods cannot be reached. Thus, by considering individual molecules (e.g., TCRs) on the cell surface as opposed to single cells captured by LLSM, we are able to provide sufficient data points in a high-throughput manner to utilize these methods.

By enacting the paradigm shift from single cell to single molecule in LLSM, the utility of data produced can be maximized.

In addition to surface molecules, this pipeline can still be used to track intracellular or recycling molecules. The analysis pipeline presented here could simply be used to address different questions. For example, LLSM has already been used to image and track T-cell granzymes, which are cell-death-inducing molecules packaged in intracellular vesicles and delivered to the surface of a target cell[305]. Our analysis pipeline therefore could analyze the tracks of these vesicles and address trafficking rates or directions. While any molecular label can be easily implemented, a large number of fluorescently detectable molecular units must exist in the biological system, highlighting an important consideration in implementing LaMDA for other studies.

Finally, LaMDA was intentionally designed to utilize existing tools and algorithms to perform complex multi-dimensional analysis, thereby enabling easy accessibility for any biologist without the need for familiarity with data science techniques. Moreover, while we used Imaris for feature extraction and an XGboost classifier for machine learning, this pipeline is highly modular and can be adapted for a variety of biological systems by incorporating alternative algorithms. Other machine learning techniques with different algorithms (e.g., XGboost or supportive vector machine), architectures (e.g., decision trees or neural network) or tasks (e.g., classification, segmentation, or detection) can be incorporated to substitute the XGboost classifier. In addition, multiple dimensionality reduction techniques are applicable (e.g., tSNE, PCA, etc.), and should be chosen based on desired purpose. Similarly, features do not have to be pre-defined and extracted from a software such as Imaris; rather, researchers could engineer context-specific features useful

to their own research questions. Therefore, we anticipate broad usage of LaMDA to maximize biological understanding from LLSM data.

Material and Methods

Mice

5C.C7 TCR-transgenic RAG2 knockout mice in B10.A background aged 8-9 weeks were used in this study. Animals of both sexes were used and the influence of sex was not considered in the analysis of the data. All experiments were conducted according to protocols approved by the Institutional Animal Care and Use Committee of the University of Chicago.

Cell Culture

Cytosolic mCherry-transduced CH27 cells were a gift generously donated by Dr. Enfu Hui. To obtain blasting T cells, 5C.C7 mouse spleen was harvested and run through a 70 μm cell strainer with warm complete RPMI. Splenocytes were resuspended in 5 mL of RBC Lysis Buffer (Life Technologies) for 5 min, washed three times, and resuspended in 5 mL of complete RPMI. MCC peptide (ANERADLIAYLKQATK; 10 μM) was added to stimulate T-cell proliferation, and recombinant mouse IL-2 (100 U/mL, Sigma-Aldrich) was added the following day. Blasting T cells were used on days 6–10 after peptide pulsing. mCherry-CH27 cells and 5C.C7 blasting T cells were both maintained in complete medium (RPMI 1640 supplemented with 10% [v/v] FBS, 1% [v/v] Pen/Strep, L-glutamine [2 mM], 2-mercaptoethanol [50 μM]).

Cell Preparation

The mCherry-CH27 cells were used as antigen presenting cells for imaging experiments. Dead mCherry-CH27 cells were first removed by Ficoll-Paque Plus (GE Healthcare, cat: 17-1440-03) density gradient centrifugation (centrifuged at 930 × g for 10 min at 4 °C, acc/dec: SLOW/SLOW). Live mCherry-CH27 cells were washed three times with complete medium, and then incubated with 10 μM K5 (ANERADLIAYFKAATKF), MCC (ANERADLIAYLKQATK) or 102S (ANERADLIAYLKQASK) for 3 h at 37 °C and 5% CO_2. Peptide-pulsed mCherry-CH27 cells were then washed three times and resuspended in imaging media (phenol red-free Leibovitz's L-15 medium supplemented with 10% [v/v] FBS, 1% [v/v] Pen/Strep, L-glutamine [2 mM]) for use.

Naïve T cells were prepared by negative isolation using 5C.C7 transgenic mouse spleens. Mouse splenocytes were prepared by Ficoll-Paque Plus (GE Healthcare, cat: 17-1440-03) density gradient centrifugation (centrifuged at 930 × g for 10 min at 4 °C, acc/dec: SLOW/SLOW) to remove dead cells. Naive $CD4^+$ T cells were then harvested by negative isolation (MojoSort Mouse CD4 T Cell Isolation Kit, BioLegend, cat: 480033). Cells were washed three times with complete medium and incubated with Alexa488-labeled anti-TCRβ (2 μg; clone H-57, Biolegend) antibody Fab for 30 min at 37 °C and 5% CO_2. After three washes, naïve T cells were resuspended in imaging media for use.

Day 6-10 blasting T cells were prepared by Ficoll-Paque Plus (GE Healthcare, cat: 17-1440-03) density gradient centrifugation (centrifuged at 930 × g for 10 min at 4 °C, acc/dec: SLOW/SLOW) to remove dead cells. Live blasting T cells were washed three times with complete medium and incubated with Alexa488-labeled anti-TCRβ (2 μg; clone H-57, Biolegend) antibody

Fab for 30 min at 37 °C and 5% CO_2. After three washes, blasting T cells resuspended in imaging media for use.

CD3ζ-GFP Transduction

Primary 5C.C7 T cells were retrovirally transduced with CD3ζ-GFP according to a previously published method[337]. Briefly, calcium phosphate precipitation was used to transfect MIG-CD3ζ-GFP vector into ecotropic platinum-E retroviral packaging cells. Supernatant containing virus was harvested after 48 and 72 h and filtered through 0.2 μm cellulose acetate membrane. Splenocytes isolated from 5C.C7 mice cultured in complete RPMI were stimulated with anti-CD3ε mAb (5 μg/mL; Clone 145-2C11, University of Chicago Monoclonal Antibody Facility), anti-CD28 mAb (0.5 μg/mL; Clone 37.51, Biolegend), and recombinant human IL-2 (40 U/mL; Peprotech). A 6-well plate was coated with Retronectin (12.5 μg/mL; Clontech) in PBS at 4 °C overnight, then centrifuged for 90 min at 3,000 x g with 2 mL of viral supernatant. Day 1 activated splenocytes were transferred to plate with viral supernatant, protamine sulfate (4 μg/mL) was added, and plate was centrifuged at 800 × g for 90 min. After 24 h, medium was replaced with fresh viral supernatant containing protamine sulfate (4 μg/mL), and plate was centrifuged at 800 × g for 90 min. After 16 h, transduction efficiency was determined by examining GFP fluorescence using flow cytometry.

Fab Preparation

All Fabs used in this study were prepared using a Micro Fab Preparation Kit (Thermo Fisher Scientific). Briefly, anti-TCRβ, anti-CD4, and anti-CD28 whole antibodies were prepared with desalting column, then digested with papain, a nonspecific thiol-endopeptidase, for 6 h on tabletop shaker at 37 °C. Digested Fabs were purified according to kit instructions. Fab purification was confirmed with SDS-PAGE gel electrophoresis.

Signaling Perturbation Assays

For PP2 assay, day 6-10 blasting T cells were pre-incubated with PP2 (10 μM; 4-Amino-5-(4-chlorophenyl)-7-(t-butyl)pyrazolo[3,4-d]pyrimidine; Sigma-Aldrich) for 1 h at 37 °C and 5% CO_2. Pre-treated cells were imaged at 37 °C and 5% CO_2 in the presence of 10 μM PP2.

For CD4 and CD28 blockade, an anti-CD4 (clone Gk1.5, BioLegend) or anti-CD28 (clone 37.51, BioLegend) Fab was prepared and purified with Micro Fab Preparation Kit (Thermo Fisher Scientific). Day 6-10 T cells were pre-incubated for 1 h at 37 °C and 5% CO_2 with anti-CD4 Fab (2 μg/mL) or anti-CD28 Fab (4 μg/mL) then imaged in the presence of 2 μg/mL Fab.

Lattice Light-sheet Microscopy

Version 2 of the Lattice Light Sheet Microscope (3i) was used for 4D imaging experiments. The LLSM was aligned daily according to manufacturer's instructions. LLSM bath was filled with imaging media and warmed to 37 °C. Round coverslips (5 mm) were prepared prior to imaging by incubation with 0.1% [w/v] solution of Poly-L-Lysine for 30 min. Poly-L-Lysine was aspirated, and coverslips were allowed to dry. To prepared coverslips were added 200,000 peptide-pulsed mCherry-CH27 cells. Cells were allowed to settle for 10 min before adhering to sample holder and placing in LLSM bath. T cells (200,000) were added dropwise to the LLSM bath above the

coverslip and imaged immediately. Imaging was conducted with dither set to 3 and 10 ms exposures. Z-steps (60) were collected with a 0.4 μm step size. Cells were imaged for no more than one hour before exchanging with fresh cells.

Image Processing

Data were deskewed and deconvolved using LLSpy (cudaDeconv) software, used under license from Howard Hughes Medical Institute, Janelia Research Campus; collected point spread functions (PSF, collected under imaging conditions) were used to deconvolve, and 20 iterations were conducted. Bleach correction was conducted in ImageJ using histogram matching. Deskewed, deconvolved, and debleached videos were imported into Imaris (Bitplane) for rendering and tracking. Tracking TCR microclusters was conducted with the "surfaces" module using the Autoregressive Motion Expert tracking method; any two consecutive surfaces were allowed to move a maximum distance of 0.5 μm and disappear for no more than 3 frames to be considered on one track; watershed was not enabled. The surfaces module was also used to create a surface of the antigen presenting cell, and a distance map was made extending outward from the antigen presenting cell surface. Surface module statistics were exported for every TCR surface. Thus, for all following sections of this Methods section, "surfaces" or "surface modules" refer to the segmentation and tracking of TCR microclusters conducted by Imaris.

Data Preprocessing

For each cell, the raw surface module statistics from the TCR microcluster tracks were imported and processed in R as follows. All non-numerical or missing statistics values were removed. All statistical variables were then combined into a dataframe with rows as unique surface modules and columns as the different statistical variables, including track-specific variables. For the initial parameter extraction from Imaris, 36 variables out of 134 available parameters were collected (see Data Sharing and Availability for complete list). These variables were selected by refraining from fitting a certain shape to the surface unit, thus removing all shape-specific features such as "BoundingBoxAA Length" or "Ellipsoid axes". Additionally, only first-order measurements were included where first and second order are available, so as to minimize errors in tracking; for example, displacement XYZ was included, but not displacement^2, as any error in connecting the two surface units along the track would be propagated. As a separate example, all acceleration measurements were also excluded, as they depend on velocity measurements, which depend on the displacement. This would compound any "displacement error". Redundant features were preserved (such as standard deviation vs. variation of speed along track), as the XGboost classifier would only select one parameter to analyze at a time, thereby self-excluding the effect of redundant parameters. To normalize the dynamic range, all intensity-related variables (e.g., Mean Intensity, Max Intensity, Min Intensity, etc.) and all size-related variables (e.g., Area, Volume, Number of Voxels, etc.) were then log transformed.

XGboost Decision Trees Ensemble Binary Classifier

To build the classifier and make predictions, all data were further processed as follows. For each cell, 19 out of the 36 variables were selected as features of interest (see Data Sharing and Availability for complete list). In this subset, all track-specific (as opposed to surface-specific) parameters were removed in order to focus the decision tree on each surface (or TCR microcluster), rather than each track, as an individual unit. The only exception was Track Duration, which is important for lifetime distribution studies. For a series of surface modules that were assigned to the same track by Imaris, the track duration feature values were defined as the temporal length of that track (i.e., the same value for surface modules of the same track). In addition, position XYZ was removed, as the position of a surface is always relative to the orientation of the cell in that particular video, therefore irrelevant to compare across videos. The chosen statistics from all microclusters on all cells were then pooled for each group to avoid bias on the biological variation from cell to cell. Surface modules with Mean Intensity (I_{mean}) values lower than mean(I_{mean})-1.5×IQR(I_{mean}) were removed as they are likely to be noise rather than a true surface module. Then, all statistics were standardized:$\frac{x-\bar{x}}{SD(x)}$.

The processed data were then used to train a binary classifier to differentiate between microclusters from resting cells and microclusters from stimulated cells as follows. Of the four groups of cell types, microclusters from stimulated blast T cells and resting naïve T cells were chosen to train the classifier, which was later tested on the stimulated naïve cells and resting blast cells. This division of cell types was chosen for two reasons: (1) to avoid any strong correlation within the same cells; and (2) to test against the confounding effect of "past stimulation" (i.e., naïve vs. blast). To build the classifier, a "train-test-validate" approach was used to avoid

overfitting (Supplementary Figure S5.1I). Thus, the processed data of stimulated blast T cells and resting naïve T cells were divided into a training set (75%; 117,066 surface modules), validation set (17.5%; 27,316 surface modules) and internal test set (7.5%; 11,705 surface modules). The binary classifier was built as a XGBoost logistic regression decision tree ensemble, using the caret package (v.6.0) in R, with 5-fold cross-validation and the following parameters: nrounds = 150; max_depth = 3; eta = 0.4; gamma = 0; colsample_bytree = 0.8; min_child_weight = 1; and subsample = 1. Feature importance was then assessed using the SHAP values extracted from the xgb.plot.shap function. All subsequent plots were constructed using the ggplot2 (v.3.2.0) and the ggpubr (v.0.2.1) package in R. Finally, the model was tested on every surface module (i.e., TCR microclusters) from new data (stimulated naïve and resting blast categories) to predict whether these TCR microclusters were from stimulated blast T cells or from resting naïve T cells. The final label was assigned to be the one with greater predicted probability (i.e., either stimulated blasting T cells or resting naïve T cells).

XGboost Decision Tree Ensemble Softmax Classifier

To differentiate multiple ligands, a separate multi-class XGboost classifier was constructed as follows. For each cell, the same 19 out of 36 variables as used for the initial XGboost Decision Tree Ensemble Binary Classifier were selected; however, a 20^{th} variable, average distance to antigen presenting cell, was added. The last feature was not applicable to the binary classifier above because the resting cells do not encounter antigen presenting cells, but it is applicable to all classes here. Surface modules with Mean Intensity (I_{mean}) values lower than mean(I_{mean})-

1.5×IQR(I_{mean}) were removed since they are likely to be noise rather than a true surface module. Then, all statistics were standardized: $\frac{x-\bar{x}}{SD(x)}$.

The processed data of microclusters from blasting T cells stimulated by MCC, 102S, and K5 were pooled and divided into a training set (75%; 126,254 surface modules), validation set (17.5%; 29,460 surface modules) and test set (7.5%; 12,623 surface modules). The softmax classifier was built as a XGBoost decision tree ensemble using softmax probabilities as the objective, using the caret package (v.6.0) in R, with 5-fold cross-validation and the following parameters: nrounds = 500; max_depth = 6; eta = 0.3; gamma = 0; colsample_bytree = 1; min_child_weight = 1; and subsample = 1. The final model was used to re-predict the probability of every surface module from the blasting T cells stimulated by MCC, 102S, and K5. The final label was assigned to be the peptide (MCC, 102S, or K5) with greatest predicted probability. Results were plotted using the ggplot2 (v.3.2.0) and the ggpubr (v.0.2.1) package in R.

Weibull Distribution

Weibull distribution[297] is a commonly used statistical distribution that describes lifetime distribution. The three-parameter Weibull model we used is of the form:

$$f(T) = \frac{\beta}{\eta}\left(\frac{T-T_0}{\eta}\right)^{\beta-1} \exp\left(-\frac{T-T_0}{\eta}\right)^{\beta}$$

where:

$$T \geq 0;\ \eta > 0;\ \beta > 0$$

A Weibull distribution was used to fit the raw track duration values for each group using the WeibullR (v.1.0.10) package in R. The location (T_0), shape (β), and scale (η) parameters were extracted directly from the model. The mean life-time values were calculated as:

$$\bar{T} = T_0 + \eta \cdot \Gamma\left(\frac{1}{\beta} + 1\right)$$

UMAP

Pre-processed data were separately processed as an independent validation of the XGboost classifier. For each cell, the same 19 out of the 37 variables as used for the XGboost classifier were selected as features of interest (see Data Sharing and Availability for complete list). The chosen statistics from all cells of all groups were then pooled for each group to avoid bias on the biological variation from cell to cell. Surface modules with Mean Intensity (I_{mean}) values lower than mean (I_{mean})-1.5×IQR (I_{mean}) were removed since they are likely to be noise rather than a true surface module. Then, all statistics were then standardized: $\frac{x-\bar{x}}{SD(x)}$. Then data of the groups of interest were then pooled and used to build a UMAP using the uwot (v.0.1.3) package in R with parameters: min_dist = 0.1; n_neighbors = 25; all others set to default. All subsequent plots were constructed using the ggplot2 (v.3.2.0) and the ggpubr (v.0.2.1) package in R.

Diffusion Maps

To construct the diffusion maps, pre-processed data were further processed similarly to the above UMAP, except that only seven variables were selected (see link in Data and Software Availability for complete list). These seven variables were selected from the distinguishing features identified by both the supervised XGboost classifier and the unsupervised UMAP. First,

the six features from the SHAP value analysis were chosen. However, “Time since Track Start” and “Track Duration” are partially redundant. While “Track Duration” indicates the stability of the microcluster on the surface of the T cell, “Time since Track Start” simply indicates the order the microclusters appear along this track, and therefore has less physical meaning. Thus, “Track Duration”, was selected and “Time since Track Start” was excluded. In addition, “Mean Intensity” was selected over “Median Intensity”, since the former is a more common analysis metric. Finally, “Area” and “Speed” were also selected, as they encode independent information that have not been captured by the initial 5 features; for instance, surface units with the same volume can have different shapes, thereby lending to different surface area.

To conserve computational resources, a subset of 8,000 surface modules were randomly sampled from every group. Data from four core groups, including stimulated blast T cells, stimulated naïve T cells, resting naïve T cells, and resting blast T cells, were pooled and used to build the diffusion maps using the diffusionMap (v.1.1.0.1) package in R with default parameters. Nystorm out-of-sample extension was then used to estimate the diffusion coordinates of surface modules of other groups. All subsequent plots were constructed using the ggplot2 (v.3.2.0) and the ggpubr (v.0.2.1) package in R.

Pseudo-Energy Plot

For each group, the diffusion coordinates of the 8,000 sampled surface modules were used to build its pseudo-energy plot. The local probability density, $\hat{p}(\vec{x})$, at each data point on the diffusion map was estimated using the ks-package (1.11.5) in R. The density values from the diffusion map embedding coordinates can be used to derive Free Energy Surface as follows:

$$\beta G(\vec{x}) = -\ln \hat{p}(\vec{x}) + const$$

where $\beta = \frac{1}{k_B T}$; G is the Gibbs free energy, and $\hat{p}(\vec{x})$ is the estimated local probabilistic density on the diffusion map. Here, we used $-\ln \hat{p}(\vec{x})$ as a pseudo-energy since we cannot experimentally determine the constant term, but this pseudo-energy should be linearly related to the Gibbs free energy and thus gives a good representation of the free energy surface[275].

Supplementary Figures

Supplementary figures for this chapter are included in Appendix A: Supplementary Figures.

CHAPTER 6: CONCLUSION AND FUTURE RECOMMENDATION

Prior to the work in this thesis, fundamental questions in T cell differentiation remain unanswered: 1) EZH2's role in Tfh fate commitment; 2) differences of T_{EX} differentiation in Tumor vs Chronic infection; 3) CAR-T cells post-infusion differentiation. In addition, there was a lack of high-dimensional quantitative analysis and machine learning pipeline for TCR signaling studies. This thesis work answered those questions and addressed the issue by taking advantage of advanced bulk/single-cell sequencing and imaging technologies.

Firstly, we found that EZH2-mediated H3K27me3 modifications regulate Tfh differentiation preferentially in the early fate commitment, but not late Tfh differentiation or memory Tfh maintenance. We also revealed that EZH2 regulation was achieved through stabilization of the chromatin accessibility of a cluster of genes that are important for T_{FH} fate commitment, particularly the master transcription factor Bcl6.

Secondly, through single-cell RNA-sequencing and single-cell ATAC-sequencing, we revealed that terminal effector-associated gene signatures/loci accessibility were preferentially enriched in Tex cells in chronic viral infection, while stem/memory-related genes and their loci accessibility more enriched in Tex subsets in tumor. In addition, we demonstrated that the disease-specific T cell exhaustion programs lead to different responsiveness of Tex cells to PD-L1 immune checkpoint blockade (ICB).

Thirdly, through longitudinal, single-cell multi-omic sequencing of YESCARTA CAR-T cells in patients with DLBCL, we characterized the heterogeneity of both CAR-T and endogenous T cells. We also found post infusion differential enrichment of effector memory/effector $CD8^+$ CAR-T cells and regulatory CAR-T cells in responders and non-responders respectively.

Moreover, we discovered using deep learning an additional IRF7-centered regulatory module that can successfully predict clinical response.

Furthermore, we developed an end-to-end pipeline that combines high resolution 4D LLSM data with machine learning and dimensionality reduction techniques to analyze TCR microcluster dynamics and predict T-cell signaling states. We accurately differentiated TCR microclusters on stimulated cells from unstimulated cells, precisely predicted attenuation of T-cell signaling after CD4 and CD28 receptor blockades, and reliably discriminated between structurally similar TCR ligands.

Taken together, the work in this thesis also demonstrates the power of single-cell sequencing and imaging technologies in enabling complex, dynamic, and deep-learning-based research on the molecular regulatory mechanisms and signaling patterns involved in T cell differentiation. These technologies continue to evolve rapidly, and they will be increasingly indispensable in future research on T cell differentiation.

To further the understanding of the differentiation and signaling processes discussed in this thesis work, future studies could focus on the following questions/issues with the following recommended single-cell sequencing/imaging approaches:

1. How is the Tfh/Th1 differentiation bifurcation and fate commitment process spatiotemporally regulated in the lymph nodes during acute infections? Single-cell RNA-seq and single-cell ATAC-seq can be applied on longitudinal samples of $CD4^+$ antigen-specific T cells in an acute infection model such as LCMV Armstrong. Spatial transcriptomic profiling technology can also be coupled for further validations.

2. Can the differences in transcriptional regulatory network in Tex cells from tumor vs chronic infection also affect other checkpoint blockade immunotherapies (CBI) such as anti-CTLA4 and Anti-LAG3? If so, how do these differences inform on disease-specific modifications to the CBI? Single-cell RNA-seq and single-cell ATAC-seq combined with deep learning could be applied with a study design similar to that of Chapter 3 but using other CBI.
3. Is the YESCARTA (CD28-costimulation) CAR-T cells heterogeneity and differentiation generalizable to the KYMRIAH (4-1BB-costimulation) CAR-T cells? How about third generation CAR-T cells? Paired single-cell RNA-seq, single-cell TCR-seq and total-seq combined with deep learning could be applied with a study design similar to that of Chapter 4 but using KYMRIAH CAR-T cells.
4. Do CARs form microclusters on CAR-T cell surface? If so, do they lead to tonic signaling in the CAR-T cell? Do CARs interact with the TCR microclusters on the CAR-T cell surface? Will potential CAR-TCR interactions augment or diminish CAR signaling? Lattice light-sheet microscopy and super-resolution microscopy can be combined to image CAR-T cells with and without targets. Multidimensional analyses and deep-learning based segmentation, tracking, and analyses can then be applied to the images taken.

REFERENCES

1 Janeway, C. A., Jr. & Medzhitov, R. Innate immune recognition. *Annu Rev Immunol* **20**, 197-216, doi:10.1146/annurev.immunol.20.083001.084359 (2002).

2 Matzinger, P. The Danger Model: A Renewed Sense of Self. *Science* **296**, 301-305, doi:10.1126/science.1071059 (2002).

3 Nemazee, D. Receptor selection in B and T lymphocytes. *Annu Rev Immunol* **18**, 19-51, doi:10.1146/annurev.immunol.18.1.19 (2000).

4 Gray, D. Immunological Memory. *Annual Review of Immunology* **11**, 49-77, doi:10.1146/annurev.iy.11.040193.000405 (1993).

5 Farber, D. L., Netea, M. G., Radbruch, A., Rajewsky, K. & Zinkernagel, R. M. Immunological memory: lessons from the past and a look to the future. *Nature Reviews Immunology* **16**, 124-128, doi:10.1038/nri.2016.13 (2016).

6 Zhang, N. & Bevan, M. J. CD8(+) T cells: foot soldiers of the immune system. *Immunity* **35**, 161-168, doi:10.1016/j.immuni.2011.07.010 (2011).

7 Wong, P. & Pamer, E. G. CD8 T CELLRESPONSES TOINFECTIOUSPATHOGENS. *Annual Review of Immunology* **21**, 29-70, doi:10.1146/annurev.immunol.21.120601.141114 (2003).

8 Swain, S. L., Mckinstry, K. K. & Strutt, T. M. Expanding roles for CD4+ T cells in immunity to viruses. *Nature Reviews Immunology* **12**, 136-148, doi:10.1038/nri3152 (2012).

9 Zhu, J. & Paul, W. E. CD4 T cells: fates, functions, and faults. *Blood* **112**, 1557-1569, doi:10.1182/blood-2008-05-078154 (2008).

10 Vignali, D. A. A., Collison, L. W. & Workman, C. J. How regulatory T cells work. *Nature Reviews Immunology* **8**, 523-532, doi:10.1038/nri2343 (2008).

11 Sallusto, F. Heterogeneity of Human CD4+T Cells Against Microbes. *Annual Review of Immunology* **34**, 317-334, doi:10.1146/annurev-immunol-032414-112056 (2016).

12 Kanno, Y., Vahedi, G., Hirahara, K., Singleton, K. & O'Shea, J. J. Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. *Annu Rev Immunol* **30**, 707-731, doi:10.1146/annurev-immunol-020711-075058 (2012).

13 Zhu, J., Yamane, H. & Paul, W. E. Differentiation of Effector CD4 T Cell Populations. *Annual Review of Immunology* **28**, 445-489, doi:10.1146/annurev-immunol-030409-101212 (2010).

14 Ivanov, I. I. *et al.* The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory IL-17+ T Helper Cells. *Cell* **126**, 1121-1133, doi:10.1016/j.cell.2006.07.035 (2006).

15 Glimcher, L. H. Trawling for treasure: tales of T-bet. *Nature Immunology* **8**, 448-450, doi:10.1038/ni0507-448 (2007).

16 Szabo, S. J. *et al.* A Novel Transcription Factor, T-bet, Directs Th1 Lineage Commitment. *Cell* **100**, 655-669, doi:10.1016/s0092-8674(00)80702-3 (2000).

17 Zheng, W.-P. & Flavell, R. A. The Transcription Factor GATA-3 Is Necessary and Sufficient for Th2 Cytokine Gene Expression in CD4 T Cells. *Cell* **89**, 587-596, doi:10.1016/s0092-8674(00)80240-8 (1997).

18 Nurieva, R. I. *et al.* Bcl6 Mediates the Development of T Follicular Helper Cells. *Science* **325**, 1001-1005, doi:10.1126/science.1176676 (2009).

19 Yu, D. *et al.* The Transcriptional Repressor Bcl-6 Directs T Follicular Helper Cell Lineage Commitment. *Immunity* **31**, 457-468, doi:10.1016/j.immuni.2009.07.002 (2009).

20 Johnston, R. J. *et al.* Bcl6 and Blimp-1 Are Reciprocal and Antagonistic Regulators of T Follicular Helper Cell Differentiation. *Science* **325**, 1006-1010, doi:10.1126/science.1175870 (2009).

21 Crotty, S. A brief history of T cell help to B cells. *Nature Reviews Immunology* **15**, 185-189, doi:10.1038/nri3803 (2015).

22 Kim, C. H. *et al.* Subspecialization of Cxcr5+ T Cells. *Journal of Experimental Medicine* **193**, 1373-1382, doi:10.1084/jem.193.12.1373 (2001).

23 Schaerli, P. *et al.* Cxc Chemokine Receptor 5 Expression Defines Follicular Homing T Cells with B Cell Helper Function. *Journal of Experimental Medicine* **192**, 1553-1562, doi:10.1084/jem.192.11.1553 (2000).

24 Breitfeld, D. *et al.* Follicular B Helper T Cells Express Cxc Chemokine Receptor 5, Localize to B Cell Follicles, and Support Immunoglobulin Production. *Journal of Experimental Medicine* **192**, 1545-1552, doi:10.1084/jem.192.11.1545 (2000).

25 Choi, S., Youn *et al.* ICOS Receptor Instructs T Follicular Helper Cell versus Effector Cell Differentiation via Induction of the Transcriptional Repressor Bcl6. *Immunity* **34**, 932-946, doi:10.1016/j.immuni.2011.03.023 (2011).

26 Choi, Y. S. *et al.* Bcl6 Expressing Follicular Helper CD4 T Cells Are Fate Committed Early and Have the Capacity To Form Memory. *The Journal of Immunology* **190**, 4014-4026, doi:10.4049/jimmunol.1202963 (2013).

27 Ditoro, D. *et al.* Differential IL-2 expression defines developmental fates of follicular versus nonfollicular helper T cells. *Science* **361**, eaao2933, doi:10.1126/science.aao2933 (2018).

28 Tubo, N. J. *et al.* Single Naive CD4+ T Cells from a Diverse Repertoire Produce Different Effector Cell Types during Infection. *Cell* **153**, 785-796, doi:10.1016/j.cell.2013.04.007 (2013).

29 Fazilleau, N., Mcheyzer-Williams, L. J., Rosen, H. & Mcheyzer-Williams, M. G. The function of follicular helper T cells is regulated by the strength of T cell antigen receptor binding. *Nature Immunology* **10**, 375-384, doi:10.1038/ni.1704 (2009).

30 Vinuesa, C. G., Linterman, M. A., Yu, D. & Maclennan, I. C. M. Follicular Helper T Cells. *Annual Review of Immunology* **34**, 335-368, doi:10.1146/annurev-immunol-041015-055605 (2016).

31 Choi, J. *et al.* Bcl-6 is the nexus transcription factor of T follicular helper cells via repressor-of-repressor circuits. *Nature Immunology* **21**, 777-789, doi:10.1038/s41590-020-0706-5 (2020).

32 Xie, M. M. *et al.* Bcl6 promotes follicular helper T-cell differentiation and PD-1 expression in a Blimp1-independent manner in mice. *European Journal of Immunology* **47**, 1136-1141, doi:10.1002/eji.201747034 (2017).

33 Dupont, C., Armant, D. & Brenner, C. Epigenetics: Definition, Mechanisms and Clinical Perspective. *Seminars in Reproductive Medicine* **27**, 351-357, doi:10.1055/s-0029-1237423 (2009).

34 Wang, Z., Yin, H., Lau, C. & Lu, Q. Histone Posttranslational Modifications of CD4+ T Cell in Autoimmune Diseases. *International Journal of Molecular Sciences* **17**, 1547, doi:10.3390/ijms17101547 (2016).

35 Rodriguez, R. M., Lopez-Larrea, C. & Suarez-Alvarez, B. Epigenetic dynamics during CD4+ T cells lineage commitment. *The International Journal of Biochemistry & Cell Biology* **67**, 75-85, doi:10.1016/j.biocel.2015.04.020 (2015).

36 Laugesen, A. & Helin, K. Chromatin Repressive Complexes in Stem Cells, Development, and Cancer. *Cell Stem Cell* **14**, 735-751, doi:10.1016/j.stem.2014.05.006 (2014).

37 Bernstein, B. E., Meissner, A. & Lander, E. S. The Mammalian Epigenome. *Cell* **128**, 669-681, doi:10.1016/j.cell.2007.01.033 (2007).

38 Karantanos, T., Chistofides, A., Barhdan, K., Li, L. & Boussiotis, V. A. Regulation of T Cell Differentiation and Function by EZH2. *Frontiers in Immunology* **7**, doi:10.3389/fimmu.2016.00172 (2016).

39 Gray, S. M., Amezquita, R. A., Guan, T., Kleinstein, S. H. & Kaech, S. M. Polycomb Repressive Complex 2-Mediated Chromatin Repression Guides Effector CD8 + T Cell Terminal Differentiation and Loss of Multipotency. *Immunity* **46**, 596-608, doi:10.1016/j.immuni.2017.03.012 (2017).

40 Leignadier, J., Hardy, M.-P., Cloutier, M., Rooney, J. & Labrecque, N. Memory T-lymphocyte survival does not require T-cell receptor expression. *Proceedings of the National Academy of Sciences* **105**, 20440-20445, doi:10.1073/pnas.0806289106 (2008).

41 Murali-Krishna, K. Persistence of Memory CD8 T Cells in MHC Class I-Deficient Mice. *Science* **286**, 1377-1381, doi:10.1126/science.286.5443.1377 (1999).

42 Becker, T. C. *et al.* Interleukin 15 Is Required for Proliferative Renewal of Virus-specific Memory CD8 T Cells. *Journal of Experimental Medicine* **195**, 1541-1548, doi:10.1084/jem.20020369 (2002).

43 Schluns, K. S., Williams, K., Ma, A., Zheng, X. X. & Lefrançois, L. Cutting Edge: Requirement for IL-15 in the Generation of Primary and Memory Antigen-Specific CD8 T Cells. *The Journal of Immunology* **168**, 4827-4831, doi:10.4049/jimmunol.168.10.4827 (2002).

44 Sarkar, S. *et al.* Functional and genomic profiling of effector CD8 T cell subsets with distinct memory fates. *Journal of Experimental Medicine* **205**, 625-640, doi:10.1084/jem.20071641 (2008).

45 Shipkova, M. & Wieland, E. Surface markers of lymphocyte activation and markers of cell proliferation. *Clin Chim Acta* **413**, 1338-1349, doi:10.1016/j.cca.2011.11.006 (2012).

46 Masopust, D. Preferential Localization of Effector Memory Cells in Nonlymphoid Tissue. *Science* **291**, 2413-2417, doi:10.1126/science.1058867 (2001).

47 Sallusto, F., Lenig, D., Förster, R., Lipp, M. & Lanzavecchia, A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708-712, doi:10.1038/44385 (1999).

48 Wherry, E. J. & Ahmed, R. Memory CD8 T-Cell Differentiation during Viral Infection. *Journal of Virology* **78**, 5535-5545, doi:10.1128/jvi.78.11.5535-5545.2004 (2004).

49 Gattinoni, L. *et al.* A human memory T cell subset with stem cell–like properties. *Nature Medicine* **17**, 1290-1297, doi:10.1038/nm.2446 (2011).

50 Wakim, L. M., Woodward-Davis, A. & Bevan, M. J. Memory T cells persisting within the brain after local infection show functional adaptations to their tissue of residence. *Proceedings of the National Academy of Sciences* **107**, 17872-17879, doi:10.1073/pnas.1010201107 (2010).

51 Gebhardt, T. *et al.* Different patterns of peripheral migration by memory CD4+ and CD8+ T cells. *Nature* **477**, 216-219, doi:10.1038/nature10339 (2011).

52 Jiang, X. *et al.* Skin infection generates non-migratory memory CD8+ TRM cells providing global skin immunity. *Nature* **483**, 227-231, doi:10.1038/nature10851 (2012).

53 Masopust, D. *et al.* Dynamic T cell migration program provides resident memory within intestinal epithelium. *Journal of Experimental Medicine* **207**, 553-564, doi:10.1084/jem.20090858 (2010).

54 Sheridan, B. S. & Lefrançois, L. Regional and mucosal memory T cells. *Nature Immunology* **12**, 485-491, doi:10.1038/ni.2029 (2011).

55 Galletti, G. *et al.* Two subsets of stem-like CD8+ memory T cell progenitors with distinct fate commitments in humans. *Nature Immunology* **21**, 1552-1562, doi:10.1038/s41590-020-0791-5 (2020).

56 Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: generation, compartmentalization and homeostasis. *Nature Reviews Immunology* **14**, 24-35, doi:10.1038/nri3567 (2014).

57 Obar, J. J. & Lefrançois, L. Memory CD8+ T cell differentiation. *Annals of the New York Academy of Sciences* **1183**, 251-266, doi:10.1111/j.1749-6632.2009.05126.x (2010).

58 Chen, Y., Zander, R., Khatun, A., Schauder, D. M. & Cui, W. Transcriptional and Epigenetic Regulation of Effector and Memory CD8 T Cell Differentiation. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.02826 (2018).

59 Chang, J. T., Wherry, E. J. & Goldrath, A. W. Molecular regulation of effector and memory T cell differentiation. *Nature Immunology* **15**, 1104-1115, doi:10.1038/ni.3031 (2014).

60 Akondy, R. S. *et al.* Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362-367, doi:10.1038/nature24633 (2017).

61 Youngblood, B. *et al.* Effector CD8 T cells dedifferentiate into long-lived memory cells. *Nature* **552**, 404-409, doi:10.1038/nature25144 (2017).

62 Mclane, L. M., Abdel-Hakeem, M. S. & Wherry, E. J. CD8 T Cell Exhaustion During Chronic Viral Infection and Cancer. *Annual Review of Immunology* **37**, 457-495, doi:10.1146/annurev-immunol-041015-055318 (2019).

63 Doering, A., Travis *et al.* Network Analysis Reveals Centrally Connected Genes and Pathways Involved in CD8+ T Cell Exhaustion versus Memory. *Immunity* **37**, 1130-1144, doi:10.1016/j.immuni.2012.08.021 (2012).

64 Schietinger, A. & Greenberg, P. D. Tolerance and exhaustion: defining mechanisms of T cell dysfunction. *Trends in Immunology* **35**, 51-60, doi:10.1016/j.it.2013.10.001 (2014).

65 Zajac, A. J. *et al.* Viral Immune Evasion Due to Persistence of Activated T Cells Without Effector Function. *Journal of Experimental Medicine* **188**, 2205-2213, doi:10.1084/jem.188.12.2205 (1998).

66 Gallimore, A. *et al.* Induction and Exhaustion of Lymphocytic Choriomeningitis Virus–specific Cytotoxic T Lymphocytes Visualized Using Soluble Tetrameric Major Histocompatibility Complex Class I–Peptide Complexes. *Journal of Experimental Medicine* **187**, 1383-1393, doi:10.1084/jem.187.9.1383 (1998).

67 Day, C. L. *et al.* PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* **443**, 350-354, doi:10.1038/nature05115 (2006).

68 Kostense, S. *et al.* High viral burden in the presence of major HIV-specific CD8+ T cell expansions: evidence for impaired CTL effector function. *European Journal of Immunology* **31**, 677-686, doi:10.1002/1521-4141(200103)31:3<677::aid-immu677>3.0.co;2-m (2001).

69 Shankar, P. *et al.* Impaired function of circulating HIV-specific CD8+ T cells in chronic human immunodeficiency virus infection. *Blood* **96**, 3094-3101, doi:10.1182/blood.v96.9.3094 (2000).

70 Goepfert, P. A. *et al.* A Significant Number of Human Immunodeficiency Virus Epitope-Specific Cytotoxic T Lymphocytes Detected by Tetramer Binding Do Not Produce Gamma Interferon. *Journal of Virology* **74**, 10249-10255, doi:10.1128/jvi.74.21.10249-10255.2000 (2000).

71 Lechner, F. *et al.* Analysis of Successful Immune Responses in Persons Infected with Hepatitis C Virus. *Journal of Experimental Medicine* **191**, 1499-1512, doi:10.1084/jem.191.9.1499 (2000).

72 Gruener, N. H. *et al.* Sustained Dysfunction of Antiviral CD8+ T Lymphocytes after Infection with Hepatitis C Virus. *Journal of Virology* **75**, 5550-5558, doi:10.1128/jvi.75.12.5550-5558.2001 (2001).

73 Luxenburger, H., Neumann-Haefelin, C., Thimme, R. & Boettler, T. HCV-Specific T Cell Responses During and After Chronic HCV Infection. *Viruses* **10**, 645, doi:10.3390/v10110645 (2018).

74 Ye, B. *et al.* T-cell exhaustion in chronic hepatitis B infection: current knowledge and clinical significance. *Cell Death & Disease* **6**, e1694-e1694, doi:10.1038/cddis.2015.42 (2015).

75 Baitsch, L. *et al.* Exhaustion of tumor-specific CD8+ T cells in metastases from melanoma patients. *Journal of Clinical Investigation* **121**, 2350-2360, doi:10.1172/jci46102 (2011).

76 Huang, A. C. *et al.* T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* **545**, 60-65, doi:10.1038/nature22079 (2017).

77 Fraietta, J. A. *et al.* Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nature Medicine* **24**, 563-571, doi:10.1038/s41591-018-0010-1 (2018).

78 Riches, J. C. *et al.* T cells from CLL patients exhibit features of T-cell exhaustion but retain capacity for cytokine production. *Blood* **121**, 1612-1621, doi:10.1182/blood-2012-09-457531 (2013).

79 Matsuzaki, J. *et al.* Tumor-infiltrating NY-ESO-1–specific CD8+T cells are negatively regulated by LAG-3 and PD-1 in human ovarian cancer. *Proceedings of the National Academy of Sciences* **107**, 7875-7880, doi:10.1073/pnas.1003345107 (2010).

80 Angelosanto, J. M., Blackburn, S. D., Crawford, A. & Wherry, E. J. Progressive Loss of Memory T Cell Potential and Commitment to Exhaustion during Chronic Viral Infection. *Journal of Virology* **86**, 8161-8170, doi:10.1128/jvi.00889-12 (2012).

81 Yao, C. *et al.* Single-cell RNA-seq reveals TOX as a key regulator of CD8+ T cell persistence in chronic infection. *Nature Immunology* **20**, 890-901, doi:10.1038/s41590-019-0403-4 (2019).

82 Scott, A. C. *et al.* TOX is a critical regulator of tumour-specific T cell differentiation. *Nature* **571**, 270-274, doi:10.1038/s41586-019-1324-y (2019).

83 Alfei, F. *et al.* TOX reinforces the phenotype and longevity of exhausted T cells in chronic viral infection. *Nature* **571**, 265-269, doi:10.1038/s41586-019-1326-9 (2019).

84 Khan, O. *et al.* TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* **571**, 211-218, doi:10.1038/s41586-019-1325-x (2019).

85 Hudson, W. H. *et al.* 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.e1044, doi:10.1016/j.immuni.2019.11.002 (2019).

86 Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439**, 682-687, doi:10.1038/nature04444 (2006).

87 Hargadon, K. M., Johnson, C. E. & Williams, C. J. Immune checkpoint blockade therapy for cancer: An overview of FDA-approved immune checkpoint inhibitors. *International Immunopharmacology* **62**, 29-39, doi:10.1016/j.intimp.2018.06.001 (2018).

88 Blank, C. U. *et al.* Defining 'T cell exhaustion'. *Nature Reviews Immunology* **19**, 665-674, doi:10.1038/s41577-019-0221-9 (2019).

89 Albinger, N., Hartmann, J. & Ullrich, E. Current status and perspective of CAR-T and CAR-NK cell therapy trials in Germany. *Gene Therapy*, doi:10.1038/s41434-021-00246-w (2021).

90 June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S. & Milone, M. C. CAR T cell immunotherapy for human cancer. *Science* **359**, 1361-1365, doi:10.1126/science.aar6711 (2018).

91 Hwu, P. *et al.* In vivo antitumor activity of T cells redirected with chimeric antibody/T-cell receptor genes. *Cancer Res* **55**, 3369-3373 (1995).

92 Hwu, P. *et al.* Lysis of ovarian cancer cells by human lymphocytes redirected with a chimeric gene composed of an antibody variable region and the Fc receptor gamma chain. *J Exp Med* **178**, 361-366, doi:10.1084/jem.178.1.361 (1993).

93 Eshhar, Z. *et al.* The T-body approach: potential for cancer immunotherapy. *Springer Semin Immunopathol* **18**, 199-209, doi:10.1007/BF00820666 (1996).

94 Stancovski, I. *et al.* Targeting of T lymphocytes to Neu/HER2-expressing cells using chimeric single chain Fv receptors. *J Immunol* **151**, 6577-6582 (1993).

95 Kershaw, M. H. *et al.* A phase I study on adoptive immunotherapy using gene-modified T cells for ovarian cancer. *Clin Cancer Res* **12**, 6106-6115, doi:10.1158/1078-0432.CCR-06-1183 (2006).

96 Vandenberghe, P. *et al.* Antibody and B7/BB1-mediated ligation of the CD28 receptor induces tyrosine phosphorylation in human T cells. *J Exp Med* **175**, 951-960, doi:10.1084/jem.175.4.951 (1992).

97 Krause, A. *et al.* Antigen-dependent CD28 signaling selectively enhances survival and proliferation in genetically modified activated human primary T lymphocytes. *J Exp Med* **188**, 619-626, doi:10.1084/jem.188.4.619 (1998).

98 Kalos, M. *et al.* T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci Transl Med* **3**, 95ra73, doi:10.1126/scitranslmed.3002842 (2011).

99 Brentjens, R. J. *et al.* Safety and persistence of adoptively transferred autologous CD19-targeted T cells in patients with relapsed or chemotherapy refractory B-cell leukemias. *Blood* **118**, 4817-4828, doi:10.1182/blood-2011-04-348540 (2011).

100 Porter, D. L., Levine, B. L., Kalos, M., Bagg, A. & June, C. H. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N Engl J Med* **365**, 725-733, doi:10.1056/NEJMoa1103849 (2011).

101 Hay, K. A. *et al.* Kinetics and biomarkers of severe cytokine release syndrome after CD19 chimeric antigen receptor-modified T-cell therapy. *Blood* **130**, 2295-2306, doi:10.1182/blood-2017-06-793141 (2017).

102 Neelapu, S. S. *et al.* Chimeric antigen receptor T-cell therapy - assessment and management of toxicities. *Nat Rev Clin Oncol* **15**, 47-62, doi:10.1038/nrclinonc.2017.148 (2018).

103 Gust, J., Taraseviciute, A. & Turtle, C. J. Neurotoxicity Associated with CD19-Targeted CAR-T Cell Therapies. *CNS Drugs* **32**, 1091-1101, doi:10.1007/s40263-018-0582-9 (2018).

104 Chen, J. *et al.* NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530-534, doi:10.1038/s41586-019-0985-x (2019).

105 Fraietta, J. A. *et al.* Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nat Med* **24**, 563-571, doi:10.1038/s41591-018-0010-1 (2018).

106 Finney, O. C. *et al.* CD19 CAR T cell product and disease attributes predict leukemia remission durability. *J Clin Invest* **129**, 2123-2132, doi:10.1172/JCI125423 (2019).

107 Kawalekar, O. U. *et al.* Distinct Signaling of Coreceptors Regulates Specific Metabolism Pathways and Impacts Memory Development in CAR T Cells. *Immunity* **44**, 712, doi:10.1016/j.immuni.2016.02.023 (2016).

108 Porter, D. L. *et al.* Chimeric antigen receptor T cells persist and induce sustained remissions in relapsed refractory chronic lymphocytic leukemia. *Sci Transl Med* **7**, 303ra139, doi:10.1126/scitranslmed.aac5415 (2015).

109 van der Stegen, S. J., Hamieh, M. & Sadelain, M. The pharmacology of second-generation chimeric antigen receptors. *Nat Rev Drug Discov* **14**, 499-509, doi:10.1038/nrd4597 (2015).

110 Maude, S. L. *et al.* Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N Engl J Med* **378**, 439-448, doi:10.1056/NEJMoa1709866 (2018).

111 Neelapu, S. S. *et al.* Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N Engl J Med* **377**, 2531-2544, doi:10.1056/NEJMoa1707447 (2017).

112 Wetterstrand, K. A. *NA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*, <www.genome.gov/sequencingcostsdata> (2020).

113 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467, doi:10.1073/pnas.74.12.5463 (1977).

114 Waern, K., Nagalakshmi, U. & Snyder, M. RNA sequencing. *Methods Mol Biol* **759**, 125-132, doi:10.1007/978-1-61779-173-4_8 (2011).

115 Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**, 651-657, doi:10.1038/nmeth1068 (2007).

116 Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019-1031, doi:10.1016/j.cell.2009.06.049 (2009).

117 Crawford, G. E. *et al.* Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res* **16**, 123-131, doi:10.1101/gr.4074106 (2006).

118 Soon, W. W., Hariharan, M. & Snyder, M. P. High-throughput sequencing for biology and medicine. *Molecular Systems Biology* **9**, 640, doi:10.1038/msb.2012.61 (2013).

119 Method of the Year 2013. *Nature Methods* **11**, 1-1, doi:10.1038/nmeth.2801 (2014).

120 Kashima, Y. *et al.* Single-cell sequencing techniques from individual to multiomics analyses. *Experimental & Molecular Medicine* **52**, 1419-1427, doi:10.1038/s12276-020-00499-2 (2020).

121 Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science* **358**, 69-75, doi:10.1126/science.aan6826 (2017).

122 Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration. *BioEssays* **42**, 1900221, doi:10.1002/bies.201900221 (2020).

123 Waylen, L. N., Nim, H. T., Martelotto, L. G. & Ramialison, M. From whole-mount to single-cell spatial assessment of gene expression in 3D. *Communications Biology* **3**, doi:10.1038/s42003-020-01341-1 (2020).

124 Burgess, D. J. Spatial transcriptomics coming of age. *Nature Reviews Genetics* **20**, 317-317, doi:10.1038/s41576-019-0129-z (2019).

125 Miragaia, R. J. *et al.* Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. *Immunity* **50**, 493-504.e497, doi:10.1016/j.immuni.2019.01.001 (2019).

126 Dong, F. *et al.* Differentiation of transplanted haematopoietic stem cells tracked by single-cell transcriptomic analysis. *Nature Cell Biology* **22**, 630-639, doi:10.1038/s41556-020-0512-1 (2020).

127 Buenrostro, J. D. *et al.* Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell* **173**, 1535-1548.e1516, doi:10.1016/j.cell.2018.03.074 (2018).

128 Pellin, D. *et al.* A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications* **10**, doi:10.1038/s41467-019-10291-0 (2019).

129 Lönnberg, T. *et al.* Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves TH1/TFHfate bifurcation in malaria. *Science Immunology* **2**, eaal2192, doi:10.1126/sciimmunol.aal2192 (2017).

130 Chen, B.-C. *et al.* Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science* **346**, 1257998, doi:10.1126/science.1257998 (2014).

131 Moen, E. *et al.* Deep learning for cellular image analysis. *Nature Methods* **16**, 1233-1246, doi:10.1038/s41592-019-0403-1 (2019).

132 Shashkova, S. & Leake, C., Mark. Single-molecule fluorescence microscopy review: shedding new light on old problems. *Bioscience Reports* **37**, BSR20170031, doi:10.1042/bsr20170031 (2017).

133 Cho, J.-H. & Sprent, J. TCR tuning of T cell subsets. *Immunological Reviews* **283**, 129-137, doi:10.1111/imr.12646 (2018).

134 Gaud, G., Lesourne, R. & Love, P. E. Regulatory mechanisms in T cell receptor signalling. *Nature Reviews Immunology* **18**, 485-497, doi:10.1038/s41577-018-0020-8 (2018).

135 Pettmann, J., Santos, A. M., Dushek, O. & Davis, S. J. Membrane Ultrastructure and T Cell Activation. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.02152 (2018).

136 Yokosuka, T. *et al.* Newly generated T cell receptor microclusters initiate and sustain T cell activation by recruitment of Zap70 and SLP-76. *Nature Immunology* **6**, 1253-1262, doi:10.1038/ni1272 (2005).

137 Campi, G., Varma, R. & Dustin, M. L. Actin and agonist MHC–peptide complex–dependent T cell receptor microclusters as scaffolds for signaling. *Journal of Experimental Medicine* **202**, 1031-1036, doi:10.1084/jem.20051182 (2005).

138 Bunnell, S. C. *et al.* T cell receptor ligation induces the formation of dynamically regulated signaling assemblies. *Journal of Cell Biology* **158**, 1263-1275, doi:10.1083/jcb.200203043 (2002).

139 Sherman, E., Barr, V. & Samelson, L. E. Super-resolution characterization of TCR-dependent signaling clusters. *Immunological Reviews* **251**, 21-35, doi:10.1111/imr.12010 (2013).

140 Hu, Y. S., Cang, H. & Lillemeier, B. F. Superresolution imaging reveals nanometer- and micrometer-scale spatial distributions of T-cell receptors in lymph nodes. *Proceedings of the National Academy of Sciences* **113**, 7201-7206, doi:10.1073/pnas.1512331113 (2016).

141 Harrison, D. L., Fang, Y. & Huang, J. T-Cell Mechanobiology: Force Sensation, Potentiation, and Translation. *Frontiers in Physics* **7**, doi:10.3389/fphy.2019.00045 (2019).

142 Cai, E. *et al.* Visualizing dynamic microvillar search and stabilization during ligand detection by T cells. *Science* **356**, eaal3118, doi:10.1126/science.aal3118 (2017).

143 Crotty, S. Follicular helper CD4 T cells (TFH). *Annu Rev Immunol* **29**, 621-663, doi:10.1146/annurev-immunol-031210-101400 (2011).

144 Victora, G. D. & Nussenzweig, M. C. Germinal centers. *Annu Rev Immunol* **30**, 429-457, doi:10.1146/annurev-immunol-020711-075032 (2012).

145 Linterman, M. A. *et al.* IL-21 acts directly on B cells to regulate Bcl-6 expression and germinal center responses. *Journal of Experimental Medicine* **207**, 353-363, doi:10.1084/jem.20091738 (2010).

146 Reinhardt, R. L., Liang, H.-E. & Locksley, R. M. Cytokine-secreting follicular T cells shape the antibody repertoire. *Nature Immunology* **10**, 385-393, doi:10.1038/ni.1715 (2009).

147 Wang, Y. *et al.* Germinal-center development of memory B cells driven by IL-9 from follicular helper T cells. *Nature Immunology* **18**, 921-930, doi:10.1038/ni.3788 (2017).

148 Liu, D. *et al.* T–B-cell entanglement and ICOSL-driven feed-forward regulation of germinal centre reaction. *Nature* **517**, 214-218, doi:10.1038/nature13803 (2015).

149 Good-Jacobson, K. L. *et al.* PD-1 regulates germinal center B cell survival and the formation and affinity of long-lived plasma cells. *Nature Immunology* **11**, 535-542, doi:10.1038/ni.1877 (2010).

150 Gunn, M. D. *et al.* A B-cell-homing chemokine made in lymphoid follicles activates Burkitt's lymphoma receptor-1. *Nature* **391**, 799-803, doi:10.1038/35876 (1998).

151 Ansel, K. M., Mcheyzer-Williams, L. J., Ngo, V. N., Mcheyzer-Williams, M. G. & Cyster, J. G. In Vivo–Activated Cd4 T Cells Upregulate Cxc Chemokine Receptor 5 and Reprogram Their Response to Lymphoid Chemokines. *Journal of Experimental Medicine* **190**, 1123-1134, doi:10.1084/jem.190.8.1123 (1999).

152 Xu, L. *et al.* The transcription factor TCF-1 initiates the differentiation of TFH cells during acute viral infection. *Nature Immunology* **16**, 991-999, doi:10.1038/ni.3229 (2015).

153 Crotty, S. T Follicular Helper Cell Differentiation, Function, and Roles in Disease. *Immunity* **41**, 529-542, doi:10.1016/j.immuni.2014.10.004 (2014).

154 Yu, D. *et al.* The transcriptional repressor Bcl-6 directs T follicular helper cell lineage commitment. *Immunity* **31**, 457-468, doi:10.1016/j.immuni.2009.07.002 (2009).

155 Nurieva, R. I. *et al.* Bcl6 mediates the development of T follicular helper cells. *Science* **325**, 1001-1005, doi:10.1126/science.1176676 (2009).

156 Johnston, R. J. *et al.* Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* **325**, 1006-1010, doi:10.1126/science.1175870 (2009).

157 Choi, Y. S., Eto, D., Yang, J. A., Lao, C. & Crotty, S. Cutting Edge: STAT1 Is Required for IL-6–Mediated Bcl6 Induction for Early Follicular Helper Cell Differentiation. *The Journal of Immunology* **190**, 3049-3053, doi:10.4049/jimmunol.1203032 (2013).

158 Nurieva, R. I. *et al.* Generation of T Follicular Helper Cells Is Mediated by Interleukin-21 but Independent of T Helper 1, 2, or 17 Cell Lineages. *Immunity* **29**, 138-149, doi:10.1016/j.immuni.2008.05.009 (2008).

159 Schmitt, N. *et al.* Human Dendritic Cells Induce the Differentiation of Interleukin-21-Producing T Follicular Helper-like Cells through Interleukin-12. *Immunity* **31**, 158-169, doi:10.1016/j.immuni.2009.04.016 (2009).

160 Betz, B. C. *et al.* Batf coordinates multiple aspects of B and T cell function required for normal antibody responses. *Journal of Experimental Medicine* **207**, 933-942, doi:10.1084/jem.20091548 (2010).

161 Bollig, N. *et al.* Transcription factor IRF4 determines germinal center formation through follicular T-helper cell differentiation. *Proceedings of the National Academy of Sciences* **109**, 8664-8669, doi:10.1073/pnas.1205834109 (2012).

162 Auderset, F. *et al.* Notch Signaling Regulates Follicular Helper T Cell Differentiation. *The Journal of Immunology* **191**, 2344-2350, doi:10.4049/jimmunol.1300643 (2013).

163 Stone, E. L. *et al.* ICOS Coreceptor Signaling Inactivates the Transcription Factor FOXO1 to Promote Tfh Cell Differentiation. *Immunity* **42**, 239-251, doi:10.1016/j.immuni.2015.01.017 (2015).

164 Wu, T. *et al.* TCF1 Is Required for the T Follicular Helper Cell Response to Viral Infection. *Cell Reports* **12**, 2099-2110, doi:10.1016/j.celrep.2015.08.049 (2015).

165 Choi, Y. S. *et al.* LEF-1 and TCF-1 orchestrate TFH differentiation by regulating differentiation circuits upstream of the transcriptional repressor Bcl6. *Nature Immunology* **16**, 980-990, doi:10.1038/ni.3226 (2015).

166 Steinke, F. C. & Xue, H.-H. From inception to output, Tcf1 and Lef1 safeguard development of T cells and innate immune cells. *Immunologic Research* **59**, 45-55, doi:10.1007/s12026-014-8545-9 (2014).

167 Busslinger, M. & Tarakhovsky, A. Epigenetic control of immunity. *Cold Spring Harb Perspect Biol* **6**, doi:10.1101/cshperspect.a019307 (2014).

168 Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome Regulation by Polycomb and Trithorax Proteins. *Cell* **128**, 735-745, doi:10.1016/j.cell.2007.02.009 (2007).

169 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349, doi:10.1038/nature09784 (2011).

170 Dupage, M. *et al.* The Chromatin-Modifying Enzyme Ezh2 Is Critical for the Maintenance of Regulatory T Cell Identity after Activation. *Immunity* **42**, 227-238, doi:10.1016/j.immuni.2015.01.007 (2015).

171 Jacob, E., Hod-Dvorai, R., Schif-Zuck, S. & Avni, O. Unconventional Association of the Polycomb Group Proteins with Cytokine Genes in Differentiated T Helper Cells. *Journal of Biological Chemistry* **283**, 13471-13481, doi:10.1074/jbc.m709886200 (2008).

172 Tumes, J., Damon *et al.* The Polycomb Protein Ezh2 Regulates Differentiation and Plasticity of CD4+ T Helper Type 1 and Type 2 Cells. *Immunity* **39**, 819-832, doi:10.1016/j.immuni.2013.09.012 (2013).

173 Yang, X.-P. *et al.* EZH2 is crucial for both differentiation of regulatory T cells and T effector cell expansion. *Scientific Reports* **5**, 10643, doi:10.1038/srep10643 (2015).

174 Zhang, Y. *et al.* The polycomb repressive complex 2 governs life and death of peripheral T cells. *Blood* **124**, 737-749, doi:10.1182/blood-2013-12-544106 (2014).

175 Tong, Q. *et al.* Ezh2 Regulates Transcriptional and Posttranslational Expression of T-bet and Promotes Th1 Cell Responses Mediating Aplastic Anemia in Mice. *The Journal of Immunology* **192**, 5012-5022, doi:10.4049/jimmunol.1302943 (2014).

176 Hale, S., J. *et al.* Distinct Memory CD4+ T Cells with Commitment to T Follicular Helper- and T Helper 1-Cell Lineages Are Generated after Acute Viral Infection. *Immunity* **38**, 805-817, doi:10.1016/j.immuni.2013.02.020 (2013).

177 Marshall, D., Heather *et al.* Differential Expression of Ly6C and T-bet Distinguish Effector and Memory Th1 CD4+ Cell Properties during Viral Infection. *Immunity* **35**, 633-646, doi:10.1016/j.immuni.2011.08.016 (2011).

178 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).

179 Shaw, L. A. *et al.* Id2 reinforces TH1 differentiation and inhibits E2A to repress TFH differentiation. *Nature Immunology* **17**, 834-843, doi:10.1038/ni.3461 (2016).

180 Liu, X. *et al.* Transcription factor achaete-scute homologue 2 initiates follicular T-helper-cell development. *Nature* **507**, 513-518, doi:10.1038/nature12910 (2014).

181 Chen, T. & Dent, S. Y. R. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature Reviews Genetics* **15**, 93-106, doi:10.1038/nrg3607 (2014).

182 Knutson, S. K. *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proceedings of the National Academy of Sciences* **110**, 7922-7927, doi:10.1073/pnas.1303800110 (2013).

183 Ogbe, A. *et al.* Early Growth Response Genes 2 and 3 Regulate the Expression of Bcl6 and Differentiation of T Follicular Helper Cells. *Journal of Biological Chemistry* **290**, 20455-20465, doi:10.1074/jbc.m114.634816 (2015).

184 Weber, J. P. *et al.* ICOS maintains the T follicular helper cell phenotype by down-regulating Krüppel-like factor 2. *Journal of Experimental Medicine* **212**, 217-233, doi:10.1084/jem.20141432 (2015).

185 Dobenecker, M.-W. *et al.* Signaling function of PRC2 is essential for TCR-driven T cell responses. *Journal of Experimental Medicine* **215**, 1101-1113, doi:10.1084/jem.20170084 (2018).

186 Li, F. *et al.* Ezh2 programs TFH differentiation by integrating phosphorylation-dependent activation of Bcl6 and polycomb-dependent repression of p19Arf. *Nature Communications* **9**, doi:10.1038/s41467-018-07853-z (2018).

187 Shen, H. *et al.* Recombinant Listeria monocytogenes as a live vaccine vehicle for the induction of protective anti-viral cell-mediated immunity. *Proceedings of the National Academy of Sciences* **92**, 3987-3991, doi:10.1073/pnas.92.9.3987 (1995).

188 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10, doi:10.14806/ej.17.1.200 (2011).

189 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

190 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

191 Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology* **11**, R119, doi:10.1186/gb-2010-11-12-r119 (2010).

192 Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**, W160-W165, doi:10.1093/nar/gkw257 (2016).

193 Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

194 Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics* **5**, 1752-1779, doi:10.1214/11-aoas466 (2011).

195 Zhu, L. J. *et al.* ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237, doi:10.1186/1471-2105-11-237 (2010).

196 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, doi:10.1186/s13059-014-0550-8 (2014).

197 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29**, 24-26, doi:10.1038/nbt.1754 (2011).

198 Ye, L., Zeng, R., Bai, Y., Roopenian, D. C. & Zhu, X. Efficient mucosal vaccination mediated by the neonatal Fc receptor. *Nature Biotechnology* **29**, 158-163, doi:10.1038/nbt.1742 (2011).

199 He, R. *et al.* Follicular CXCR5-expressing CD8+ T cells curtail chronic viral infection. *Nature* **537**, 412-416, doi:10.1038/nature19317 (2016).

200 Rasheed, M. A. U. *et al.* Interleukin-21 Is a Critical Cytokine for the Generation of Virus-Specific Long-Lived Plasma Cells. *Journal of Virology* **87**, 7737-7746, doi:10.1128/jvi.00063-13 (2013).

201 Hashimoto, M. *et al.* CD8 T Cell Exhaustion in Chronic Infection and Cancer: Opportunities for Interventions. *Annu Rev Med* **69**, 301-318, doi:10.1146/annurev-med-012017-043208 (2018).

202 Wherry, E. J. T cell exhaustion. *Nature Immunology* **12**, 492-499, doi:10.1038/ni.2035 (2011).

203 Thommen, D. S. & Schumacher, T. N. T Cell Dysfunction in Cancer. *Cancer Cell* **33**, 547-562, doi:10.1016/j.ccell.2018.03.012 (2018).

204 Mognol, G. P. *et al.* Exhaustion-associated regulatory regions in CD8+ tumor-infiltrating T cells. *Proceedings of the National Academy of Sciences* **114**, E2776-E2785, doi:10.1073/pnas.1620498114 (2017).

205 Topalian, L., Suzanne, Drake, G., Charles & Pardoll, M., Drew. Immune Checkpoint Blockade: A Common Denominator Approach to Cancer Therapy. *Cancer Cell* **27**, 450-461, doi:10.1016/j.ccell.2015.03.001 (2015).

206 Ribas, A. & Wolchok, J. D. Cancer immunotherapy using checkpoint blockade. *Science* **359**, 1350-1355, doi:10.1126/science.aar4060 (2018).

207 Dyck, L. & Mills, K. H. G. Immune checkpoints and their inhibition in cancer and infectious diseases. *European Journal of Immunology* **47**, 765-779, doi:10.1002/eji.201646875 (2017).

208 Wykes, M. N. & Lewin, S. R. Immune checkpoint blockade in infectious diseases. *Nature Reviews Immunology* **18**, 91-104, doi:10.1038/nri.2017.112 (2018).

209 Beltra, J.-C. *et al.* Developmental Relationships of Four Exhausted CD8+ T Cell Subsets Reveals Underlying Transcriptional and Epigenetic Landscape Control Mechanisms. *Immunity* **52**, 825-841.e828, doi:10.1016/j.immuni.2020.04.014 (2020).

210 Miller, B. C. *et al.* Subsets of exhausted CD8+ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nature Immunology* **20**, 326-336, doi:10.1038/s41590-019-0312-6 (2019).

211 Im, S. J. *et al.* Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417-421, doi:10.1038/nature19330 (2016).

212 Leong, Y. A. *et al.* CXCR5+ follicular cytotoxic T cells control viral infection in B cell follicles. *Nature Immunology* **17**, 1187-1196, doi:10.1038/ni.3543 (2016).

213 Martinez, G. J. *et al.* The Transcription Factor NFAT Promotes Exhaustion of Activated CD8 + T Cells. *Immunity* **42**, 265-278, doi:10.1016/j.immuni.2015.01.006 (2015).

214 Liu, X. *et al.* Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. *Nature* **567**, 525-529, doi:10.1038/s41586-019-0979-8 (2019).

215 Chen, J. *et al.* NR4A transcription factors limit CAR T cell function in solid tumours. *Nature* **567**, 530-534, doi:10.1038/s41586-019-0985-x (2019).

216 Trautmann, L. *et al.* Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. *Nature Medicine* **12**, 1198-1202, doi:10.1038/nm1482 (2006).

217 Pauken, K. E. *et al.* Epigenetic stability of exhausted T cells limits durability of reinvigoration by PD-1 blockade. *Science* **354**, 1160-1165, doi:10.1126/science.aaf2807 (2016).

218 Wu, T. D. *et al.* Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* **579**, 274-278, doi:10.1038/s41586-020-2056-8 (2020).

219 Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nature Medicine* **25**, 1251-1259, doi:10.1038/s41591-019-0522-3 (2019).

220 Lee, T. C. & Ziff, E. B. Mxi1 Is a Repressor of the c-myc Promoter and Reverses Activation by USF. *Journal of Biological Chemistry* **274**, 595-606, doi:10.1074/jbc.274.2.595 (1999).

221 He, R. *et al.* Follicular CXCR5- expressing CD8(+) T cells curtail chronic viral infection. *Nature* **537**, 412-428, doi:10.1038/nature19317 (2016).

222 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet* **17**, 3, doi:10.14806/ej.17.1.200 (2011).

223 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915, doi:10.1038/s41587-019-0201-4 (2019).

224 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

225 Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**, 1650-1667, doi:10.1038/nprot.2016.095 (2016).

226 Liao, Y., Smyth, G. K. & Shi, W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**, e47, doi:10.1093/nar/gkz114 (2019).

227 Yao, C. *et al.* Single-cell RNA-seq reveals TOX as a key regulator of CD8(+) T cell persistence in chronic infection. *Nat Immunol* **20**, 890-901, doi:10.1038/s41590-019-0403-4 (2019).

228 Yost, K. E. *et al.* Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat Med* **25**, 1251-1259, doi:10.1038/s41591-019-0522-3 (2019).

229 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

230 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

231 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

232 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

233 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

234 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

235 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849, doi:10.1093/bioinformatics/btw313 (2016).

236 Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* **9**, 5233, doi:10.1038/s41598-019-41695-z (2019).

237 Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936, doi:10.1038/s41587-019-0206-z (2019).

238 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).

239 Hudson, W. H. *et al.* 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 e1044, doi:10.1016/j.immuni.2019.11.002 (2019).

240 Im, S. J. *et al.* Defining CD8+ T cells that provide the proliferative burst after PD-1 therapy. *Nature* **537**, 417-421, doi:10.1038/nature19330 (2016).

241 Leong, Y. A. *et al.* CXCR5(+) follicular cytotoxic T cells control viral infection in B cell follicles. *Nat Immunol* **17**, 1187-1196, doi:10.1038/ni.3543 (2016).

242 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

243 Hamilton, N. E. & Ferry, M. ggtern: Ternary Diagrams Using ggplot2. *2018* **87**, 17, doi:10.18637/jss.v087.c03 (2018).

244 Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

245 Stuart, T. H., Paul; Satija, Rahul. *Signac*, <https://satijalab.org/signac> (

246 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).

247 Schuster, S. J. *et al.* Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine* **380**, 45-56, doi:10.1056/nejmoa1804980 (2019).

248 Neelapu, S. S. *et al.* Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *New England Journal of Medicine* **377**, 2531-2544, doi:10.1056/nejmoa1707447 (2017).

249 Weber, E. W., Maus, M. V. & Mackall, C. L. The Emerging Landscape of Immune Cell Therapies. *Cell* **181**, 46-62, doi:10.1016/j.cell.2020.03.001 (2020).

250 Larson, R. C. & Maus, M. V. Recent advances and discoveries in the mechanisms and functions of CAR T cells. *Nature Reviews Cancer* **21**, 145-161, doi:10.1038/s41568-020-00323-z (2021).

251 Amancha, P. K. *et al.* Distribution of Functional CD4 and CD8 T cell Subsets in Blood and Rectal Mucosal Tissues. *Scientific Reports* **9**, doi:10.1038/s41598-019-43311-6 (2019).

252 Salter, A. I. *et al.* Phosphoproteomic analysis of chimeric antigen receptor signaling reveals kinetic and quantitative differences that affect cell function. *Science Signaling* **11**, eaat6753, doi:10.1126/scisignal.aat6753 (2018).

253 Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nature Reviews Immunology* **15**, 486-499, doi:10.1038/nri3862 (2015).

254 Youngblood, B., Davis, C. W. & Ahmed, R. Making memories that last a lifetime: heritable functions of self-renewing memory CD8 T cells. *International Immunology* **22**, 797-803, doi:10.1093/intimm/dxq437 (2010).

255 Guedan, S. *et al.* Single residue in CD28-costimulated CAR-T cells limits long-term persistence and antitumor durability. *Journal of Clinical Investigation* **130**, 3087-3097, doi:10.1172/jci133215 (2020).

256 Cherkassky, L. *et al.* Human CAR T cells with cell-intrinsic PD-1 checkpoint blockade resist tumor-mediated inhibition. *Journal of Clinical Investigation* **126**, 3130-3144, doi:10.1172/jci83092 (2016).

257 Deng, Q. *et al.* Characteristics of anti-CD19 CAR T cell infusion products associated with efficacy and toxicity in patients with large B cell lymphomas. *Nature Medicine* **26**, 1878-1887, doi:10.1038/s41591-020-1061-7 (2020).

258 Curtsinger, J. M. & Mescher, M. F. Inflammatory cytokines as a third signal for T cell activation. *Current Opinion in Immunology* **22**, 333-340, doi:10.1016/j.coi.2010.02.013 (2010).

259 Ali, S. *et al.* Sources of Type I Interferons in Infectious Immunity: Plasmacytoid Dendritic Cells Not Always in the Driver's Seat. *Frontiers in Immunology* **10**, doi:10.3389/fimmu.2019.00778 (2019).

260 Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nature Reviews Immunology* **14**, 36-49, doi:10.1038/nri3581 (2014).

261 Zhao, Z. *et al.* Structural Design of Engineered Costimulation Determines Tumor Rejection Kinetics and Persistence of CAR T Cells. *Cancer Cell* **28**, 415-428, doi:10.1016/j.ccell.2015.09.004 (2015).

262 Kochenderfer, J. N. *et al.* Construction and Preclinical Evaluation of an Anti-CD19 Chimeric Antigen Receptor. *Journal of Immunotherapy* **32**, 689-702, doi:10.1097/cji.0b013e3181ac6138 (2009).

263 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

264 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296, doi:10.1038/s41592-019-0619-0 (2019).

265 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

266 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740, doi:10.1093/bioinformatics/btr260 (2011).

267 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).

268 Abadi, M. i. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed

Systems. (2016).

269 Ribeiro, M. T., Singh, S. & Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. (2016).

270 Kahan, S. M., Wherry, E. J. & Zajac, A. J. T cell exhaustion during persistent viral infections. *Virology* **479-480**, 180-193, doi:10.1016/j.virol.2014.12.033 (2015).

271 Janeway, C. *Immunobiology : the immune system in health and disease*. 5th edn, (Garland Pub., 2001).

272 Kumar, B. V., Connors, T. J. & Farber, D. L. Human T Cell Development, Localization, and Function throughout Life. *Immunity* **48**, 202-213, doi:10.1016/j.immuni.2018.01.007 (2018).

273 Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* (ACM).

274 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension

Reduction. (2020).

275 Ferguson, A. L., Panagiotopoulos, A. Z., Debenedetti, P. G. & Kevrekidis, I. G. Systematic determination of order parameters for chain dynamics using diffusion maps. *Proceedings of the National Academy of Sciences* **107**, 13597-13602, doi:10.1073/pnas.1003293107 (2010).

276 Coifman, R. R. & Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis* **21**, 5-30, doi:10.1016/j.acha.2006.04.006 (2006).

277 Van Der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nature Reviews Immunology* **11**, 47-55, doi:10.1038/nri2887 (2011).

278 Chakraborty, A. K. & Weiss, A. Insights into the initiation of TCR signaling. *Nature Immunology* **15**, 798-807, doi:10.1038/ni.2940 (2014).

279 Yi, J., Balagopalan, L., Nguyen, T., Mcintire, K. M. & Samelson, L. E. TCR microclusters form spatially segregated domains and sequentially assemble in calcium-dependent kinetic steps. *Nature Communications* **10**, doi:10.1038/s41467-018-08064-2 (2019).

280 Wang, Q.-L. *et al.* T Cell Receptor (TCR)-Induced PLC-γ1 Sumoylation via PIASxβ and PIAS3 SUMO E3 Ligases Regulates the Microcluster Assembly and Physiological Function of PLC-γ1. *Frontiers in Immunology* **10**, doi:10.3389/fimmu.2019.00314 (2019).

281 Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T Cell Receptor-Proximal Signals Are Sustained in Peripheral Microclusters and Terminated in the Central Supramolecular Activation Cluster. *Immunity* **25**, 117-127, doi:10.1016/j.immuni.2006.04.010 (2006).

282 Taylor, M. J., Husain, K., Gartner, Z. J., Mayor, S. & Vale, R. D. A DNA-Based T Cell Receptor Reveals a Role for Receptor Clustering in Ligand Discrimination. *Cell* **169**, 108-119.e120, doi:10.1016/j.cell.2017.03.006 (2017).

283 Smoligovets, A. A., Smith, A. W., Wu, H.-J., Petit, R. S. & Groves, J. T. Characterization of dynamic actin associations with T-cell receptor microclusters in primary T cells. *Journal of Cell Science* **125**, 735-742, doi:10.1242/jcs.092825 (2012).

284 Schamel, W. W. A. *et al.* Coexistence of multivalent and monovalent TCRs explains high sensitivity and wide range of response. *Journal of Experimental Medicine* **202**, 493-503, doi:10.1084/jem.20042155 (2005).

285 Roh, K.-H., Lillemeier, B. F., Wang, F. & Davis, M. M. The coreceptor CD4 is expressed in distinct nanoclusters and does not colocalize with T-cell receptor and active protein tyrosine kinase p56lck. *Proceedings of the National Academy of Sciences* **112**, E1604-E1613, doi:10.1073/pnas.1503532112 (2015).

286 Murugesan, S. *et al.* Formin-generated actomyosin arcs propel T cell receptor microcluster movement at the immune synapse. *Journal of Cell Biology* **215**, 383-399, doi:10.1083/jcb.201603080 (2016).

287 Lillemeier, B. F. *et al.* TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. *Nature Immunology* **11**, 90-96, doi:10.1038/ni.1832 (2010).

288 Lewis, J. B. *et al.* ADAP is an upstream regulator that precedes SLP-76 at sites of TCR engagement and stabilizes signaling microclusters. *Journal of Cell Science* **131**, jcs215517, doi:10.1242/jcs.215517 (2018).

289 Hui, E. *et al.* T cell costimulatory receptor CD28 is a primary target for PD-1–mediated inhibition. *Science* **355**, 1428-1433, doi:10.1126/science.aaf1292 (2017).

290 Huang, J. *et al.* A Single Peptide-Major Histocompatibility Complex Ligand Triggers Digital Cytokine Secretion in CD4+ T Cells. *Immunity* **39**, 846-857, doi:10.1016/j.immuni.2013.08.036 (2013).

291 Hashimoto-Tane, A. *et al.* Dynein-Driven Transport of T Cell Receptor Microclusters Regulates Immune Synapse Formation and T Cell Activation. *Immunity* **34**, 919-931, doi:10.1016/j.immuni.2011.05.012 (2011).

292 Hashimoto-Tane, A. *et al.* Micro–adhesion rings surrounding TCR microclusters are essential for T cell activation. *Journal of Experimental Medicine* **213**, 1609-1625, doi:10.1084/jem.20151088 (2016).

293 Gagnon, E., Schubert, D. A., Gordo, S., Chu, H. H. & Wucherpfennig, K. W. Local changes in lipid environment of TCR microclusters regulate membrane binding by the CD3ε cytoplasmic domain. *Journal of Experimental Medicine* **209**, 2423-2439, doi:10.1084/jem.20120790 (2012).

294 Crites, T. J. *et al.* TCR Microclusters Pre-Exist and Contain Molecules Necessary for TCR Signal Transduction. *The Journal of Immunology* **193**, 56-67, doi:10.4049/jimmunol.1400315 (2014).

295 Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles. (2019).

296 Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. (2017).

297 Lawless, J. F. *Statistical models and methods for lifetime data*. 2nd edn, (Wiley-Interscience, 2003).

298 Kumari, S., Curado, S., Mayya, V. & Dustin, M. L. T cell antigen receptor activation and actin cytoskeleton remodeling. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838**, 546-556, doi:10.1016/j.bbamem.2013.05.004 (2014).

299 Tsopoulidis, N. *et al.* T cell receptor–triggered nuclear actin network formation drives CD4+T cell effector functions. *Science Immunology* **4**, eaav1987, doi:10.1126/sciimmunol.aav1987 (2019).

300 Valitutti, S., Dessing, M., Aktories, K., Gallati, H. & Lanzavecchia, A. Sustained signaling leading to T cell activation results from prolonged T cell receptor occupancy.

Role of T cell actin cytoskeleton. *Journal of Experimental Medicine* **181**, 577-584, doi:10.1084/jem.181.2.577 (1995).

301 Roy, N. H. & Burkhardt, J. K. The Actin Cytoskeleton: A Mechanical Intermediate for Signal Integration at the Immunological Synapse. *Frontiers in Cell and Developmental Biology* **6**, doi:10.3389/fcell.2018.00116 (2018).

302 Fritzsche, M. *et al.* Cytoskeletal actin dynamics shape a ramifying actin network underpinning immunological synapse formation. *Science Advances* **3**, e1603032, doi:10.1126/sciadv.1603032 (2017).

303 Bunnell, S. C., Kapoor, V., Trible, R. P., Zhang, W. & Samelson, L. E. Dynamic Actin Polymerization Drives T Cell Receptor–Induced Spreading. *Immunity* **14**, 315-329, doi:10.1016/s1074-7613(01)00112-1 (2001).

304 Billadeau, D. D., Nolz, J. C. & Gomez, T. S. Regulation of T-cell activation by the cytoskeleton. *Nature Reviews Immunology* **7**, 131-143, doi:10.1038/nri2021 (2007).

305 Ritter, A. T. *et al.* Actin depletion initiates events leading to granule secretion at the immunological synapse. *Immunity* **42**, 864-876, doi:10.1016/j.immuni.2015.04.013 (2015).

306 Irvine, D. J., Purbhoo, M. A., Krogsgaard, M. & Davis, M. M. Direct observation of ligand recognition by T cells. *Nature* **419**, 845-849, doi:10.1038/nature01076 (2002).

307 Kersh, G. J., Kersh, E. N., Fremont, D. H. & Allen, P. M. High- and Low-Potency Ligands with Similar Affinities for the TCR. *Immunity* **9**, 817-826, doi:10.1016/s1074-7613(00)80647-0 (1998).

308 Huang, J. *et al.* The kinetics of two-dimensional TCR and pMHC interactions determine T-cell responsiveness. *Nature* **464**, 932-936, doi:10.1038/nature08944 (2010).

309 Alam, S. M. *et al.* T-cell-receptor affinity and thymocyte positive selection. *Nature* **381**, 616-620, doi:10.1038/381616a0 (1996).

310 Corse, E., Gottschalk, R. A., Krogsgaard, M. & Allison, J. P. Attenuated T Cell Responses to a High-Potency Ligand In Vivo. *PLoS Biology* **8**, e1000481, doi:10.1371/journal.pbio.1000481 (2010).

311 Janeway, C. A. *et al.* CD4+ T Cells: Specificity and Function. *Immunological Reviews* **101**, 39-80, doi:10.1111/j.1600-065x.1988.tb00732.x (1988).

312 Chen, L. & Flies, D. B. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nature Reviews Immunology* **13**, 227-242, doi:10.1038/nri3405 (2013).

313 Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosine-protein kinase p56lck. *Cell* **55**, 301-308, doi:10.1016/0092-8674(88)90053-0 (1988).

314 Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. The CD4 receptor is complexed in detergent lysates to a protein-tyrosine kinase (pp58) from human T lymphocytes. *Proceedings of the National Academy of Sciences* **85**, 5190-5194, doi:10.1073/pnas.85.14.5190 (1988).

315 Janeway, C. A., Jr. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation. *Annu Rev Immunol* **10**, 645-674, doi:10.1146/annurev.iy.10.040192.003241 (1992).

316 van der Merwe, P. A. & Davis, S. J. Molecular interactions mediating T cell antigen recognition. *Annu Rev Immunol* **21**, 659-684, doi:10.1146/annurev.immunol.21.120601.141036 (2003).

317 Esensten, H., Jonathan, Helou, A., Ynes, Chopra, G., Weiss, A. & Bluestone, A., Jeffrey. CD28 Costimulation: From Mechanism to Therapy. *Immunity* **44**, 973-988, doi:10.1016/j.immuni.2016.04.020 (2016).

318 Mueller, D. L., Jenkins, M. K. & Schwartz, R. H. An accessory cell-derived costimulatory signal acts independently of protein kinase C activation to allow T cell proliferation and prevent the induction of unresponsiveness. *J Immunol* **142**, 2617-2628 (1989).

319 Jenkins, M. K., Ashwell, J. D. & Schwartz, R. H. Allogeneic non-T spleen cells restore the responsiveness of normal T cell clones stimulated with antigen and chemically modified antigen-presenting cells. *J Immunol* **140**, 3324-3330 (1988).

320 Linsley, P. S. & Ledbetter, J. A. The role of the CD28 receptor during T cell responses to antigen. *Annu Rev Immunol* **11**, 191-212, doi:10.1146/annurev.iy.11.040193.001203 (1993).

321 Boise, L. H. *et al.* CD28 costimulation can promote T cell survival by enhancing the expression of Bcl-xL. Immunity. 1995. 3: 87-98. *J Immunol* **185**, 3788-3799 (2010).

322 Rabinowitz, J. D. *et al.* Altered T Cell Receptor Ligands Trigger a Subset of Early T Cell Signals. *Immunity* **5**, 125-135, doi:10.1016/s1074-7613(00)80489-6 (1996).

323 Li, Q.-J. *et al.* CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse. *Nature Immunology* **5**, 791-799, doi:10.1038/ni1095 (2004).

324 Reay, P. A., Kantor, R. M. & Davis, M. M. Use of global amino acid replacements to define the requirements for MHC binding and T cell recognition of moth cytochrome c (93-103). *J Immunol* **152**, 3946-3957 (1994).

325 Li, W., Joshi, M., Singhania, S., Ramsey, K. & Murthy, A. Peptide Vaccine: Progress and Challenges. *Vaccines* **2**, 515-536, doi:10.3390/vaccines2030515 (2014).

326 Hos, B. J., Tondini, E., Van Kasteren, S. I. & Ossendorp, F. Approaches to Improve Chemically Defined Synthetic Peptide Vaccines. *Frontiers in Immunology* **9**, doi:10.3389/fimmu.2018.00884 (2018).

327 O'Shaughnessy, E. C. *et al.* Software for lattice light-sheet imaging of FRET biosensors, illustrated with a new Rap1 biosensor. *Journal of Cell Biology* **218**, 3153-3160, doi:10.1083/jcb.201903019 (2019).

328 Mir, M. *et al.* Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. *eLife* **7**, doi:10.7554/elife.40497 (2018).

329 Gao, R. *et al.* Cortical column and whole-brain imaging with molecular contrast and nanoscale resolution. *Science* **363**, eaau8302, doi:10.1126/science.aau8302 (2019).

330 Mir, M. *et al.* in *Methods in Molecular Biology* 541-559 (Springer New York, 2018).

331 Phillips, J. K. *et al.* Characterization of neurite dystrophy after trauma by high speed structured illumination microscopy and lattice light sheet microscopy. *Journal of Neuroscience Methods* **312**, 154-161, doi:10.1016/j.jneumeth.2018.12.005 (2019).

332 Mcarthur, K. *et al.* BAK/BAX macropores facilitate mitochondrial herniation and mtDNA efflux during apoptosis. *Science* **359**, eaao6047, doi:10.1126/science.aao6047 (2018).

333 Ellefsen, K. L. & Parker, I. Dynamic Ca2+ imaging with a simplified lattice light-sheet microscope: A sideways view of subcellular Ca2+ puffs. *Cell Calcium* **71**, 34-44, doi:10.1016/j.ceca.2017.11.005 (2018).

334 David, A. F. *et al.* Augmin accumulation on long-lived microtubules drives amplification and kinetochore-directed growth. *Journal of Cell Biology* **218**, 2150-2168, doi:10.1083/jcb.201805044 (2019).

335 Condon, N. D. *et al.* Macropinosome formation by tent pole ruffling in macrophages. *Journal of Cell Biology* **217**, 3873-3885, doi:10.1083/jcb.201804137 (2018).

336 Thorn, K. A quick guide to light microscopy in cell biology. *Molecular Biology of the Cell* **27**, 219-222, doi:10.1091/mbc.e15-02-0088 (2016).

337 Leisegang, M. *et al.* Eradication of Large Solid Tumors by Gene Therapy with a T-Cell Receptor Targeting a Single Cancer-Specific Point Mutation. *Clinical Cancer Research* **22**, 2734-2743, doi:10.1158/1078-0432.ccr-15-2361 (2016).

APPENDIX A: SUPPLEMENTARY FIGURES

Supplementary Figures for Chapter 2

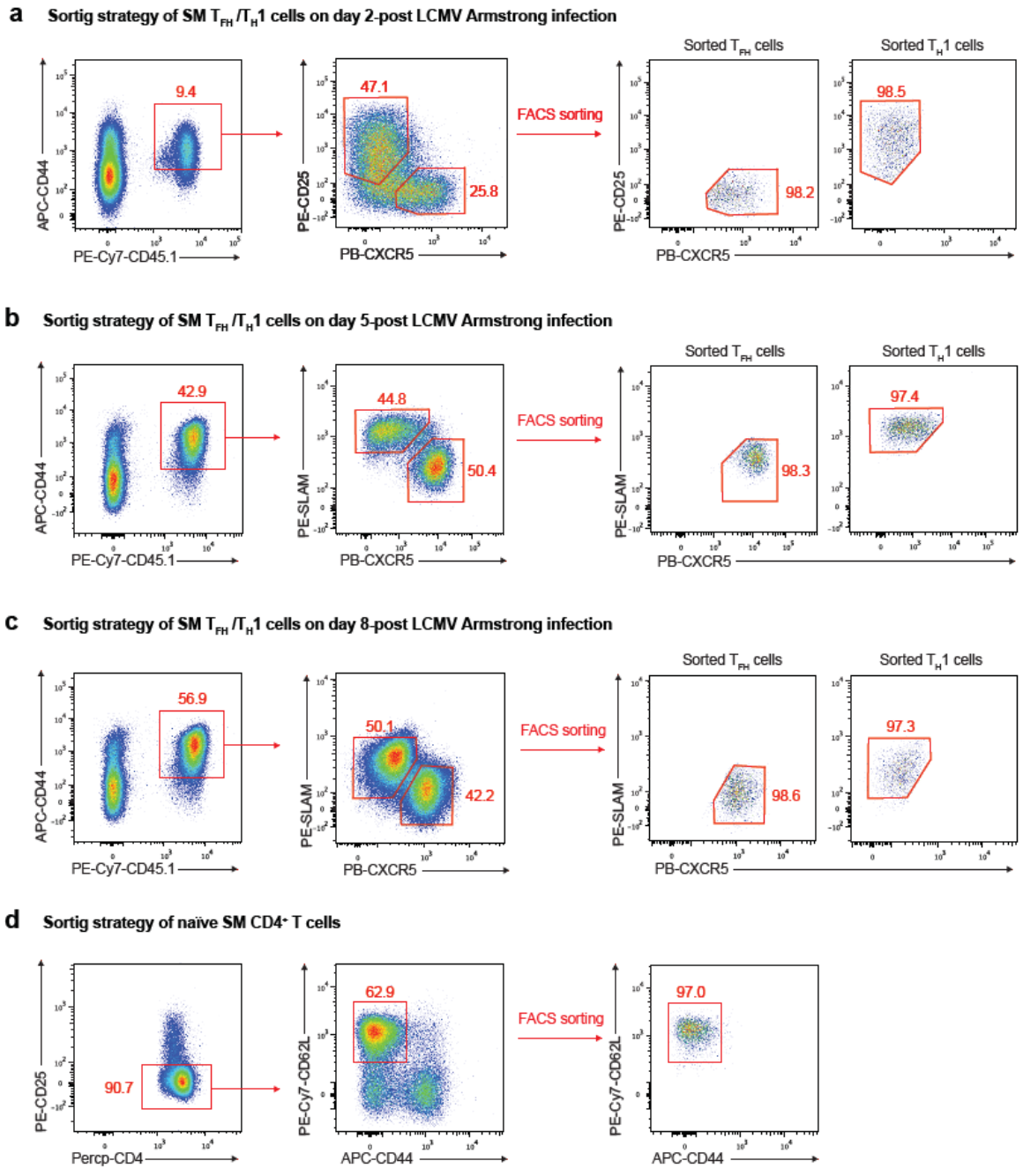


Figure S2. 1 FACS sorting strategies of naïve CD4⁺ T cells, T_{FH} cells and T_H1 cells in the ATAC-Seq profiles.

Figure S2. 1, Continued

SMARTA (SM) T_{FH} cells or SM T_H1 cells were sorted from the spleens of $CD45.2^+$ recipients that underwent adoptive transfer of $CD45.1^+$ SM cells on day 2 (**a)**, day 5 (**b**) or day 8 (**c**) after LCMV Armstrong infection. (**d**) Naïve $CD4^+$ T cells ($CD4^+CD25^-CD62L^+CD44^{lo}$) were sorted from naïve SM mice.

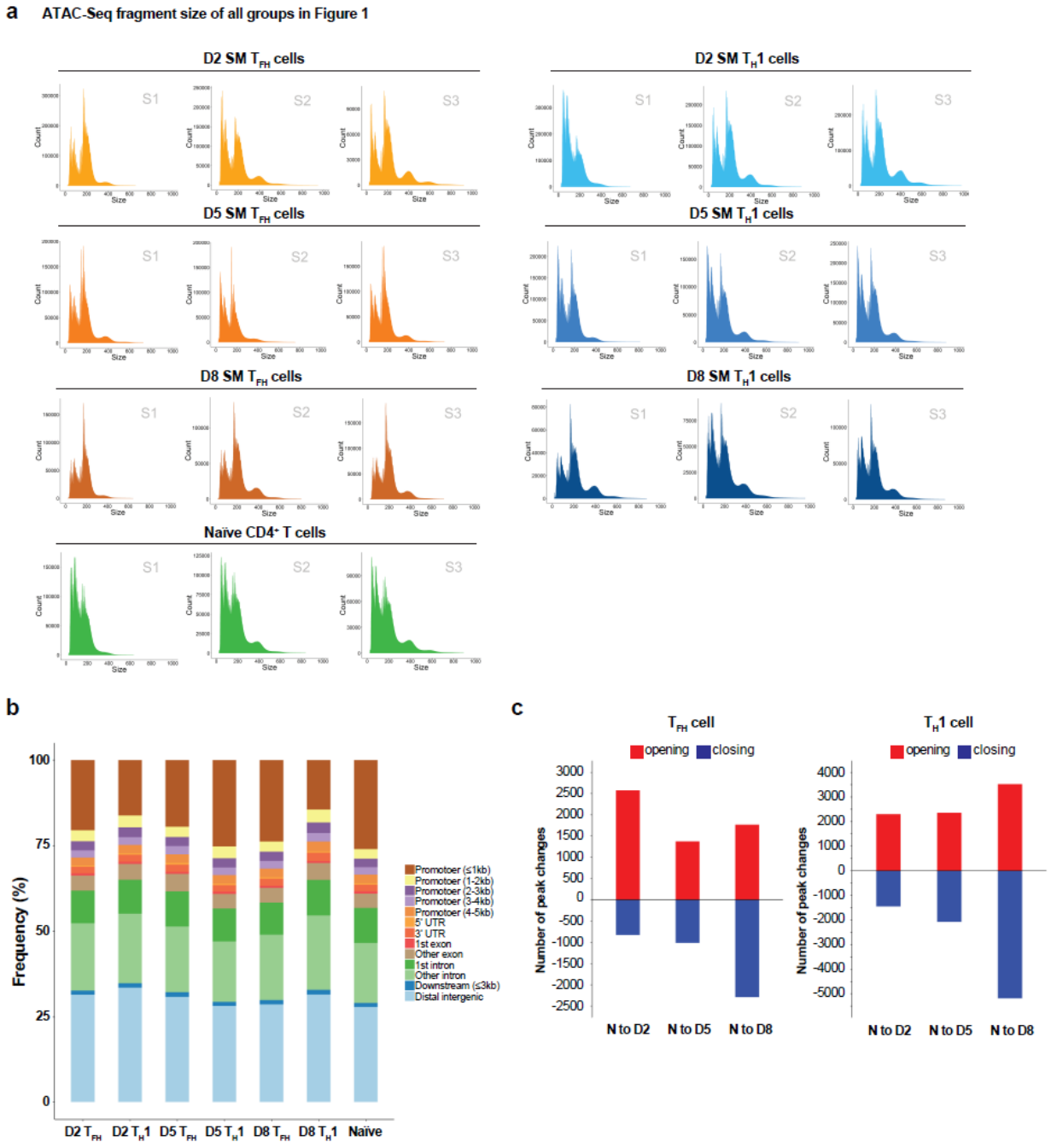


Figure S2. 2 Quality control of ATAC-Seq profiles of naïve CD4$^+$ T cells, T_{FH} cells and T_H1 cells.

a ATAC-Seq fragment sizes from all groups in Fig. 2.1. **b** Frequencies of chromatin-accessible regions in naïve CD4$^+$ T cells (Naïve), SMARTA (SM) T_{FH} cells and SM T_H1 cells partitioned into overlapping promoter regions (≤ 1 kb, 1-2 kb, 2-3 kb, 3-4 kb and 4-5 kb), 5' untranslated regions (UTRs), 3' UTRs, 1st exons, other exons, 1st introns, and other introns, and downstream

Figure S2. 2, Continued

(≤ 3 kb) and distal intergenic areas as indicated. **c** Numbers of chromatin peaks with differential accessibility (FDR < 0.05; FC > 1.5) between naïve $CD4^+$ T cells (N) and SM T_{FH} cells (left) or SM T_H1 cells (right) at indicated time-post infection.

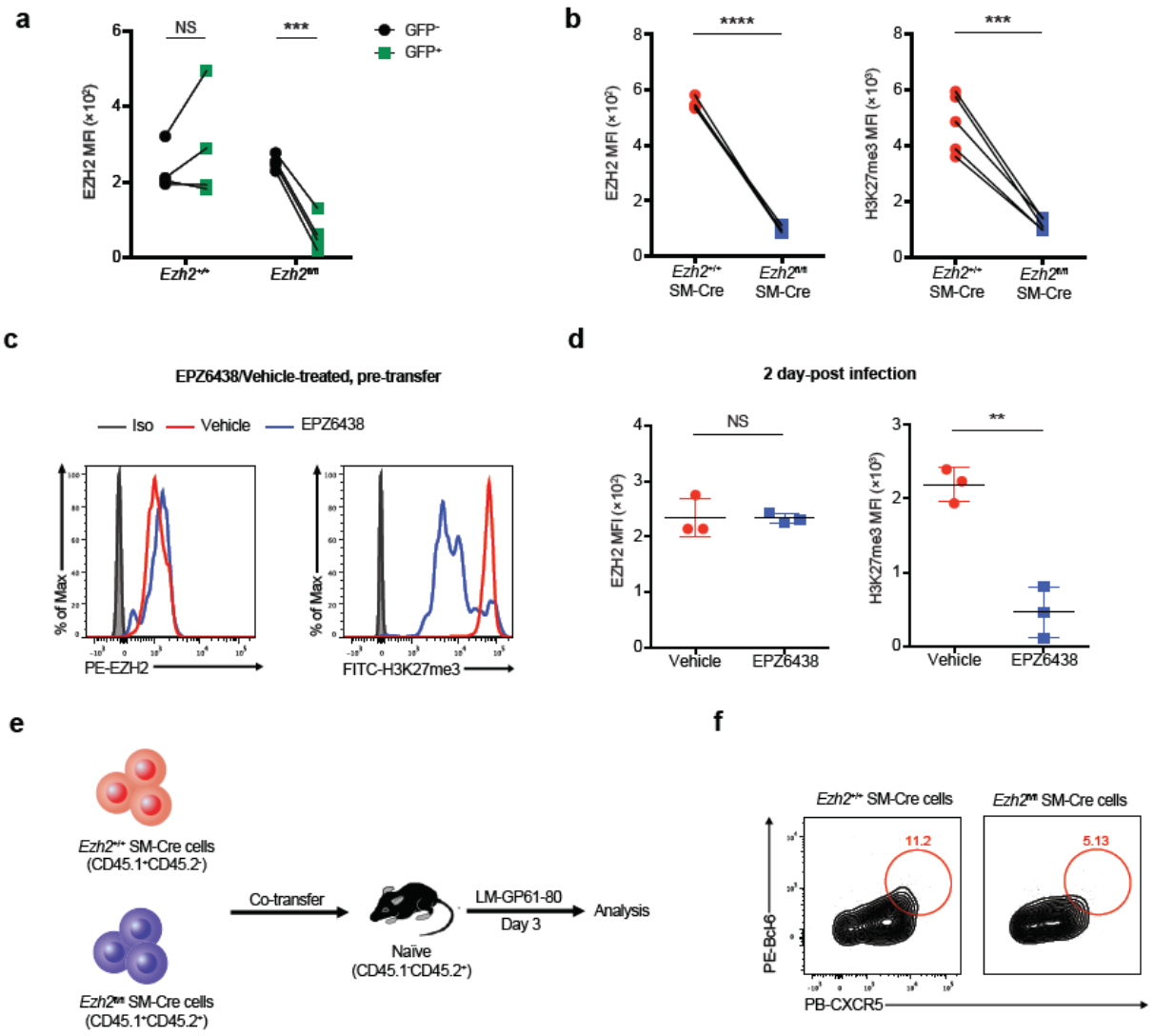


Figure S2. 3 EZH2 is essential for early commitment to T_{FH} differentiation.

a Expression of EZH2 in *Ezh2*$^{+/+}$ SMARTA (SM) cells and *Ezh2*$^{fl/fl}$ SM cells that were transduced with a retrovirus expressing iCre (GFP^{+}) or not transduced (GFP^{-}) on day 2 after infection of the host with LCMV Armstrong. **b** EZH2 and H3K27me3 expression levels in CD25^{lo}CXCR5^{+} SMARTA T_{FH} cells from the *Ezh2*$^{+/+}$*Cd4*-Cre or *Ezh2*$^{fl/fl}$*Cd4*-Cre lines as described in Fig. 2.3e. **c** Quantitation of EZH2 and H3K27me3 expression levels in SM cells treated for 3 days with 2 μM EPZ6438 or vehicle. **d** The expression levels of EZH2 and H3K27me3 in the CD25^{lo}CXCR5^{+} TFH cells described in Fig. 2.3j. **e** Experimental setup. *Ezh2*$^{+/+}$*Cd4*-Cre SM cells (*Ezh2*$^{+/+}$ SM-Cre; CD45.1^{+}CD45.2^{-}) and *Ezh2*$^{fl/fl}$*Cd4*-Cre SM cells (*Ezh2*$^{fl/fl}$ SM-Cre; CD45.1^{+}CD45.2^{+}) were co-transferred into WT recipients (CD45.1^{-}CD45.2^{+}) that were assessed on day 3 after infection with Listeria monocytogenes expressing GP61-80 (LM-GP61-80).

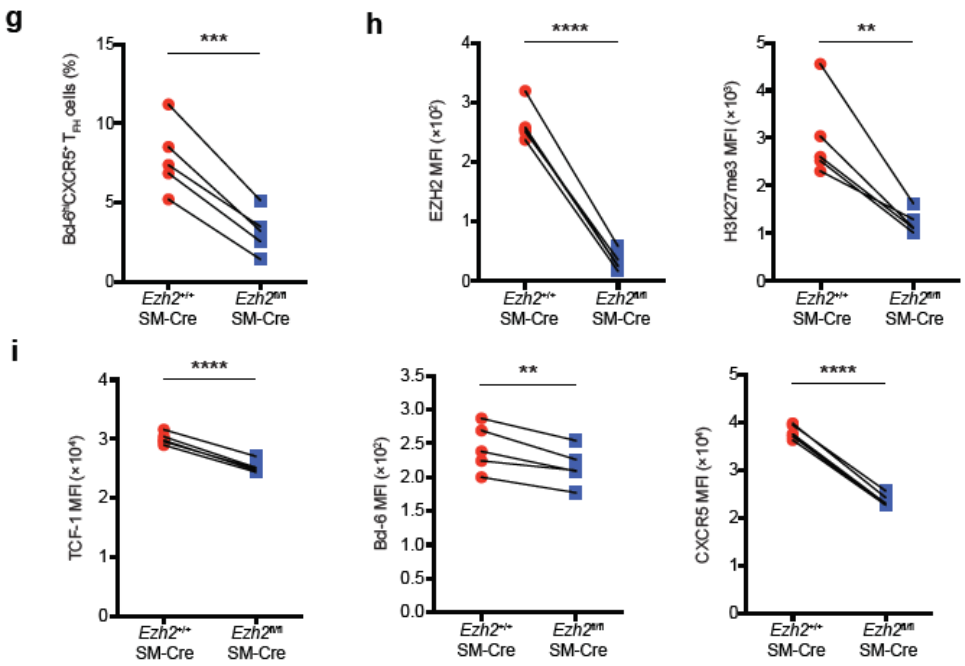


Figure S2. 3, Continued

f Flow cytometry analysis of the *Ezh2*$^{+/+}$ SM-Cre and *Ezh2*$^{fl/fl}$ SM-Cre SM cells in **e** on day 3 post-LM-GP61 infection. The numbers adjacent to the outlined areas indicate the percentages of Bcl-6^{hi}CXCR5^{+} T_{FH} cells, which are summarized in **g**. **h** EZH2 and H3K27me3 expression levels in the *Ezh2*$^{+/+}$ SM-Cre and *Ezh2*$^{fl/fl}$ SM-Cre SM cells described in **e**. **i** Quantification of the TCF-1, Bcl-6 and CXCR5 expression levels in the *Ezh2*$^{+/+}$ SM-Cre and *Ezh2*$^{fl/fl}$ SM-Cre SM cells described in **e**. NS, not significant; *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001 (paired two-tailed t-test (**a, b, g, h, i**) or unpaired two-tailed t-test (**d**)). The data are representative of two independent experiments (**a, b, d, g, h, i**) with at least three mice per group (**d**; error bar, s.d.).

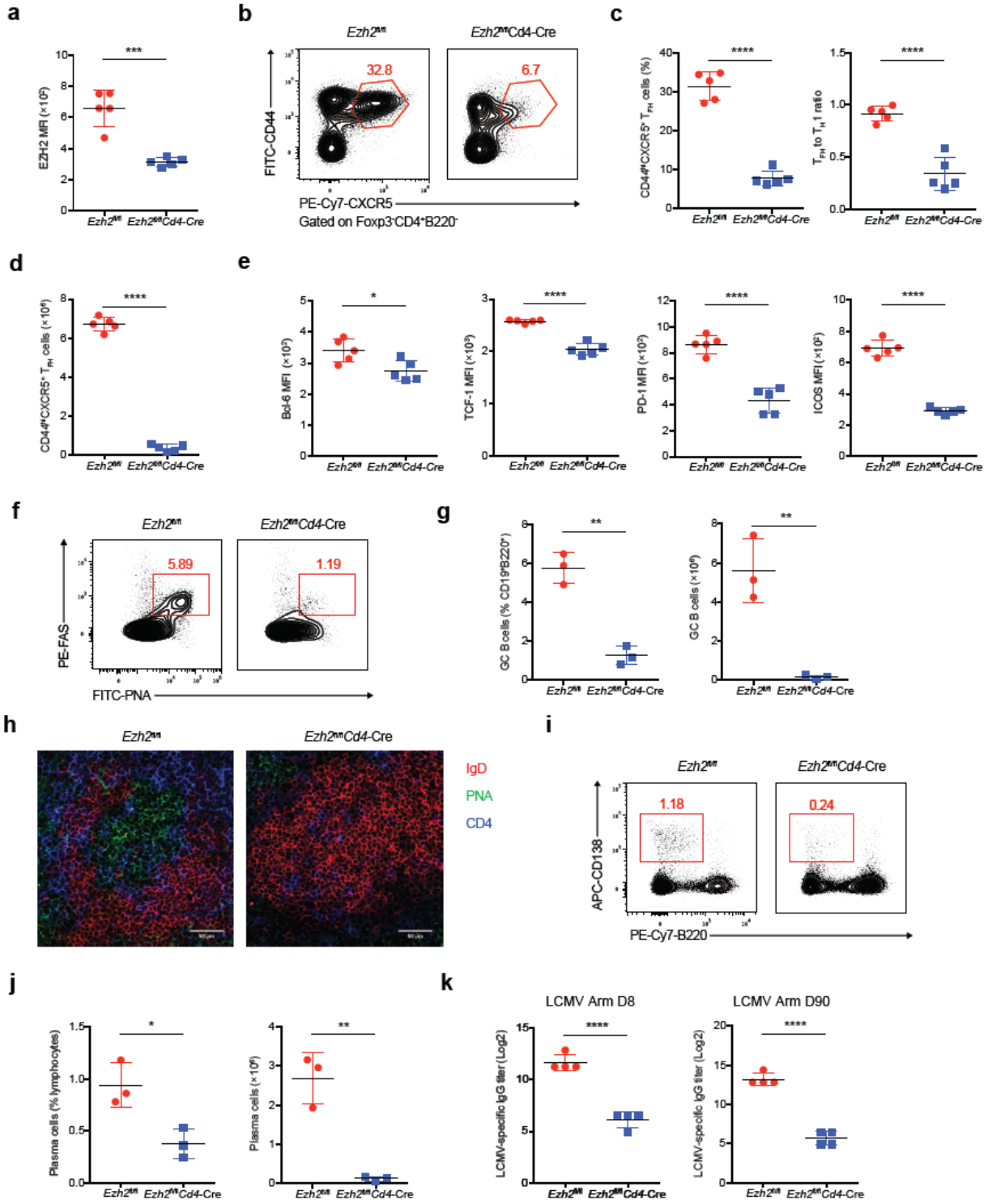


Figure S2. 4 Requirement of EZH2 expression for endogenous bulk T_{FH} cell differentiation and T_{FH} cell effector function.

Figure S2. 4 Continued

a Expression of EZH2 in activated virus-specific bulk $CD44^{hi}CXCR5^{+}$ T_{FH} cells from the spleens of *Ezh2*$^{fl/fl}$*Cd4*-Cre or *Ezh2*$^{fl/fl}$ mice on day 8 post-LCMV Armstrong infection. **b** Flow cytometry analysis of $Foxp3^{-}CD4^{+}$ cells from the spleens described in **a**. The numbers adjacent to the outlined areas indicate the proportions of $CD44^{hi}CXCR5^{+}$ T_{FH} cells. **c** Summary of $CD44^{hi}CXCR5^{+}$ T_{FH} cell proportions (left) and the ratios of $CD44^{hi}CXCR5^{+}$ T_{FH} cells to $CD44^{hi}CXCR5^{-}$ T_H1 cells (right) in (b). **d** Numbers of $CD44^{hi}CXCR5^{+}$ T_{FH} cells in **b**. **e** Quantification of Bcl-6, TCF-1, PD-1 and ICOS in the $CD44^{hi}CXCR5^{+}$ T_{FH} cells in **b**. **f** Flow analysis of $B220^{+}CD19^{+}$ B cells from the spleens of *Ezh2*$^{fl/fl}$*Cd4*-Cre or *Ezh2*$^{fl/fl}$ mice on day 8 post-LCMV infection. The numbers adjacent to the outlined areas indicate the proportions of $FAS^{+}PNA^{+}$ GC B cells. **g** The percentages (left) and numbers (right) of GC B cells in (f). **h** Immunofluorescence analysis of the GCs from the spleens of *Ezh2*$^{fl/fl}$*Cd4*-Cre or *Ezh2*$^{fl/fl}$ mice on day 8 post-LCMV infection. Scale bar, 50 μm. **i** Flow cytometry analysis of lymphocytes in the spleens of *Ezh2*$^{fl/fl}$*Cd4*-Cre or *Ezh2*$^{fl/fl}$ mice on day 8 post-LCMV infection. The numbers adjacent to the outlined areas indicate the percentages of $CD138^{hi}B220^{lo}$ plasma cells. **j** The percentages (left) and numbers (right) of plasma cells in **i**. (**k**) LCMV-specific IgG in the sera of the *Ezh2*$^{fl/fl}$*Cd4*-cre or *Ezh2*$^{fl/fl}$ mice was measured by ELISA on days 8 (left) and 90 (right) after LCMV Armstrong infection. NS, not significant; *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001 (unpaired two-tailed t-test (a, c, d, e, g, j, k)). The data are representative of three independent experiments with at least three mice per group (**a, c, d, e, g, j, k)** (error bars (**a, c, d, e, g, j, k**), s.d.).

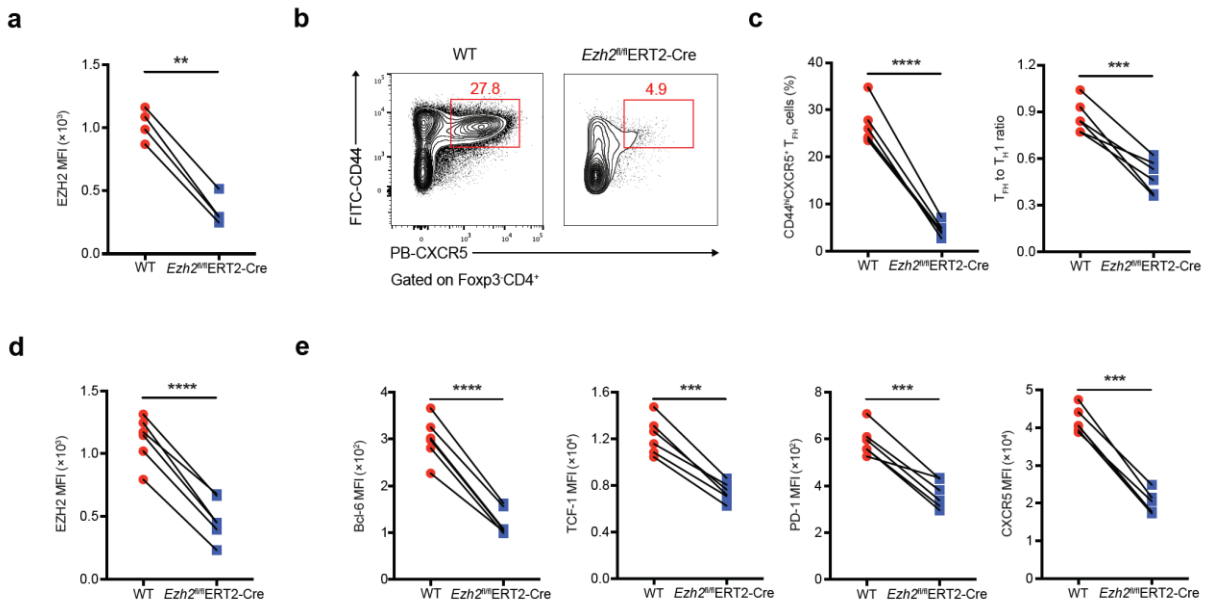


Figure S2. 5 The role of cell-autonomous EZH2 in the regulation of endogenous activated virus-specific bulk T_{FH} differentiation.

a EZH2 expression levels in bulk $CD44^{hi}CXCR5^{+}$ T_{FH} cells of WT and *Ezh2*$^{fl/fl}$ERT2-Cre origin mice on day 8 after the infection of bone marrow chimera (BMC) recipients (Fig. 2.4d). **b** Flow cytometry analysis of $CD4^{+}Foxp3^{-}$ cells of the WT and *Ezh2*$^{fl/fl}$ERT2-Cre origin spleens of the BMC mice described in Fig. 2.4d on day 8 after LCMV Armstrong infection. The numbers adjacent to the outlined areas indicate the percentages of $CD44^{hi}CXCR5^{+}$ T_{FH} cells. **c** The proportions of $CD44^{hi}CXCR5^{+}$ T_{FH} cells (left) and the ratio of $CD44^{hi}CXCR5^{+}$ T_{FH} cells to $CD44^{hi}CXCR5^{-}$ T_H1 cells (right) in (b). (**d** and **e**) The expression levels of EZH2 (d), Bcl-6, TCF-1, PD-1 and ICOS (e) in the $CD44^{hi}CXCR5^{+}$ T_{FH} cells in **b**. $^{**}P < 0.01$, $^{***}P < 0.001$ and $^{****}P < 0.0001$ (paired two-tailed t-test (**a, c, d, e**)). The data are representative of two independent experiments with at least four mice per group (**a, c, d, e**).

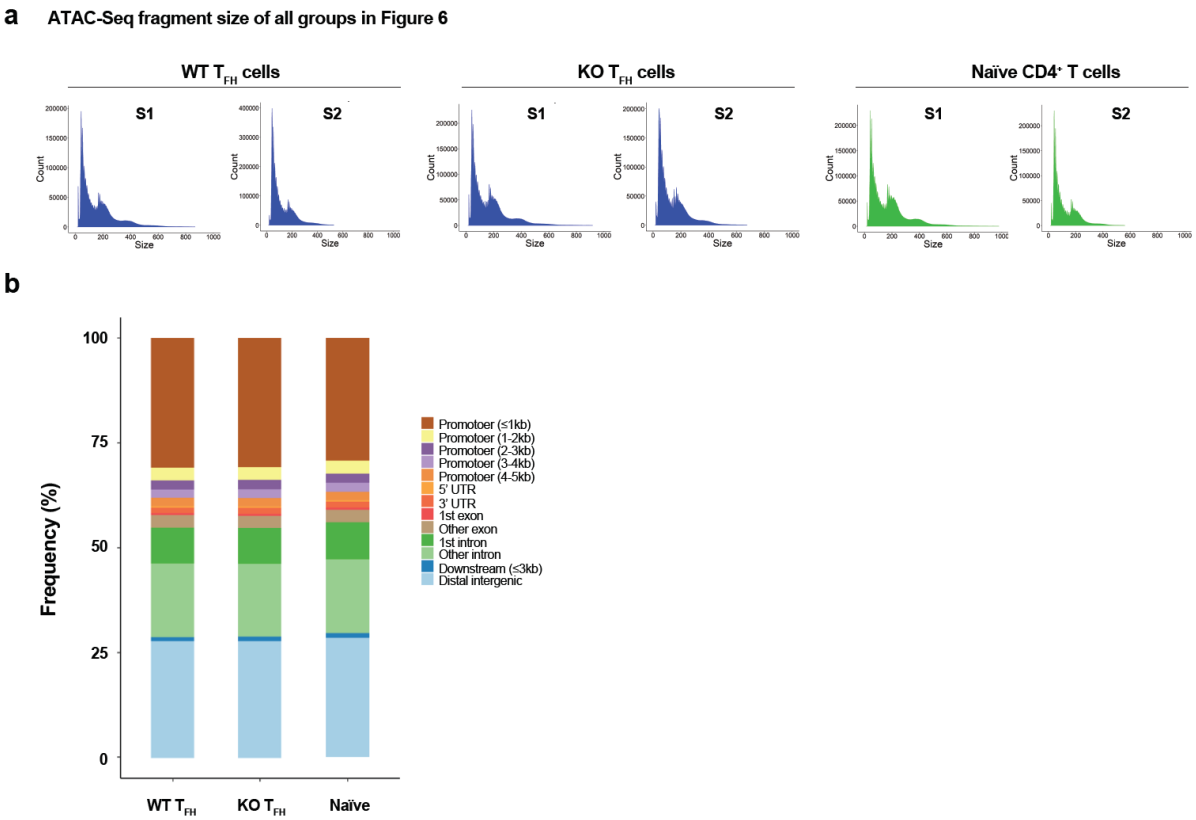


Figure S2. 6 Quality control of ATAC-Seq profiles of *Ezh2*-WT and *Ezh2*-KO bulk T_{FH} cells.

a ATAC-Seq fragment sizes from WT T_{FH} cells or KO T_{FH} cells or naïve $CD4^+$ T cells. **b** Frequencies of the chromatin-accessible regions of groups in **a** partitioned into overlapping promoter regions (≤ 1 kb, 1-2 kb, 2-3 kb, 3-4 kb and 4-5 kb), 5' UTRs, 3' UTRs, 1st exons, other exons, 1st introns, other introns and downstream (≤ 3 kb) and distal intergenic areas as indicated.

Supplementary Figures for Chapter 3

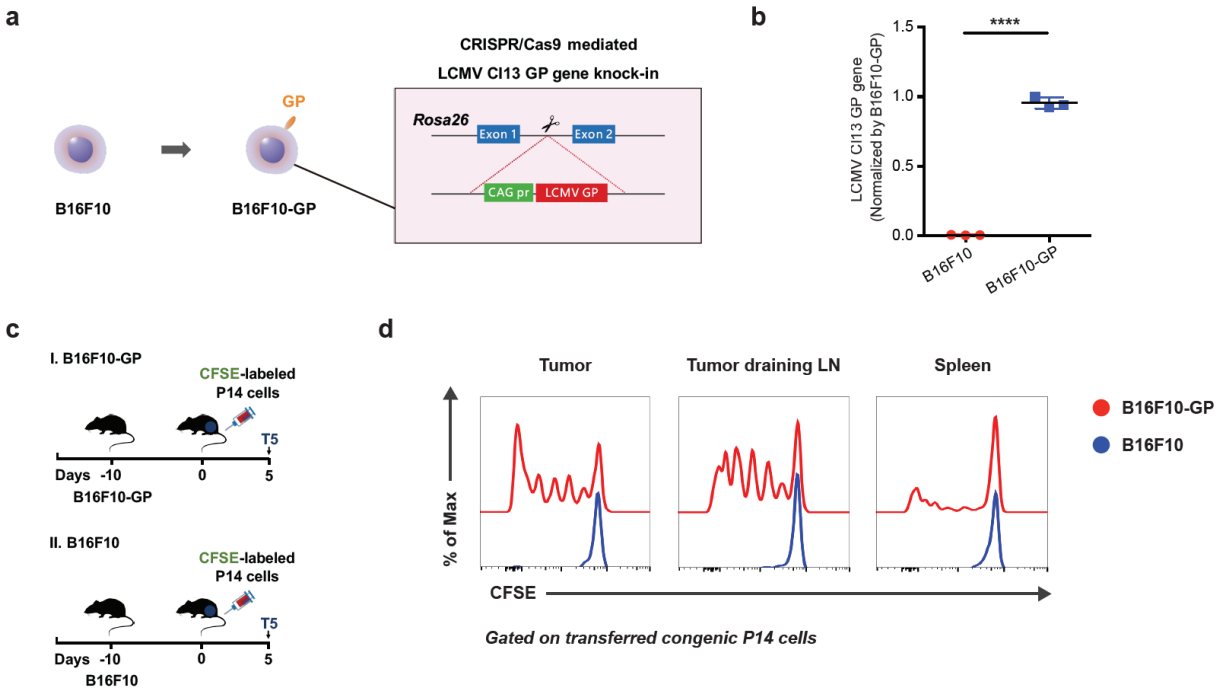


Figure S3. 1 Functional assays of B16F10 cell line expressing LCMV Cl13 GP.

a, Schematic diagram showing the construction of LCMV Cl13 GP-expressing B16F10 cell line. **b**, Quantitative RT-PCR of LCMV Cl13 GP gene mRNA in B16F10 and B16F10-GP cell lines. **c**, Experimental scheme of proliferation assay of CFSE-labeled P14 $CD8^+$ T cells. **d**, Flow cytometry analysis of CFSE-labeled congenic P14 $CD8^+$ T cells in tumor, tumor draining lymph node (LN) and spleen of B16F10 or B16F10-GP engrafted recipients.

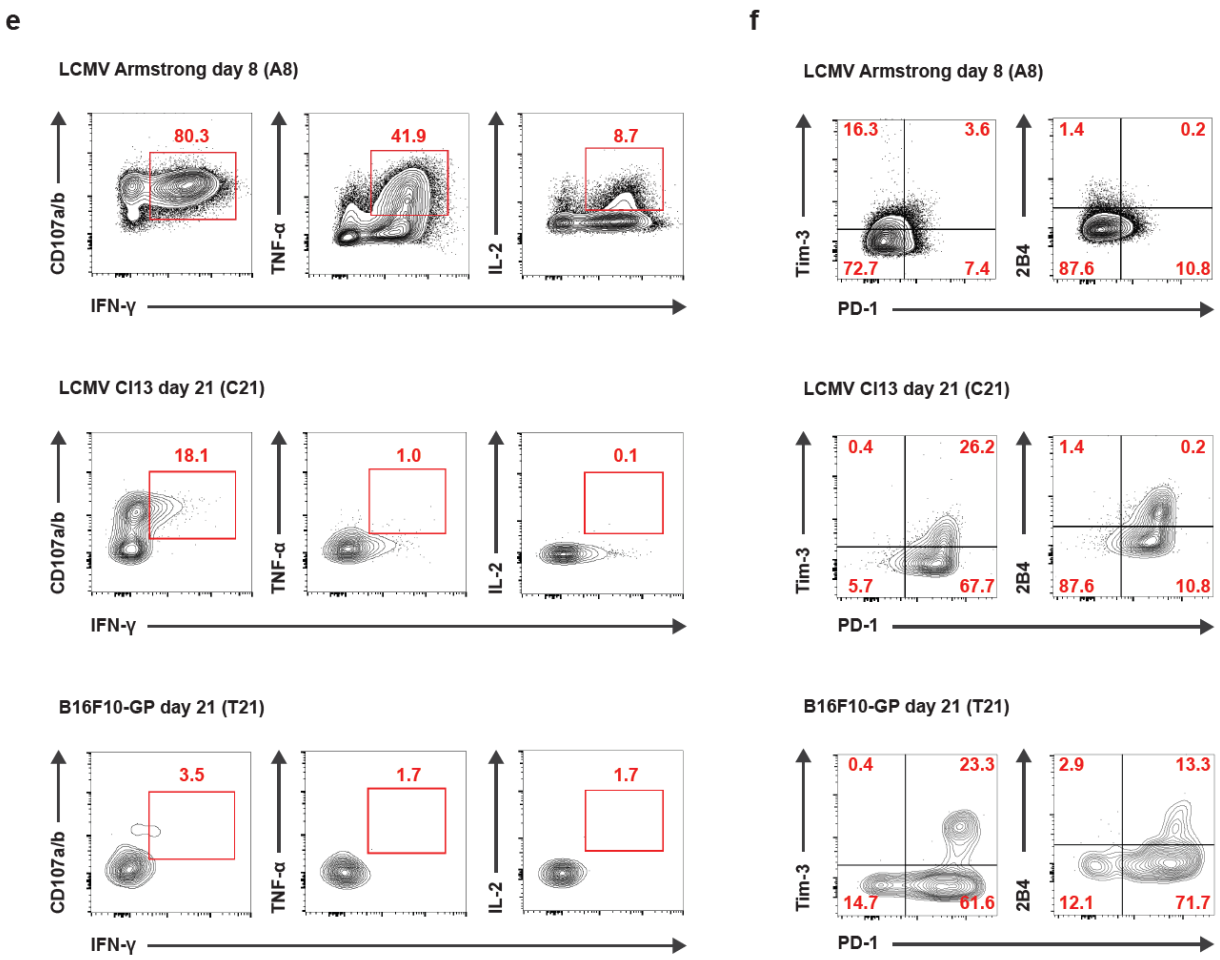


Figure S3. 1, Continued

e, f, Flow cytometry analysis of cytokine productions (IFN-γ, TNF-α and IL-2) and CD107a/b expression (**e**) and inhibitory receptors (PD-1, Tim-3 and 2B4) (**f**) of transferred congenic P14 $CD8^+$ T cells from the spleens of LCMV Armstrong-infected recipients or the spleens of LCMV Cl13-infected recipients or the tumor tissues of B16F10-GP cell line engrafted recipients at indicated time point. The data are representative of at least three independent experiments (**b, d, e, f**), and were analyzed by two-tailed unpaired *t*-test (**b**). ****$P < 0.0001$. Error bar (**b**) denotes s.d.

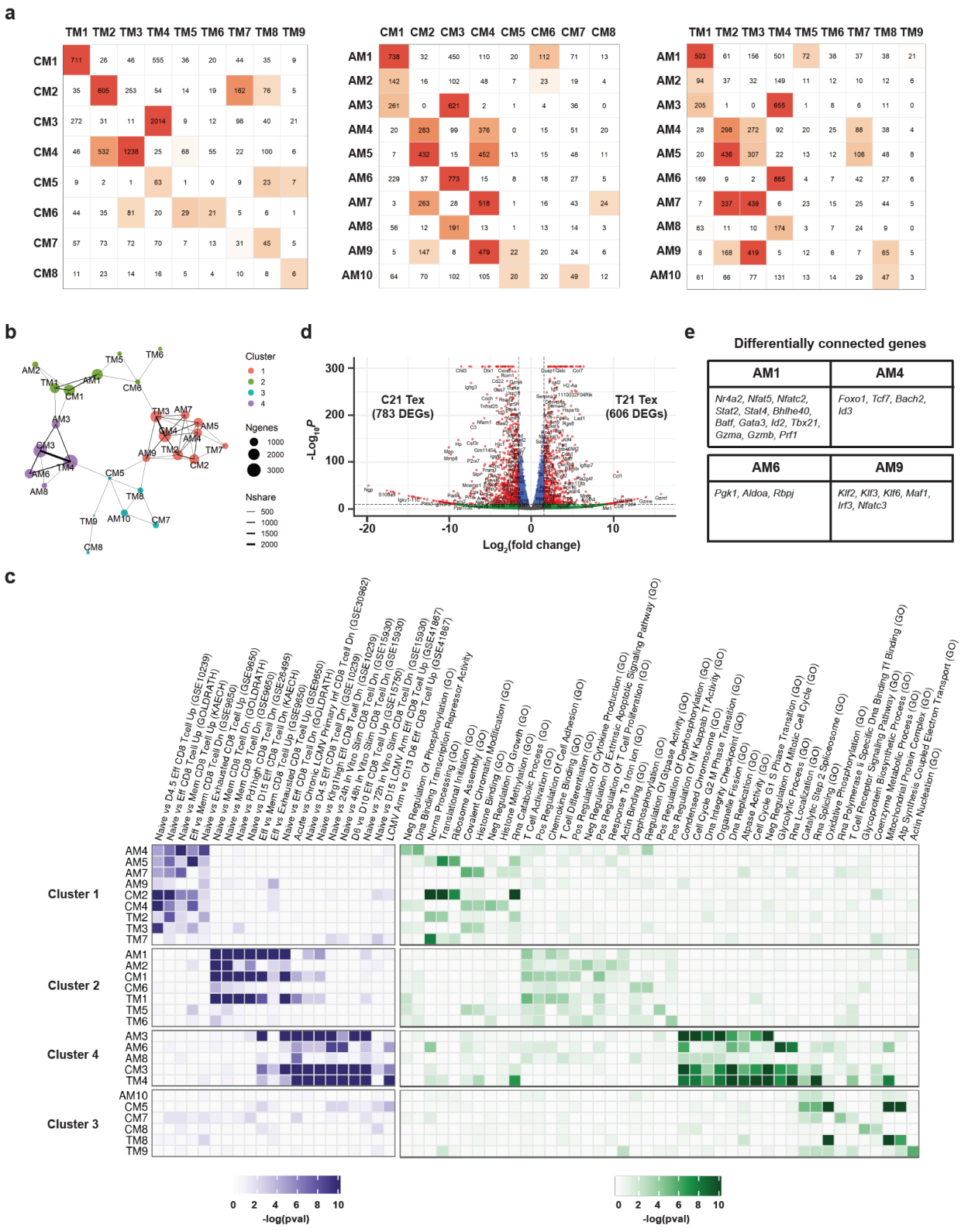


Figure S3. 2 WGCNA modules of disease-specific CD8$^+$ T cells.

Figure S3. 2, Continued

a, Correlations of modules between CM/TM (left), AM/CM (middle) and AM/TM (right). The numbers of shared genes between all pairings are shown. **b**, Clustering of AM, CM and TM modules. Modules are grouped into 4 color-coded clusters and sized by node size. Each significant overlap between two modules is represented as the connected nodes with an edge showing sharing genes. **c**, Genome browser view of the correlations of all the modules to the selected datasets (left) and the corresponding gene ontology (GO) enrichment analysis (right). **d**, Volcano plot showing the DEGs between C21-Tex and T21-Tex. **e**, Genes in AM1, AM4, AM6 and AM9 known to be crucial in T cell biology are listed.

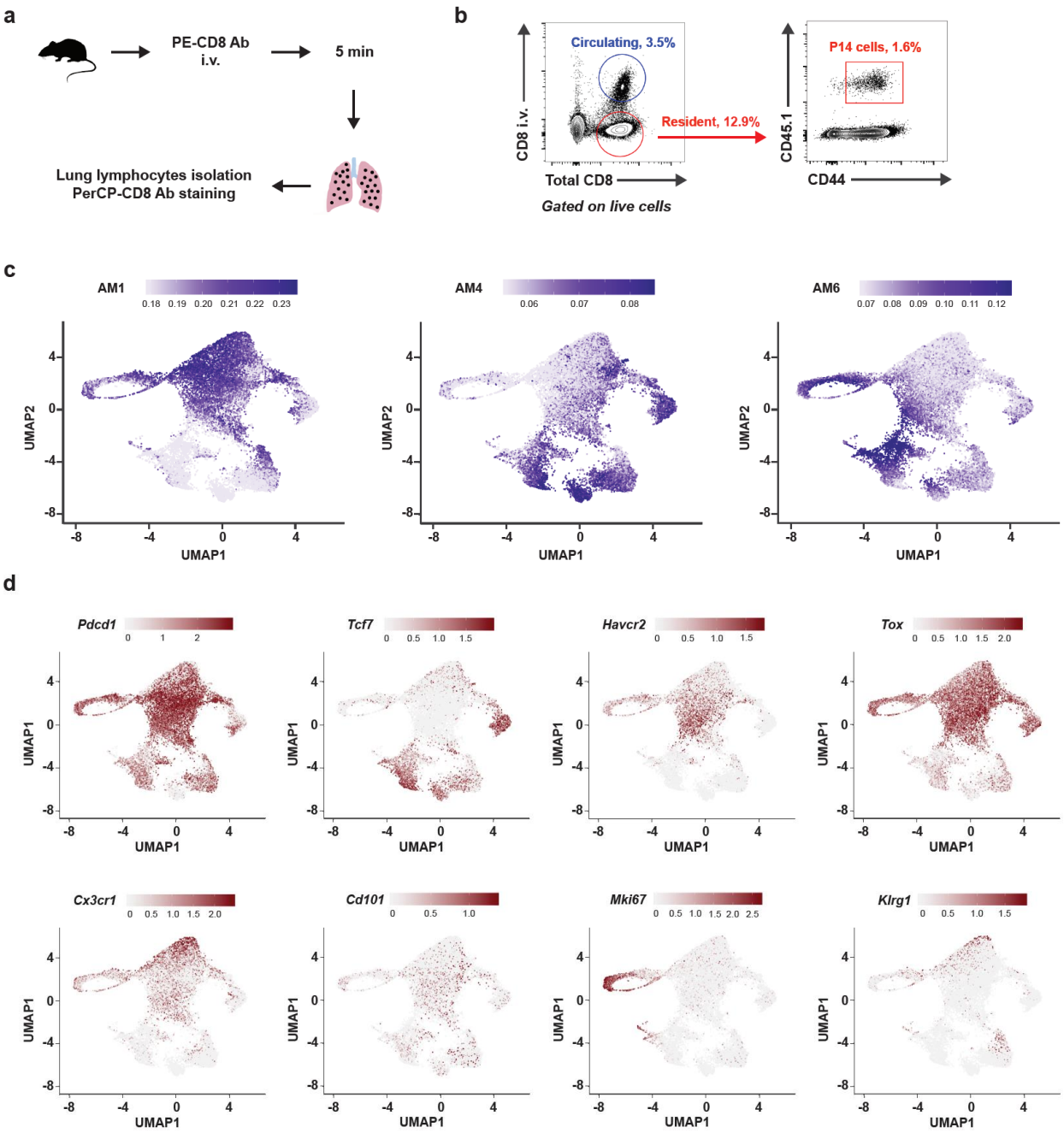
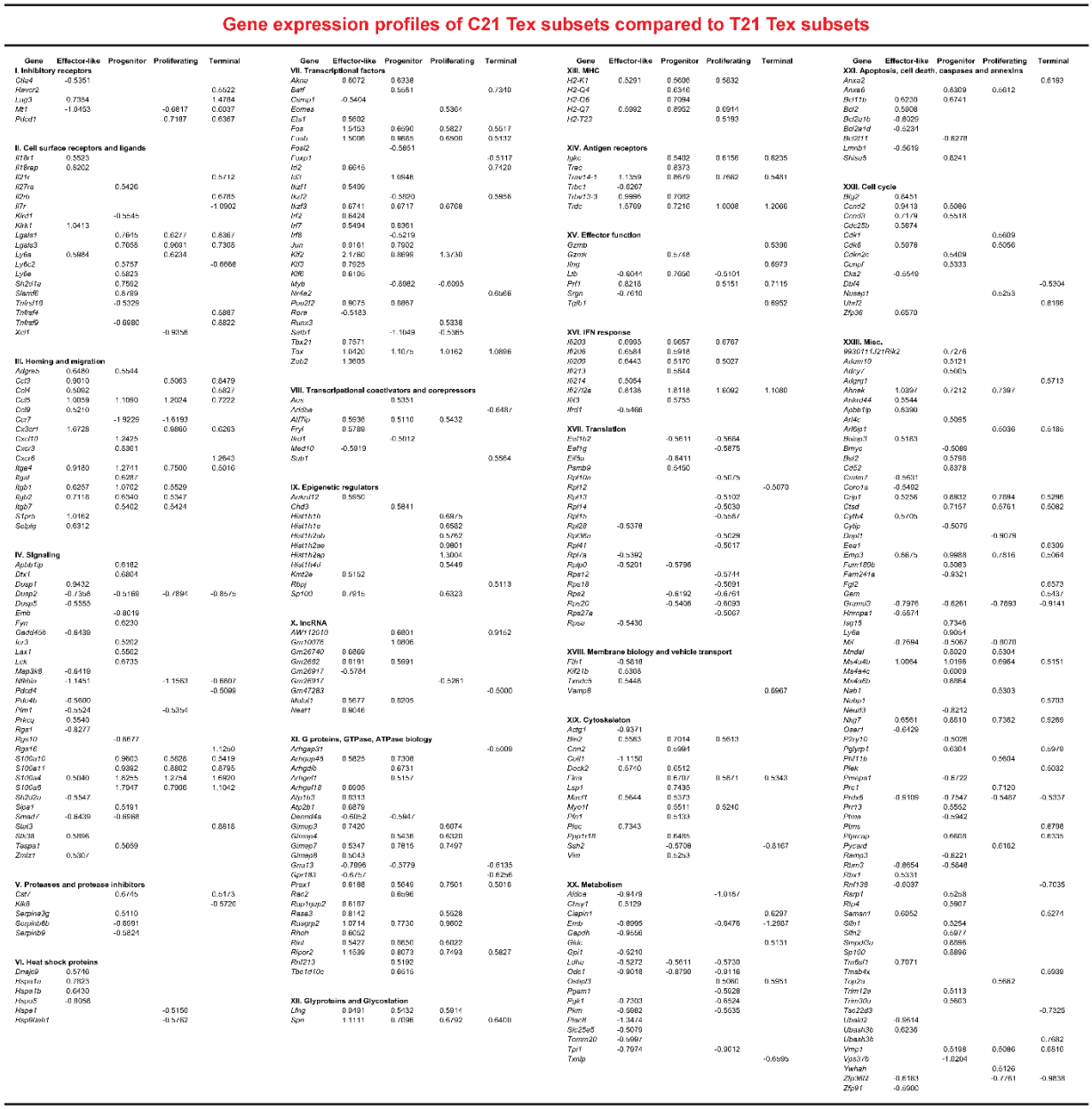


Figure S3. 3 Subset annotations of disease-specific exhausted CD8$^+$ T cells by single-cell RNA-seq.

a, Experimental scheme oflung-resident CD8$^+$ T cell isolation. **b**, Flow cytometry gating strategy of lung-resident congenic P14 CD8$^+$ T cells. **c**, UMAP showing the scores of AM1, AM4, and AM6 modules from bulk RNA-seq in each Tex cell from the scRNA-seq profiles. **d**, UMAP of all Tex cells colored according to marker gene expression.



Average LogFC = C21 Tex subset / T21 Tex subset, > 0.5 or < -0.5.

Figure S3. 4 Differential gene expression profiles of diseaseI specific exhausted T cell subsets.

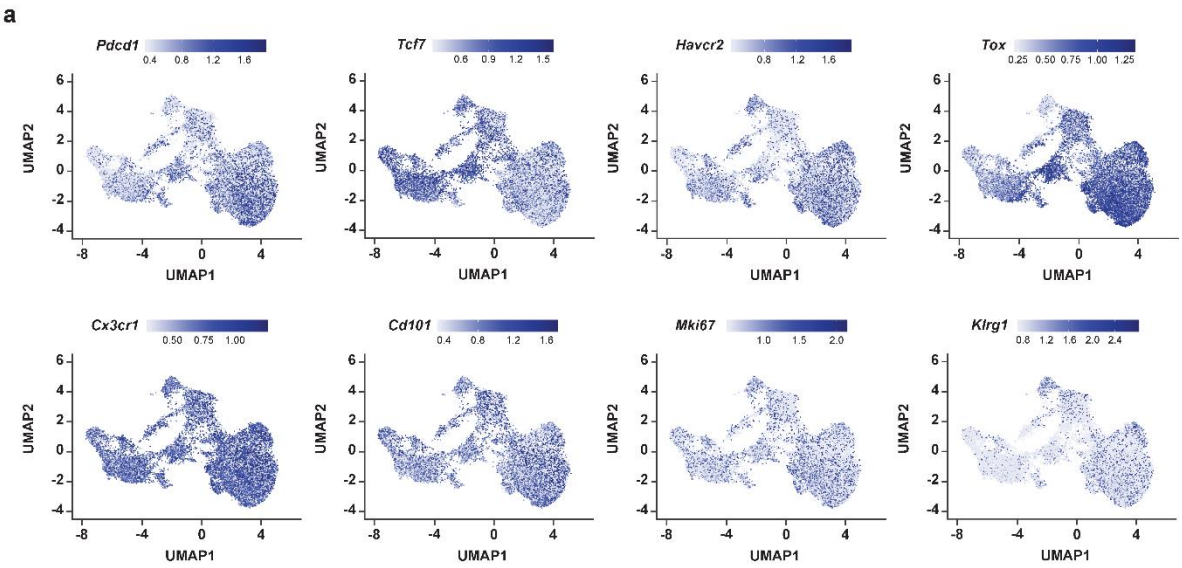


Figure S3. 5 Tex subset annotations of disease-specific exhausted CD8$^+$ T cells by single-cell ATAC-seq.

a, UMAP showing all Tex cells colored according to marker gene accessibility.

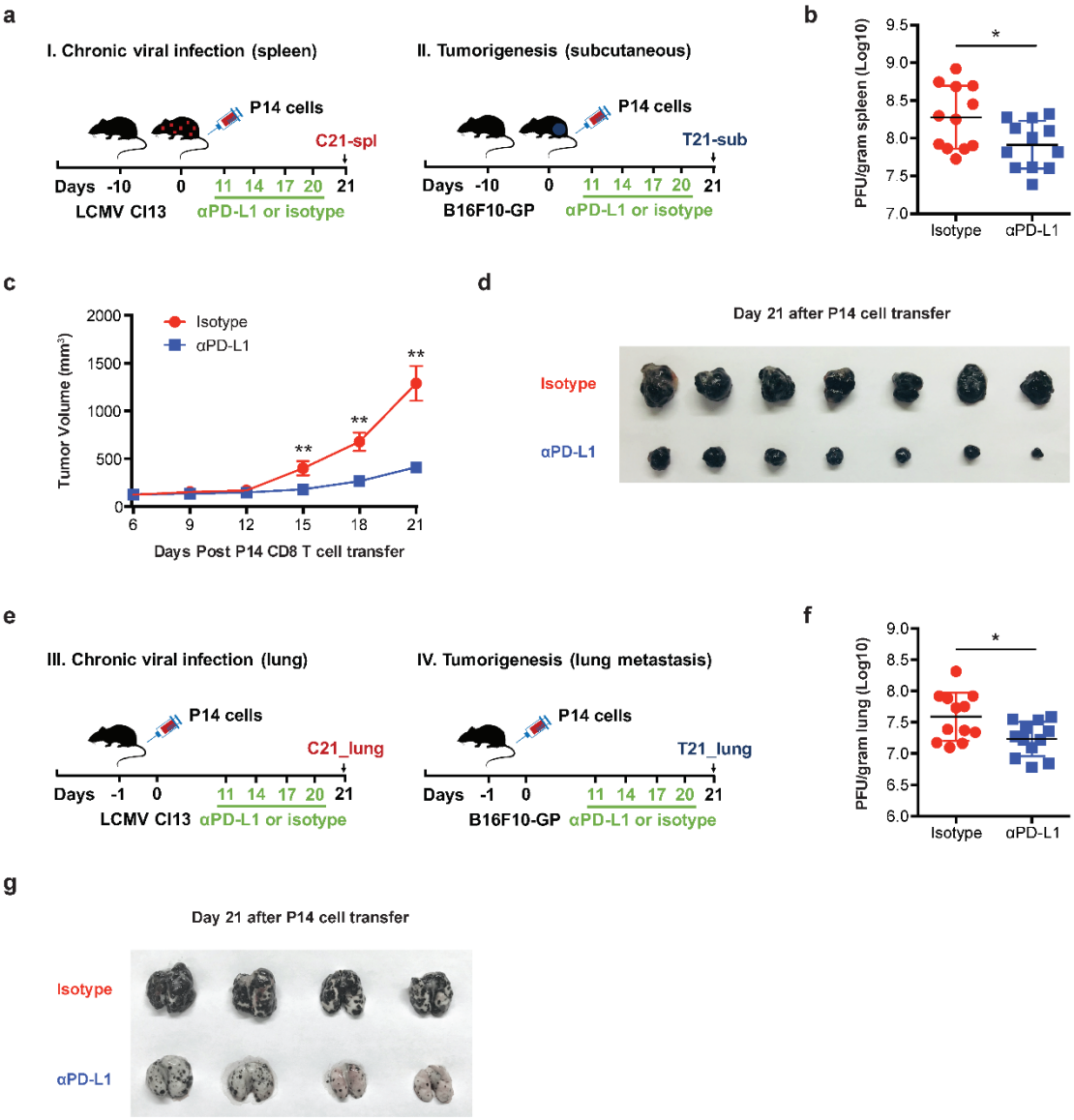


Figure S3. 6 Anti-PD-L1 treatment curtails tissue viral loads in chronic virus-infected mice and blunts tumor progression in mouse models of tumors.

a, Experimental scheme. **b**, Viral titration of LCMV Cl13 in the spleens from the mice treated with isotype or αPD-L1 antibody described in **a** (left panel, I). **c**, Tumor size measurement of the mice treated with isotype or αPD-L1 antibody described in **a** (right panel, II). **d**, Tumors of the mice described in **a** (right panel, II) at day 21 after P14 $CD8^+$ T cell transfer. **e**, Experimental scheme. **f**, Viral titration of LCMV Cl13 in the lungs from the mice treated with isotype or αPD-L1 antibody described in **e** (left panel, III).

Figure S3. 6, Continued

g, Metastatic lungs of the mice described in **e** (right panel, IV) at day 21 after P14 $CD8^+$ T cell transfer. The data are representative of at least three independent experiments (**b, c, d, f, g**), and were analyzed by two-tailed unpaired *t*-test (**b, c, f**). $*P < 0.05$, $**P < 0.01$. Error bars (**b, c, f**) denote s.d.

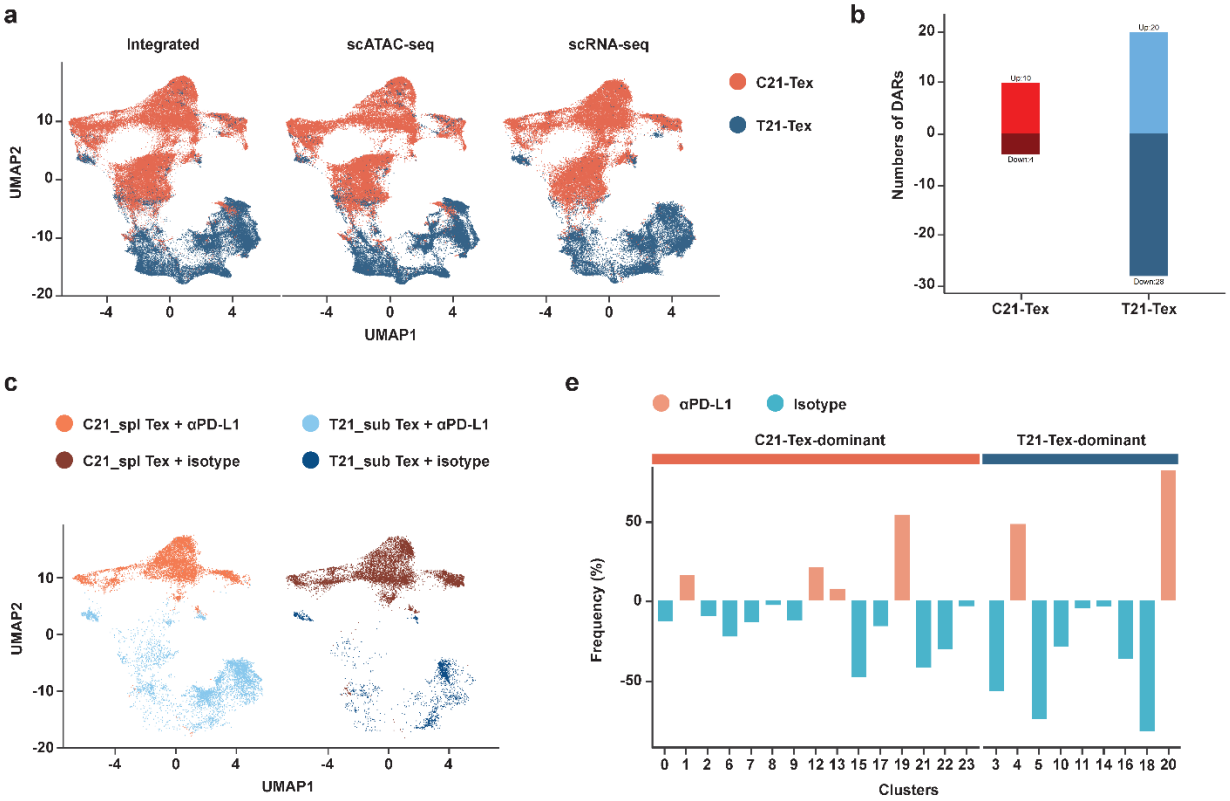


Figure S3. 7 Annotations of exhausted T cell dynamics upon PD-L1 blockade.

a, UMAP showing the integrated scRNA-seq and scATAC- seq profiles of isotype- and αPD-L1-treated Tex cells. Clusters are denoted by color labeled with inferred disease type (Red represents C21-Tex, and blue represents T21-Tex). Left: integrated scRNA-seq and scATAC-seq profiles. Middle: scRNA-seq profiles. Right: scATAC-seq profiles. b, Numbers of DARs between αPD-L1 treatment and isotype treatment in C21-Tex or T21-Tex. c, UMAP of C21_spl Tex and T21_sub Tex with αPD-L1 treatment or isotype treatment. Clusters of αPD-L1- and isotype-treated C21_spl Tex are denoted as light red and dark red, respectively. Clusters of αPD-L1- and isotype-treated T21_sub Tex are denoted as light blue and dark blue, respectively.

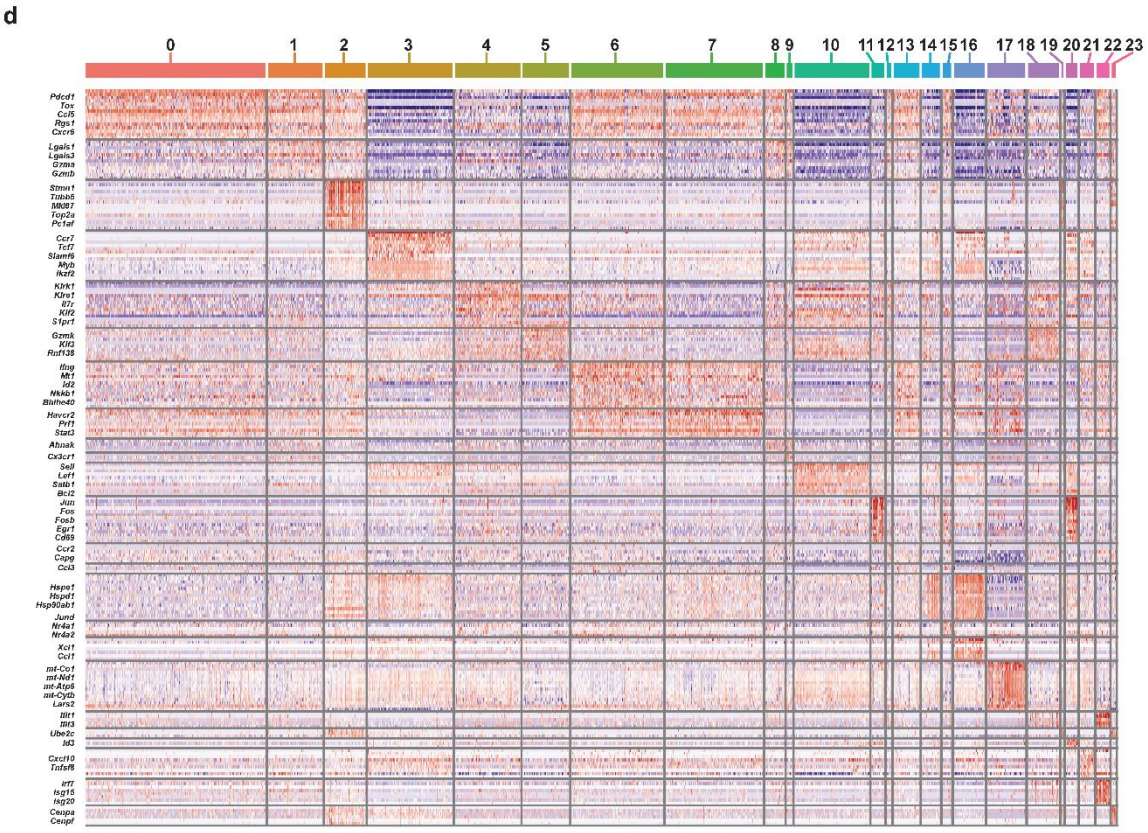


Figure S3. 7, Continued

d, Heatmap showing the representative genes of the 24 clusters. **e**, Frequency change of each cluster of total C21-Tex cells or T21-Tex cells upon αPD-L1 treatment. The frequency change value of each cluster is calculated as: % of αPD-L1 treatment in one cluster - % of isotype treatment in one cluster.

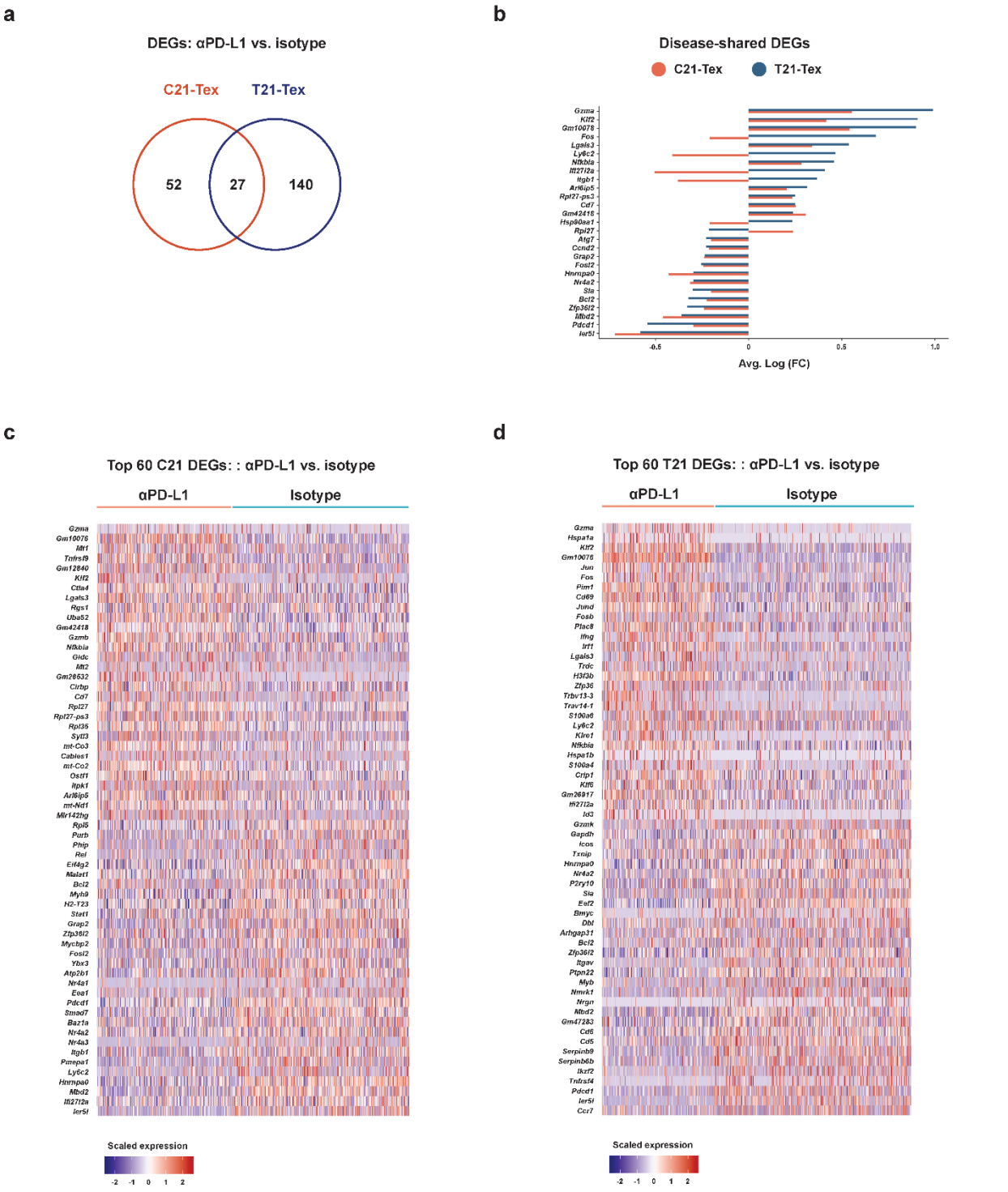


Figure S3. 8 Differential expression genes in disease-specific exhausted T cells by PD-L1 blockade.

a, Venn diagram representing the numbers of disease-specific and -shared αPD-L1-responding DEGs between C21-Tex and T21-Tex.

Figure S3. 8, Continued

b, The 27 shared αPD-L1-responding DEGs between C21-Tex and T21-Tex. **c, d**, Heatmap showing the top 60 αPD-L1-responding DEGs of C21-Tex (**c**) or T21-Tex (**d**), respectively.

Figure S3. 9 Identifications of disease-specific regulons responding to PD-L1 blockade in exhausted T cells.

a, Workflow of the regulon analysis. Briefly, the αPD-L1-responding DEGs were identified by scRNA-seq. Then, a pool of TFs was screened out by using enriched TF motifs analysis in the accessible gene loci of the αPD-L1-responding DEGs. Only these TFs whose mRNAs were also upregulated stayed in the final putative TF pool. Next, coexpression analysis identifies possible target genes of each putative TF. These target genes were further filtered by the corresponding putative TF motif enrichment in their accessible gene loci, which generateed the direct target genes whose loci contain accessible genes with that TF's consensus DNA binding motif. Thus, each regulon consisted of a differentially upregulated central TF and its putative target genes. All the regulons were then connected via co-targeted genes to generate a regulon network. By the same token, isotype- specific regulons were generated. **b**, Regulon correlations between C21-Tex-isotype(CI)/T21-Tex-isotype (TI), C21-Tex-αPD-L1 (CP)/T21-Tex-αPD-L1 (TP), CI/TP and CP/TI. The number denotes shared target genes between all pairings of indicated regulons.

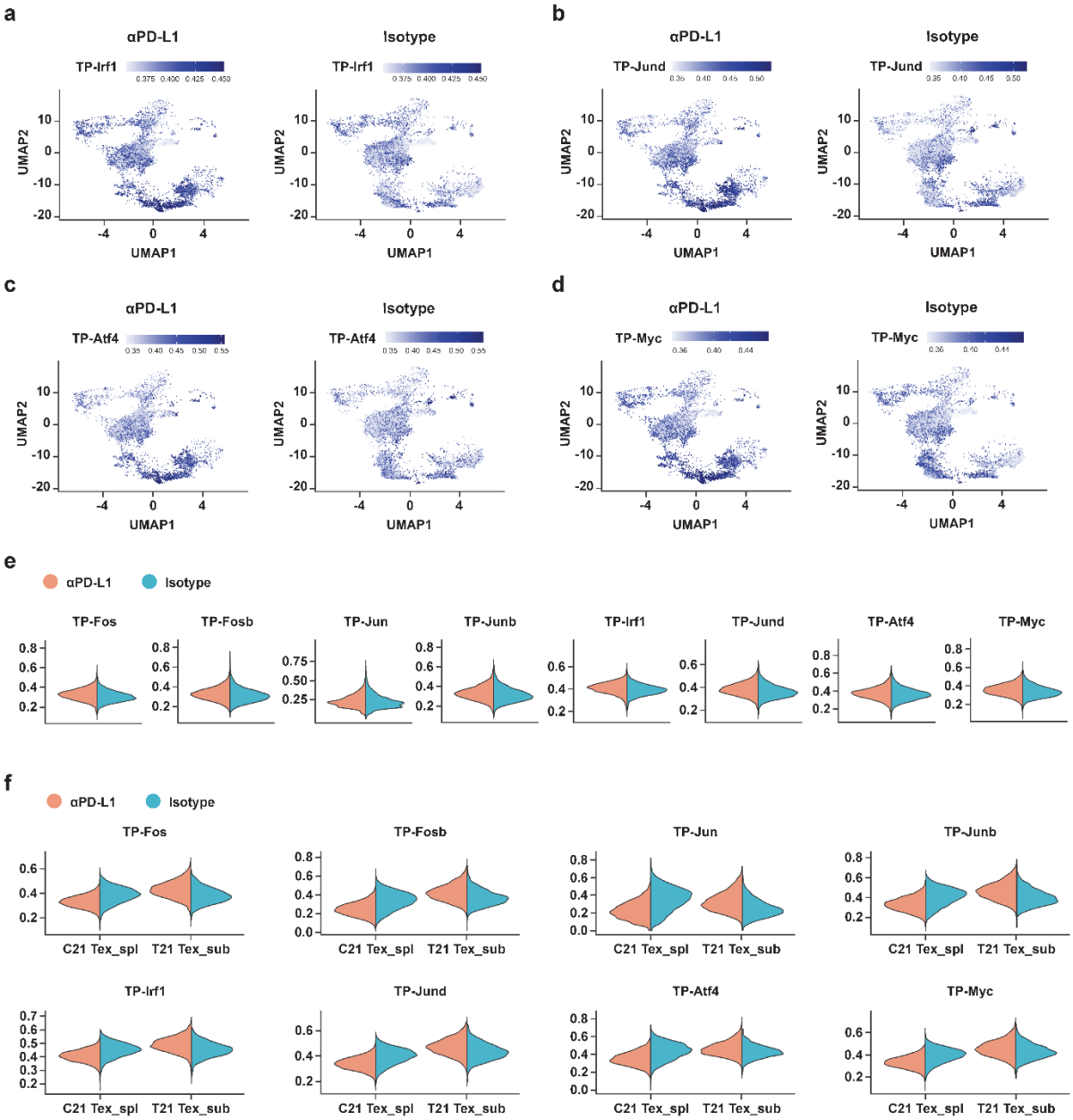


Figure S3. 10 Regulon analysis of tumor-specific exhausted T cells following αPD-L1 treatment.

a-d, UMAP showing the scores of PD-L1 ICB T21-Tex-derived Irf1- (**a**), Jund- (**b**), Atf4- (**c**) and Myc-centred regulons (**d**) in lung-resident Tex receiving αPD-L1 treatment (left) or isotype treatment (right). **e**, Scores of the regulon networks of αPD-L1-treated T21-Tex, including Fos-, Fosb-, Jun-, Junb-, Irf1-, Jund-, Atf4- and Myc-centred regulons, in lung- resident C21-Tex with isotype or PD-L1 ICB treatment. **f**, Scores of the regulon networks of αPD-L1-treated T21-Tex, including Fos-, Fosb-, Jun-, Junb-, Irf1-, Jund-, Atf4- and Myc-centred regulons, in C21_spl Tex and T21_sub Tex with isotype or PD-L1 ICB treatment.

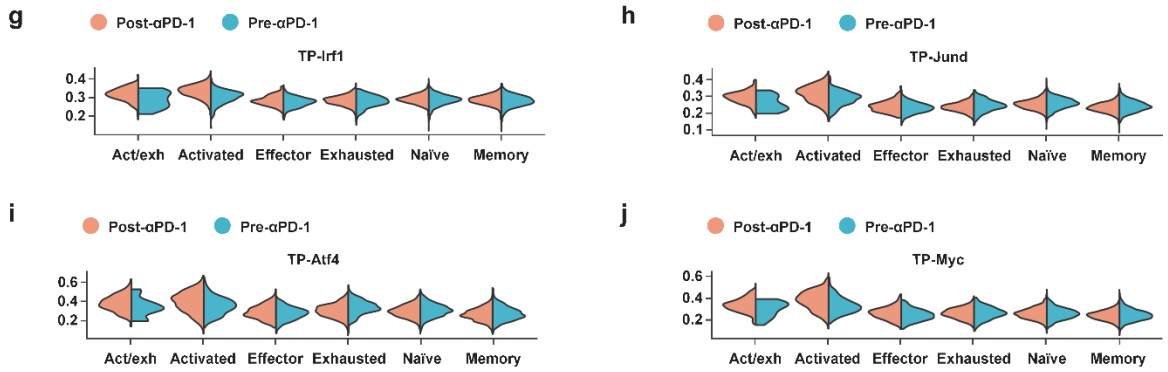


Figure S3. 10, Continued

g-j, Scores of the regulon networks of αPD- L1-treated lung-resident T21-Tex, including Irf1- (**g**), Jund- (**h**), Atf4- (**i**) and Myc-centred regulons (**j**), in indicated human BCC TIL subsets pre- and post-PD-1 ICB treatment.

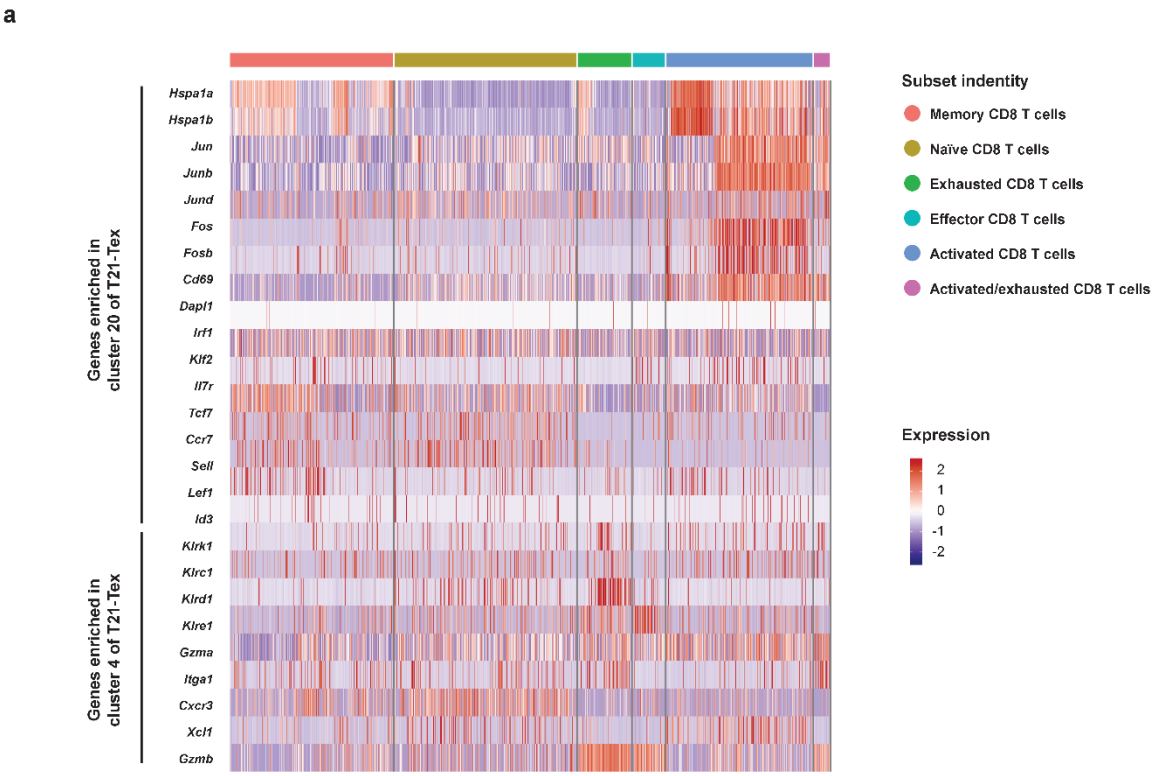


Figure S3. 11 Gene signatures of tumor-infiltrating CD8$^+$ T cell subsets of BCC patients.

a, Heatmap showing the expression levels of represented genes of T21-Tex clusters 4/20 in human BCC tumor-infiltrating CD8$^+$ T cell subsets.

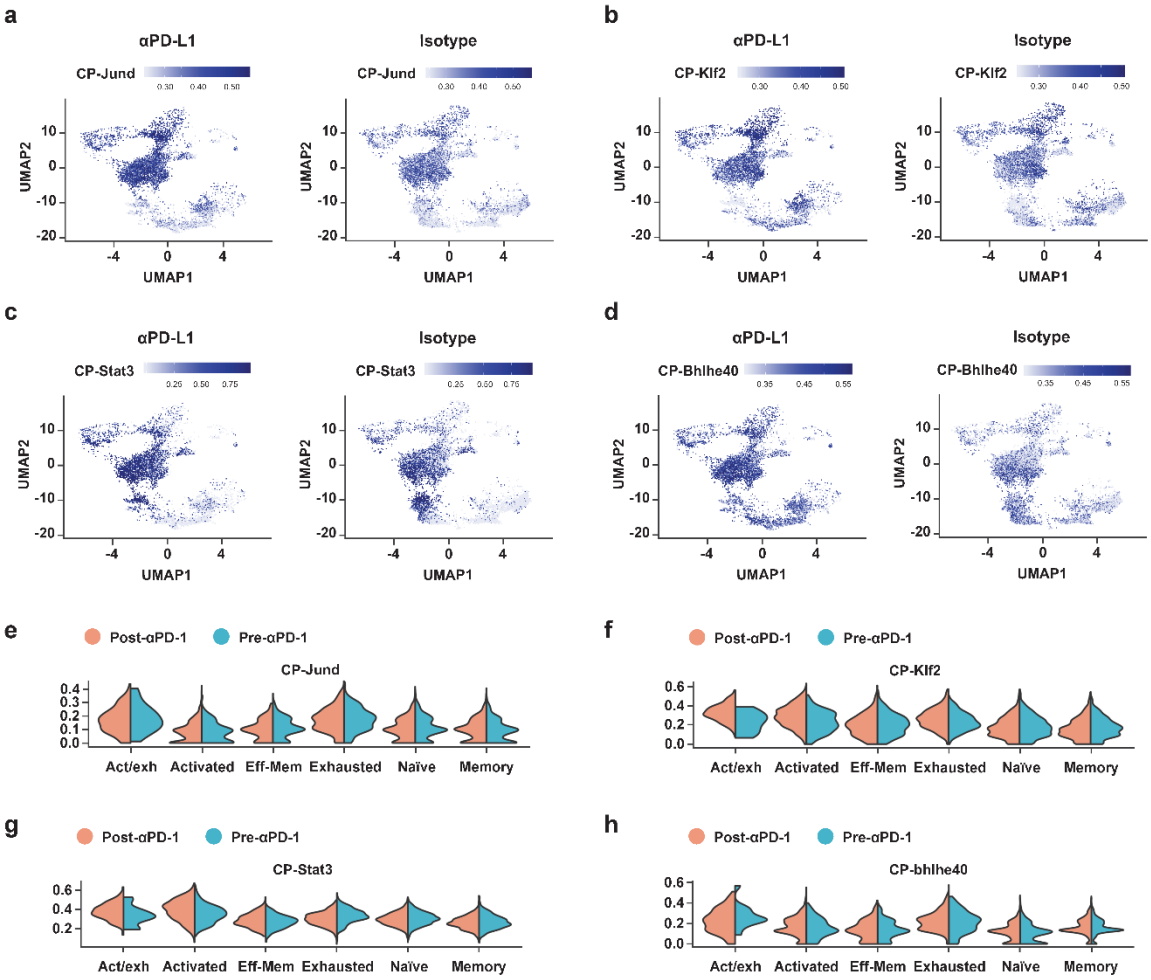
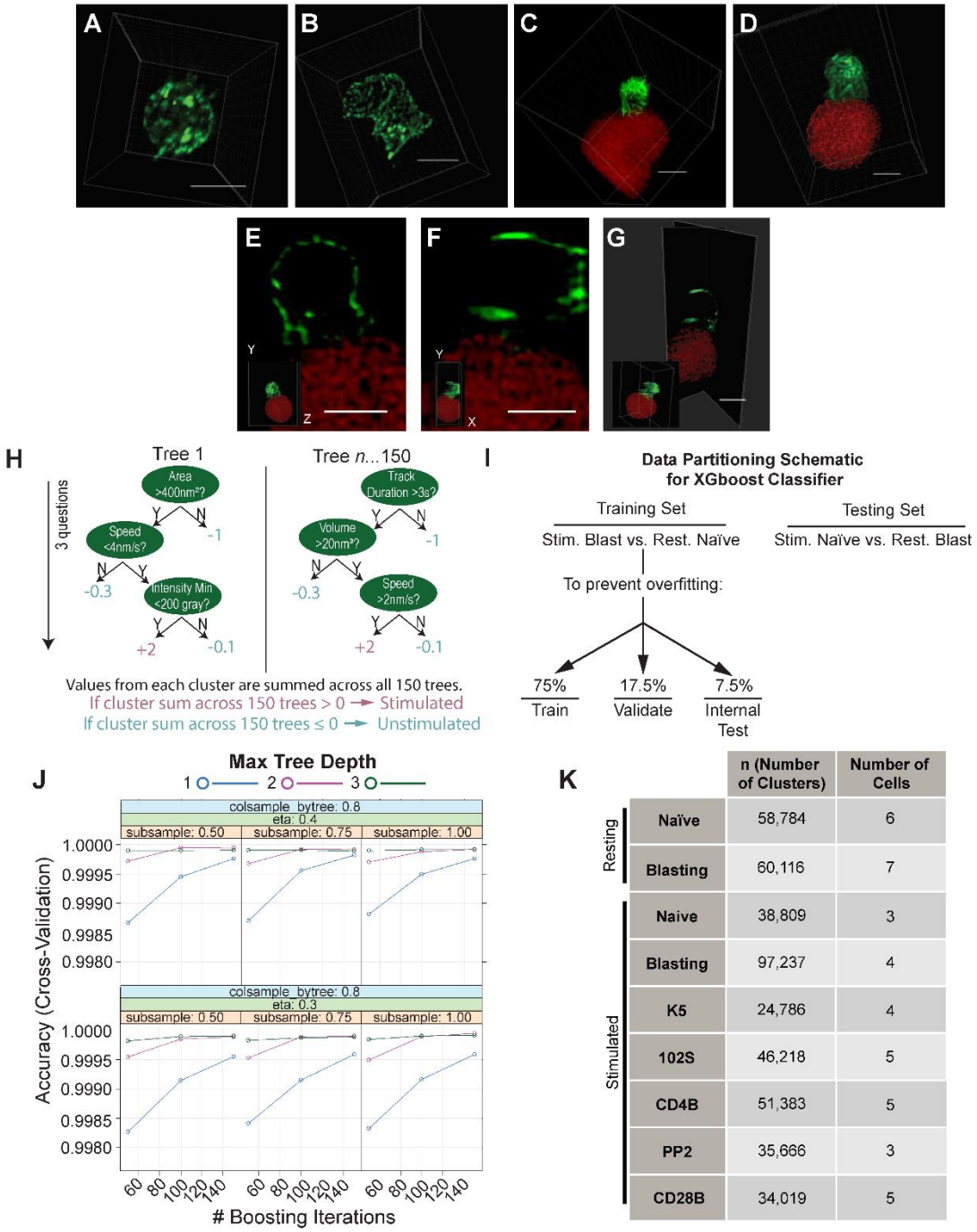


Figure S3. 12 Regulon analysis of chronic viral infection-specific exhausted T cells following αPD-L1 treatment.

a-d, UMAP showing the scores of PD-L1 ICB C21-Tex-derived Jund- (**a**), Klf2- (**b**), Stat3- (**c**) and Bhlhe40-centred regulons (**d**) in lung-resident Tex receiving αPD-L1 treatment (left) or isotype treatment (right). **e-h**, Scores of the regulon networks of αPD- L1-treated lung-resident C21-Tex, including Jund- (**e**), Klf2- (**f**), Stat3- (**g**) and Bhlhe40-centred regulons (**h**), in indicated human BCC TIL subsets pre- and post-PD-1 ICB treatment.

Supplementary Figures for Chapter 5

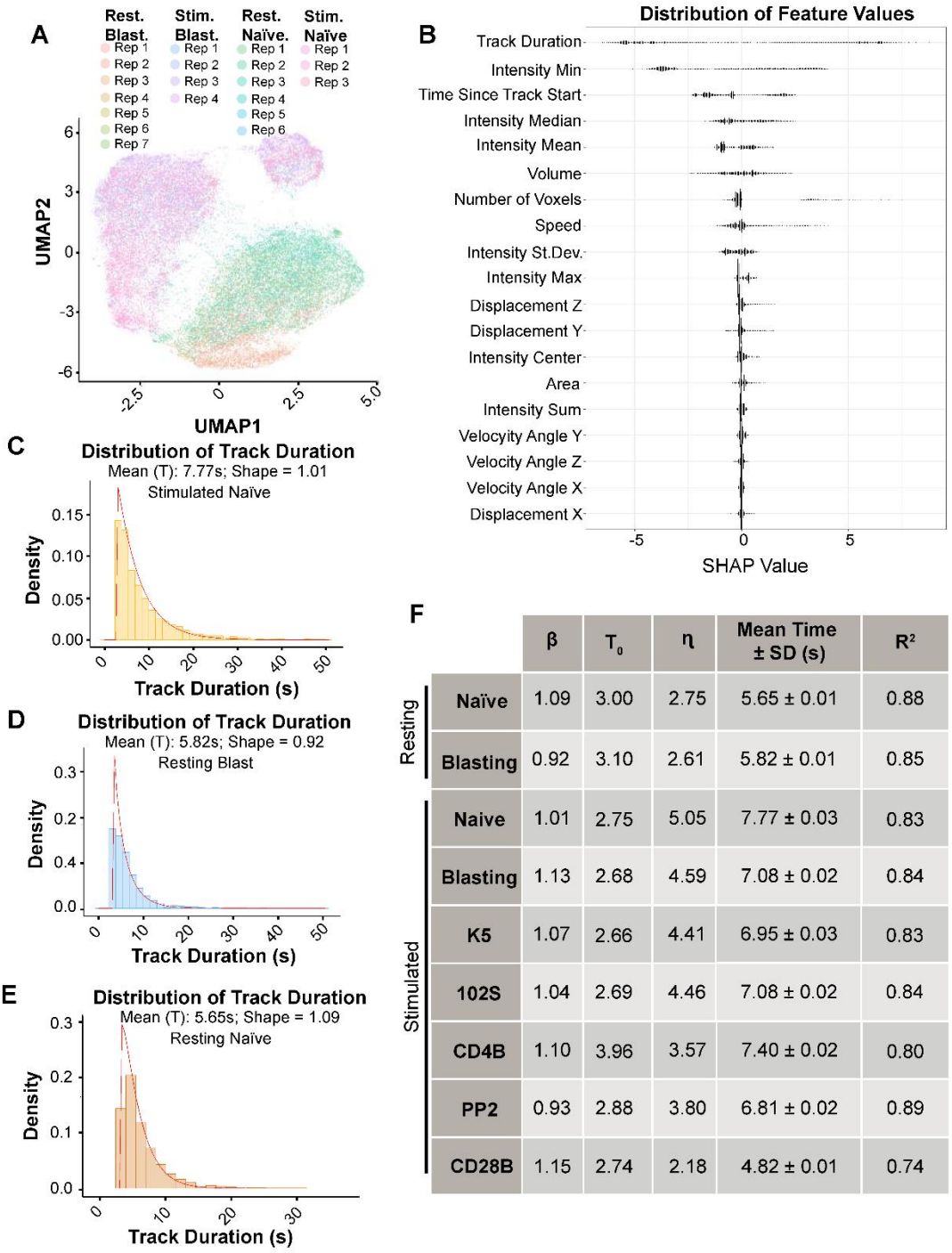


		n (Number of Clusters)	Number of Cells
Resting	Naïve	58,784	6
	Blasting	60,116	7
Stimulated	Naive	38,809	3
	Blasting	97,237	4
	K5	24,786	4
	102S	46,218	5
	CD4B	51,383	5
	PP2	35,666	3
	CD28B	34,019	5

Figure S5. 1 Training and validation of machine learning using high-dimensional TCR microcluster imaging data, related to Fig. 5.2

Figure S5. 1, Continued

(**A**) Three-dimensional rendering of a naïve resting $CD4^+$ T cell labeled with Alexa Fluor 488 (AF488)-conjugated anti-TCRβ Fab. Scale bar = 5 μm. (**B**) Three-dimensional rendering of a blasting resting $CD4^+$ T cell labeled with Alexa Fluor 488 (AF488)-conjugated anti-TCRβ Fab. Scale bar = 5 μm. (**C**) Three-dimensional rendering of a naïve $CD4^+$ T cell labeled with Alexa Fluor 488 (AF488)- conjugated anti-TCRβ Fab as it encounters an mCherry-CH27 cell loaded with agonist MCC peptide. Scale bar = 5 μm. (**D**) Three-dimensional rendering of a blasting $CD4^+$ T cell labeled with Alexa Fluor 488 (AF488)-conjugated anti-TCRβ Fab as it encounters an mCherry-CH27 cell loaded with agonist MCC peptide. Scale bar = 5 μm. (**E**-**G**) XY (**E**), YX (**F**), or dual (**G**) orthogonal slices of a representative $CD4^+$ T cell labeled with Alexa Fluor 488 (AF488)-conjugated anti-TCRβ Fab as it encounters an mCherry-CH27 cell loaded with agonist MCC peptide. Inset is a reference frame of the whole cell. Scale bars = 5 μm. (**H**) Cartoon describing decision tree ensemble structure and prediction algorithm used in XGboost binary classifier. (**I**) Schematic detailing data partitioning for XGboost classifier. (**J**) Diagram showing tuning of hyper-parameters during the training of the XGboost model. (**K**) Numbers of TCR microclusters from each cell used for XGboost classification and UMAP production.



F

		β	T_0	η	Mean Time ± SD (s)	R^2
Resting	Naïve	1.09	3.00	2.75	5.65 ± 0.01	0.88
	Blasting	0.92	3.10	2.61	5.82 ± 0.01	0.85
Stimulated	Naive	1.01	2.75	5.05	7.77 ± 0.03	0.83
	Blasting	1.13	2.68	4.59	7.08 ± 0.02	0.84
	K5	1.07	2.66	4.41	6.95 ± 0.03	0.83
	102S	1.04	2.69	4.46	7.08 ± 0.02	0.84
	CD4B	1.10	3.96	3.57	7.40 ± 0.02	0.80
	PP2	0.93	2.88	3.80	6.81 ± 0.02	0.89
	CD28B	1.15	2.74	2.18	4.82 ± 0.01	0.74

Figure S5. 2 UMAP, SHAP values, and Weibull distribution fitting, related to Fig. 5.2.

(A) UMAP from Fig. 5.2H color-coded by cell replicate. (B) Distribution of SHAP values of microclusters from Fig. 5.2J. (C-E) Weibull distribution fitting of microcluster track duration from stimulated naïve cells

Figure S5. 2, Continued

(C), resting blasting cells (D), and resting naïve cells (E). See also Fig. 5.2L. (F) Table of β, T_0, η, and mean duration obtained by Weibull distribution fittings of microcluster track durations from each cell group. Duration are presented as mean ± standard deviation (SD) and the goodness of fit was indicated by R2 values.

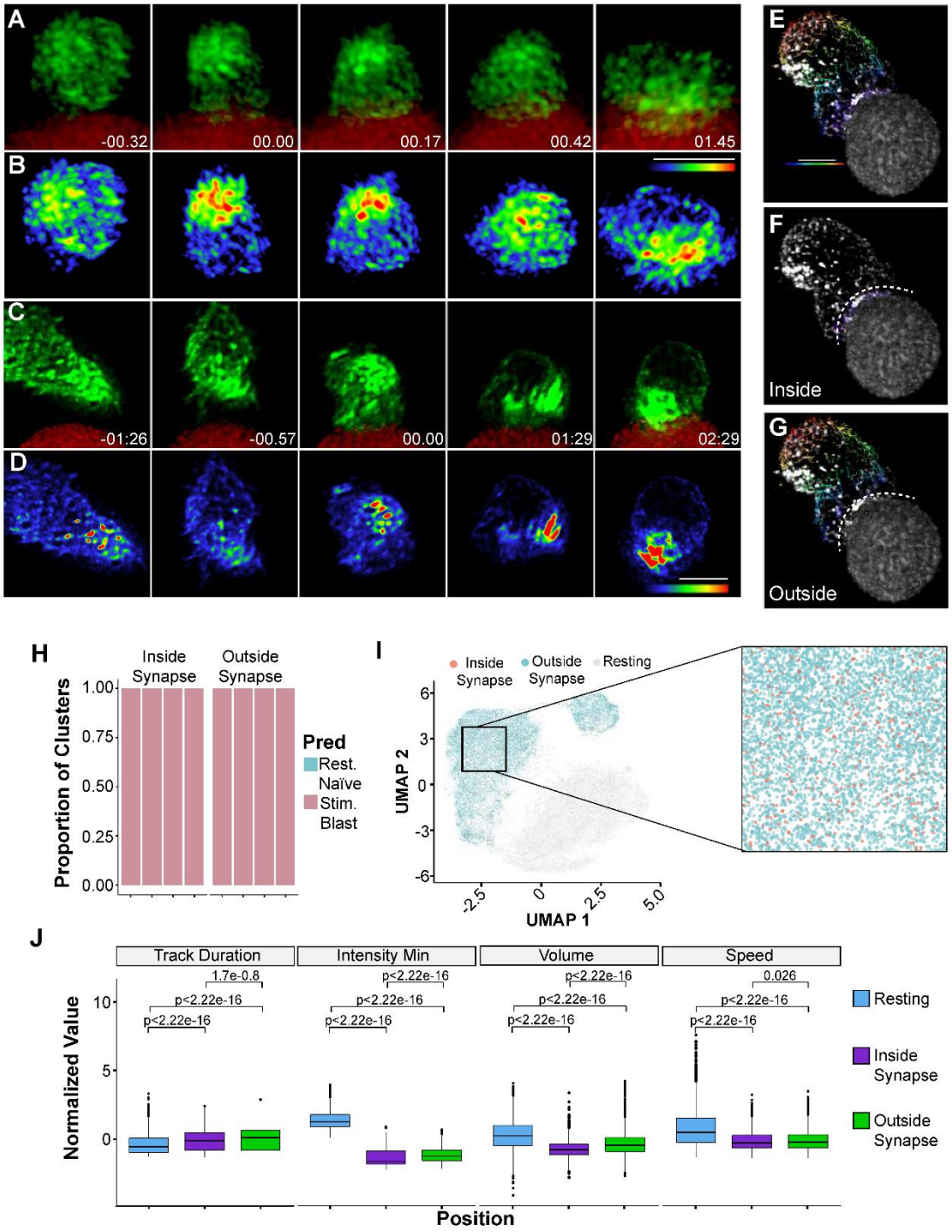


Figure S5. 3 LaMDA identifies TCR microcluster global change, related to Fig. 5.3.

(**A**) Time-lapse of a three-dimensional rendering of a naïve CD4^{+} T cell labelled with anti-TCRβAF488 Fab as it encounters a mCherry-CH27 cell loaded with MCC peptide. Time is shown as min:s with 00:00 as initial contact time. (**B**) Time lapse from (**A**) pseudo-colored by intensity from cold (0 gray) to hot (85 gray).

Figure S5. 3, Continued

(**C**) Time lapse of a three-dimensional rendering of a blasting $CD4^+$ T cell labelled with antiTCRβ-AF488 Fab as it encountered an mCherry-CH27 cell loaded with agonist MCC peptide. Time is shown as min:s. (**D**) Time lapse (from **C**) pseudo-colored by intensity from cold (0 gray) to hot (78 gray). Scale bar = 5 μm. (**E**-**G**) A distance map outward from the antigen presenting cell surface was created to determine the distance of each microcluster from the synapse. With a z-step size of 0.4 μm (representing maximum z-resolution) all microclusters ≤ 0.4 μm from the antigen presenting cell surface were labeled as "inside synapse", and those > 0.4 μm from the antigen presenting cell surface were labeled as "outside synapse". Stimulated blasting cell overlaid with dragon tails showing particle positions over the previous eight frames are color coded to show particle distance from antigen presenting cell. Cell is shown with all tracks (**E**), tracks inside synapse only, as defined by distance ≤ 0.4 μm (**F**), and tracks outside synapse only, as defined by distance > 0.4 μm (**G**). (**H**) XGboost prediction results on microclusters from stimulated blasting cells inside and outside of the synapse. Each bar represents an independent cell. (**I**) UMAP from Figure 5.2H colored by microcluster location on cell comparing inside synapse versus outside synapse. Inset shows enlarged region. (**J**) Box plots showing the distribution of track duration, intensity minimum, volume, and speed of microclusters from resting, inside synapse, and outside synapse. Center line represents median; box limits represent upper and lower quartiles; whiskers represent 1.5× interquartile range; points represent outliers. P values determined by Wilcoxon signed-rank test. Scale bars = 5 μm.

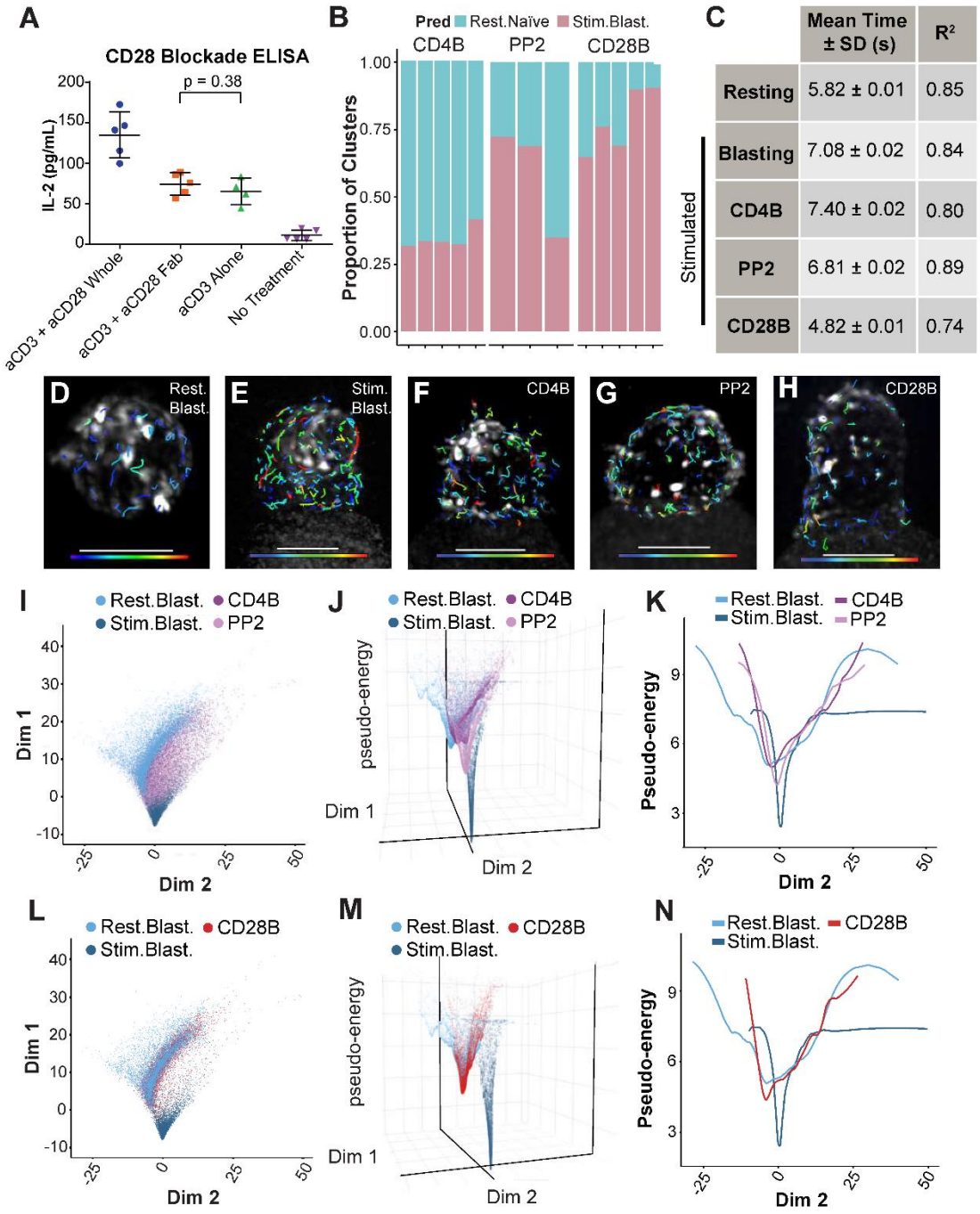


Figure S5. 4 LaMDA reveals changes in cell signaling states induced by perturbations, related to Fig. 5.4.

(A) CD28 blockade by anti-CD28 Fab diminishes T cell cytokine production. T-cell IL-2 cytokine production measured by ELISA experiment. 5C.C7 T cell blasts are stimulated by anti-CD3 mAb and anti-CD28 mAb, anti-CD3 mAb and anti-CD28 Fab, anti-CD3 mAb alone, and culture medium. Data are presented as mean ± standard deviation (SD).

Figure S5.4, Continued

(**B**) XGboost prediction results on microclusters from stimulated blasting cells treated with CD4 blockade (n = 51,383 microclusters), PP2 (n = 35,666 microclusters), and CD28 blockade (n = 34,019 microclusters). Each bar represents an independent cell. See also Fig. S5.1K. (**C**) Table of mean duration obtained by Weibull distribution fittings of microcluster track durations from each cell group. Data are presented as mean ± standard deviation (SD) and the goodness of fit is indicated by R2 values. See also Fig. S5.2F. (**D**-**H**) Three-dimensional renderings of resting blasting cell (**D**), stimulated blasting cell (**E**), stimulated blasting cell treated with CD4 blockade (**F**), stimulated blasting cell treated with PP2 (**G**), and stimulated blasting cell treated with CD28 blockade (**H**). All images are overlaid with dragon tails showing particle positions over the previous eight frames are color coded to show particle track duration (0-25 s). Scale bars = 5 µm. (**I**) Diffusion map built from the same selected features as Figure 5.3A, with colors indicating cell group. (**J**) 3D Pseudo-energy map of (**I**) of resting blasting cells, stimulated blasting cells, stimulated blasting cells with CD4 blockade (CD4B) and stimulated blasting cells treated with PP2 with colors indicating each cell group. (**K**) Projection of pseudo-energy map from (**J**) along dimension 2. (L) Diffusion map built from the same selected features as Figure 5.3A, with colors indicating cell group. (**M**) 3D pseudo-energy map of (**L**) of resting blasting cells, stimulated blasting cells, and stimulated blasting cells with CD28 blockade (CD28B) with colors indicating each cell group. (**N**) Projection of pseudo-energy map from (**M**) along dimension 2.