ETH zürich

DISS. ETH. NO. 27331

Differentiation trajectories of CD8$^+$ T cells towards exhaustion in chronic LCMV infection

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
Dario Cerletti
MSc Biologie, Systembiologie, ETH Zürich
born 31.10.1990
Citizen of Cham ZG

Zürich, 2021

accepted on the recommendation of:

Supervisor: Prof. Dr. M. Claassen

Co-supervisor: Prof. Dr. A. Oxenius

Referees: Prof. Dr. B. Becher, Prof. Dr. R. Gunawan

Institute of Molecular Systems Biology &
Institute of Microbiology, D-BIOL, ETH Zürich

English summary

CD8 T cells are an integral part of the adaptive immune response towards viral infections. These cells specifically identify virus-infected cells and eliminate them through cytotoxic mechanisms, and secrete cytokines to further hinder viral spread in neighboring cells and to recruit other immune cells. In an acute infection these cells proliferate and differentiate into multiple functional subsets, forming effector cells that lead to viral clearance and memory cells that persist after resolved infection and allow for quick re-expansion and acquisition of effector function upon reinfection. In chronic viral infections though, persistent T cell receptor (TCR) stimulation of CD8 T cells leads to increased and sustained expression of inhibitory surface receptors, leading to a continuous loss of effector functions, a process termed *exhaustion*. The Lymphocytic Choriomeningitis Virus (LCMV) infection model has been used to study exhausted of T cells in mice.

Recent findings revealed substantial heterogeneity in the CD8 T cell pool during chronic viral infections, describing subsets with either effector-like phenotypes, memory-like or terminally exhausted phenotypes, with memory-like cells exhibiting molecular and functional similarity with memory-cells and effector-like cells revealing similarity to effector cells formed after acute infection. The aim of this thesis is to resolve and characterize the sequence of cellular states during the differentiation processes that lead to cellular heterogeneity during the course of a chronic infection using single-cell RNA sequencing and computational tools for lineage inference.

We first investigated the extent to which exhausted CD8 T cells are able to sense and transmit signals via their TCR, both *in vitro* and *in vivo*. While *in vitro* stimulation of isolated CD8 T cells from chronically infected mice revealed strong transcriptional activation after TCR engagement, measured by RNAseq, *in vivo* TCR signaling during established chronic infection was strongly impaired and almost absent as a result of inhibition by highly expressed co-receptors on the surface of these cells in the *in vivo* tissue context.

To study progression of CD8 T cells towards this exhausted state, we generated a high-resolution time series scRNAseq dataset, comprising samples spanning from the beginning of chronic infection until development of highly exhausted cells around day 21 post infection with high dose LCMV Clone-13. For data driven lineage inference across all samples, we developed Cytopath that uses RNA velocity to simulate likely cell transition based on information of spliced and unspliced RNA counts and aligning them to generate global trajectories. We identified a terminally exhausted cell state with high expression of co-inhibitory receptors, and a TCF1 expressing memory-like cell state as endpoints of differentiation. Lineage inference with Cytopath revealed two main developmental trajectories towards either the terminally exhausted or the memory-like endpoint, that bifurcate around day five post infection.

We identified surface protein markers that allowed us to isolate cells from both trajectories before and after the bifurcation point. Adoptive transfer of cells before the branch point into infection matched hosts showed phenotypic plasticity of these cells towards both the memory-like and the exhausted end-point, confirming initially shared differentiation paths. Transfer of cells isolated after the branch had committed to the exhausted end point maintained their exhausted phenotype after transfer. After transfer of memory-like cells we recovered both terminally exhausted as well as memory-like cells, indicating transitions between these cell states, that we did not observe in our single-cell RNA sequencing time series data. We further investigated potential drivers of these transitions by transferring cells from the three branches into hosts infected with a Clone-13 mutant that does not induce TCR stimulation in P14 cells. Without antigenic stimulation, recovered cells after transfer of memory-like cells largely retained their phenotype, indicating that their transition to terminally exhausted cells is antigen driven. Cells isolated before the branching point also retained their uncommitted phenotype, highlighting antigenic TCR stimulation as a major driver of differentiation in chronic infection.

Our findings complement recent studies of CD8 T cell differentiation and add to a comprehensive model of CD8 T cells lineage relations in chronic infections.

Zusammenfassung

CD8 T Zellen sind ein integraler Bestandteil der adaptiven Immunantwort auf Virusinfektionen. Diese Zellen identifizieren spezifisch virusinfizierte Zellen und eliminieren sie durch zytotoxische Mechanismen und sezernieren Zytokine, um die Virusausbreitung in Nachbarzellen weiter zu hemmen und weitere Immunzellen zu rekrutieren. Bei einer akuten Infektion proliferieren diese Zellen und differenzieren sich in mehrere funktionelle Subpopulationen, wobei Effektorzellen gebildet werden, die zur Viruselimination führen und Gedächtniszellen, die nach einer überwundenen Infektion fortbestehen und nach einer Reinfektion rasch expandieren und Effektorfunktion erwerben. Bei chronischen Virusinfektionen hingegen führt die persistierende Stimulation der CD8-T-Zellen durch den T-Zell-Rezeptor (TCR) zu einer erhöhten und anhaltenden Expression von inhibitorischen Oberflächenrezeptoren, was zum kontinuierlichen Verlust der Effektorfunktionen führt. Dieser Prozess wird als *Erschöpfung* bezeichnet. Das Infektionsmodell des Lymphozytären Choriomeningitis-Virus (LCMV) wurde zur Untersuchung der Erschöpfung von T -Zellen in Mäusen verwendet.

Neuere Befunde zeigten eine beträchtliche Heterogenität im CD8-T Zell-Pool während chronischer Virusinfektionen, wobei Subpopulationen mit entweder effektorähnlichen Phänotypen, gedächtnisähnlichen oder terminal erschöpften Phänotypen
beschrieben wurden, wobei gedächtnisähnliche Zellen molekulare und funktionelle Ähnlichkeit mit Gedächtniszellen und effektorähnliche Zellen Ähnlichkeit mit Effektorzellen aufweisen, die nach einer akuten Infektion gebildet wurden. Ziel dieser Arbeit ist die Aufschlüsselung und Charakterisierung zellulärer Zustände während der Differenzierungsprozesse, die im Verlauf einer chronischen Infektion zu zellulärer Heterogenität führen, mit Hilfe von Einzelzell-RNA-Sequenzierung und eines Algorithmus zur Abschätzung von Differenzierungslinien.

Wir untersuchten zunächst, inwieweit erschöpfte CD8 T Zellen in der Lage sind, Signale über ihren TCR, sowohl *in vitro* als auch *in vivo*, zu erfassen und zu übertragen. Während die *in vitro*-Stimulation isolierter CD8 T-Zellen aus chronisch infizierten Mäusen eine starke Transkriptionsaktivierung nach TCR-Stimulation zeigte, gemessen mittels RNAseq, zeigte sich der TCR-Signalweg *in vivo* während einer etablierten chronischen Infektion stark beeinträchtigt und fehlte fast völlig, was auf die Hemmung durch hochexprimierte Co-Rezeptoren auf der Oberfläche dieser Zellen im *in vivo*-Gewebekontext zurückzuführen ist.

Um die Entwicklung von CD8 T Zellen in Richtung dieses erschöpften Zustands zu untersuchen, generierten wir einen hochauflösenden scRNAseq-Zeitreihendatensatz, der Proben vom Beginn der chronischen Infektion bis zur Entwicklung terminal erschöpfter Zellen um den Tag 21 nach der Infektion mit hochdosiertem LCMV-Klon 13 herum umfasste. Für die datenbasierte Konstruktion von Differenzierungslinien über alle Proben hinweg entwickelten wir *Cytopath*, das die *RNA-Geschwindigkeit* zur Simulation wahrscheinlicher Zellübergänge auf der Grundlage von Informationen über gespleißte und ungespleißte RNA-Transkripte verwendet und diese zeitlich ausrichtet zur Generierung globaler Trajektorien. Wir identifizierten einen terminal erschöpften Zellzustand mit hoher Expression von Koinhibitionsrezeptoren und einen gedächtnisähnlichen Zellzustand, erkennbar durch TCF1 Expression, als Endpunkte der Differenzierung. Die von *Cytopath* geschätzten Differenzierungslinien ergaben zwei Hauptentwicklungspfade, entweder in Richtung des terminal erschöpften oder des gedächtnisähnlichen Endpunkts. Diese Pfade gabelten die sich um den fünften Tag nach der Infektion in einem Bifurkationspunkt.

Wir identifizierten Oberflächenproteinmarker, die es uns erlaubten, Zellen aus beiden Pfaden vor und nach dem Bifurkationspunkt zu isolieren. Der adoptive Transfer von Zellen vor dem Bifurkationspunkt in infektionsangepasste Wirtsmäuse zeigte eine phänotypische Plastizität dieser Zellen sowohl in Richtung des gedächtnisähnlichen als auch des erschöpften Endpunktes, was die anfänglich gemeinsamen Differenzierungspfade bestätigte. Die transferierten Zellen, die isoliert wurden, nachdem der Zweig auf den erschöpften Endpunkt zusteuerte, behielten ihren erschöpften Phänotyp nach dem Transfer bei. Nach dem Transfer gedächtnisähnlicher Zellen fanden wir sowohl terminal erschöpfte als auch gedächntisähnliche Zellen, was auf Übergänge zwischen diesen Zellzuständen hinweist, die wir in unseren Zeitreihendaten der Einzelzell-RNA-Sequenzierung nicht beobachtet hatten. Wir untersuchten mögliche Treiber dieser Übergänge, indem wir Zellen aus den drei Zweigen in Wirtsmäuse transferierten, die mit einer Klon-13-Mutante infiziert waren, die in P14-Zellen

keine TCR-Stimulation induziert. Ohne Antigen-Stimulation behielten die nach dem Transfer von gedächtnisähnlichen Zellen gefundenen Zellen ihren Phänotyp weitgehend bei, was darauf hindeutet, dass ihr Übergang zu terminal erschöpften Zellen antigeninduziert ist. Auch Zellen, die vor dem Verzweigungspunkt isoliert wurden, behielten ihren undifferenzierten Phänotyp bei, was darauf hinweist, dass die antigene TCR-Stimulation ein wichtiger Faktor für die Differenzierung bei chronischen Infektionen ist.

Unsere Ergebnisse ergänzen die jüngsten Studien zur CD8-T Zell-Differenzierung und erweitern ein umfassendes Modell der CD8-T Zell-Populationsbeziehungen bei chronischen Infektionen.

Contents

Chapter 1

Introduction

1.1 Biological Background

1.1.1 Role of CD8 T cells in immune response to viral infection

CD8 T cells are in integral part of the adaptive immune system that protects our body from foreign pathogens and also eliminates tumor cells. Intracellular pathogens like viruses reproduce strictly inside cells where they are shielded from most immune cells and effector molecules such as antibodies. CD8 T cells have the ability to specifically detect viral peptides presented on the surface of infected cells in combination with self MHC molecules and upon recognition respond by killing the target cells after having undergone differentiation and maturation to effector cells. Additional to this cytotoxic function, CD8 T cells also signal through the secretion of cytokines to induce antiviral responses in neighboring cells and to directly restrict viral replication in infected cells. Cytotoxicity is crucial to eliminate viruses from the host but requires tight regulation, since excessive killing of host cells can lead to immune-pathological damage and in severe cases even death of the host.

The process of viral infection of a host cells is initiated by binding of the virus (usually via specific viral spike proteins) to specific host cell receptors, followed by either direct membrane fusion of enveloped viruses or endocytosis, followed by viral envelope fusion with the endosomal membrane. These fusion events liberate the viral core complexes and allow the viral genomes in form of DNA or RNA to enter the host cell cytoplasm where replication of viral genomes and production of viral proteins in initiated. During this process, infected cells will sense the presence of viral nucleic acids by pattern recognition receptors (PRR). These include membrane bound toll-like receptors (TLRs) which survey the endoplasmic compartment and Rig-I, MDA-5, cGAS or AIM that survey the cytoplasm [Bowie, 2007]. These receptors activate signaling networks that lead to expression of pro-inflammatory cytokines that recruit immune cells, including CD8 T cells. Additionally, type I & III interferons are secreted that activate "antiviral states" in neighboring cells [Carty et al., 2020].

During the process of viral protein synthesis, some of these *de novo* synthesized proteins will be degraded (usually via the proteasome) and resulting peptides are actively transported into the ER where they can associate with newly synthesized MHC class I molecules. These peptide / MHC class I complexes will then be transported to the plasma membrane where the will be presented to CD8 T cells. Naïve CD8 T cells that carry a T cell receptor (TCR) specific for a given peptide MHC complex, present in secondary lymphoid organs, will engage with these peptide MHC class I complexes on the surface of activated antigen presenting cells (APCs). TCR stimulation, co-stimulation via CD80/86 to CD28 and signaling through pro-inflammatory cytokines will activate the naïve T cell. This process entails massive proliferation (clonal expansion) and differentiation into functional effector cells. This differentiation process is accompanied by transcriptional, epigenetic and metabolic changes [Böttcher and Knolle, 2015]. Effector T cells exhibit strong cytotoxic activity and the ability to produce and secrete pro-inflammatory cytokines such as IFNγ and TNF. They shift metabolism towards glycolysis and are short lived. Another subset that arises early upon activation are memory-progenitor cells that maintain oxidative phosphorylation as their main metabolic pathway, and they show little cytotoxic activity and divide less compared to cells that differentiate into effector cells.

After clearance of the virus, around day 10 after infection, effector cells undergo apoptosis and the whole CD8 T cell pool contracts in numbers. The memory progenitors form the pool of long-lived memory cells that survive independent of antigen but depend an IL-7 and IL-15

signaling. Upon reinfection with the same pathogen, they can readily re-expand and acquire effector functions.

The exact lineage relations of effector and memory CD8 T cells is still under debate. There is evidence, that high antigen exposure and strong TCR stimulation lead to an effector phenotype, whereas low TCR stimulation would result in memory cells.

Single cell transcriptomics experiments in acute infection of lymphocytic choriomenengitis virus (LCMV) indicate that effector and memory phenotypes may already be determined after the first cell division of the naïve CD8 T cell [Kakaradov et al., 2017].

1.1.2 CD8 T cell exhaustion in chronic infection

Some viruses induce chronic infections, like HIV and HCV in humans or high doses of LCMV Clone-13 in mice. Chronic infections, by definition, involve continued viral replication and inflammation over prolonged periods of time. This results in sustained high antigenic exposure of T cells and in up-regulation of a multitude of co-inhibitory surface receptors. These receptors dampen TCR stimulation and signaling through ligand competition, direct interaction with the TCR or its downstream signaling as well as transcriptional modulation. Over time, epigenetic modifications in promotor regions of cytokine genes will manifest. The combined effect of inhibited TCR signaling as well as epigenetic down-regulation results in these cells having diminished abilities to produce and secrete cytokines, they are less proliferative *in vitro* and are collectively called "exhausted" T cells [Moskophidis et al., 1993].

T cell exhaustion is a gradual process that develops over the first three weeks after the start of chronic infection [Wherry et al., 2007] during which the secretion if cytokines INFγ and TNF is decreasing and apoptosis is more prevalent. Cytotoxic capacity of these cells is debated, since they show reduced killing activity *in vivo* but can kill target cells *in vitro* [Agnellini et al., 2007] or transferred target cells that do not express ligands for the co-inhibitory receptors [Graw et al., 2011, Sandu et al., 2020a]. This loss of effector functions is accompanied by an increase in the expression of co-inhibitory surface receptors such as PD-1, LAG-3, CD244 and CD160 in response to continued stimulation of the TCR. Despite these receptors also being transiently upregulated during acute infection, in chronic infection their expression is maintained at high levels and strongly dampens TCR signaling.

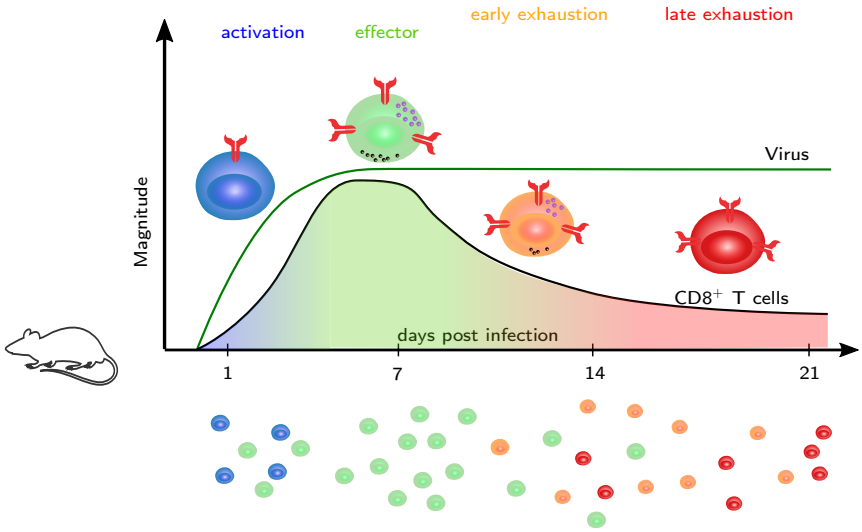


Figure 1.1: Overview of the CD8 T cell response in infections with persistent viral replication. After the initial activation phase with formation of effector cells, exhaustion increasingly reduces effector functions and cell number decrease. The T cell pool is heterogeneous at any time-point of the infection due to asynchronous activation and different history of the cells.

The profound regulation of CD8 T cell function in the context of chronic viral infections serves to attenuate the detrimental effect of excessive cytotoxic activity. Abolishing the signaling of these co-inhibitory receptors through administration of blocking antibodies, mainly through anti-PD-1 and anti-CTLA4, releases the inhibition of signaling and results in excessive cytotoxicity. This uncontrolled killing by CD8 T cells in chronic infection, where a huge number of cells are infected, can lead to lethal immuno-pathology [Frebel and Oxenius, 2013a] and death of the host.

Although single-cell studies recently being prevalent in immunology, many characteristic features of exhausted T cells have been identified using bulk transcriptional measurements first. A comprehensive molecular analysis has revealed a multitude of transcriptional differences between effector cells from acute infection and exhausted cells from chronic infection with LCMV [Wherry et al., 2007]. These include vast changes in expression of inhibitory & surface receptors (Pdcd1, Ctla4, Lag3), molecules associated with homing and migration (Ccl5, Cxcr3), survival and proliferation (Bcl2, Casp3) as well as metabolism (Acas2l, Pdha1).

Studies using multiple viral strains to modulate the amount of antigen stimulus CD8 T cells receive have shown that excessive antigen stimulation is the source of exhausted features of these cells [Utzschneider et al., 2016b]. Transcriptional profiling of T cells with high and low antigen stimulation, respectively, have revealed the transcription factor TOX as a major transcription factor required for the exhausted phenotype in chronic infection. It is indispensable in generation of the exhausted phenotype and drives epigenetic remodeling, affecting expression of IFNγ, T cell homing and NFAT signaling [Alfei et al., 2019]. Genetic deletion of TOX restores effector functions, namely secretion of INFγ and TNF.

Flow cytometry experiments using high affinity antibodies for PD-1 discovered that expression of PD-1 is not uniform across the pool of CD8 T cells in chronic infection but that there exists a sub-population of cells that expresses only intermediate levels [Blackburn et al., 2008a].

1.1.3 Phenotypic cell states in exhaustion

First indications of multiple stages of exhaustion came from heterogeneous responses to anti-PD-L1 treatment, where only a subset expressing lower amounts of co-inhibitory receptors would expand and (re-)acquire effector functions [Blackburn et al., 2008a]. These intermediately exhausted cells (PD-1^{int}) were separate to the terminally exhausted cells (PD-1^{high}) that do not respond to anti-PD-L1 treatment.

The terminally exhausted T cells make up the largest fraction in the T cell pool and show the highest degree of exhaustion in terms of number of different co-inhibitory receptors they express and the abundance of these. Cytokine production and secretion is reduced in these cells compared to effector cells from acute infection and cytotoxicity is not apparent *in vivo*. Whereas cytokine regulation seems to be inhibited on the epigenetic level [Sen et al., 2016] and post-transcriptional modification [Mackerness et al., 2010], reduced TCR signaling through strong activity of co-inhibitory receptors blocks cytotoxicity [Sandu et al., 2020a].

Advances in single-cell measurement technology have further helped to structure the exhausted T cell pool in chronic infections using mass cytometry [Bengsch et al., 2018] and single-cell transcriptomics [Chen et al., 2019, Zander et al., 2019].

An important subset of virus-specific T cells was described to show memory-like features, even in chronic infections [Utzschneider et al., 2016a]. These cells express the transcription factor TCF1 and share properties with central memory cells in acute infection, such as high proliferative capacity, higher longevity and ability to differentiate into effector-like and exhausted cells. Opposed to memory cells developing in the context of acutely resolved infection, these memory-like cells in the context of chronic infections additionally show many features of exhausted cells, like expression of the co-inhibitory receptors PD-1 and LAG-3.

Further, a subset was described that maintains an effector-like phenotype even in chronic infection [Hudson et al., 2019]. These cells express the surface receptor CX3CR1 and show less exhausted properties and more effector-like features. Namely, they had higher expression of proliferative markers (*Mki67*) and Granzyme B and after *in vitro* stimulation secreted more cytokines (IFNγ). Additionally these cells seemed to reside preferentially in in the lung as well as in proximity to vasculature [Sandu et al., 2020b].

1.1.4 Lineage relationships of Exhausted CD8 T cells

It is still under debate how the different CD8 T cell subsets arise and differentiate in chronic LCMV infection.

Multiple studies have established a connection between the memory-like and the terminally exhausted populations through transfer experiments, where memory-like cells give rise to terminally exhausted cells [Utzschneider et al., 2016a, Wu et al., 2016]. However the time-point at which the memory-like population forms and how it is connected to the terminally exhausted lineage, remains unclear.

The origin of the described effector-like CX3CR1 population is unclear so far. Adoptive transfer studies of TCF1$^+$ cells into infection-matched hosts showed appearance of this population transiently, before differentiating into terminally exhausted cells. They were therefore placed as an intermediate population between memory-like cells and terminally exhausted cells [Hudson et al., 2019, Beltra et al., 2020]. Another study adoptively transferred TCF1 defective cells into chronically infected hosts but still observed appearance of a CX3CR1$^+$ cell population, indicating that they compose an independent lineage [Raju et al., 2020].

Several studies have investigated early time-points during chronic LCMV infection [Zander et al., 2019, Yao et al., 2019] using singe-cell transcriptome data but none has connected it towards late time-points. This makes it difficult to fully capture the lineage relation between effector-like, memory-like and terminally exhausted cells. A longitudinal data-set at high detail single-cell resolution, fit for this purpose, has not yet been available.

1.2 Single-cell RNA sequencing

Studying differentiation processes poses experimental challenges, as they involve changes in many genes and proteins simultaneously and, in the case of CD8 T cell differentiation, are asynchronous. High dimensional single-cell technologies, like flow and mass cytometry, have expanded the possibilities to study complex developmental processes on the single cell level. Single-cell transcriptome sequencing has the additional advantage of being untargeted and genome wide, making it the current prime choice for discovering novel differentiation mechanisms.

Dropping cost of genome sequencing has led to a vast increase in available transcriptome sequencing technology. Single-cell RNA sequencing allows genome-wide quantification of mRNA transcripts to characterize cellular states with high detail. Capturing transcriptional heterogeneity between individual cells adds a new level of depth to the understanding of biological processes. Identifying transcripts from single requires cell lysis & mRNA capture, reverse transcription, 2nd strand synthesis, cDNA amplification and library construction. The sequencing of the library is done using next-generation sequencing technology.

Well-based technologies like Smart-seq2 capture individual cells using microfluidic chips in microliter volume chambers and perform lysis, reverse transcription and 2nd strand synthesis in these micro-wells. These methods offer high sensitivity, full-length reads but have limited throughput of cells, typically in the hundreds. This results also in higher costs per sequenced cell but allows better coverage of the transcriptome.

10x Genomics Chromium is a droplet-based sequencing technology that encapsulates single cells in water droplets in an oil emulsion using microfluidics (Fig. 1.2). Oligo-dT primer tagged beads carry a bead-specific barcode as well as a unique molecular identifier (UMI) for each tag. Lysis buffer, beads and cells are encapsulated in droplets at a specific ratio to ensure the majority of droplets carry a maximum of one cell and one bead. Lysis, reverse transcription and 2nd strand synthesis takes place in these droplets directly on the microfluidic device. The barcoded beads allow multiplexing of the subsequent cDNA amplification and library preparation. This method allows capturing a high number of cells (usually thousands) and is very cost efficient. The droplet bases reaction however reduces the number of mRNA molecules captured and therefore also the number of detected genes per cell. 10x genomics detects between 1000 and 4000 genes, compared to 4000 to 8000 genes with Smart-seq2. Additionally it is not possible to obtain full length reads [Ziegenhain et al., 2017].

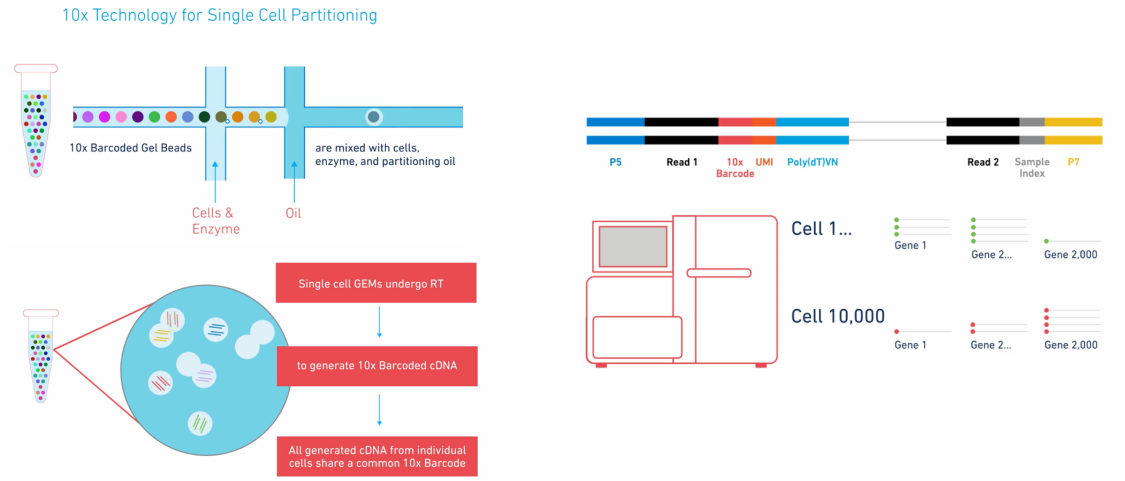


Figure 1.2: 10x genomics is a droplet based technology, that uses barcoded beads to capture mRNA from lysed single-cells in oil-emulsion droplets. Sequencing is preformed using next-generation sequencing technology.

For the purpose of studying T cells in chronic infection, 10x Genomics has the advantage of being more high-throughput, allowing for analyzing more cells and therefore higher chance of detecting rare or transient cell states. Usually full-length sequencing is not required, as genes can easily be mapped to a reference genome. However, the capture rate of mRNA transcripts is relatively low in 10x Genomics (5-20%) therefore many transcripts will not be sequenced at all.

Pre-processing of scRNAseq data usually includes filtering, normalization and dimensionality reduction.

Filtering addresses the removal of cells of low-sequencing quality where lysis and mRNA capture failed. These cells are identified by a low number of genes detected as well as a high fraction of mitochondrial genes. Additionally, genes that are very rarely detected are excluded as they probably do not contribute to the process or are not expressed in this cell type.

Normalization is required to compensate for variation in lysis and capture efficiency between cells. Total transcript counts can be drastically different and therefore the absolute counts per gene are mainly a result of technical capture efficiency. For this, cells are scaled by their total transcript count to compute transcript frequencies per cell.

Dimensionality reduction is required to make the transcript space, that usually holds thousands of dimensions, more amenable to statistical analysis. The major challenge lies in conserving high dimensional data structure such as distinct populations or phenotypic differences, to the lower dimension. A first step usually includes reducing the number of genes based on their variance, as genes with low variance do not change across the data and therefore might not

include information about the process of interest. Subsequent principle component analysis (PCA) further de-noises the data by removing features not contributing to overall heterogeneity. To conserve significant high dimensional variability, 20 to 50 principle components are kept for further analysis. To make data structure visually interpretable but maintain most of the high-dimensional structure, non-linear methods are applied that aim at conserving neighborhood relationships in lower dimensional space. Commonly used techniques include t-distributed stochastic neighbor embedding (tSNE) [Van Der Maaten and Hinton, 2008] or the graph based uniform manifold approximation and projection (UMAP) [Becht et al., 2018].

A major challenge in the analysis of scRNAseq is the problem of not detected genes. Since only a fraction of all mRNA transcripts are captured from a cell during the lysis and reverse transcription, some gene transcripts will be sampled more likely and some lowly abundant transcripts will not be sampled at all. Those lowly abundant transcripts that appear as false zeros are termed *dropouts*. Dropouts are more prone to happen for lowly expressed genes but can also affect medium to highly expressed genes, although to a lower extent. As these false zeros can make up a substantial portion of the data, they pose an issue for subsequent downstream analysis like clustering and dimensionality reduction. It needs mentioning that comparative analysis of control studies has indicated that observed dropouts are consistent with real biological variation between cells [Svensson, 2019] and are not the result of technical shortcomings. Nonetheless, these dropouts would mask meaningful interpretation of the data, since the number of dropped out genes would be the major source of variation.

The main approach to deal with dropouts is approximating zeros present in the data with values from neighboring cells, a procedure termed *imputation*. There are numerous methods of varying complexity to impute missing data locally or globally from similar cells [Zhang and Zhang, 2019]. Performance of these algorithms depends on properties of the dataset and it is not straightforward to avoid introduction of additional artifacts by the imputation.

1.2.1 Single-cell RNA seq based reconstruction of differentiation processes

Asynchronous development leads to heterogeneity in transcriptional states of cells. This heterogeneity is captured by scRNAseq and can be used to reconstruct the underlying dynamic process through connecting the different cell states and order them along the differentiation axis. Ordering cells on a single dimension is termed pseudotime inference. Assigning cells to different orderings and allowing for bifurcations, is considered lineage inference.

A popular concept in the analysis of single-cell data is pseudotime [Tritschler et al., 2019]. Due to the destructive nature of the single-cell RNA seq measurement, it is not possible to trace cell development over time. However, since many biological processes are asynchronous across a cell population, this developmental heterogeneity within the sample can be leveraged to infer the multiple states that cells occupy in the course of the process. Essentially, pseudotime inference is the construction of a one-dimensional ordering of cells such that the start of the processes, e.g. undifferentiated stem cells, are on one end and the end of the process, like fully differentiated cells, on the other end. Intermediate states are ordered along this line according to their developmental time.

The basis for the ordering is the similarity of expression profiles between the cells. Euclidean distance is not a useful measures of similarity in high dimensional space due to distortion of space and distance. Non-linear transformations of Euclidean distance using Gaussian processes [Angerer et al., 2015], spanning trees [Trapnell et al., 2014] or graph based measures like diffusion maps [Haghverdi et al., 2016] are used to provide more appropriate measures of similarity.

All these methods assume that expression profile similarity corresponds to developmental proximity and additionally that all relevant genes and gene expression events of the process of interest are captured in the data. Missing cell states can disconnect parts of a lineage, and introduce artificial end-points. Missing gene information could hinder separation of two biologically distinct states. For the interpretation of the resulting pseudotime it is important to bear in mind

that it does not equate with the real time coordinate of the dynamic process, but is just an ordering of the cells that is meant to *correlate* with real time.

Biological development is not restricted to linear processes but also includes bifurcation through polarization of cell populations or asymmetric cell division. Going beyond the one-dimensional pseudotime representation requires the inference of specific lineages and branch-points where fates start to diverge from one another.

Many methods are available to accomplish this task and their performance depends strongly on the complexity and structure of the underlying data to infer lineages on [Saelens et al., 2018]. Usually a lower dimensional representation is used as input to infer clusters and then construct lineages by connecting the clusters to a graph [Wolf et al., 2019] or fitting curves to the cell state profiles [Street et al., 2018]. Assignment of cells to lineages and time-points can reveal molecular mechanisms involved in the process and indicate events of interest at branch-points.

The interpretation of the resulting lineages requires accounting for prior knowledge since none of these methods allow to assign directionality to the process. Prior knowledge helps to orient the process and assign start and end points, if such information is available. The use of time-series data with coarse time labels can alleviate this issue by providing a prior on expected order of events.

Recently, the analysis of scRNA seq data has been extended to infer local directionality of differentiation. Specifically, RNA velocity [Manno et al., 2018] leverages gene-wise ratios between spliced and unspliced mRNA transcripts to identify whether a gene is on the verge of being turned on or off. This information allows predicting a likely future state of each cell and allows informed predictions of transitions between cell states.

1.3 Thesis goals and contributions

Despite the recent discoveries of novel phenotypes and lineage relations in CD8 T cell exhaustion, there is no comprehensive understanding of the differentiation process of CD8 T cells in the setting of chronic LCMV infection.

After investigating the extent of TCR signaling in exhausted T cells, using bulk transcriptional profiling and an *in vivo* reporter of TCR signaling, we made use of an infection time course experiment to study the population dynamics leading to exhaustion. Single-cell transcriptome analysis provided the level of detail required to characterize subtle changes in phenotypic states and identify multiple populations reliably. High-resolution time-series data provided the necessary biological prior on timing to draw informed conclusion about the structure of the obtained data. In considering the complete time span from the beginning of infection until formation of terminally exhausted T cells, we aimed for a comprehensive picture about origin and fate of relevant cell subsets.

We built on established scRNAseq analysis methods and aspired to improve lineage inference by leveraging RNA velocity to infer complete differentiation directionality in a data-driven but biologically founded fashion. The final goal was to discover new properties of the differentiation paths in chronic infection and identify relevant intermediate cell states and associated molecular markers.

Following inference of trajectories, we wanted to find stainable protein markers, that allowed us to isolate cells from different lineages and test their differentiation potential using transfer experiments. We also aimed to identify potential cues, that shape the development of T cells, including potential effects of antigenic TCR stimulation.

Finally, we put our results in context to recent studies of CD8 T cell differentiation in chronic LCMV infection, critically comparing computational inference and subsequent validation experiment to our own study.

Chapter 2

Exhausted CD8 T cells exhibit low and strongly inhibited TCR signaling during chronic LCMV infection

Ioana Sandu[1,2], Dario Cerletti[1,2], Manfred Claassen[2,&] and Annette Oxenius[1]

[1]Institute of Microbiology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland

[2]Institute of Molecular Systems Biology, ETH Zurich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland

[&]Current address: Internal Medicine I, University Hospital Tübingen, Faculty of Medicine, University of Tübingen, Otfried-Müller-Straße 10, 72076 Tübingen, Germany

I.S., D.C., M.C., A.O. designed the experiments; I.S., D.C. carried out the experiments, I.S., D.C., M.C., A.O. analyzed the experiments; I.S., D.C., M.C., A.O. wrote the manuscript.

The thesis author conducted and analyzed the bulk RNA sequencing experiment in this work, contributing the results in section 2.2.1 and provided support in conducting the remaining experiments. The experiments and analysis of results in section 2.2.2 to 2.2.7, discussion and supplementary results were contributed by I. Sandu.

Abstract

Chronic viral infections are often associated with impaired CD8 T cell function, referred to as exhaustion. While the molecular and cellular circuits involved in CD8 T cell exhaustion have been well defined, with sustained presence of antigen being one important parameter, it remains unclear how much T cell receptor signaling is actually ongoing in vivo during established chronic infection. In this study, we characterize the in vivo T cell receptor signaling of virus-specific exhausted CD8 T cells in a mouse model, leveraging T cell receptor signaling reporter mice in combination with transcriptomics. We show that the in vivo signaling in exhausted cells is low, in contrast to their in vitro signaling potential, and despite antigen being abundantly present. Both checkpoint blockade and adoptive transfer of naïve target cells increase T cell receptor signaling, demonstrating that engagement of co-inhibitory receptors profoundly curtails CD8 T cell signaling and function in vivo.

2.1 Introduction

LCMV (Lymphocytic Choriomeningitis Virus) infection in mice is a well-established model for chronic infection, leading to T cell exhaustion following chronic infection in mice [Moskophidis et al., 1993]. Compared to effector CD8 T cells generated following an acute infection, exhausted CD8 T cells have impaired effector functions and are transcriptionally different [Wherry et al., 2007]. An important feature of exhausted CD8 T cells is the co-expression of multiple co-inhibitory receptors (such as PD-1, CTLA-4, LAG-3, TIGIT, CD39, TIM-3), which dampen T cell activation [Barber et al., 2006, Blackburn et al., 2008b, Gupta et al., 2015], [Richter et al., 2009, Jin et al., 2010, Johnston et al., 2014], [McLane et al., 2019, Anderson et al., 2016] by various mechanisms. These include limiting co-stimulation by receptor competition (CTLA-4, TIGIT [Anderson et al., 2016]), direct inhibition of signal transduction downstream of T cell receptor (TCR) engagement by limiting the phosphorylation of signaling molecules such as CD3, ZAP70 and PCK (PD-1 [Parry et al., 2005], TIM-3 [Huang et al., 2014]), restraining metabolic changes [Are et al., 2016, Patsoukis, 2015], changes at the transcriptional level (PD-1 [Quigley,

2010]), interfering with proliferation (LAG-3 [Cook and Whitmire, 2016]), or suppressing inflammatory cues (CD39 [Gupta et al., 2015]). Significant effort has been invested to enhance the effector functions of exhausted cells; indeed, checkpoint inhibitors, targeting various of the above mentioned co-inhibitory receptors, are very efficient in improving CD8 T cell numbers and effector function of exhausted CD8 T cells in both cancer and chronic infections [McLane et al., 2019]. The phenotypic and functional landscape of exhausted cells is very diverse [Beltra et al., 2020, Miller et al., 2019, Zander et al., 2019, Hudson et al., 2019] translating into differential responsiveness to checkpoint inhibition. A specific subset of non-terminally exhausted cells, termed memory-like and characterized by expression of TCF1 and SLAMF6 [Utzschneider et al., 2016a], was shown to replenish the pool of terminally exhausted cells and to respond to checkpoint blockade by proliferation and differentiation into more effector-like and eventually terminally exhausted cells [Beltra et al., 2020, Utzschneider et al., 2016a, Im et al., 2016]. Despite the compelling evidence that CD8 T cell function is impaired in chronic LCMV infection and that continued exposure to antigen significantly contributes to exhaustion [Utzschneider et al., 2016b], there is little insight into how much TCR signaling is actually ongoing in exhausted CD8 T cells in vivo during established chronic infection.

Here we characterize in vivo TCR signaling in virus-specific CD8 T cells in the setting of chronic LCMV infection. To this end, we use virus-specific TCR transgenic CD8 T cells expressing the Nr4a1-GFP reporter as proxy for TCR signaling in adoptive transfer experiments. We show that, despite abundant availability of antigen in form of peptide-MHC class I complexes, there is very limited TCR signaling ongoing in exhausted CD8 T cells during chronic infection, evidenced by low expression of the GFP reporter and by RNAseq analysis of TCR signaling associated genes. We observe enhanced TCR signaling after in vivo blocking of PD-1/PD-L1 interaction or in vivo exposure of exhausted CD8 T cells to antigen on naïve target cells, which have not been exposed to the inflammatory milieu of chronic infection. This observation suggests that the engagement of co-inhibitory receptors, such as PD-1, exerts a pronounced inhibition of TCR signaling in vivo.

2.2 Results

2.2.1 RNAseq indicates low TCR signaling in chronic infection

Previous studies showed that, compared to classical effector CD8 T cells formed in acute infection, exhausted virus-specific CD8 T cells are transcriptionally distinct and fail to exert several effector functions such as production of the inflammatory cytokines IFN-γ and TNF-α [Wherry et al., 2007]. To test whether this defect is due to intrinsically impaired TCR signaling, we set out to investigate whether ex vivo restimulation of these cells in absence of the chronic environment leads to upregulation of new transcriptional modules. CD45.1+ TCR-transgenic P14 CD8 T cells recognizing the LCMV-derived gp33-41 peptide were adoptively transferred into CD45.2+ hosts one day prior infection with a low (200 ffu) or high ($2x10^6$ ffu) dose of LCMV clone 13 to induce an acute or chronic infection. Two weeks later, P14 cells were isolated from the spleens and restimulated in vitro with plate-bound anti-CD3 and anti-CD28 antibodies.

As described previously [Wherry et al., 2007], cells isolated from chronic infection have a distinct transcriptional profile compared to the ones isolated from an acute infection (Fig. 2.1a). Upon in vitro restimulation in presence of anti-CD3 and anti-CD28 antibodies, there were significant changes in CD8 T cells isolated from both infection conditions, especially regarding inhibitory receptor genes, such as Pdcd1 (encoding PD-1) and Cd160, and cytokine production-related transcripts like Ifng, Tnf, Il10 and Xcl1, demonstrating that exhausted CD8 T cells are not inherently unable to transmit TCR signals into a transcriptional output. In fact, 429 genes showed similar changes in expression in both CD8 T cells isolated from acute infection and from chronic infection. Surprisingly, only few genes (35) were differentially expressed in exhausted cells when compared to effector/memory cells after stimulation (Fig. 2.1b), most of which include inhibitory receptor genes that were already found at higher levels before stimulation and

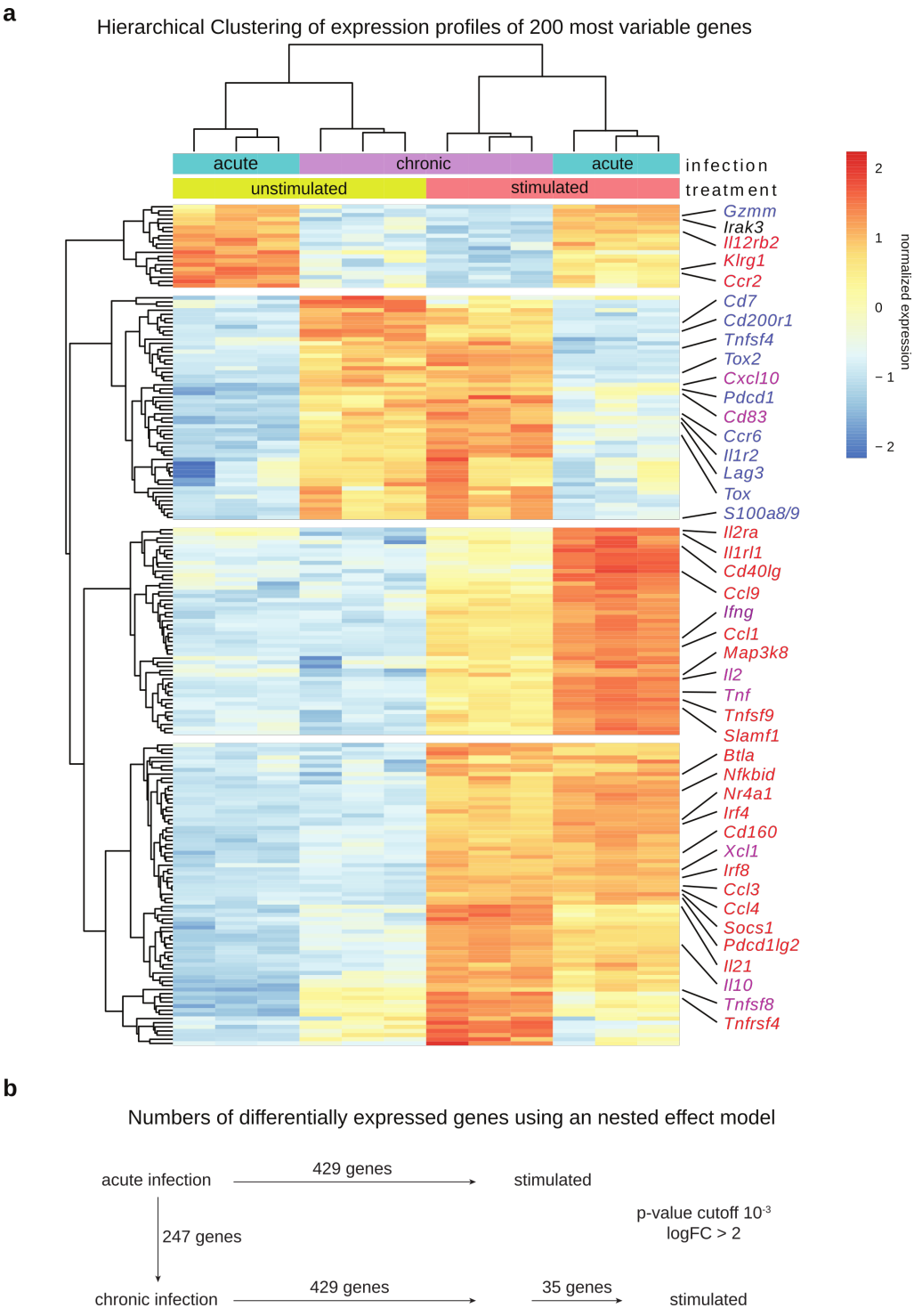


Figure 2.1: *Transcriptional profiling of functional or exhausted P14 cells with or without restimulation.* P14 cells were adoptively transferred into mice 1 day prior high or low dose LCMV clone 13 infection. Animals were sacrificed after 14 days. CD8+ P14 cells were stimulated with anti-CD3 and anti-CD28 for 4 hours. RNA was extracted and sequenced. **a** Heatmap of the 200 most variable gene profiles was generated using hierarchical clustering (k=4). Selected genes are annotated. Heatmap colors indicate normalized gene expression ranging from high (red) to low (blue). Gene color highlights differentially expressed genes between groups indicated in b. **b** Number of differentially expressed genes between the conditions are indicated. The nested effect model disentangles the effect of chronic infection and stimulation and contains an interaction term for stimulation in chronic infection. Color code of gene names indicates the group of the differential expression analysis (see Methods for details). Data from one experiment are shown.

were therefore not increased as strongly as in effector/memory cells. Moreover, genes involved in TCR signaling such as Nr4a1 and TCR-induced genes Ifng and Tnf had a low expression in chronic infection ex vivo, which increased after antibody stimulation, suggesting that the cells were either not properly activated and/or strongly inhibited in vivo.

2.2.2 NUR77 is rapidly downregulated in P14 cells in vivo

In order to investigate the TCR signaling in an in vivo setting, we took advantage of a transgenic mouse model, in which GFP is expressed under the control of the Nr4a1 promoter [Moran et al., 2011]. NUR77, encoded by Nr4a1, is triggered by antigen stimulation in T cells and has been used as a proxy for TCR signaling [Moran et al., 2011, Ryseck et al., 1989, Au-Yeung et al., 2014]. In vitro stimulation of transgenic CD8 T cells labeled with a cell proliferation dye (CPD) with plate-bound anti-CD3 and anti-CD28 antibodies for 20 hours led to fast upregulation of GFP. After the stimulus was removed, the signal was rapidly lost: after approximately 50 hours, the GFP levels were similar to the ones observed in naïve cells (Supplementary Fig. 1a). Moreover, GFP was downregulated before cells started dividing (29 hours after stimulation), suggesting that the initial drop in the observed signal was due to transcriptional regulation and protein degradation, but not proliferation-related dilution. We further sought to characterize the TCR signaling in vivo in chronic LCMV infection. Nr4a1-GFP mice were crossed to the P14 strain to generate P14-Nr4a1-GFP mice. CD45.1$^+$ P14-Nr4a1-GFP CD8 T cells were adoptively transferred into wild type CD45.2$^+$ C57BL/6J mice one day prior infection with a high dose of 2x10^6 ffu LCMV clone 13. Mice were sacrificed at different time points after infection to investigate TCR signaling ex vivo by measuring GFP in P14 cells. Only one day after chronic infection, P14-Nr4a1-GFP cells isolated from the spleen were highly activated (Fig. 2.2a). Consistent with this observation, the activation marker and inhibitory receptor PD-1 was also upregulated within 24 hours after infection (Fig. 2.2a). However, the GFP levels were rapidly downregulated in vivo already three days post-infection and kept decreasing with time, despite the maintenance of high PD-1 levels (Fig. 2.2a-c). Since LCMV has a broad tropism and induces a systemic infection [Matloubian et al., 1993], P14-Nr4a1-GFP cells isolated from various infected tissues (lymph nodes, bone marrow, spleen, blood, liver, lung, and kidney) were also analyzed three weeks post LCMV infection. Nr4a1-GFP levels were extremely low in all screened tissues three weeks into chronic infection compared to in vivo primed cells one day after infection (Fig. 2.2d, Supplementary Fig. 1b-d).

2.2.3 In vitro stimulation of P14 cells increases TCR signaling

There is extensive work showing that exhausted CD8 T cells have a dysfunctional transcriptional network compared to functional cells developed upon an acute infection [Wherry et al., 2007]. To test whether TCR signaling can be increased in exhausted cells, P14-Nr4a1-GFP cells were isolated three weeks post chronic or acute LCMV infection, sorted, and stimulated in vitro with anti-CD3 and anti-CD28 antibodies for 6 hours. There was a strong upregulation of Nr4a1-GFP in cells that responded to stimulation (CD107a$^+$) compared to the ones that did not (CD107a$^-$) for both P14 cells isolated from acute or chronic infection (Fig. 2.3). However, the extent to which the Nr4a1-GFP signal could be induced in cells isolated from chronically infected mice, which also expressed high levels of PD-1, was significantly lower compared to cells isolated from acutely infected mice. Similar observations were made in virus-specific cells isolated from various tissues such as LN, blood, BM, lung, liver, and kidney (Supplementray Fig. 2). Given the fact that chronic LCMV infection is characterized by high viral titers, the difference between in vivo and in vitro stimulated virus-specific exhausted CD8 T regarding Nr4a1-GFP signal cells was surprising.

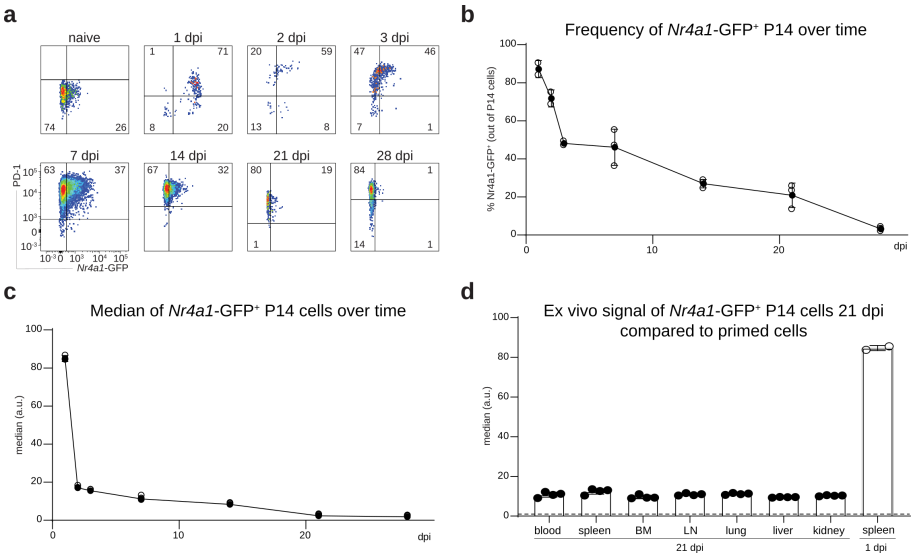


Figure 2.2: *In vivo kinetics of TCR activation during chronic infection.*
P14-Nr4a1-GFP cells were adoptively transferred into mice one day prior high dose LCMV clone 13 infection. Animals were sacrificed at various time points. **a** Examples of flow cytometry plots showing expression of Nr4a1-GFP and PD-1 in P14 cells (gated on alive single CD8$^+$ CD45.1$^+$ cells) isolated from the spleen of naïve or chronically infected mice at various days post-infection (dpi). Kinetics of GFP frequency (**b**) and GFP median (**c**) of P14-Nr4a1 GFP$^+$ cells. **d** Nr4a1-GFP expression in P14 cells isolated from different tissues 21 dpi compared to the signal observed in activated cells isolated from the spleen one dpi. Full circles (b-c) and bar plots (d) represent mean ± SD. Empty circles (b-c) show individual data points. One of two experiments is shown (n=2 mice for 1 to 3 dpi and 3 mice for 7 to 28 dpi). See also Supplementary Fig. 1.

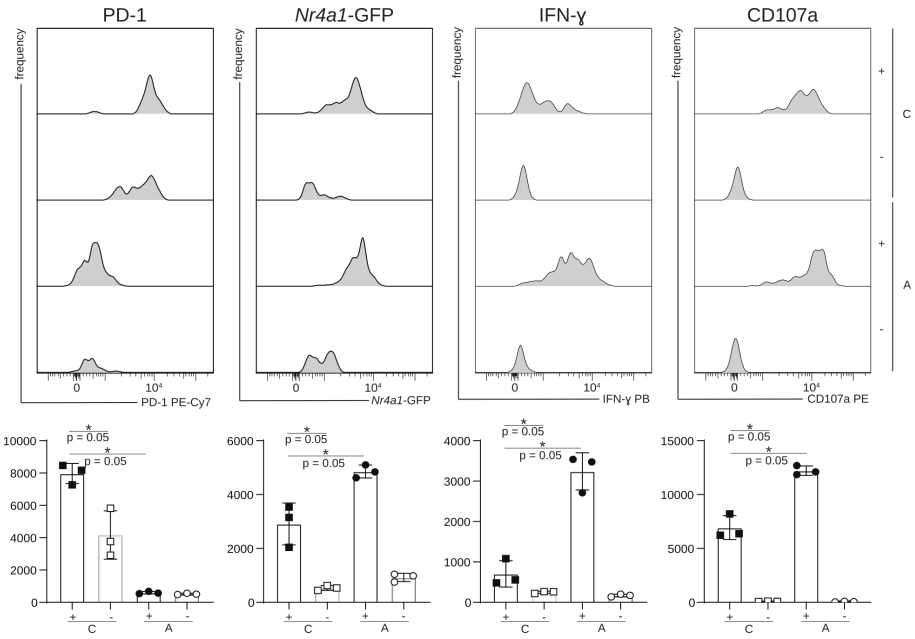


Figure 2.3: *In vitro restimulation of sorted P14 cells of acutely or chronically infected mice.*
CD45.1$^+$ P14-Nr4a1-GFP cells were adoptively transferred into CD45.2$^+$ mice one day prior infection with a high (2x10^6 ffu for chronic (C) infection) or low (200 ffu to induce an acute (A) infection) dose of LCMV clone 13. Three weeks post-infection, P14 cells were sorted from the spleen and stimulated in vitro for six hours with plate-bound antibodies. Various markers were quantified (bar plots represent average of medians ± SD) and a representative histogram per group (n=3 mice) is shown for cells that responded (CD107a$^+$), denoted by plus signs, and cells that did not (CD107a$^-$), denoted by minus signs. * represents p-value = 0.05 (Mann-Whitney one tailed-test without adjustment for multiple comparisons). Each dot represents an individual mouse. One of two experiments is shown. See also Supplementary Fig. 2.

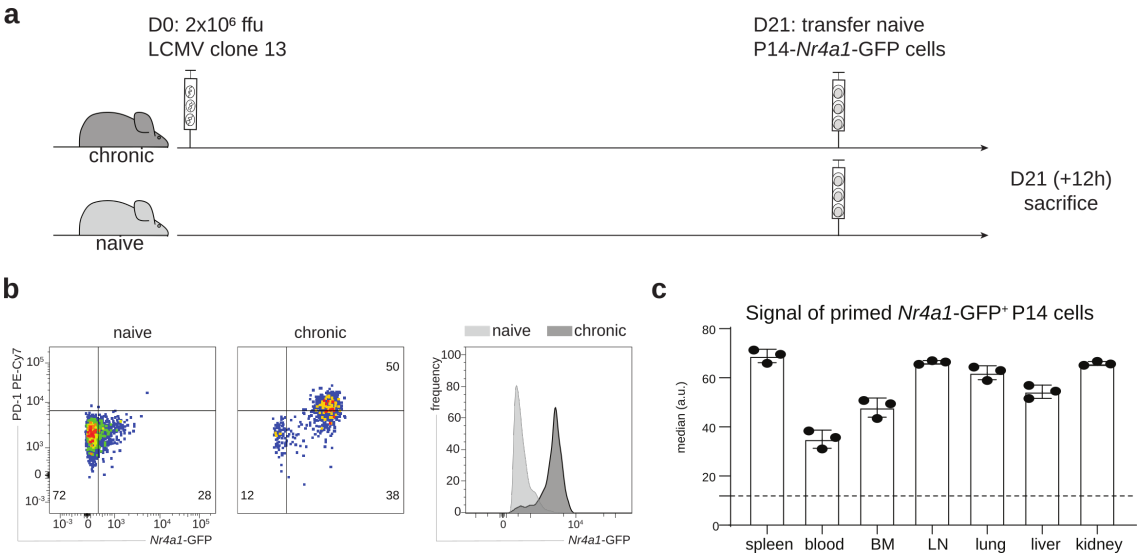


Figure 2.4: Antigen is still present three weeks into chronic infection. **a** Naïve P14-Nr4a1-GFP were adoptively transferred into chronically infected or naïve hosts. Nr4a1-GFP expression in P14 cells (gated on alive single CD8$^+$ CD45.1$^+$ cells) isolated from various tissues was assessed 12 hours later. **b** Representative flow cytometry plots and histograms (of P14 isolated from the spleen) show the difference between the two groups (chronically infected in dark gray and naïve in light gray). **c** P14 cells isolated from all screened tissues have increased GFP levels (gated on alive single CD8$^+$ CD45.1$^+$GFP$^+$ cells) compared to the signal observed in naïve hosts (dotted line). Bar plots show average medians (mean ± SD). Each dot represents an individual mouse. One of three experiments is shown (n=3 mice).

2.2.4 Antigen is present three weeks into chronic LCMV infection

The in vitro data suggested that the TCR signaling was still functional to some extent, documented by increased GFP expression after restimulation of exhausted P14-Nr4a1-GFP cells. Several hypotheses could explain this difference: i) the GFP signal is diluted in vivo due to proliferation, ii) there is little antigen (gp33-41/Db) available in vivo despite abundance of infectious viral particles three weeks post chronic infection, and iii) there is strong inhibition of TCR signaling in vivo. Since the proliferation rate of exhausted cells is low in vivo [Zajac, 1998], we decided to focus on the last two hypotheses. To test whether antigen sensed by P14 CD8 T cells is present 3 weeks post chronic LCMV infection, naïve P14-Nr4a1-GFP cells were transferred into three weeks chronically infected or naïve mice and the GFP signal was assessed 12 hours later (Fig. 2.4a). Nr4a1-GFP was highly upregulated in the transferred P14 cells isolated from all screened tissues, suggesting that antigen was abundantly present in mice with established chronic infection, leading to potent activation of naïve gp33-41-specific CD8 T cells (Fig. 2.4b-c).

2.2.5 P14 cells are PD-1$^+$, while most infected cells are PD-L1$^+$

There are numerous reports documenting the role of multiple co-inhibitory receptors on exhausted CD8 T cells, such as CTLA-4, PD-1, LAG3, CD39, TIM-3, 2B4 [Barber et al., 2006, Gupta, 2015, Richter et al., 2010, Jin et al., 2010, Johnston et al., 2014, Blackburn, 2010, Frebel, 2012, Menner, 2015]. Terminally exhausted cells were shown to co-express multiple co-inhibitory receptors [Blackburn et al., 2010] that dampen TCR signaling by various mechanisms [McLane et al., 2019]. PD-1 is perhaps the most important inhibitory receptor because its absence from the onset of chronic LCMV infection leads to death caused by T cell-related immunopathology [Frebel, 2012], while its blocking at later stages reinvigorates exhausted CD8 T cells [Barber et al., 2006]. One of its ligands, PD-L1, is expressed on most infected cells (denoted by intracellular staining for the viral nucleoprotein using VL4 antibody) which are found in all screened tissues (Supplementary Fig. 3a, b). Expression of several inhibitory receptors (PD-1, CD39, TIM-3) on P14 cells was high (∼50%) during chronic infection (Supplementary Fig. 3c, d); notably, PD-1 is ex-

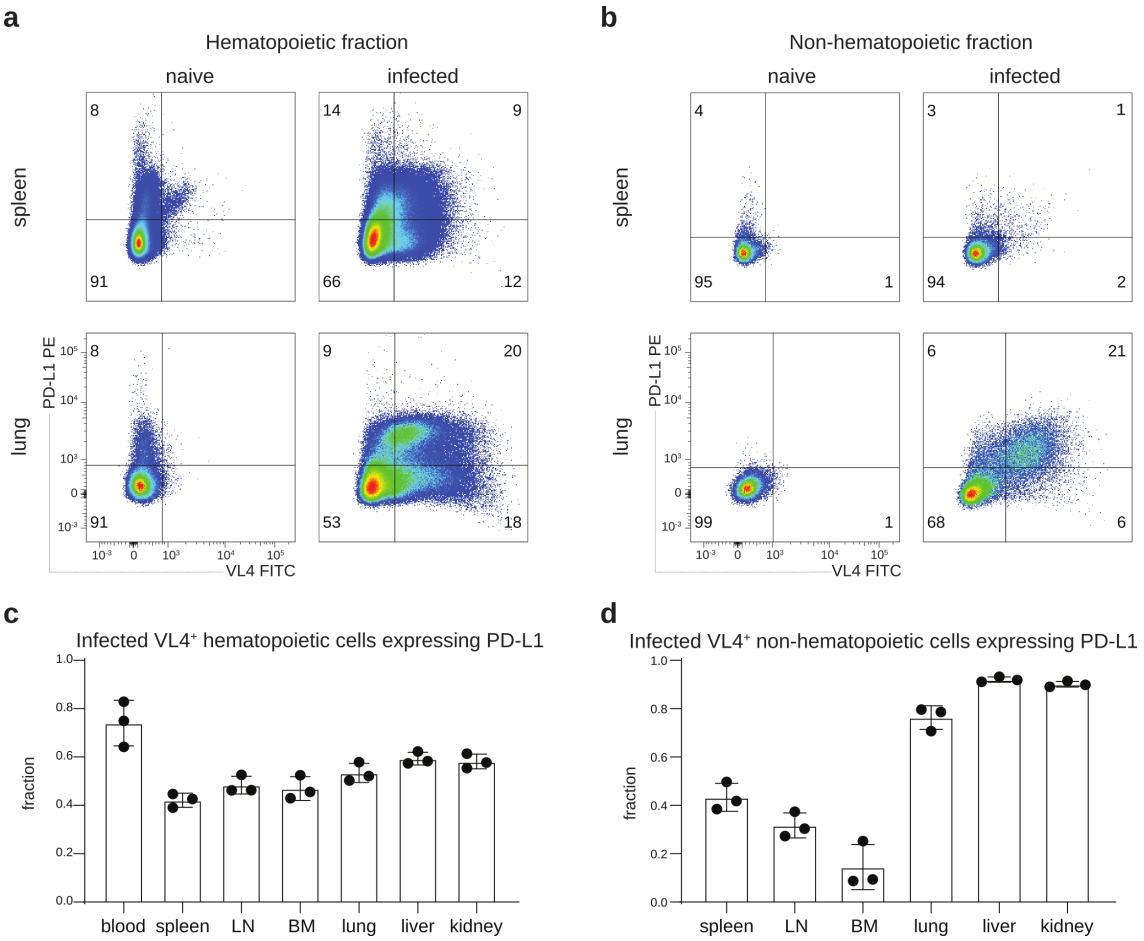


Figure 2.5: *PD-L1 expression on infected cells isolated three weeks into chronic infection.*
Seven tissues from chronically LCMV infected mice were analyzed for infected (VL4+) cells 21 days post-infection. Representative flow cytometry plots show PD-L1 and VL4 expression in hematopoietic (spleen) and peripheral tissues (lung) in the hematopoietic fraction (**a**) (gated on alive $CD45^{+}CD45.2^{+}$ cells) or in the non-hematopoietic fraction (**b**) (gated on alive $CD45^{-}CD45.2^{-}$ cells). Fraction of infected hematopoietic (**c**) or non-hematopoietic (**d**) cells (gated on $VL4^{+}$) expressing PD-L1 is quantified in the bar plots showing mean ± SD. One of two experiments is shown (n=3 mice). Each dot represents an individual mouse. See also Supplementary Fig. 3.

pressed in over 90% of P14 cells isolated from various tissues. Moreover, the fraction of infected hematopoietic cells expressing PD-L1 was quite similar across tissues (Fig. 2.4a, c), whereas the fraction of PD-L1+ infected non-hematopoietic cells was higher in peripheral tissues compared to hematopoietic organs (Fig. 2.5).

2.2.6 PD-L1 blockade leads to increased TCR signaling in vivo

In order to assess the effect of PD-1/PD-L1 interaction on TCR signaling in vivo, we performed PD-L1 blockade [Barber et al., 2006] in mice that had been chronically infected for three weeks. Three hours after administration of the blocking antibody, various tissues were harvested and P14 cells were analyzed (Fig. 2.6a). Blockade efficiency was confirmed by failure of in vitro counterstaining with the same clone (Supplementary Fig. 4a). TCR signaling, read out by GFP expression, was significantly increased in the treated group (Fig. 2.6b-d, Supplementary Fig. 4b-c). Moreover, PD-1 levels were also higher in P14 cells isolated from anti-PD-L1-treated mice compared to the control mice (Fig. 2.6e, Supplementary Fig. 4d). Since PD-1 is also an activation marker, this increase is likely a result of increased stimulation. Additional short-term co-blockade of other inhibitory receptors (LAG-3, TIM-3, CTLA4, 2B4, TIGIT) did not signifi-

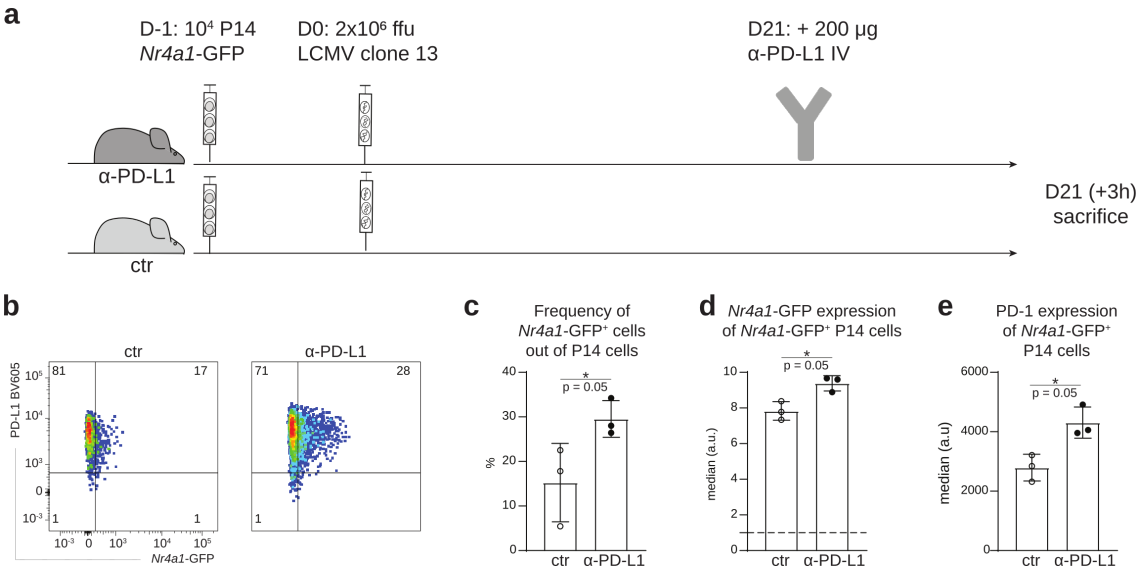


Figure 2.6: *Increase in TCR signaling following short-term PD-L1 blockade.*
a P14-Nr4a1-GFP were adoptively transferred into naïve hosts one day prior high dose LCMV clone 13 infection. 21 days post-infection, one group received one dose of anti-PD-L1 blocking antibody intravenously. The mice were sacrificed three hours after treatment. **b** Representative plots showing Nr4a1-GFP signal in two groups, control (ctr) and treated (α-PD-L1) are shown. Fraction of Nr4a1-GFP$^+$ cells (gated on alive single CD8$^+$ CD45.1$^+$ P14 cells) (**c**), median of GFP (**d**) and PD-1 (**e**) (gated on alive single CD8$^+$ CD45.1$^+$ CD45.1 GFP$^+$ cells) are shown for cells isolated from the spleen. Bar plots represent mean $\pm$ SD. * represents p-value = 0.05 (Mann-Whitney one tailed-test). One of three experiments is shown (n=3 mice). Each dot represents an individual mouse. See also Supplementary Fig. 4-5.

cantly enhance TCR signaling compared to PD-L1 blockade (Supplementary Fig. 5), suggesting that PD-1/PD-L1 interaction is the main inhibitory signal.

2.2.7 In vivo killing of naïve targets enhances TCR signaling

Previous work showed that, despite being less functional than virus-specific cells generated upon an acute infection, exhausted CD8 T cells are capable of efficient in vivo killing of target cells isolated from a naïve host [Graw et al., 2011]. To corroborate that gp33-41-pulsed target cells originating from naïve mice and transferred into recipients with established chronic LCMV infection leads to their killing, and test whether such recognition and killing would be associated with in vivo TCR signaling in exhausted CD8 T cells, peptide-pulsed or unpulsed naïve splenocytes were transferred into chronically infected hosts 21 days post infection (Fig. 2.7a). Gp33-41-pulsed targets were killed very efficiently only three hours after transfer (Fig. 2.7b-c, Supplementary Fig. 6a). Moreover, TCR signaling was increased in cells isolated from the recipients that received the pulsed targets compared to the control group, which received unpulsed splenocytes (Fig. 2.7d-f, Supplementary Fig. 6b, c). Given the fact that the virus is not cleared in vivo, the killing efficiency of target cells originating from naïve mice might seem surprising. However, there are several factors which might explain these observations. First of all, the splenocyte suspension is mainly composed of lymphocytes, which are not normally infected by LCMV. Second, antigen expression levels might be different in vivo. Third, likely of highest relevance, the pulsed targets did not express high levels of PD-L1, one of the two ligands that can bind PD-1 in vivo. Indeed, PD-L1 expression on infected cells (denoted by VL4 staining) was several fold higher compared to basal expression on naïve splenocytes (Fig. 2.7g). Hence this last explanation seems the most likely one and would indicate remarkably potent negative regulation of TCR signaling in vivo in exhausted CD8 T cells during established chronic infection. The pool of exhausted cells is very heterogeneous and includes at least three subsets characterized by different phenotypic and functional properties: memory-like expressing TCF1 [Utzschneider et al., 2016a, Im et al., 2016], CX3CR1^{hi} effector-like and terminally exhausted

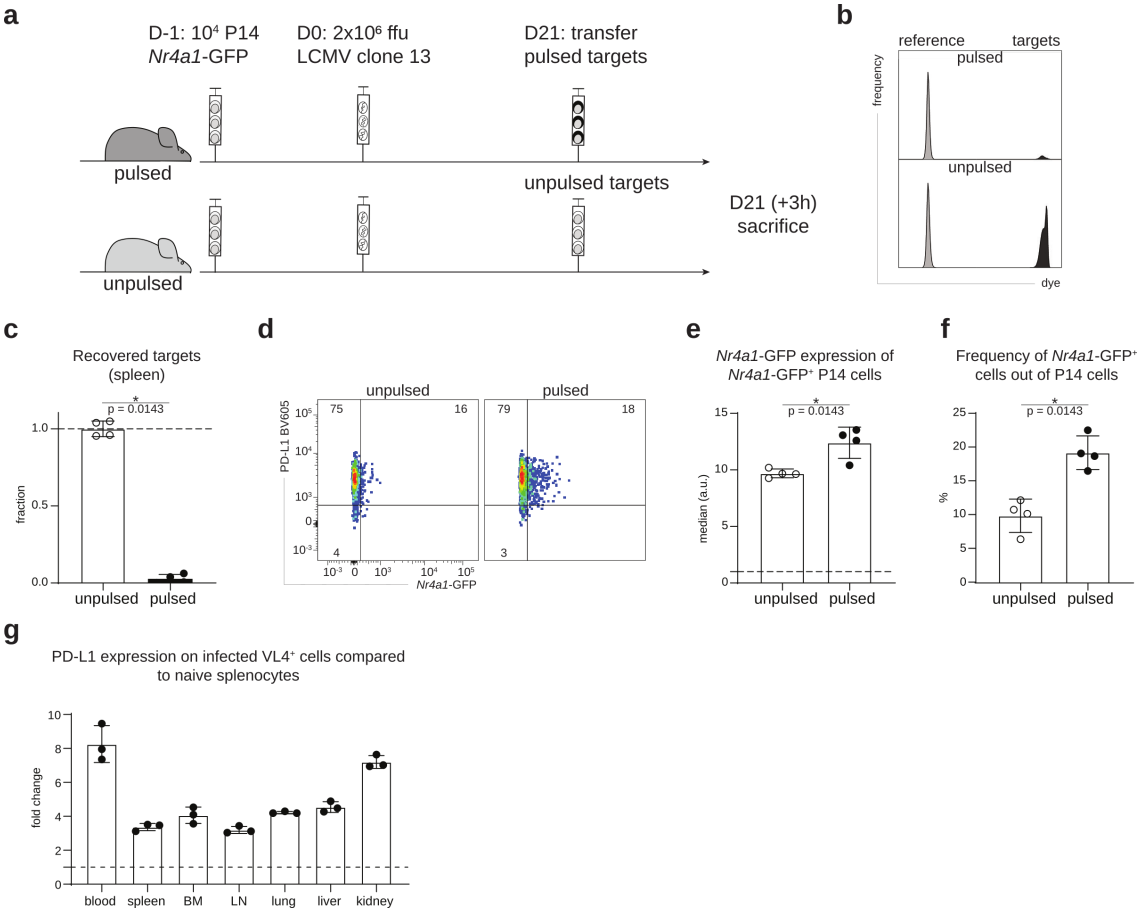


Figure 2.7: *In vivo cytotoxicity of exhausted cells CD8 T cells three weeks into chronic LCMV infection.*
a CD45.1$^+$ P14-Nr4a1-GFP cells were adoptively transferred into naïve CD45.1$^+$ mice one day prior to high dose LCMV clone 13 infection. 21 days post infection, the hosts received gp33-41-pulsed or unpulsed labeled targets together with a reference population. Three hours later the mice were sacrificed. **b** Representative plots of transferred targets isolated from the spleen. **c** Fraction of recovered targets (calculated based on the reference population) in spleens isolated from unpulsed or pulsed groups is shown. **d** Representative flow cytometry plots showing ex vivo Nr4a1-GFP expression in P14 cells (gated on alive single CD8$^+$ CD45.1$^+$ cells) from the spleens of mice which received unpulsed or pulsed targets. Medians of GFP (gated on alive single CD8$^+$ CD45.1$^+$ GFP$^+$ cells) (**e**) and frequency (**f**) of GFP$^+$ cells (out of P14 cells) are shown. **g** Fold change of PD-L1 expression between infected cells (gated on alive single VL4$^+$ cells) isolated from different tissues three weeks into chronic LCMV infection and naïve splenocytes (dotted line). Bar plots represent mean $\pm$ SD. * denotes a significant p-value $\leq$ 0.05 (Mann-Whitney one tailed-test). One of three experiments is shown (n=4 (c-f), 3 (g) mice). Each dot represents an individual mouse. See also Supplementary Fig. 6.

PD-1^{hi}CXCR6^{hi} cells [Miller et al., 2019, Zander et al., 2019, Hudson et al., 2019]. In vitro killing assays performed with sorted P14 cells and pulsed targets in presence or absence of blocking PD-L1 antibody confirmed the inhibitory role of PD-1/PD-L1 axis with respect to cytotoxicity, irrespective of the exhausted subset (Supplementary Fig. 7).

2.3 Discussion

We have characterized TCR signaling in virus-specific exhausted CD8 T cells on the transcriptional level and in an in vivo setting by using the Nr4a1-GFP reporter mouse in the context of chronic LCMV infection. The transcriptomic data shows that there are significant differences in virus-specific CD8 T cells from either acutely or chronically infected mice ex vivo without stimulation. These include many co-inhibitory receptors as well as the transcription factor Tox, which has been recently shown to be specifically expressed in exhausted cells [Alfei et al., 2019, Khan

et al., 2019, Yao et al., 2019]. However, there are many genes upregulated after in vitro restimulation in both conditions with hardly any difference between the chronic and acute infection, indicating that even CD8 T cells from chronic infection are capable of transcriptionally responding to stimulation.

The low TCR signaling observed on the transcriptional level was confirmed in vivo, by measuring the activity of the Nr4a1 promoter used as a proxy for TCR signaling [Moran et al., 2011]. There was a strong signal induced after initial priming which was rapidly downregulated in vivo. The fast decrease of the signal could be attributed, at least at this stage (1-5 days) post infection, to signal dilution due to proliferation and/or downregulation of Nr4a1-GFP. Nevertheless, in vitro data showed that the signal is rapidly downregulated before first cell division, suggesting a tight regulation of Nr4a1, consistent with its short half-life [Ryseck et al., 1989].

Surprisingly, despite abundant antigen being present three weeks post infection, TCR signaling is very low in vivo. At this later stage in chronic infection, the low signal is unlikely to be due to proliferation-related dilution because exhausted CD8 T cells do not divide at a fast rate [Wherry et al., 2003]. TCR signaling could be increased by in vitro stimulation of exhausted cells, but not to the same extent as observed in functional CD8 T cells (generated upon an acute infection). Together with the RNAseq data and previous studies [Wherry et al., 2007], this suggests that TCR signaling is strongly regulated in exhausted cells, as maximum signaling levels are not reached in vivo, despite the antigen being abundantly present. Additionally, recent work showed Nr4a1 transcription is not induced by NFAT alone [Jennings et al., 2019] and there is evidence for ERK signaling mediated AP-1 induction being involved in Nr4a1 transcription [Stocco et al., 2002]. In chronic LCMV infection, the formation of NFAT/AP-1 dimers is impaired [Martinez et al., 2015], implying that Nr4a1 does not report the full extent of TCR signaling in this setting.

IFN-γ secretion and degranulation were also significantly lower in exhausted cells compared to functional cells (generated upon acute LCMV infection), as previously shown [Zajac, 1998, Agnellini, 2007] (Fig. 2.3, Supplementary Fig. 3). Not surprisingly, exhausted virus-specific CD8 T cells co-expressed a multitude of inhibitory receptors which dampen TCR signaling [Blackburn, 2009]. Indeed, both short-term PD-L1 blockade and adoptive transfer of pulsed target cells isolated from naïve mice led to increased Nr4a1-GFP and PD-1 expression in vivo, linked to efficient killing of these target cells as shown previously [Garcia et al., 2015]. The highest increase in TCR signaling was observed in P14-Nr4a1 cells isolated from spleen and lungs after adoptive transfer of pulsed target cells isolated from naïve mice, probably due to the nature and delivery of targets. The pulsed cells were splenocytes, mainly composed of naive lymphocytes, which are primarily in circulation and home to secondary lymphoid tissues. Additionally, due to the intravenous delivery, most targets would initially reach the lungs where there are many P14 cells [Frebel, 2012] that could kill the pulsed targets specifically, resulting in fewer pulsed targets reaching other peripheral organs. Importantly, the adoptively transferred target cells from naïve mice expressed lower levels of PD-L1 compared to $VL4^+$ LCMV-infected cells in chronically infected hosts, thus lowering negative regulation of TCR signaling in exhausted CD8 T cells. This difference might explain why naïve targets are recognized and eliminated, while most endogenous infected targets are not [Matloubian et al., 1994]. Altogether, these results suggest that TCR signaling is strongly inhibited in vivo.

Compared to PD-L1 blockade alone, short-term co-blockade of several inhibitory receptors (PD-1, LAG-3, CTLA-4, TIM-3, TIGIT) did not show a significant increase of Nr4a1-GFP expression in exhausted LCMV-specific CD8 T cells in vivo (Supplementary Fig. 5), suggesting that the PD-1/PD-L1 axis exerts the strongest inhibition on TCR signaling. This observation is consistent with previous work, which showed that PD-1/PD-L1 interaction protects mice from CD8-mediated lethal immunopathology [Frebel, 2012]. Several studies proved that co-blockade of PD-L1 and another inhibitory receptor had synergistic effects, showing benefits over PD-L1 blockade alone [McLane et al., 2019]. Moreover, in some cases, blocking an inhibitory receptor is only efficient when combined with PD-L1 blockade. For example, blocking LAG-3 in

a chronic LCMV infection has no effect on the function of exhausted virus-specific CD8 T cells [Richter et al., 2010], while co-blockade of PD-L1 and LAG-3 leads to their reinvigoration [Blackburn, 2009]. This suggests that blocking LAG-3 is only efficient when PD-1/PD-L1 interaction is inhibited. The fact that there was no increase of TCR signaling compared to single PD-L1 short-term blockade suggests that the direct inhibition of TCR signaling transduction is mainly due to the PD-1/PD-L1 axis. Nevertheless, other inhibitory receptors are able to interfere with signaling cascade downstream the TCR as well. For example, CTLA-4 can inhibit AKT [Parry et al., 2005], while TIGIT ligation can lead to downregulation of TCR components [Joller, 2011]. However, unlike PD-1 whose ligand is widely expressed [Frebel and Oxenius, 2013b], the ligand expression of CTLA-4 and TIGIT is restricted to APCs [Anderson et al., 2016, Parry et al., 2005]. It is possible that the short-term blockade did not allow for a substantial fraction of exhausted cells to interact with target cells expressing ligands of these inhibitory receptors. Our results suggest that, out of the screened blocking antibodies (PD-L1, LAG-3, CTLA-4, TIM-3, TIGIT), PD-1 is the main inhibitory receptor dampening TCR signaling. Once this axis is hampered, blocking other receptors that have complimentary inhibitory mechanisms (such as effects on proliferation, cytokine production or interference with co-stimulatory pathways [McLane et al., 2019]) could lead to a synergistic effect.

In summary, exertion of effector functions of exhausted CD8 T cells is strongly and actively inhibited in vivo by potent inhibition of TCR signaling via engagement of co-inhibitory receptors. In their absence, exhausted CD8 T cells are able to resume function and, above all, their cytotoxic potential.

2.4 Material & Methods

RNAseq

P14 cells were sorted (BD FACS Aria, BD Biosciences) from animals with either acute or chronic LCMV infection. After 4 hours of in vitro restimulation in presence of anti-CD3 (clone 145-2C11, Biolegend) and anti-CD28 (clone 37.51, Biolegend) antibodies, RNA was extracted using TRIzol (Thermofisher Scientific) according to the manufacturer's protocol and processed for sequencing at the Functional Genomics Center Zürich (FGCZ). The extracted RNA was selected for mRNA transcripts using polyA-tail capture. The enriched mRNA was reverse transcribed and sequenced on the Illumnia HiSeq 4000 platform. Resulting transcript counts were analyzed and differential gene expression was analyzed using the DEseq2 package [Love et al., 2014] in R [core team, 2019].

Mice

Wild-type male CD45.2 C57BL/6J mice purchased from Janvier Elevage, Nr4a1-GFP mice expressing GFP under the control of the NUR77 promoter [Moran et al., 2011], Tcf7-GFP mice expressing GFP under the control of Tcf7 (encoding TCF1) promoter [Utzschneider et al., 2016b], P14 transgenic (CD45.1) mice expressing a TCR specific for LCMV peptide gp33–41 [Pircher, 1990] were housed at 24°C and 50% humidity and bred under specific pathogen–free conditions at the ETH Phenomics Center Hönggerberg. Mice were exposed to a 12:12 h light-dark cycle with unrestricted access to water and food. All mice used in experiments had between 6–16 weeks. P14-Nr4a1-GFP or P14-Tcf7-GFP were generated by crossing Nr4a1-GFP or Tcf7-GFP mice to P14 mice. All animal experiments were conducted according to the Swiss federal regulations and were approved by the Cantonal Veterinary Office of Zürich (Animal experimentation permissions 147/2014, 115/2017).

Virus

LCMV clone 13 was propagated on baby hamster kidney 21 cells (BHK21 [C13] (ATCC® CCL10). Viral titers of virus stocks were determined as described previously by using MC57G (ATCC®

CRL-2295) cells [Battegay et al., 1991].

Infections

10^4 transgenic cells (P14 or P14-Nr4a1-GFP) were adoptively transferred 1 day prior LCMV clone 13 intravenous (IV) infection with 2 x 10^6 or 200 ffu/mouse to induce a chronic or acute infection, respectively. For analysis at 1, 2, and 3 days post-infection, 5x10^5 P14-Nr4a1-GFP cells were transferred.

Cell isolation from tissues

Mice were sacrificed with carbon dioxide. Blood was taken from the heart or vena cava. The mice were then perfused with 20 mL PBS, and organs (spleen, bone marrow, lymph nodes, lung, liver, and kidney) were isolated. Lungs, liver, and kidney were cut into pieces and digested in complete RPMI (Bioconcept) (RPMI-1640 containing 10% fetal bovine serum (Omnilab), 2 mM L-glutamine (Gibco), 1% penicillin-streptomycin (Gibco), 1 mM sodium pyruvate (Gibco), 50 nM betamercapthoethanol, 0.1 mM non-essential (glycine, L-alanine, L-asparagine, L-aspartic acid, L-glutamic acid, L-proline, L-serine) amino acids (Gibco), 20 mM HEPES (Gibco)) supplemented with 2.4 mg/ml collagenase type I (Gibco), and 0.2 mg/ml DNase I (Roche Diagnostics, Rotkreuz, Switzerland) for 30 minutes at 37°C. Spleens and lymph nodes were mashed through 70 µm filter with a syringe (1 mL) plunger. Cell suspension from lungs, liver, and kidney were enriched for lymphocytes by centrifugation over Percoll density centrifugation (30% (v/v) Percoll in PBS) at 4°C (500 g) for 30 minutes. Cell suspensions were filtered (70 µm) and treated with ammonium-chloride-potassium buffer (150 mM NH4Cl, 10 nM KHCO3, 0.1 mM EDTA in water) to lyse erythrocytes for 5 min at room temperature.

Cell sorting

Spleen samples were depleted of CD4 and B cells by incubating splenocyte suspensions in enrichment buffer (PBS, 1%FCS, 2 mM EDTA) with biotinylated α-CD4 and α-B220 antibodies (1/100 dilution) at room temperature for 20 min, followed by incubation with streptavidin-conjugated beads (Mojo, Biolegend) (4%) for 5 min at room temperature. Cells were then placed on a magnetic separator (StemCell) for 3 min at room temperature, followed by collection of supernatant. Enriched samples from the spleen or cell suspensions from other tissues of interest (see "Cell isolation from tissues") were stained with α-CD8 and α-CD45.1 to sort P14 cells (ARIA cell sorter, BD Biosciences).

In vitro restimulations

Restimulations of sorted virus-specific P14-Nr4a1-GFP cells were performed in 96 well flat-bottom plates pre-coated with anti CD3 and anti CD28 at 1 μg/mL. Sorted cells were incubated in complete RPMI at 37° C for 6 hours at 37° C in presence of 2μM monensin A, 10 μg/mL brefeldin, and α-CD107a antibody.

Flow cytometric analysis

Surface staining was performed at room temperature for 30 minutes in FACS buffer (2% FCS, 1% EDTA in PBS). LIVE/DEAD™Fixable Near-IR Dead (Thermo Fisher) was used to discriminate alive from dead cells. Fluorophore-conjugated antibodies used for flow cytometry were purchased from eBioscience (San Diego, California, USA) (α-IFNγ PE XMG1.2), and BioLegend (Lucerna Chem AG, Luzern, Switzerland) (α-IFNγ Pacific Blue XMG1.2, α-CD107a PE 1D4R, α-CD44 BV510 IM7; α-CD45.1 Pacific Blue A20; α-CD45.1 APC A20; α-CD8 PerCP 53-6.7; α-CD39 AF647 A7R34; α-TIM-3 PE RMT3-23; α-PD-1 PE-Cy7 29F.1A12, α-PD-L1 PE 10F.9G2, α-CD45.2 Pacific Blue 104), BD Biosciences (Allschwil, Switzerland) (α-CD45 PerCP 30-F11) or

self-produced (VL4-FITC). Intracellular staining was performed according to the manufacturer's protocol (Biolegend (intracellular cytokine staining kit). Data was acquired LSR II and LSR II Fortessa using Diva software (BD Biosciences, Allschwil, Switzerland) and analyzed in FlowJo (BD Biosciences, Allschwil, Switzerland). Plots were generated and statistical analysis was performed in GraphPad Prism 8 (La Jolla, California, USA). Gating strategy is shown in Supplementary Fig. 7.

In vivo blocking

One dose of blocking antibodies (200 μg of α-PD-L1 (clone 10F.9G2), α-TIM-3 (clone RMT3-23), LAG-3 (clone C9B7W), α-CTLA-4 (clone UC10-4F10-11), α-TIGIT (clone 1B4) [Dixon, 2018]) were administered IV three hours prior sacrifice.

In vivo killing assay

Splenocytes were incubated at 37°C for 1 hour with or without 0.1 µg/mL gp33-41 in RPMI, labeled with cell proliferation dye (eFluor 450 or Cell Trace Yellow, Thermo Fisher) according to manufacturer's protocols. The reference population and the target population (pulsed or unpulsed) were mixed (15 x 10^6 splenocytes per population) and transferred IV into chronically infected mice. After three hours, the animals were sacrificed and target cell abundance was quantified by flow cytometry.

In vitro killing assay

CD45.1$^+$ P14-Tcf7-GFP (transgenic P14 cells expressing GFP under the control of the TCF1 promoter, encoded by Tcf7) cells were transferred into CD45.2$^+$ hosts one day prior chronic LCMV clone 13 infection. P14 cells were sorted from organs (lungs, liver, and spleen) of chronically infected mice 21 dpi based on CX3CR1, CXCR6, and TCF1-GFP into different subsets. TargeT cells (EL4 cells) were incubated with IFN-γ at 100 U/mL (PeproTech) for 6 hours prior pulsing to induce PD-L1 upregulation and pulsed with gp33-41 peptide at 37°C for 1 h, then washed with medium. For one condition, PD-L1 was blocked with 30 µg/mL α-PD-L1 (clone 10F.9G2) for one hour at 37°C. Sorted populations were incubated with unpulsed or pulsed (in presence or absence of α-PD-L1). EL4 targeT cells at 5:1 E/T ratio. Counting beads (CaliBRITE, BD Biosciences) were added to the samples stained for flow cytometry.

Statistical Analysis

Graphpad prism 8.2.0 software or R was used to calculate significance between samples. p-values ≤ 0.05 were considered significant. Statistical test is indicated in each figure.

Acknowledgments

We thank Prof. Roman Spörri for providing Nr4a1-GFP reporter mice. We would like to thank Nathalie Oetiker and Franziska Wagen for excellent technical assistance. We are grateful to the members of the Claassen, Oxenius, Joller (University of Zurich), and Sallusto groups for helpful discussions and feedback. Funding: This work was supported by the ETH Zurich (grant no. ETH-39 14-2 to MC and AO).

Competing interest: The authors declare that they have no competing interests.

Supplementary Figures

Supplementary figures are found in the appendix (p. 69).

Chapter 3

Cytopath: Simulation based inference of differentiation trajectories from RNA velocity fields

Revant Gupta[1,†,&], Dario Cerletti[1,2,†], Gilles Gut[1,†], Annette Oxenius[2] and Manfred Claassen[1,&]

[1]Institute of Molecular Systems Biology, ETH Zurich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland

[2]Institute of Microbiology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland

[&]Current address: Internal Medicine I, University Hospital Tübingen, Faculty of Medicine, University of Tübingen, Otfried-Müller-Straße 10, 72076 Tübingen, Germany

[†]these authors contributed equally to this work

R.G., D.C., G.G., M.C. designed the method; G.G., R.G. implemented the method, R.G., D.C., G.G., A.O., M.C. tested and applied the method; R.G., D.C., M.C. wrote the manuscript.

Abstract

Unsupervised trajectory inference from single cell RNA sequencing data bears the potential to systematically reconstruct complex differentiation processes but remains challenging in spite of the many available solutions. In general, trajectory and pseudotime inference methods so far suffer from the ambiguity of the static single-cell transcriptome snapshots lacking a measure of directionality of transcriptional activity. We report Cytopath, a method for trajectory inference that takes advantage of transcriptional activity information from RNA velocity analyses to perform trajectory inference. Cytopath performs this task by defining a Markov chain model, simulation of an ensemble of possible differentiation trajectories and subsequent statistical analysis to extract the topological and molecular characteristics of the studied differentiation process. We demonstrate the capability of Cytopath to reconstruct differentiation trajectories with varying bifurcated and circular topology studied in single-snapshot as well as time-series single-cell RNA sequencing experiments. Comparison to state-of-the-art trajectory inference approaches demonstrate superior and enabling capability to reconstruct the considered differentiation trajectories. Trajectory inference constitutes a frequent crucial step in interpreting single-cell RNA sequencing studies. We expect cytopath to enable researchers to tap the directionality information present in single-cell RNA sequencing data to achieve trajectory inference for possibly complex lineages at an unprecedented precision and resolution.

3.1 Introduction

Biological processes such as hematopoiesis [Povinelli et al., 2018], tumor evolution [Ren et al., 2018] and immune responses [Stubbington et al., 2017] are defined by a temporal sequence of coordinated phenotypic state changes in the context of, possibly heterogeneous, cell populations. Such phenotypic states can be characterized by e.g. epigenetic, transcriptional and proteomic cell profiles. Furthermore, biological processes are often asynchronously triggered and might give rise to bi- and multifurcated state sequences.

This situation requires single-cell approaches to measure and ultimately investigate these biological processes. The repertoire of suitable technologies to monitor different types of molecular

profiles has increased dramatically over the last years, including capabilities to study the single-cell proteome [Yang et al., 2019], transcriptome [Kulkarni et al., 2019] or metabolome [Heinemann and Zenobi, 2011] and in particular includes single-cell RNA sequencing (scRNAseq) that has gained in widespread use due to the generic applicability of sequencing technology. While these measurements are information-rich, their analysis and interpretation is challenged by high-dimensionality, measurement noise and their destructive nature only yielding snapshots of the whole process.

Different computational approaches have been proposed to reconstruct cell state sequences of biological processes from scRNAseq data. This reconstruction involves a computational inference step, also known as pseudotime ordering [Tritschler et al., 2019]. This procedure aims at ordering the measured molecular profiles consistent with the state sequence of the underlying process and is typically guided by the assumption that phenotypic similarity reflects temporal proximity and therefore utilize similarity of gene expression profiles to estimate adjacency in a graph of cell state transitions. Available methods to achieve this include minimum spanning trees [Trapnell et al., 2014], diffusion maps [Haghverdi et al., 2016] and quite a few more variants [Saelens et al., 2019]. The interpretation of pseudotime ordering for branched processes is particularly challenging and requires an additional trajectory inference step. Trajectory inference methods fit trees or ensemble of trees (TreeTop [Macnair et al., 2019], Slingshot [Street et al., 2018]) or use graph-based connectivity of cell states (PAGA [Wolf et al., 2019]).

Static expression profiles are ambiguous with respect to the directionality of potential cell state transitions. This ambiguity constitutes a major limitation of pseudotime ordering and trajectory inference, and specifically precludes data driven assignment of start and endpoints without previous knowledge about the process and the detection of cell state transitions marked by large deviations in expression patterns.

Recently, it became possible to estimate transcriptional activity from scRNAseq data via RNA velocity analysis [Manno et al., 2018] consequently allowing to infer likely transitions between different cell states in a data-driven fashion, ultimately opening the possibility to mitigate the limitations of afore reconstruction approaches.

So far trajectory inference approaches lack the incorporation of such directionality information to estimate pseudotime or construct trajectories, with the exception of a recent preprint-reported approach to improve pseudotime ordering and differentiation probability estimation by accounting for RNA velocity, however lacking the inference of prototypic trajectories relying instead on fate probabilities [Lange et al., 2020].

In this work we aim at closing this gap by proposing cytopath, a simulation based trajectory inference approach that accounts for RNA velocity. We show that cytopath achieves accurate and robust cell state trajectories of known biological process with linear, circular and bifurcated topologies in datasets and is superior to state-of-the art trajectory inference approaches.

3.2 Results

In this section we present the core features and mechanisms of Cytopath. Trajectory inference is performed on three scRNA seq datasets presenting three different scenarios. We show that Cytopath successfully recovers known biological processes in the datasets with high accuracy. Furthermore we compare the performance of Cytopath to widely used trajectory inference tools, Slingshot and Monocle3. We demonstrate that the shortcomings of these tools which do not include RNA velocity information are successfully addressed by Cytopath.

3.2.1 Simulation based trajectory inference with Cytopath

Trajectory inference with Cytopath follows on the RNA velocity analysis of a scRNA seq dataset, and is specifically based on the resulting cell transition probability matrix. The objective of

trajectory inference with Cytopath is to estimate trajectories from root to terminal cell states. Root and terminal states can either be derived from a Markov random-walk model utilizing the transition probability matrix itself, as described in [Manno et al., 2018] or can be supplied by the user, based on suitable prior knowledge.

The trajectory inference process is divided into four steps (Figure 3.1). In the first step, Markov sampling of consecutive cell state transitions is performed based on the probabilities of the transition matrix, resulting in an ensemble of simulated cell state sequences. Sampling is initialized at the predefined root states. Sampling is performed for a fixed number of steps until a sufficient number of unique cell state sequences reaching the terminal states have been generated (Figure 3.1B-C).

The generated cell state sequences are individual simulations of the biological process from the root to the terminal state. Due to the stochastic nature of the sampling process, the cell state sequences cannot be considered as aligned with respect to the cell states at each transition step. Consequently, in the second step, simulations that terminate at a common terminal state are aligned using Dynamic Time Warping [Salvador and Chan, 2007]. The procedure aligns simulations to a common differentiation coordinate such that cell states from any simulation at a particular differentiation coordinate (pseudotime) represent similar cell states (Figure 3.1C-D).

In the third step, consensus states across the steps of the aligned simulations are estimated, giving rise to the reported trajectory. Cell states at every step of the ensemble of aligned simulations are averaged and the average value is considered as the consensus state of the trajectory at the particular step. Alternatively, trajectories can be anchored to observed cell states in the data by choosing the cell state closest to the aforementioned average value.

In the final step, cell groups are assigned to each step of the inferred trajectory. Assignment is based on an alignment score evaluating for each cell both similarity of its static expression profile as well as the velocity profile with those of the consensus states. The score is used to measure the trajectory position as well as the relative association of a cell state to possibly multiple branches of a differentiation processes with complex topology.

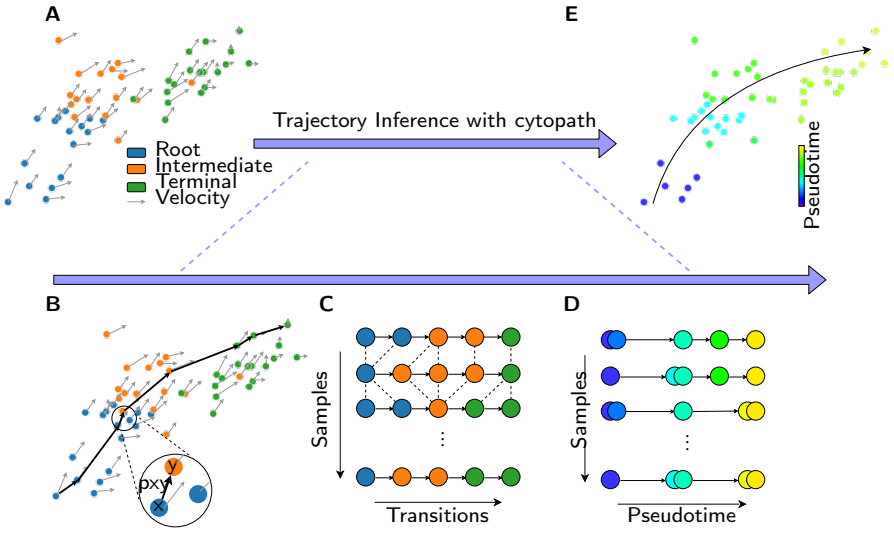


Figure 3.1: Method overview. (A) Input data for trajectory inference consisting of single cell gene expression profiles, RNA velocity profiles, root and terminal state annotation. (B) Simulation by Markov sampling is performed based on the cell to cell transition probabilities. Sampling is initialized on cells from a RNA velocity derived root state. (C) Simulation is performed for a defined number of steps. Transition steps are aligned using Dynamic Time warping. (D) Post alignment, cells at each step represent the same consensus state. (E) Cells along the inferred trajectory are assigned the aligned transition-step value as the pseudotime.

3.2.2 Reconstruction of neuronal differentiation in the developing mouse hippocampus

We assessed the capability to reconstruct developmental processes with multiple branching. To this end, we applied Cytopath, Monocle3 and Slingshot to the developing mouse hippocampus dataset, which was first used to demonstrate RNA velocity of single cells. This dataset is composed of 18,140 cells. RNA velocity was used to identify the root and terminal regions [Manno et al., 2018].

Data from the original publication including t-SNE projection, RNA velocity and transition probability matrix were utilized to for trajectory inference with Cytopath. The cell type identities and their ordering reported in [Manno et al., 2018] (Figure 3.2A) were used for performance assessement (Figure 3.2B). We also supply the known root and terminal states as supervision to Slingshot and Monocle3 (root states only). The dataset consists of five terminal regions and a common root state. The topology of the data is multifurcating with development branches arising directly from the root state, namely Astrocytes and Oligodendrocyte precursors, but also as branches from intermediate states, namely Neuroblast and Cajal-Retzius differentiation.

Trajectories and pseudotime estimated by each trajectory inference method are shown in Figure 3.2C-E. Cytopath estimates the correct number of trajectories to each terminal state while Slingshot produces a spurious trajectory despite having the terminal states as supervision. Monocle3 produces two parallel trajectories within the Granulocyte branch, wrongly dividing it into smaller parts. Monocle3 also fails to produce a connected trajectory, producing a disjoint graph and therefore unable to estimate pseudotime for a large portion of the dataset.

Spearman correlation between the pseudotime estimated by each method and the true lineage ordering of cell identities shown indicates that pseudotime estimated by Cytopath outperforms other methods in terms of recovering the expected ordering of cells (Figure 3.2B).

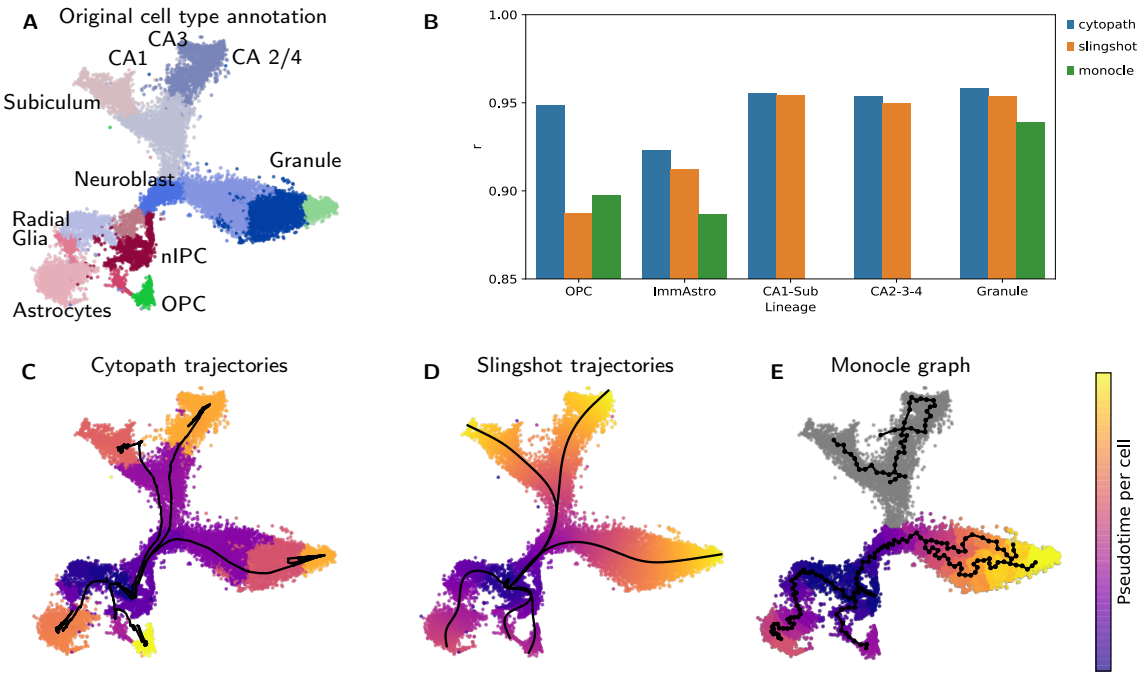


Figure 3.2: Reconstruction of neuronal differentiation in the developing mouse hippocampus. (A) t-SNE projection annotated with stages of neuronal differentiation as described in [Manno et al., 2018]. (B) Spearman correlation of pseudotime values assigned by each method to the known ordering of cell types per trajectory. (C-E) Trajectories inferred and mean pseudotime per cell by by Cytopath (C), Slingshot (D) Monocle3 (E).

3.2.3 Reconstruction of interlaced cell cycling and bifurcated differentiation in pancreatic endocrinogenesis

We further assessed trajectory inference performance for processes with multiple interlaced non-trivial topologies. To this end we considered a dataset studying pancreatic endocrinogenesis with lineages to four terminal states, alpha, beta, gamma and delta cells and dominant cell cycling at the onset of differentiation. [Byrnes et al., 2018]. Preprocessing, RNA velocity and the transition probability matrix estimation were performed with scvelo [Bergen et al., 2020].

Biological cell type annotation from [Byrnes et al., 2018] was used to provide root and terminal state supervision to Cytopath (Figure 3.3A). Trajectories estimated by cytopath to each terminal state capture the expected differentiation events of endocrinogenesis (Figure 3.3B, [Byrnes et al., 2018]).

Cell cycle scoring of cells in the root region revealed clearly distinct cell cycle states (Figure 3.3C). Figures 3.3D-F show the structure of the mean trajectory estimated by each method in the root region of this dataset. It is clear that the detailed structure of the trajectory inferred by cytopath captures the process of cell cycle. To verify this in we show the frequency of cell cycle phase as a function of the transition step of individual simulations comprising the final trajectory [Figure 3.3G]. Spearman correlation of cell cycle phase with the transition steps of each simulation indicate faithful recapitulation of the cell cycle stages at the single-simulation level (Figure 3.3H).

The simulation based approach of Cytopath ensures that even in the absence of explicit supervision, transcriptional patterns are faithfully reconstructed. In contrast, due to the absence of RNA velocity information, Slinghshot and Monocle3 consider the designated root states to be isotropic and therefore are unable to capture transcriptinonal heterogeneity within this region.

In Figure 3.4 we show the trajectory estimation with respect to the full pancreatic endocrinogenesis process. Slingshot and Monocle3 produce spurious trajectories. The correlation between the pseusotime estimated by Cytopath and the expected ordering of cells is the highest for all four lineages.

3.2.4 Reconstruction of bifurcating development of CD8 T cells from scRNA seq time series

We assessed the performance of Cytopath on a scRNAseq dataset from CD8 T cell development in chronic LCMV infection (see Chapter 4). In this infection model system T cells differentiate from early activated cells into exhausted and memory-like cells over three weeks' time. Samples were collected at four experimental time points after infection with LCMV, covering all stages of the process. Although these samples are heterogeneous snapshots of multiple populations at a particular time point they provide an approximate biological time of development. Starting from a shared early activated population, the cells start to differentiate into the two distinct terminal sub-populations five days after LCMV infection. This differentiation is characterized by strong transcriptional changes and expression of different surface markers.

We identified the root and end points of the process computing the equilibrium of the Markov process as in [Manno et al., 2018] (Fig. 3.5). The endpoints were validated by expression levels of known marker genes (see details Chapter 4) The exhausted endpoint showed high expression in co-inhibitory markers like CD39 (*Entpd1*), CD160 (*Cd160*) and PD-1 (*Pdcd1*). The memory-like endpoint had high expression of TCF1 (*Tcf7*) and IL-7R (*Il7r*). Applying Cytopath to our data resulted in two trajectories leading from a shared starting region to the two expected endpoints. The two trajectories strongly overlap in the beginning of the process but then sharply diverge at a branching point (Fig. 3.5D).

Comparing the pseudotime estimates from Cytopath with the discrete experimental time labels from the samples showed high agreement of the two (Fig. 3.6B). The experimental time-points,

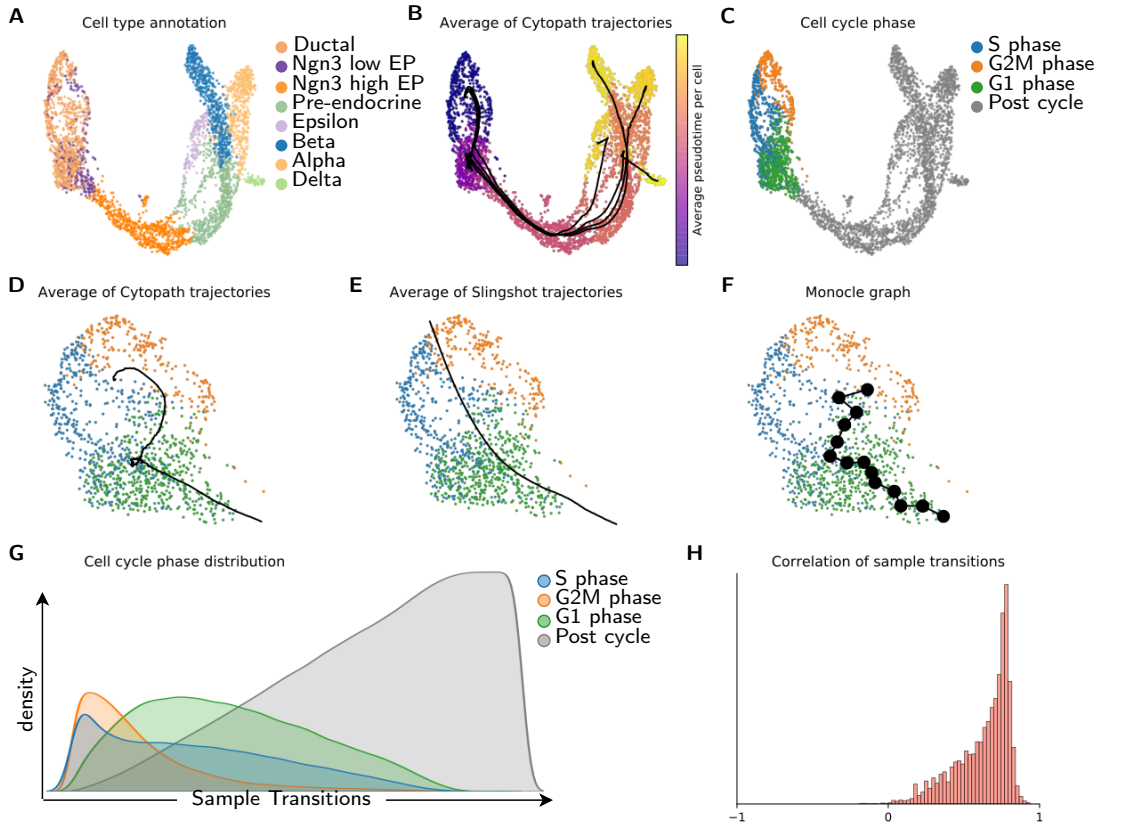


Figure 3.3: Reconstruction of interlaced cell cycling and bifurcated differentiation in pancreatic endocrinogenesis. (A) UMAP projection annotated with stages of neuronal differentiation as described in [Byrnes et al., 2018]. (B) Trajectories inferred by Cytopath and mean pseudotime per cell. (C) Estimated cell cycle phase assignment. (D-F) Cell cycle phase annotation and mean trajectory in the root region by Cytopath (D), Slingshot (E) and Monocle3 (F). (G) Cell cycle phase as a function of transition step of individual Cytopath simulations. (H) Spearman correlation of cell cycle phase with transition step of individual Cytopath simulations

which corresponds roughly to biological developmental time, were ordered correctly and with very little overlap between the time-points. Despite the non-trivial structure of the umap projection, cells are ordered correctly. Although the development speed is occasionally abrupt the relative ordering is respected with the two endpoints having the highest pseudotime. This illustrates that the average trajectory produced by Cytopath respect the experimental time ordering of the cells. Slingshot found a trajectory towards the exhausted endpoint but did not manage to infer a trajectory towards the memory-like endpoint. It generates two spurious trajectory that end half-way through the process. The pseudotime inference fits the sample labels for the exhausted branch, but does not match the experimental time of the memory-like branch. Monocle3 reconstructs the global structre of the data, but includes additional loops and branches within the exhaustion branch. Just like in Slingshot, pseudotime of the exhaustion branch is matching expectations, but the memory-like cell ordering does not recapture experimental time ordering.

We further tested the validity of the average trajectories of Cytopath by the expression profiles of known lineage marker genes in the differentiation process. The chemokine receptor CXCR6 has been shown to mark exhausted T cells in chronic LCMV infection [Sandu et al., 2020b]. The average expression of *Cxcr6* increases in the trajectory towards the exhausted cluster right before the divergence of the two branches (Fig. 3.6C), indicating that the paths are indeed governed by the exhaustion process.

Conversely, T cell Factor 1 (TCF1) and the expression of its gene *Tcf7* is an established marker

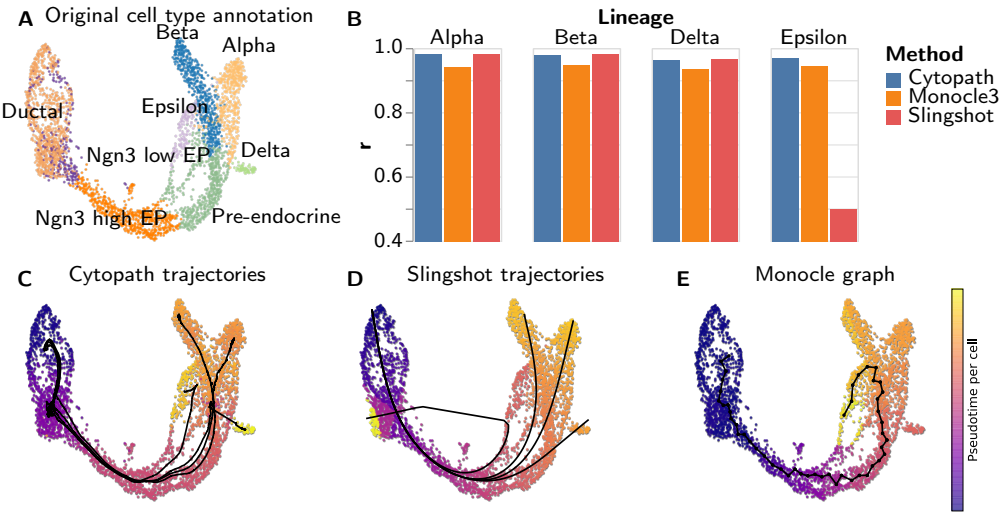


Figure 3.4: Reconstruction of differentiation in pancreatic endocrinogenesis. (A) t-SNE projection annotated with stages of pancreatic differentiation as described in [Bergen et al., 2020]. (B) Spearman correlation of pseudotime values assigned by each method to the known ordering of cell types per trajectory. (C-E) Trajectories inferred and mean pseudotime per cell by by Cytopath (C), Slingshot (D) Monocle3 (E).

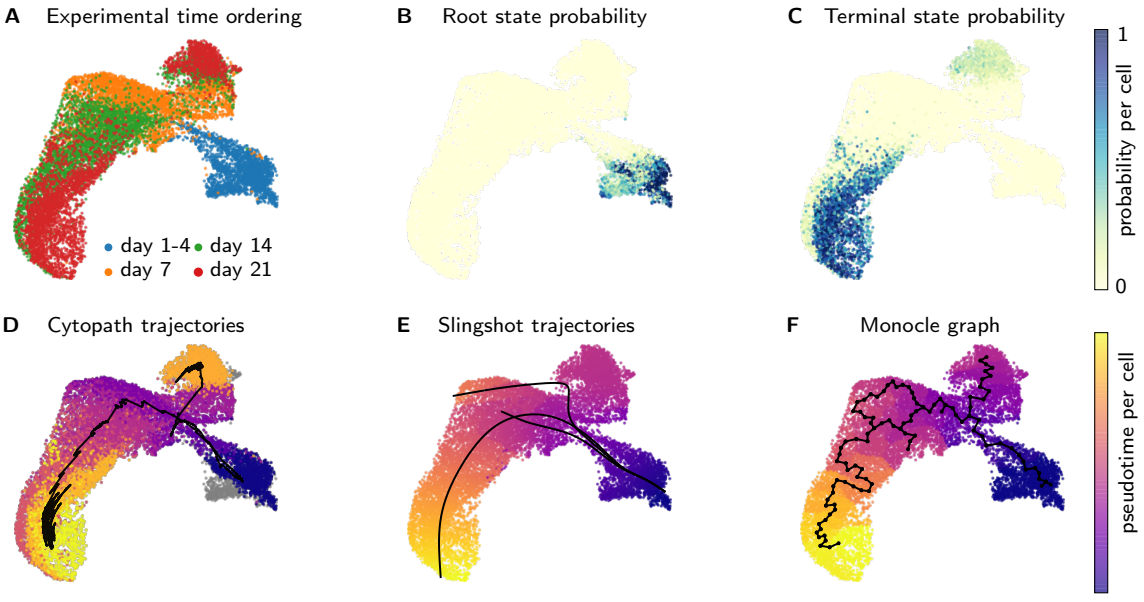


Figure 3.5: Reconstructuon of bifurcating development of CD8 T cells from scRNA seq time series. (A) Ordering of samples with respect to time post infection with LCMV C13. (B-C) Probability estimated based on RNA velocity of the cell being (B) a root state (C) a termimal state (D-F) Trajectories inferred and mean pseudotime per cell by Cytopath (C), Slingshot (D) Monocle3 (E).

of memory-like cells [Utzschneider et al., 2016c]. Expression of this gene was increased in memory-like cells just after the cells started to diverge after the bifurcation point. Towards the memory-like endpoint at late time-points, *Tcf7* expression is exclusive to the memory-like population.

An additional observation is the high expression of *Gzmb* early both branches that drops off towards later timepoints (Fig. 3.6C). The expression of *Gzmb* is a shared feature of both branches and known to decrease in both branches as the infection progresses and expression is low to-

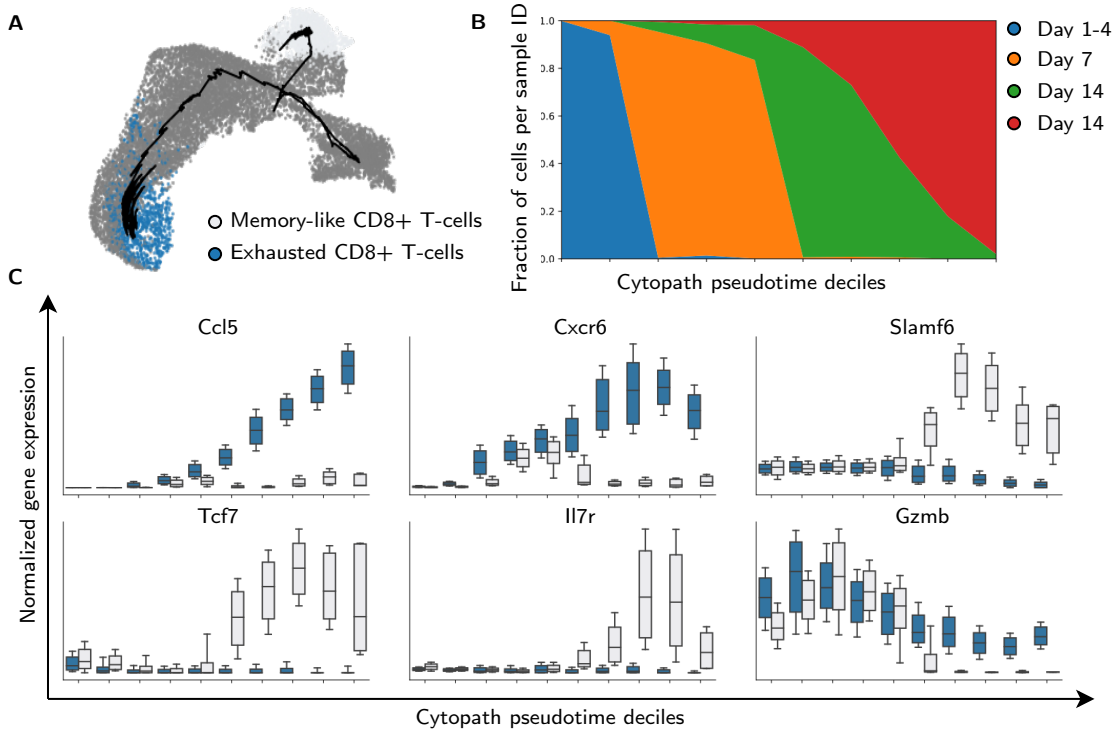


Figure 3.6: Cytopath trajectories identify exhaustion branch specific gene dynamics. (A) Biological identity of terminal states and trajectories inferred by Cytopath. (B) Fraction of cells from each experimental sample in each pseudotime decile. (C) Expression of key marker genes in each pseudotime decile.

wards late timepoints [Wherry et al., 2007].

We were able to reconstruct biologically relevant differentiation trajectories from a long-term time-series dataset. We identified correct differentiation branches of CD8 T cells in chronic infection, demonstrated by correct ordering of the experimental time labels. Further did we capture the temporal dynamics in expression of branch specific genes accurately.

3.3 Discussion

Trajectory inference is a challenging task since scRNA seq data is noisy and - until recently - has been evaluated to achieve only static expression profile snapshots. Inclusion of directional transcriptional activity estimates from RNA velocity analyses is expected to achieve more precise and sensitive trajectory inference. With cytopath we present an approach that takes advantage of this information and we demonstrate superior capability to reconstruct complex differentiation hierarchies from scRNAseq data.

Trajectory inference tools typically operate in low dimensional embeddings of scRNAseq data. Trajectory inference in lower dimensional spaces, especially two dimensional projections such as UMAP and t-SNE is particularly challenging for multifurcating datasets.

The branching structure of the data implies that the terminal states are further away from the root state however the constraint of two dimensional projection results in terminal states being either closer or further with respect to their coordinates in two dimensions relative to the differentiation process. This phenomenon results in similarity of cells in low dimensional space being unrepresentative of their transcriptional differences.

Trajectory inference with cytopath in addition to cell similarity considers the alignment of RNA velocity with the difference of the cell locations. This ensures that only biologically meaningful connections between cells are considered in the trajectories derived from these transitions.

In contrast, we found that Monocle3 is highly susceptible to projection artefacts resulting in unstable graph and pseudotime inference.

Another issue faced by conventional trajectory inference methods including Slingshot and Monocle3 is that the projection itself is typically dominated by the most prominent source of variation and often contains artefacts that mislead any inference that considers cell similarity in the projected space as a proxy for similarity of cell state.

Transitions between cells are inferred on the basis of the high dimensional expression and velocity profiles. Since Cytopath is based upon transitions that use the full expression and velocity profiles of cells, it is less prone to projection artifacts distorting expression profile similarity. In addtion, this approach specifically considers likely and discards unlikely transitions, and thereby is able to identify for instance cyclic trajectories in an apparently diffusely populated and isotropic region of expression space, such as the early stages of pancreatic endocrinogenesis. Furthermore, these hidden transcriptional patterns are made apparent by its simulation based approach. RNA velocity-based pseudotime and cell fate estimation based analytical analysis of Markov chain properties, performed by scvelo and CellRank, will not present the user this information about substructures even if the pseudotemporal ordering estimated by these tools captures these patterns.

These properties render Cytopath a prime tool to study complex differentiation processes such as the development of CD8 T cells in chronic infections. For this system several phenotypic populations and characteristic markers had been described before, but the connecting differentiation trajectories of those populations are a subject of ongoing research [Chen et al., 2019, Zander et al., 2019, Yao et al., 2019, Raju et al., 2020]. These studies provide evidence for branching in the development process, and only recently in conjunction with simulation based trajectory inference it was possible to resolve this event in more detail (see Chapter 4).

The competitor Monocle-3 produces too fine grained structures and seems to overfit the data, Slingshot produced too coarse trajectories, not capturing enough detail. Presumably, their assumptions made about granularity of the process are not met in our data. Cytopath has the advantage of velocity based transitions that can vary in across the dataset. Depending on the local properties of the underlying process, trajectories might be coarser or finer.

Validation of known markers changing monotonically along the inferred trajectory pseudotime supported their biological validity (see Chapter 4).

In addition to trajectory inference, Cytopath also returns the underlying simulations. Each simulation can be interpreted as a realisation of the biological process. Therefore these simulations can be utilized as input for time series mechanistic modelling of the biological process.

Furthermore, the velocity based alignment score of cells in a process results in each cell aligning to multiple steps of the trajectory with the degree of alignment quantified by the score. In addition to fate assignment, the distribution of steps that the cell aligns to, indicates the synchronicity of the process at that stage. In the analogy of Waddington's landscape, regions of narrow alignment could indicate low variability of cell state and correspondingly a transcriptional bottleneck with strong transcriptional directionality.

Conclusively, Cytopath is a promising tool to infer trajectories in novel biological systems of unknown structure and discover differentiation associated molecular markers. Additionally, it holds the option to further integrate and analyze the RNA velocity based simulations, opening up possibilities to characterize biological processes in greater detail.

3.4 Methods

Trajectory inference with Cytopath

Simulations are initialized at random cell states selected uniformly within the defined root state cluster and consist of a fixed number of cell to cell transitions. At each step, a single transition from the current cell state is realized based on the cell to cell transition probability matrix as described below.

The cell state c at step i of simulation j is selected as follows,

$$c_{ij} = \arg\min_k(\alpha - P_{ij}^k) \tag{3.1}$$

where $\alpha \sim U(0,1)$ and

$$P_{ij}^k = Pr(c^k | c_{(i-1)j}) \tag{3.2}$$

is the transition probability for k neighbors of $c_{(i-1)j}$

Simulations that terminate within defined terminal state clusters are considered for trajetory inference. Simulations were generated until a defined number of unique simulations have been returned.

By default a single trajectory is estimated for each root and terminal cluster pair. For each trajectory, simulations were aligned using the *fastdtw* python package. [Salvador and Chan, 2007]

The median value of the coordinates of cells at each step of the aligned simulations is the coordinate of the trajectory at that step.

For coordinate of trajectory at step i, λ_i^T with $\lambda \in \mathbb{R}^D$ where D is the dimensionality of the embedding

$$\lambda_i^T = \text{median}(\lambda_i^{(c_i)_l}) \; \forall \; l \tag{3.3}$$

For each step of the trajectory a fixed number of neighboring cells were recovered using K-dimensional tree search. For step i of the trajectory with neighboring cells l, cell type of cell

$$c_{in} = \omega_{im} \tag{3.4}$$

where $\omega_{im} \in (0...k)$ with k clusters. If the proportion of cells belonging to a particular cluster exceed a defined threshold then the cluster was considered as a compositional cluster of the trajectory.

$$\sum_0^M \mathbb{1}_k(\omega_{im} = k) > \nu \cdot M \tag{3.5}$$

where ν is the threshold and if the condition is met k is a compositional cluster. Within a trajectory, cells within a defined neighborhood at each step and belonging to a compositional cluster were assigned to the step. The average step value of a cell is the pseudotime value with repsect to the trajectory. For a dataset with N cells, pseudotime of a cell with respect to a particular trajectory, π_s is as follows,

$$\pi_s = \text{mean}(i \cdot \mathbb{1}(c_{im} = c_n)) \tag{3.6}$$

Alignment score of a cell with neighbors k to step i is the maximum of two scores. The score with respect to the trajectory segment from *i-1* to *i*, ξ_i is calculated as follows,

$$\xi_i = \frac{1}{k}\sum_k cos(\eta_k^b) \cdot exp(\gamma_k) \tag{3.7}$$

where η is the cosine angle between the section of the trajectory and all possible transition partners k of the cell. γ is the unormalized transition probability of the cell to it's neighbors. The score with respect to the trajectory segment from *i* to *i+1*, τ_i is calculated similarly,

$$\tau_i = \frac{1}{k}\sum_k cos(\eta_k^f) \cdot exp(\gamma_k) \tag{3.8}$$

The alignment score of the cell with respect to step i, D_i is

$$D_i = max(\xi_i, \tau_i) \tag{3.9}$$

Processing of the Dendate Gyrus dataset

The dataset was obtained and processed using the following notebook (https://github.com/velocyto-team/velocyto-notebooks/blob/master/python/DentateGyrus.ipynb) provided by [Manno et al., 2018].

Biological cell type annotation and RNA velocity derived root states were used to supervise trajectory inference for Cytopath, Slingshot and Monocle3 (root states only). Inference was peformed in the original T-SNE projection using default parameters for Slingshot and Monocle3. For Cytopath, in addition to the projection the original RNA velocity field was used to infer cell to cell transitions.

Processing of the pancreatic endocrinogenesis dataset

Data was downloaded and processed using scvelo [Bergen et al., 2020]. RNA velocity was estimated using the stochastic model. Biological cell type annotation was used to supervise trajectory inference for all methods. Inference was performed in the original UMAP projection using default parameters for Slingshot and Monocle3. Cell cycle scoring was performed using the scanpy analysis pipeline [Wolf et al., 2018].

Processing of the CD8 development dataset

Transgenic P14 CD8 T cells were sampled longitudinally during chronic infection with LCMV Clone-13 infection. The samples were acquired from four phases of the infection, namely activation (day 1-4), effector (day 7), early exhaustion (day 14) and late exhaustion (d21) and scRNAseq was performed using the 10x Genomics platform (see Chapter 4 for details).

Read counts were realigned and sorted for spliced and unspliced counts using the analysis pipeline from velocyto [Manno et al., 2018]. Contaminating other cell types were removed from the dataset based on outliers in diffusion components. Reads were filtered and normalized according to the Zheng recipe [Zheng et al., 2017] of the scanpy analysis pipeline [Wolf et al., 2018] retaining 5000 highly variable genes. Louvain clustering and UMAP projection were computed using standard parameters, using the first 50 principle components.

Comparison of trajectory inference

Cluster containing cells with greater than 0.95 probability of being root states were designated as the root state in all datasets. Terminal states were defined as described in the original publications. Spearman correlation of the pseudotime values generated by each method with the cell type cluster ordering for each biological lineage was used to compare the performance of the three methods.

Chapter 4

Fate trajectories of CD8$^+$ T cells in chronic LCMV infection

Dario Cerletti[1,2], Ioana Sandu[1,2], Revant Gupta[1,&], Annette Oxenius[2,*] and Manfred Claassen[1,*,&]

[1]Institute of Molecular Systems Biology, ETH Zurich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland

[2]Institute of Microbiology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland

[&]Current address: Internal Medicine I, University Hospital Tübingen, Faculty of Medicine, University of Tübingen, Otfried-Müller-Straße 10, 72076 Tübingen, Germany

[*]shared contribution

D.C., I.S., M.C., A.O. designed the experiments; D.C., I.S. carried out the experiments, D.C., R.G., M.C., A.O. analyzed the experiments; D.C., M.C., A.O. wrote the manuscript.

Abstract

In chronic infections CD8$^+$ T cells acquire a state termed "exhaustion" which is characterized by impaired effector functions and expression of co-inhibitory receptors as response to continuous TCR stimulation. Recently, the pool of exhausted T cells has been shown to harbor multiple functionally distinct populations with memory-like and effector-like features, though differentiation and lineage relations between these are unclear. In this work we present a comprehensive scRNAseq time-series analysis from beginning of infection to established exhaustion in CD8 T cells. We apply lineage inference using informed cell transitions derived from RNA velocity to identify differential start and end states and connections between them. We identify a branch region early during chronic infection where pre-committed cells separate into an exhausted and a memory-like lineage and discovered molecular markers demarcating this branch event. Adoptive transfer experiments confirmed fate-commitment of cells only after this branch point. We additionally linked the progression along developmental lineages to antigenic TCR stimulation.

4.1 Introduction

Viral infections with human immunodeficiency virus (HIV), hepatitis C virus (HCV) and in mice with lymphocytic choriomenengitis virus can result in chronic infection with ongoing viral replication and high antigenic load for weeks or months. This continuous exposure to antigen will drive CD8$^+$ T cells into a functionally distinct phenotype, termed exhaustion [Wherry et al., 2007]. This state is characterized by functional, transcriptional and epigenetic changes that result in expression of co-inhibitory receptors such as PD-1, LAG-3, 2B4 and CD160, decreased secretion of cytokines like INFγ and TNFα as well as reduced proliferation and survival. Acquisition of this exhausted phenotype is a continuous and gradual process driven by excessive TCR stimulation [Utzschneider et al., 2016b].

In this context of chronic antigen exposure, CD8 T cells undergo a differentiation program that differs markedly from the one observed during acute resolved infection. Previous studies have analyzed and inferred differentiation trajectories of virus-specific CD8 T cells using bulk or single cell transcriptomic profiling in various systems, including the model of chronic LCMV infection [Utzschneider et al., 2016a, Chen et al., 2019, Zander et al., 2019]. Most of these studies have inferred differentiation trajectories based on bulk or single cell analyses performed at one or two time points during chronic LCMV infection [Chen et al., 2019, Zander et al., 2019].

Asynchronicity in this process as well as different micro-environments that CD8$^+$ T cells experience result in a heterogeneous population of cells at a given time point of the infection. One sub-population of virus-specific T cells acquires a phenotype that shares properties with memory T cells from acute infection and has been linked to the expression and activity of T cell Factor 1 (TCF1) [Utzschneider et al., 2016a] [Miller et al., 2019]. In contrast to terminally exhausted or effector T cells, these cells retain proliferative activity and have better survival in the infected host [Chen et al., 2019]. It is not yet fully understood how and when these different cell states arise during the course of the infection and which intermediate cell states precede these.

Recent advances in sequencing technologies have made it possible to profile cells genome wide on the transcriptional level using single-cell RNA sequencing (scRNAseq). This technology allows capturing the transcriptional heterogeneity of multiple cell populations and to computationally infer orders of cell states traversed during dynamic processes such as T cell differentiation in chronic infections. When analyzing scRNAseq data-sets, cells are treated as points in transcriptome space based on their expression profile. Dimensionality reduction techniques like t-SNE [Van Der Maaten and Hinton, 2008] and UMAP [Becht et al., 2018] construct two-dimensional representations for analysis and interpretation of the high dimensional single-cell expression data. Pseudotime and lineage inference methods aim at constructing likely transitions between cell states [Saelens et al., 2019].

Recent studies aimed at reconstructing cell state sequences of CD8$^+$ T cell differentiation in chronic LCMV infection [Chen et al., 2019], [Zander et al., 2019], [Yao et al., 2019]. They discovered multiple phenotypic subsets, namely memory-like, terminally exhausted and effector-like cells and investigated likely transitions between these subset. However, these studies lack temporal resolution to reliably infer trajectories and identify potential branching events in differentiation. Samples were either generated from different infection settings at single time-points, or at far spaced time-points. Further, applied trajectory inference methods infer pseudotime and lineages based on similarity of transcripts and lack taking advantage of all the information present in the scRNA seq data. Directionality information is now – in principle – available for trajectory inference via RNA velocity analysis. RNA velocity [Manno et al., 2018] considers additional information about the ratio of un-spliced to spliced mRNA in transcript data, which serves as a measure to determine the stage (early, intermediate, late) of individual gene expressions and allows to predict the future expression state and hence to better infer the directionality towards their neighbors in the high-dimensional transcriptional space. So far no study leveraged RNA velocity in order to include this information to disambiguate the results from conventional trajectory inference.

In this work we conducted scRNAseq measurements at multiple time-points ranging from the beginning of chronic LCMV infection until manifestation of exhaustion three weeks after infection. This level of time resolution allows more detailed identification of cell states and their differentiation. We further included information from RNA velocity analysis to perform simulation based trajectory inference of differentation events leading to the different terminal CD8$^+$ T cell states observed in chronic LCMV infection. This is the first attempt to make use of RNA velocity to produce informed differentiation trajectories that connect the different cell states. This analysis allowed us to construct faithful lineage trajectories towards the two endpoints of differentiation, namely a terminally exhausted and a TCF1$^+$ cell population. We identified a potential branching point in the initially shared trajectories and validated our findings using adoptive transfer experiments of cells positioned before or after the branching point. We confirmed that

cells before the branch point gave rise to both exhausted and TCF1$^+$ cells, whereas exhausted cells after the branch point maintained their phenotype. Additionally, we demonstrated that TCF1$^+$ cells largely retained their phenotype in absence of antigen stimulation, corroborating the end-point differentiation characteristics of this population. However, if exposed to antigen stimulus, the TCF1$^+$ population has the ability to differentiate into terminally exhausted cells, in line with previous adoptive transfer experiments.

4.2 Results

We first investigated the differentiation landscape of CD8 T cells, followed by RNA velocity analysis to reveal developmental endpoints. Afterwards we identified two branching trajectories towards memory-like and exhausted cell states, respectively. We validated commitment to the branches using adaptive transfer experiments and additionally highlight the importance of antigen stimulation during development.

4.2.1 Differentation landscape of CD8 T cells during chronic LCMV infection

We acquired single cell transcriptomic data from multiple time points during chronic infection, covering the very early phases (day 1-4), peak phase (day 7), contraction phase (day 14) and late phase (day 21) (Fig. 4.1), with the aim to capture an increased spectrum of the transcriptional landscape during the course of the infection that would allow a time-resolved analysis of single cell heterogeneity and possibly more accurate inference of differentiation trajectories of virus-specific CD8 T cells. To this end, T cell receptor (TCR) transgenic (tg) LCMV gp33-41-specific CD8 T cells (P14 cells) were adoptively transferred into naïve C57BL/6 mice, followed by infection with LCMV clone 13 (Cl13). Activated and expanded P14 cells were isolated at the above indicated time points and subjected to single cell RNAseq (scRNAseq) analysis using the 10x Genomics platform.

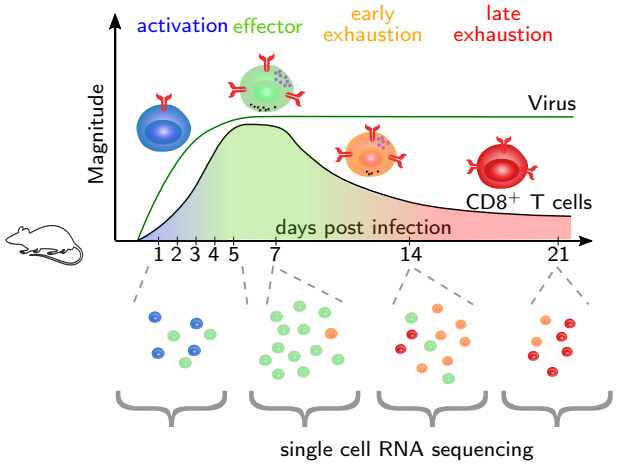


Figure 4.1: Transgenic P14 CD8 T cells were sampled longitudinally during infection. The samples were acquired from four phases of the infection activation (day 1-4), effector (day 7), early exhaustion (day 14) and late exhaustion (d21) and scRNAseq was performed using the 10x Genomics platform.

For exploratory analysis of the transcriptome data of the multiple time points, we applied commonly used filtering and scaling of the raw data [Zheng et al., 2017] and applied principal component analysis to reduce noisy signals. The resulting multidimensional data was projected into two dimensions using UMAP [Becht et al., 2018].

The UMAP projection revealed a continuous structure of the data emerging from the time-resolved samples, supportive of a developmental process (Fig. 4.2a). Louvain clustering [Blon-

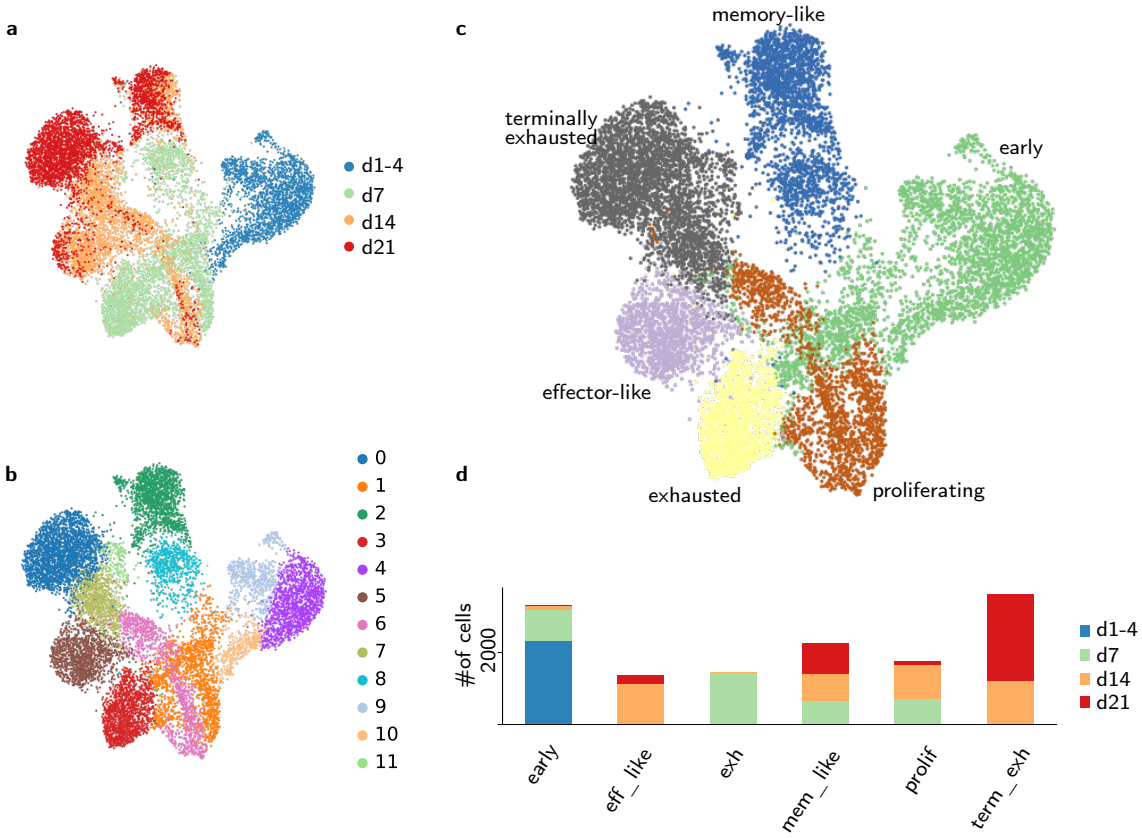


Figure 4.2: UMAP projections of the filtered and normalized transcript counts are shown. (a) Color indicates the time-point after infection, at which cells were isolated for scRNAseq (b) louvain cluster assignment based on the first 50 PCs (c) phenotypic cluster annotation based on previous marker genes and DE genes [Chen et al., 2019] (d) Cell composition of the phenotypic clusters by sample time-point.

del et al., 2008] defined 11 distinct clusters across all samples (Fig. 4.2b). Based on previously described markers for the exhausted subsets, such as CD160, CX3CR1 and TCF1 [Utzschneider et al., 2016a, Wherry et al., 2007, Chen et al., 2019], we further aggregated these cluster into 6 phenotypic groups (Fig. 4.2c). Activated cells from day 1 to 4 post infection (dpi) clustered at one peripheral region in the data, termed in the following **early group** (green in Fig. 4.2c). Differential expression revealed that the early group presented expression patterns of proliferation as well as of exhaustion, indicated by expression of *Mki67*, *Cdca3* but also *Cd160*.

Conversely, two distinct populations from the latest time point at 21 dpi clustered at the other extremes of the spectrum. Of these "late" endpoints, one population showed high expression of a number of inhibitory receptors, including *PD-1 (Pdcd1)*, *CD39 (Entpd1)*, *LAG-3 (Lag3)* and *CD160 (Cd160)* (Fig. 4.3a, b), indicating a terminally exhausted phenotype [Blackburn et al., 2008a]. This **terminally exhausted group** (grey in Fig 4.2c) was composed of clusters from d7 and 14 and had the highest expression in co-inhibitory receptors and additionally showed high expression of the transcription factor *EOMES*. The other end-point populations showed high expression of the transcription factor *Tcf7* (Fig. 4.3a, b), the memory-marker *Il7r* as well as *Slamf6*, revealing this cluster as the previously described memory-like population [Utzschneider et al., 2016a]. This **memory-like group** (blue in Fig 4.2c) was composed of two louvain clusters from day 7, 14 and 21, all having high expression of *Tcf7*.

Cells from 7 and 14 dpi were situated in between the 14 dpi and 21 dpi samples, with the 7 dpi samples being similar to cells from early time-points on one end of the spectrum, but also connecting to already splitting trajectories into exhausted and memory-like populations on dpi

14. At day 7 was one cluster identified that presented clear signatures of exhausted CD8 T cells but retained some expression of *Gzmb* but also apoptotic genes like *Anxa1*, we termed this the **exhauted group** (yellow in Fig. 4.2c).

At 14 dpi, some of the cells were still connected to the 7dpi cell states but a considerable fraction of cells had already further differentiated towards the two endpoints of 21dpi, in particular towards the memory-like endpoint. One louvain cluster expressing *Cx3cr1* exclusively was termed **effector-like group** (purple in Fig. 4.2c). Differential expression analysis (Fig. 4.3a, Supp. Fig. S2) revealed higher expression levels of *Cx3cr1* and additionally killer lectin receptor genes (*Klre1*, *Klra3*). These effector-like cells were only present in samples from day 14 and 21.

We also identified a strong cell cycle component in the two clusters presenting high expression levels of e.g. *Mki67* (cluster 1 & 6) (Fig. 4.3b). Further, we calculated scores for cell cycle and cell division genes based on the three different cell cycle stages G1 phase, S phase and G2/M phase (Supp. Fig. S3). We observed that this **proliferating group** (brown in Fig. 4.2c) presented high scores for G2/M phase and was composed of cells from the 7 and 14 dpi and to a lesser degree from the 21 dpi time-point.

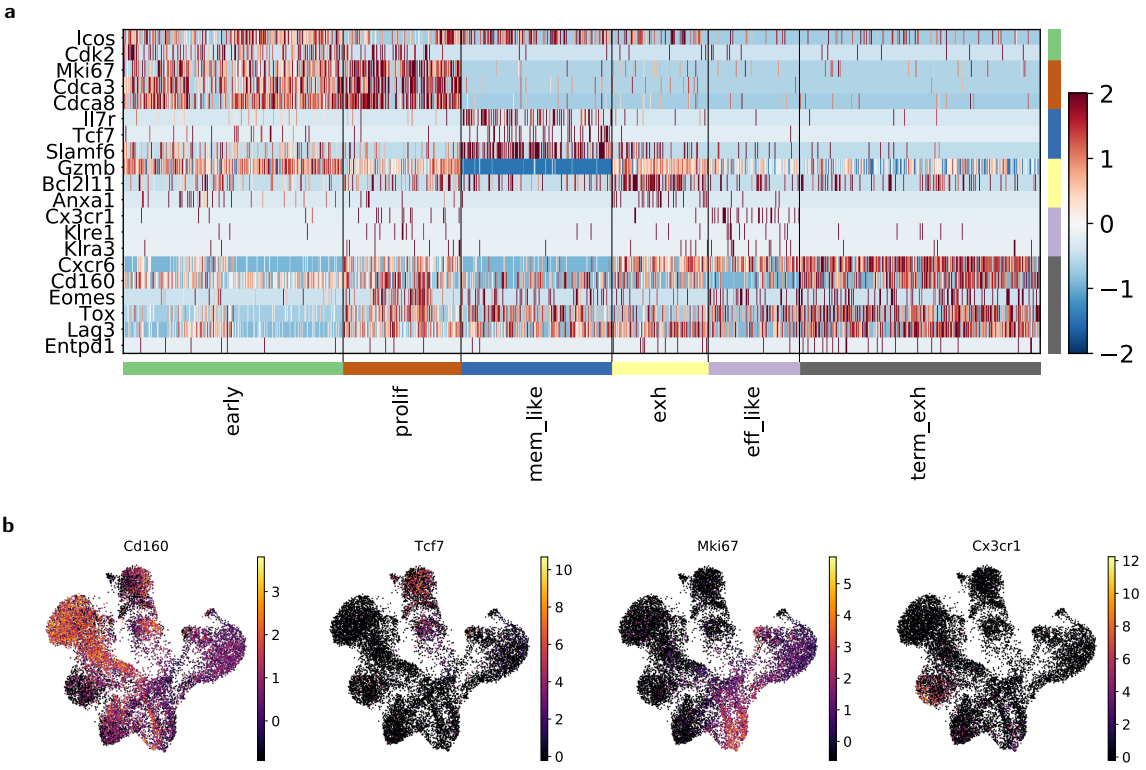


Figure 4.3: (a) Heatmap of normalized gene expression for a selection of group specific genes. Columns are individual cells arranged by phenotypic groups. The genes in the rows are grouped according to their phenotypic assignment.
(b) UMAP projection of expression pattern of identified group specific genes for the terminally exhausted (Cd160), memory-like (Tcf7), proliferating (Mki67) and effector-like group (Cx3cr1)

4.2.2 RNA velocity analysis reveals developmental end points

Having mapped serial single cell transcriptomes into a continuous landscape of cellular states, we next aimed at inferring informed differentiation trajectories into this landscape. To this end we leveraged RNA velocity [Manno et al., 2018] and applied it to our longitudinal data set, revealing a vector field demarcating likely transitions for each cell in the dataset. Applied to our dataset, this analysis clearly revealed a transition flow from early activated cells at 1-4 dpi towards early (dpi 7) and late exhausted cells (dpi 14 & 21) (Fig. 4.4a).

Using the calculated cell transitions, we defined a Markov process with cells as the states and transition probability estimates from RNA velocity. Computing the equilibrium distribution of the forward Markov process corresponded biologically to the most differentiated phenotypes as the cells transitioned to more differentiated states until they had acquired their final transcriptional state and did not differentiate further. Conversely, we inverted the transition probabilities and computed the equilibrium of the backward Markov process. Cell states thereby transitioned to their most likely previous transcriptional state resulting in the most undifferentiated state, having the highest probability.

This allowed us to assign cells that are most likely at the start and at the end of the differentiation process. The highest probably of a start region was at the edge region of the early group (Fig. 4.4b). This seemed plausible, as we would expect that the differentiation process started at the edge of the earliest sample (dpi 1-4), where there are no preceding cell states. This region showed gene signatures indicating strong DNA synthesis and cell cycle activity, that conferred an activated phenotype (Supp. Fig. S4).

The highest probability for end points was in regions from dpi 21 in the terminally exhausted group (0.5 on average, maximum 1.0). In this region many signaling related genes had changed expression, which could be a result of co-inhibitory receptor signaling (Supp. Fig. S4). Additionally, there was a local maximum in end point probability in the memory-like group from dpi 21 (0.1 on average, maximum 0.3). Differentially expressed genes in this region comprised typical genes of the memory-like signature, namely *Il7r* and *Tcf7* (Supp. Fig. S4). Both the terminally exhausted cells as well as the memory-like cells seemed to comprise an endpoint of differentiation state.

We assessed and confirmed the robustness of the velocity fields by confirming the practical equivalence of start and endpoint estimates across 574 parameter variants (Supp Fig. S5).

RNA velocity analysis additionally indicated that the process from the start to the end-points is gradual, since there were a multitude of intermediate transcriptional states between the two extremes. We observed also transitions between all these intermediate states (Fig. 4.4a).

4.2.3 Simulation based inference reveals trajectories towards exhausted and memory-like phenotypes

Based on the high-dimensional vector field resulting from the RNA velocity analysis, we used the calculated transition matrix to simulate differentiation trajectories for thousands of single cells. We aimed at understanding the developmental paths that an activated CD8$^+$ T cell could follow to acquire the two differentiated end-point phenotypes. The probability of a cell moving from one transcriptional state to another can be approximated by the transition probabilities from RNA velocity. We used the calculated root cells (Fig. 4.4b) as starting points for stochastic simulations. Each differentiation step in the 13728x13728-dimensional matrix was simulated according to the transition probability until one of the previously defined end stages (Fig. 4.4b) was reached. This sequence of steps approximated one possible path of differentiation for each cell (Supp. Fig. S6). We simulated 2000 trajectories per endpoint to sample the whole spectrum of possible differentiation trajectories. We observed a strong disbalance in preference for the endpoints. Only about 1% of the simulated sequences ended up in the memory-like cluster, whereas the remaining ones differentiated into the exhausted endpoint. We expected the ratio

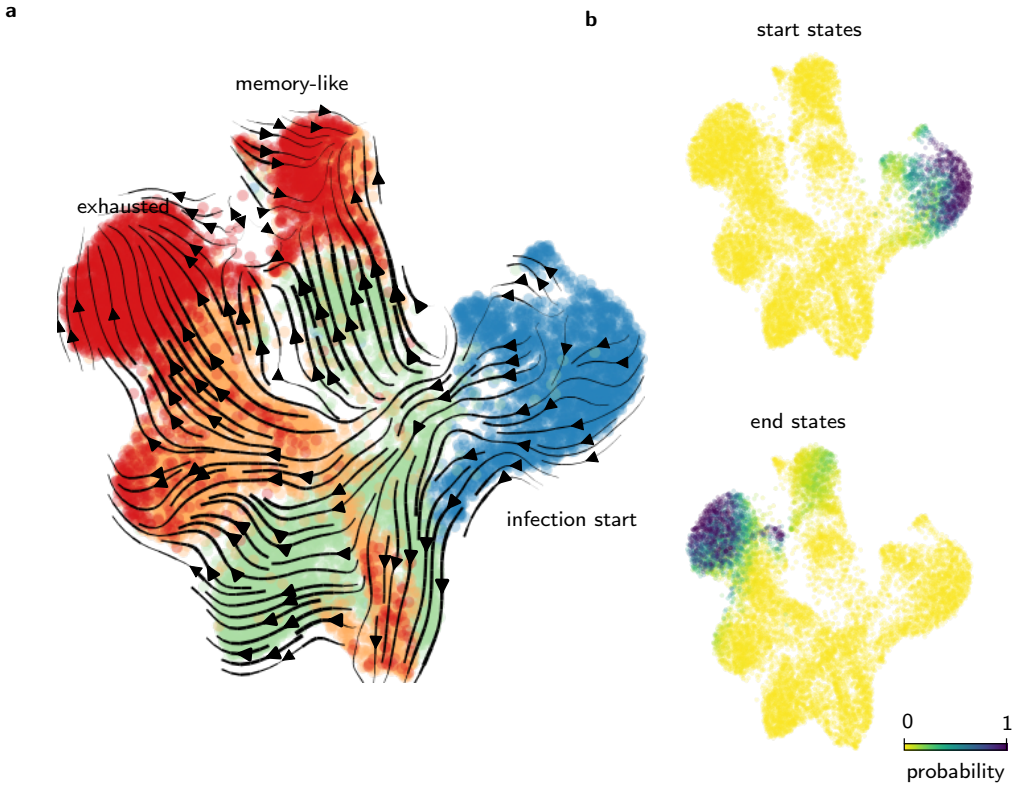


Figure 4.4: (a) stream plot visualizing likely transitions between cells inferred from RNA velocity
(b) The stationary distribution of the backward and the forward transition matrix, respectively, indicate start and end cell states.

to be shifted towards the exhausted phenotype but not to this extent, since we measured around 10% memory-like cells at day 21. We performed more simulations towards the memory-like endpoint to balance the number of trajectories. All the obtained trajectories were then aligned using dynamic time warping and clustered to generate average trajectories. Each cell was assigned to the nearest average trajectory according to the alignment score (Method section 4.4), calculated from cosine distance of its RNA velocity with the direction of the trajectory (Fig. 4.6a). This resulted in a temporal ordering of the cells in conjunction with a score to which trajectory it belongs. The detailed procedure of Cytopath is described in Chapter 3).

Our analysis revealed two main trajectories, one towards the exhausted endpoint the other to the memory-like endpoint (Fig. 4.5). Both trajectories shared the same cell populations up to a region that was composed of cells obtained from around 5 dpi. From thereon the trajectories started to diverge into the two phenotypic branches. The differentiation trajectory towards the exhausted path included the cell population with high cell cycle activity. The memory-like trajectory did not seem to pass the region of high proliferation, but diverged earlier and transitioned towards the memory-like endpoint.

Multiple genes were found to be differentially expressed between the two trajectories (Supp. Fig. S7). In the memory-like trajectory *Slamf6*, *Ccr6*, *Tnfsf8*, *Xcl1* and *Cxcl10* were expressed at higher levels. Many of them showing gradually increasing expression towards the differentiated endpoint (*Slamf6*, *Ccr6*, *Tnfsf8*). *Xcl1* was highest at the start of the trajectory and later decreased but was still maintained at much higher levels than on the exhausted branch. The exhausted branch showed increasingly higher expression of *Cxcr6*, *Ccl5* and *Nkg7* as the trajectory progressed towards the end point. The two genes *Ifngr1* and *Lgals3* were transiently upregulated in the exhausted trajectory exclusively, but decreased towards the end point (Supp. Fig. S7).

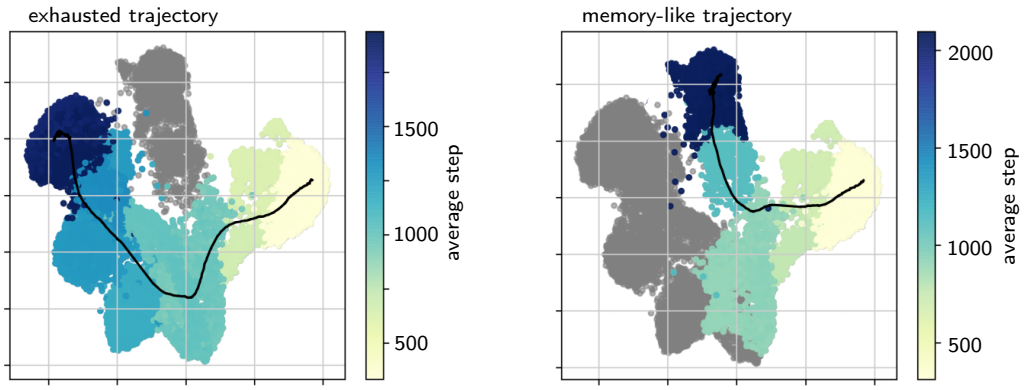


Figure 4.5: Average trajectories after simulations towards the two endpoint populations using cytopath. The average simulation steps to arrive at a cell are color coded per cluster. The coordinates of the average trajectory after alignment is depicted as black line. The shown average trajectories are based on 2000 simulations per endpoint.

4.2.4 Branching point of differentiation towards the end points

Based on the simulated mean trajectories we wanted to identify a branching point at which the trajectories of cells moving towards the exhausted and the memory-like endpoints would diverge. This branching region would demarcate the point after which a cell would be committed to only one endpoint.

We therefore computed the ratio of the alignment score for each cell to the two average trajectories (Fig 4.6a). We observed that in the unstructured early region all cells had equal scores for both the exhausted and the memory-like fate. However, between average simulation step 800 and 1200, the cells started to be uniquely assigned to only one of the two trajectories. Interestingly, we observed a region along the differentiation trajectories where some cells aligned clearly to the exhausted trajectory, some to the memory-like trajectory and some to both. It seemed that this region was where bifurcation took place. Using a threshold on the alignment score we assigned all cells to either the exhausted branch (blue), memory-like branch (orange) or pre-committed branch (green, Fig. 4.6).

To determine which biological time-point would correspond to the identified branch region, we investigated the branch composition of the four samples (Fig. 4.6c). The earliest samples were almost exclusively composed of pre-committed cells, whereas the day 7 sample contained already a large fraction of lineage-committed cells. We reasoned that bifurcation must take place between day 5 and 6 after infection.

Differential gene expression between the branching region and its immediately adjacent committed branches did not reveal any clear transcriptional signatures, that would precede or succeed bifurcation. However the trajectory assignment clearly implied that after the branching region, cells were fully committed to their lineage. Since the alignment score also considers the velocity direction of each cell, alignment to only one trajectory indicates that differentiation will take place along this path. The abrupt increase in the alignment score after the branch region suggests that cells beyond the bifurcation do exhibit coherent and significant velocity away from the bifurcation and this process is unlikely to be reversible.

4.2.5 Identification of marker genes predictive for different developmental fates

To further demarcate markers that would specify the branching point, we searched for genes that were characteristic for this bifurcation - either being expressed at divergent levels before, at, or after the branching point. We trained a classifier to predict the assigned branch label from the transcriptional profile of each cell (see Method section). If a gene was expressed in one branch but not the other, it was considered relevant for the prediction. To validate the

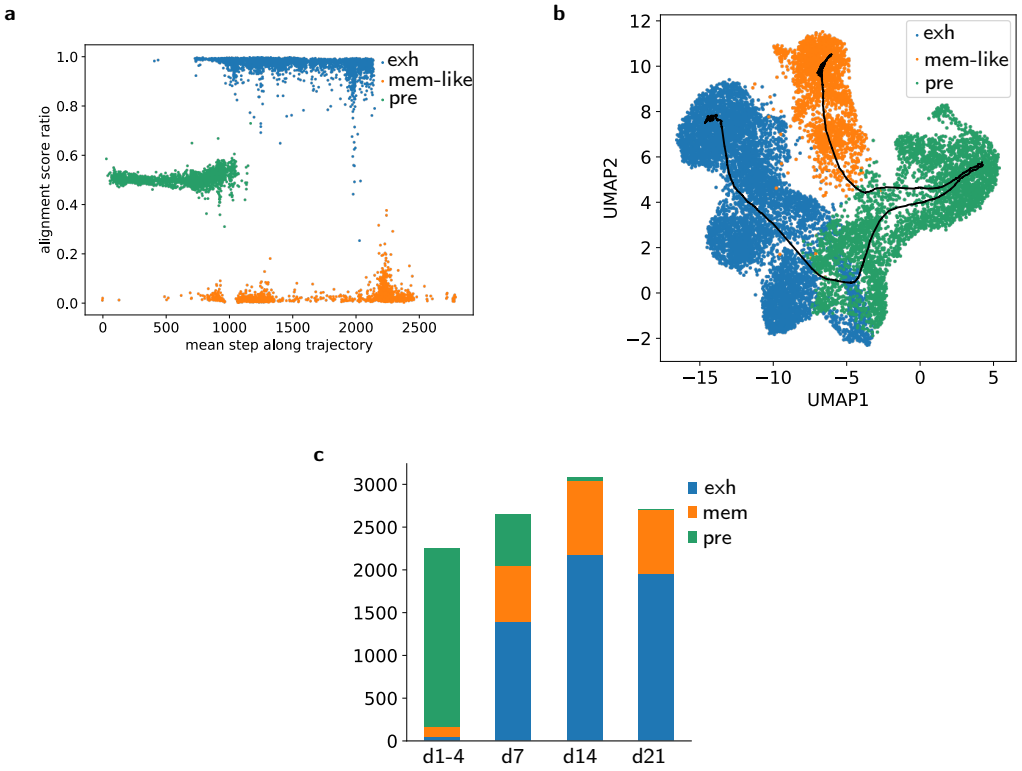


Figure 4.6: Cells are aligned to the branches of the two fates.
(a) The two average trajectories (black) towards the exhausted (blue) and the memory-like fate (orange) with an pre-commited shared part (green).
(b) Alignment score ratio along the cytopath simulation steps. High values indicated preferential alignment with the exhaustion trajectory, low values alignment with the memory-like trajectory.
(c) sample time-point composition with respect to branch assignment.

identified branching markers later on a protein level using flow cytometry, we restricted the transcriptional input to genes transcribing proteins with validated antibodies for staining. The result of this analysis variant allowed us to sort and experimentally analyze the differentiation potential of pre-committed and committed cell states in later validation experiments.

The classifier identified 12 genes that were most relevant to distinguish the three branches (Fig. 4.7). These included already described markers of the memory-like population, such as *Il7r* (IL7R), *Tcf7* (TCF1) and *Slamf6* (Ly108) [Utzschneider et al., 2016a, Chen et al., 2019], but revealed also potential new candidates *Icos* (CD278), *Ly6e* (SCA-2) and *Itgb1* (CD29) that are highly expressed in cells from the memory-like branch. Markers relevant for the exhausted branch contained *Cxcr6* (CXCR6), *Ifngr1* (IFNGR1) and *Cd3g* (CD3G) but also *Selplg* (CD162), of which only CXCR6 has been linked previously to exhaustion [Sandu et al., 2020b]. The pre-committed branch showed high expression of *Gzmb* (Granzyme B) and Mif (MIF) both of which were expressed at lower levels in the other lineages.

Predicting the branch labels using only these markers resulted in good prediction accuracy (0.85). Additionally, using only these genes as input to UMAP showed a good separation into the three branches (Supp. Fig. S8).

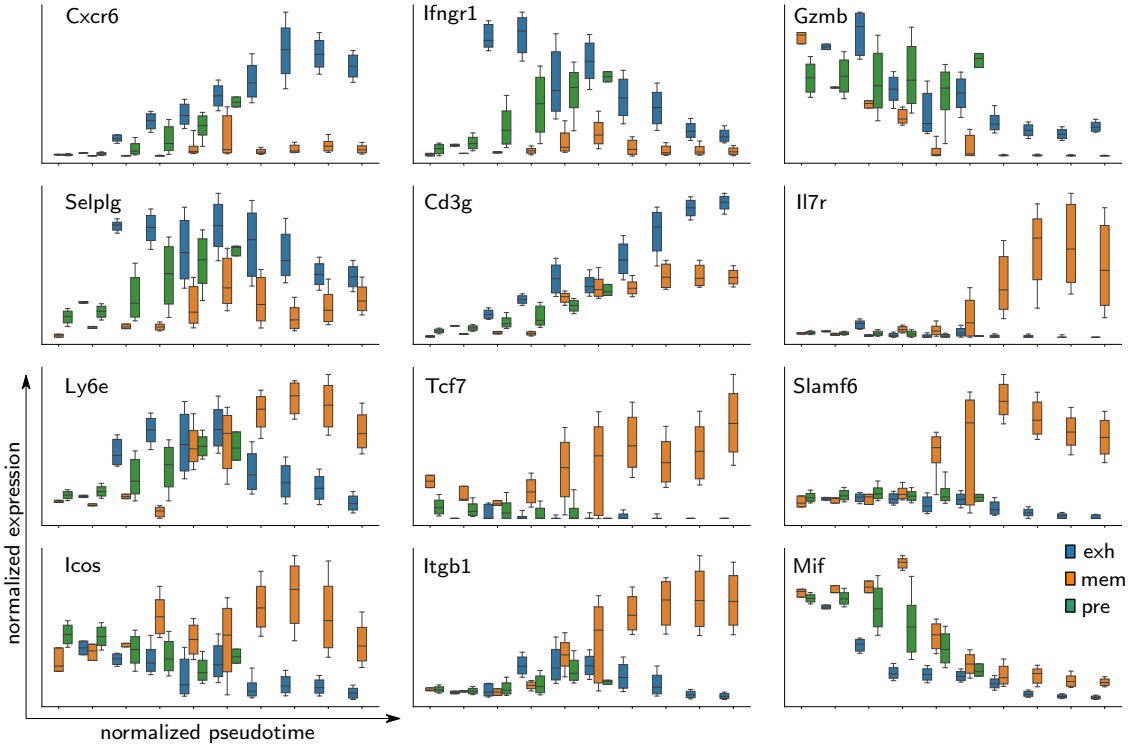


Figure 4.7: The 12 genes identified by the classifier to predict the three branch labels. Boxplot of normalized expression are shown along normalized pseudotime from cytopath. Color indicates the branch the cells were assigned to based on the Cytopath alignment score.

4.2.6 Experimental transfer of cells with a presumably pre-committed or committed phenotype show distinct differentiation potential

To experimentally validate our branch classification, we set out to identify and later sort LCMV-specific CD8 T cells with phenotypes indicative of a pre-committed state, a committed state towards the exhausted endpoint or the memory-like endpoint. For validation of phenotypic markers classifying these three cell states, we tested all 12 identified markers for their ability to discriminate early populations to then test their fate potential. Specifically, we transferred naïve P14 cells into naïve C57BL/6 mice, followed by chronic LCMV infection. 5 days post infection (when representatives of all three populations of interest had formed, i.e. the uncommitted cells and the committed exhausted and memory-like cells), P14 cells were analyzed according to the markers identified using the classifier. We first identified protein markers that showed high variance at the branching time-point. We determined CXCR6 and TCF1 as the prime candidates for sorting the branches into pre-committed (CXCR6$^-$ TCF1$^-$), memory-like (CXCR6$^-$ TCF1$^+$) and exhausted (CXCR6$^+$ TCF1$^-$) cells.

Having identified markers that allowed to distinguish between uncommitted and committed cells into the exhausted and memory-like branch, we used these markers to isolate the respective populations 5 days into chronic LCMV infection. Specifically, we transferred naïve P14 T cells expressing GFP under the TCF1 promoter into C57BL/6 mice and infected them with high dose LCMV Clone-13 (Fig. 4.8a). At dpi 5 we sorted P14 cells from the three branches according to expression of CXCR6 and TCF1 (detected by GFP) and transferred them into infection-matched hosts (Fig. 4.8b). At one week after transfer (at dpi 12 from the initial infection), we analyzed the progeny of cells originating from the three branches in the spleen (Fig. 4.8c). We observed that cells recovered after transfer of exhausted cells into infection-matched recipients retained their exhausted phenotype. Cells recovered after transfer of the memory-like branch exhibited phenotypes of both exhausted and memory-like cells, confirming previous results of differen-

tiation from memory-like into exhausted cells [Chen et al., 2019] but contradicting our finding of a memory-like endpoint. Recovered cells after transfer of pre-committed cells exhibited both a memory-like or an exhausted phenotype, confirming their differentiation potential into both memory-like and exhausted cells. However, there was a strong bias towards differentiation along the exhaustion branch, which might be explained by much more extensive proliferation of these cells compared to memory-like cells.

4.2.7 Differentiation transitions are driven by antigenic TCR stimulation

Since our RNA velocity and trajectory analysis had revealed the memory-like cells as an endpoint of differentiation, we speculated that their unexpected differentiation into terminally exhausted cell states after adoptive transfer is triggered by external cues that drastically changed their state. We investigated the possibility of TCR stimulation as such a cue by transferring cells belonging to the three branches (isolated from mice at 5 dpi following adoptive transfer of P14 cells and induction of chronic infection with Clone-13) into hosts infected with a LCMV Clone-13 strain that expresses a variant of the gp33 peptide that is not recognized by P14 cells (Fig. 4.8a).

We recovered cells from spleen and lymph nodes and analyzed their phenotype based on CXCR6 and Ly108 expression (Fig. 4.8d). The recovered cells after transfer of the exhausted branch again retained their exhausted phenotype. Surprisingly, we recovered much fewer cells with an exhausted phenotype after transfer of memory-like cells and a major fraction retained their memory-like phenotype, indicating that further differentation was largely halted in the absence of antigen. The transfer of pre-committed cells resulted in recovery of cells with a largely undifferentiated phenotype of neither terminally exhausted nor memory-like, largely retaining their pre-committed state. These results clearly pointed towards TCR stimulation being a major driver of differentiation during chronic infection for both the pre-committed state and for further differentiation of the memory-like state into fully exhausted cells.

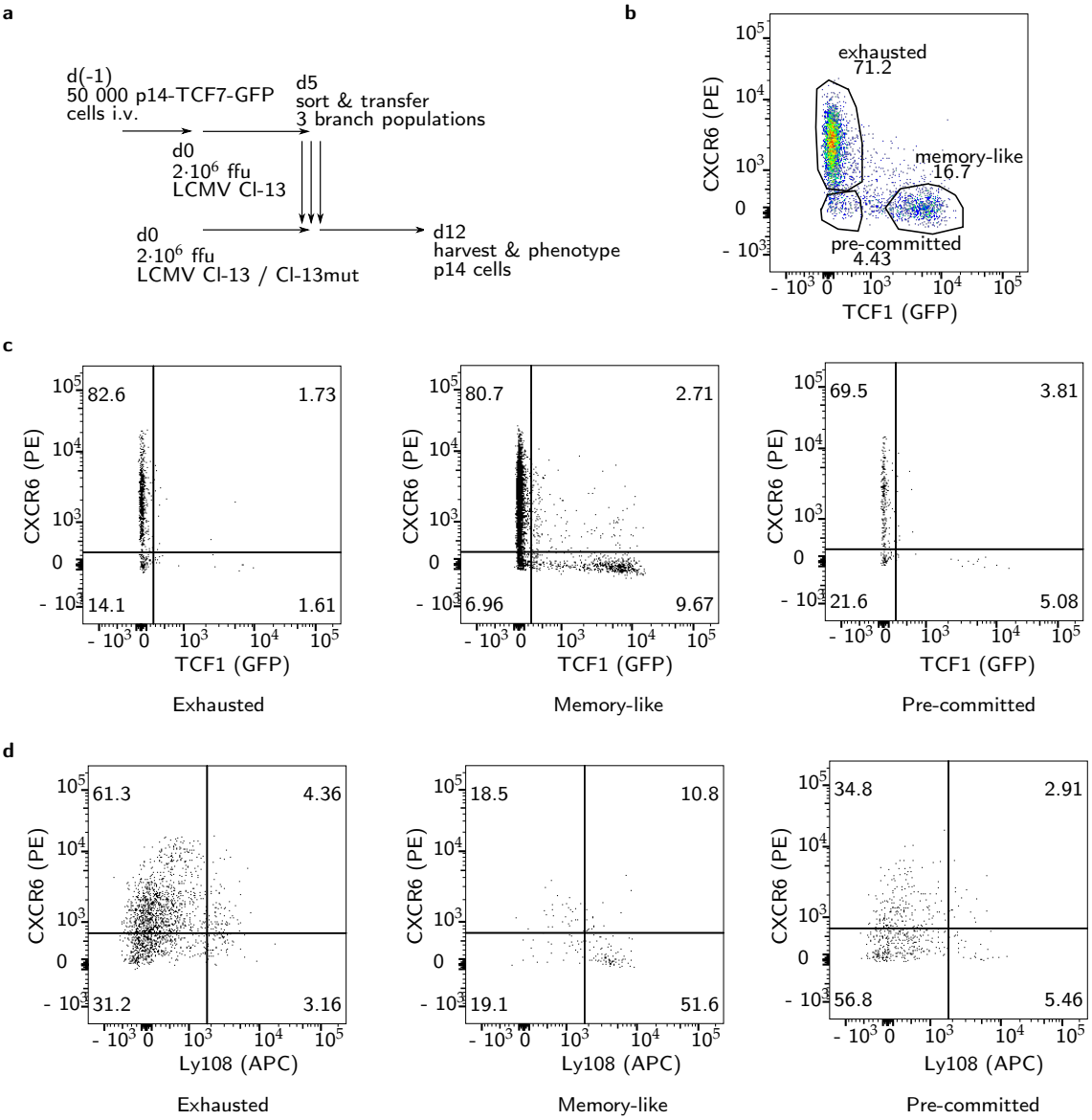


Figure 4.8: (a) The three P14 branch populations were isolated at 5 dpi from high-dose Clone-13 infected mice (that had been transferred with naive P14 cells prior to infection) and transferred into infection matched hosts and their phenotype was assessed 7 days post transfer.
(b) flow cytometry gates used to sort exhausted, memory-like and pre-committed branch
(c) phenotype of the recovered P14 cells at 12 dpi from spleens after high-dose Clone-13 infection and transfer of either exhausted, memory-like or pre-committed cell populations isolated at 5 dpi. Cells are gated on P14 cells.
(d) phenotype of the recovered cells at 12 dpi from spleens after transfer into hosts infected with Clone-13 P14 escape mutant. Naïve P14 cells were first transferred into naive C57BL6 mice, followed by Clone-13 infection. At 5 dpi, exhausted, memory-like and pre-committed populations were sorted and adoptively transferred into infection matched hosts with Clone-13 escape mutant. Recovered P14 cells are shown.

4.3 Discussion

We analyzed differentiation trajectories of virus-specific CD8$^+$ T cells during chronic LCMV infection using scRNAseq time-series data from four different time points covering activation, peak, contraction and late phase of the response. The time-resolved traversal of the transcriptional landscape revealed a continuous and bifurcating process, with early activated cells at the

beginning and both terminally exhausted as well as memory-like cells at the end of this process. We observed cells with an exhausted phenotype and an effector-like cell population as transient states during early and late stages of this process, respectively. We further observed for all time points a population of cells with transcriptional profiles of proliferation.

Applying RNA velocity analysis to our single-cell transcriptional data allowed us to estimate transitions between the cell states during the progressing immune response. We computed most likely end and start regions and identified two major differentiation paths leading to the exhausted population and the memory-like population respectively. At the early time-points until about day 5, the two average trajectories were nearly indistinguishable, but then diverged increasingly towards their respective endpoints.

We identified a branching region in the early stages of infection before day 5 post infection. Before this branching point, pre-committed cells would still have the potential to differentiate into both the exhausted and the memory-like phenotype, whereas cells that had passed the branching point would be destined to differentiate into the endpoint they had committed to.

Although it is conceivable that factors that were not revealed by transcriptional analysis might be involved to already pre-determine cell fate during early activation and regulate differentiation, this is not evident on a transcriptional and protein expression level, where branching of the two main fates manifested itself around day 5 post infection.

We derived a small set of gene markers that separate cells into the pre-committed (TCF1^{neg} CXCR6^{neg}), exhausted (TCF1^{neg} CXCR6^{hi}) and memory-like branch (TCF1^{hi} CXCR6^{neg}) by a classification model. Adoptive transfer of cells sorted according to these markers, and thereby likely belonging to the three branches, at day five post chronic LCMV infection into infection-matched new hosts, confirmed the plasticity of pre-committed cells to acquire both exhausted as well as memory-like phenotype. Conversely, progeny from committed exhausted cells largely retained their phenotypes. Although, our velocity and trajectory analysis suggested the memory-like cells to represent a developmental end-point, we still recovered both memory-like and exhausted cells after transfer of cells from the memory-branch. These transitions might be rare (and fast) events in the integral setting of an infected mouse, and thus not be represented in the scRNAseq data. Additionally, i.v. adoptive transfer of memory-like cells into circulation might expose them to different antigenic burden compared to their natural niches, thereby accelerating a differentation process.

Previously published work used scRNAseq to study CD8 T cell differentiation during chronic infection, isolating virus specific T cells at different time-points [Chen et al., 2019, Zander et al., 2019, Yao et al., 2019, Raju et al., 2020]. They made use of dimensionality reduction and computational tools for trajectory inference. Although, these studies used fewer time-points and the lineage inference included one time-point only. All these studies consistently described the memory-like and the terminally exhausted cell state, which we also identified. Some studies additionally described an effector-like CX3CR1^{+} population arising late during the infection, which we also found in our late samples from 14 & 21 dpi.

Two studies [Chen et al., 2019, Zander et al., 2019] investigated plasticity and differentiation of the memory-like cells computationally by lineage inference on a single time-point and through adoptive transfer experiments, concluding that memory-like cells partially maintain their phenotype and can give rise terminally exhausted and effector-like cells. Our lineage inference across multiple time-points suggests that the exhausted and the memory-like lineages are separate. We could not exclude, that we missed certain cellular states in our data set, since we only studied CD8 cells isolated from the spleen of infected animals. We were unable study developmental stages that are spatially restricted to lymph nodes or specific anatomical regions within secondary lymphoid organs, or specific differentiation processes that are restricted to non-lymphoid organs [Sandu et al., 2020b]. Additionally, if cell state transitions are rare or fast, it is unlikely that we would capture them in our snapshot analysis.

Our adoptive transfer experiments of memory-like cells revealed extensive transitions to terminally exhausted states, which our lineage inference did not detect. However, our transfer experiments into Clone-13 P14 escape mutant infected hosts suggested that theses transitions were strictly dependent on antigenic TCR stimulation. This could imply that we did not observe these transitions in our scRNAseq data because memory-like cells receive little TCR stimulation from their microenvironment in a natural setting. Isolation of memory-like cells by removing them from their niche and transferring them via intravenous injections could expose them to excessive antigen and trigger differentiation towards terminally-exhausted cells.

Chen et al. [Chen et al., 2019] studied early bifurcation events towards either an effector state or TCF$^+$ precursor state. However, the population they termed "effector" cells already expressed many co-inhibitory receptors like our exhausted group from 7 dpi. They found this early effector cells to be very short-lived and disappear between 8 and 12 dpi. Although our data suggested a similar bifurcation into effector and memory-like lineage, our exhausted trajectory placed the early exhausted cells as an intermediate state on the differentiation towards terminally exhausted states. Since Chen et al. used KLRG1 to identify their effector population, the disappearance of these cells could be explained by down-regulation of *Klrg1* during differentiation towards terminal exhaustion.

The velocity based endpoint analysis did not reveal either the early exhausted or the effector-like states to compose stable end-points, but that all those states differentiated into terminally exhausted cells. The velocity transitions did show some flow out of these populations though, which could indicate migration out of the tissue or apoptosis of these cells, although we did not find apoptotic signatures in our data. Considering, that apoptotic cells are cleared very fast by the phagocytes, clearance of apoptotic cells might be too fast to capture.

Our classification analysis of the branch point revealed a set of genes that discriminates between the three branches. Even though, our validation experiments confirmed that CXCR6 and TCF1 expression patterns capture the differentiation potential of CD8 T cells early during the infection, other identified genes might still be relevant in shaping this bifurcation. Both genes *Ifngr1* and *Selplg* are transiently upregulated around the birfucation point, which could imply some influence on cell fate. Considering, that withdrawing T cells from antigen stimulation practically halted differentiation of pre-committed cells and maintained their state, points towards a significant role of TCR stimulation and signaling in this bifurcation and decision process.

This work provides additional insights into the differentiation process of CD8 T cells using a combined approach of scRNAseq analysis, computational trajectory inference and adoptive transfer experiments. Our study revealed an early bifurcation event, that shaped the differentiation fate during the course of a chronic infection and additionally highlights TCR stimulation as a significant driver of this differentiation.

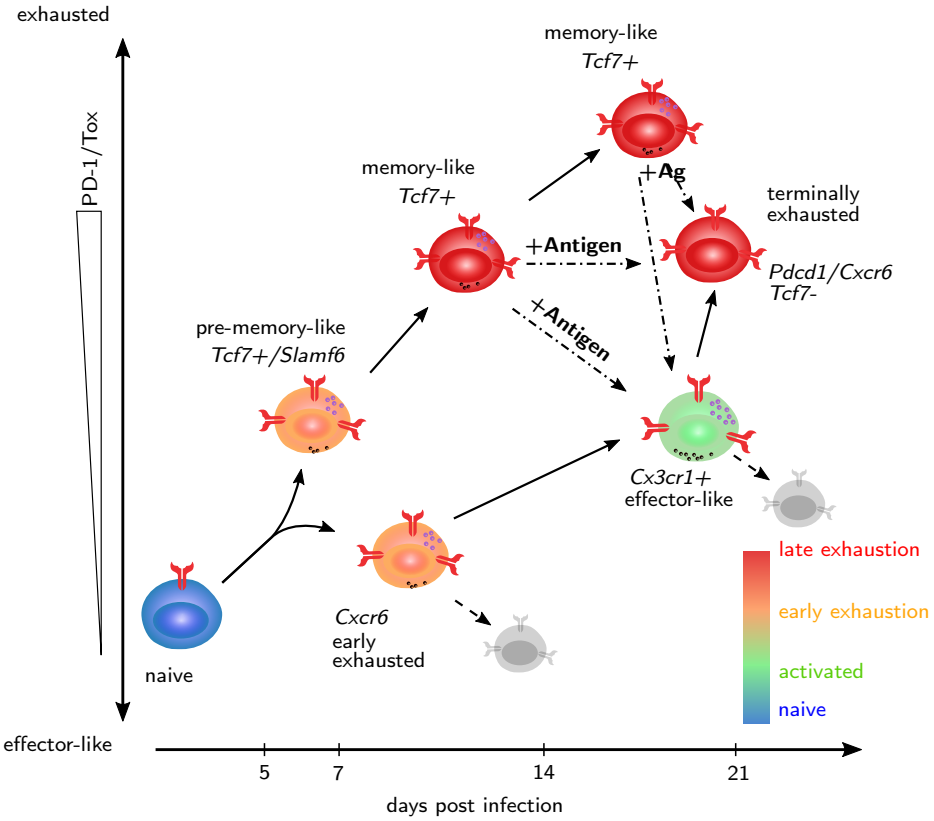


Figure 4.9: Differentiation model proposed by our lineage inference and adoptive transfer experiments. Early commitment to either effector/exhaustion or memory-like lineage results in two developmental lineages. Transitions from memory-like to exhausted lineage are driven by antigen stimulation.

4.4 Material & Methods

Infections and cell isolation

Mice Wild-type male C57BL/6J mice were purchased from Janvier Elevage. Nr4a1-GFP mice expressing GFP under the control of the NUR77 promoter [Moran et al., 2011], P14 transgenic (CD45.1) mice expressing a TCR specific for LCMV peptide gp33–41 [Pircher et al., 1989] and TCF1-GFP mice expressing GFP under the control of the Tcf7 promoter [Utzschneider et al., 2016a] were housed and bred under specific pathogen–free conditions at the ETH Phenomics Center Hönggerberg. All mice used in experiments were between 6–16 weeks of age. P14-Nr4a1-GFP mice were generated by crossing Nr4a1-GFP mice to P14 mice. P14-TCF1-GFP mice were generated by crossing TCF1-GFP mice to P14 mice. All animal experiments were conducted according to the Swiss federal regulations and were approved by the Cantonal Veterinary Office of Zürich (Animal experimentation permissions 147/2014, 115/2017).

Virus LCMV clone 13 [Ahmed et al., 1984] was propagated on baby hamster kidney 21 cells. LCMV clone 13 P14 escape mutant [Utzschneider et al., 2016b] was propagated on MC57G cells. Viral titers of virus stocks were determined as described previously [Battegay et al., 1991].

Infection 10^4 transgenic cells (P14, P14-TCF1-GFP or P14-Nr4a1-GFP) were adoptively transferred 1 day prior LCMV clone 13 intravenous (IV) infection with 2 x 10^6 ffu/mouse. For isolation at 1, 2, 3, 4 days post-infection, 10^5 P14-Nr4a1-GFP cells were transferred.

Cell isolation from tissues After 1, 2, 3, 4, 7, 14 and 21 days of chronic infection, mice were sacrificed with carbon dioxide and organs (spleen, lymph nodes) were isolated. Spleens and lymph nodes were mashed through 70 µm filters with a syringe (1 mL) plunger. Cell suspensions were filtered (70 µm) and treated with ammonium-chloride-potassium buffer (150 mM NH4Cl, 10 nM KHCO3, 0.1 mM EDTA in water) to lyse erythrocytes for 5 min at room temperature.

Cell sorting Spleen samples were depleted of CD4 and B cells by incubating splenocyte suspensions in enrichment buffer (PBS, 1%FCS, 2 mM EDTA) with biotinylated α-CD4 and α-B220 antibodies at room temperature for 20 min, followed by incubation with streptavidin-conjugated beads (Mojo, Biolegend) (4%) for 5 min at room temperature. Cells were then placed on a magnetic separator (StemCell) for 10 min at room temperature, followed by collection of supernatant. For scRNAseq samples, cell suspensions of spleens isolated from five mice were pooled in samples from day 7, 14 and 21 post infection and from three mice for samples from early timepoints at day 1, 2, 3 and 4 order to ensure the samples were representative of a population. All samples from day 1 to 4 were pooled for sorting and sequencing due to the low frequency of P14 cells in these samples. Enriched samples from the spleen or cell suspensions from lymph nodes were stained with α-CD8-PerCP and α-CD45.1-APC to sort P14 cells (ARIA cell sorter, BD Biosciences).

Single-cell RNA sequencing & analysis

Sorted P14 cells from different time-points were washed and resuspended in 0.04% BSA. The single cell sequencing was performed at the Functional Genomics Center Zurich. The cell lysis and RNA capture was performed according to the 10XGenomics protocol (Single Cell 3' v2 chemistry). The cDNA libraries were generated according to the manufacturer's protocol (Illumina) and further sequenced (paired-end) with NovaSeq technology (Illumina). The transcripts were mapped with 10Xgenomics CellRanger pipeline (version 2.0.2).

Pre-processing & Normalization Read counts were realigned and sorted for spliced and un-spliced counts using the analysis pipeline from velocyto [Manno et al., 2018]. Contaminating other cell types were removed from the dataset based on outliers in diffusion components. Reads were filtered and normalized according to the Zheng recipe [Zheng et al., 2017] of the scanpy analysis pipeline [Wolf et al., 2018] retaining 5000 highly variable genes. Louvain clustering and UMAP projection were computed using standard parameters, using the first 50 principle components.

RNA velocity RNA velocity uses the relative abundance in reads of un-spliced to spliced mRNA to infer the future state of a particular cell, with a high ratio of un-spliced / spliced mRNA being indicative of recent gene activation, a balanced ratio of un-spliced / spliced mRNA being indicative of gene expression equilibrium, and a low rate of un-spliced / spliced mRNA being indicative for terminating gene expression. Integrating the expression levels of the corresponding genes in neighboring cells allows computing likely transitions between different cellular states in our data set, revealing a vector field demarcating likely transitions.

Scvelo [Bergen et al., 2020] was used to estimate RNA velocity and infer transition probabilities between cells. The transition probabilities were used to construct a Markov process. Inference of RNA velocity relies an a set of assumptions that can be adjusted through several parameters before analysis. The type of pre-processing used, the number of principle components used as well as the neighborhood size for imputation may influence the resulting transition matrix. We conducted an extensive assessment of a many parameter combinations to validate the stability of our RNA velocity analysis. Comparing the equilibrium distributions across 574 parameter sets revealed that the global structure of the transition matrix was quite stable. The parameter set we have chosen for further analysis produced results found in an overwhelming majority of tested sets (Supp. Fig. S5).

We estimated gene moments using neighborhood connectivities using 50 principle components, 30 neighbors with the UMAP method. Velocity was inferred using "stochastic" mode.

Cytopath Cytopath is an extension of scvelo and relies on its core routines (Chapter 3). We applied scvelo's terminal states routine to compute equilibrium distributions of the forward and backward Markov process, excluding self transitions. Regions with terminal state probability higher than 0.3 were identified and the louvain clusters corresponding to these regions used as start and endpoints. Markov simulations were initialized at the start points and simulated for a maximum of 2000 steps or until they reached the endpoint. We simulated 2000 trajectories from random cell states in the starting region. All simulated paths were aligned to average trajectories from startpoint to each endpoint using dynamic time warping. Neighboring cells (2000 nearest neighbors) were aligned to the trajectories using an alignment score, which was computed based on distance and cosine distance between the cell's velocity and the direction of the trajectory. Cells that aligned to only one trajectory were assigned to the exhausted or memory-like branch, respectively. Cells at the beginning of the infection that aligned to both trajectories were assigned to the pre-commited branch.

All cells were assigned exhausted, memory-like or pre-commited fate according to their alignment score to the trajectories. Gene expression profiles of 650 genes coding for antibody stainable proteins were then used to predict these labels using L1-penalized Logistic Regression. We used cross validation to identify the optimal L1-penalty that would give a reasonably small number of genes but still good prediction accuracy at C=0.1. The resulting prediction using 12 proteins still classified most cells correctly (accuracy: 0.85). These proteins were then stained on P14 cells from chronic infection at d5 to sort the branches and adoptively transfer them into infection matched hosts.

Adoptive transfer experiments

After 5 days of chronic infection, CD8 T cell enriched samples from the spleen or cell suspensions from lymph nodes were stained with α-CD8-PerCP, α-CD45.1-APC/FITC and α-Cxcr6-PE and α-Ly108-APC to sort P14 cells into the exhausted, memory-like and pre-commited populations. (ARIA cell sorter, BD Biosciences).

Sorted cells from exhausted (10^6 cells), memory-like ($2x10^5$ cells) and pre-committed ($5x10^4$ cells) populations were transferred via intravenous (IV) injection into infection matched hosts infected with either Clone-13 or Clone-13 P14esc mutant. Cells were recovered from from spleens of these mice 12 days post infection prior to phenotypic characterization.

Flow cytometry

Surface staining was performed at room temperature for 30 minutes in FACS buffer (2% FCS, 1% EDTA in PBS). LIVE/DEAD™ Fixable Near-IR Dead (Thermo Fisher) was used to discriminate alive from dead cells. Fluorophore-conjugated antibodies used for flow cytometry were purchased from BioLegend (Lucerna Chem AG, Luzern, Switzerland) (α-CD45.1 BV711 A20; α-CD45.1 APC A20; α-CD8 PerCP 53-6.7; α-CD8 BV395 53-6.7; α-PD-1 PE-BV605 29F.1A12; α-Cxcr6 PE SA051D1; α-Ly108 APC 330-AJ; α-CD8 BV395 53-6.7). Data was acquired LSR II Fortessa using Diva software (BD Biosciences, Allschwil, Switzerland) and analyzed in FlowJo (BD Biosciences, Allschwil, Switzerland). Gating and plotting was done using FlowJo (BD Bioscience, Allschwil, Switzerland).

Supplementary Figures

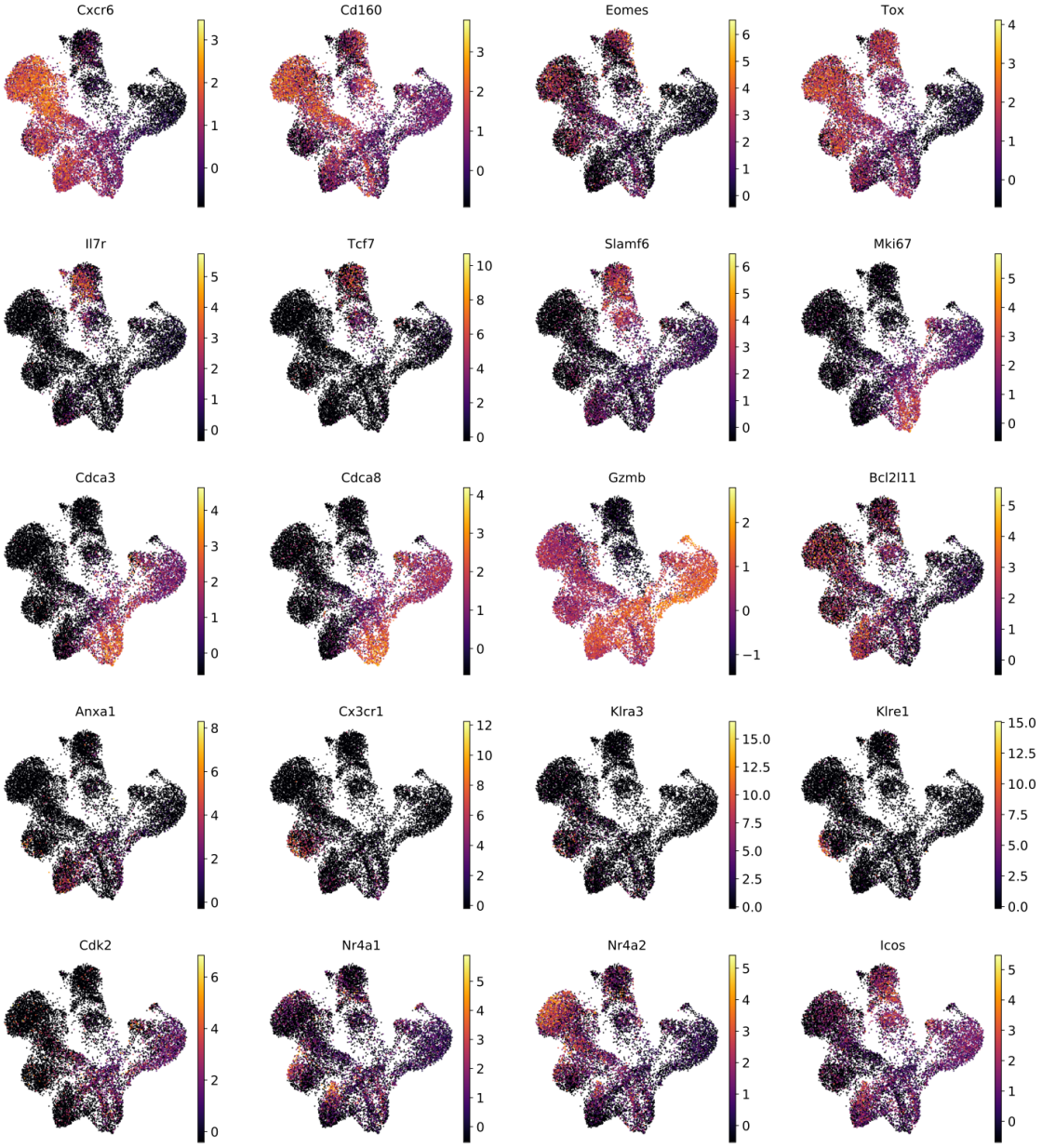


Figure S1: distribution of relevant genes

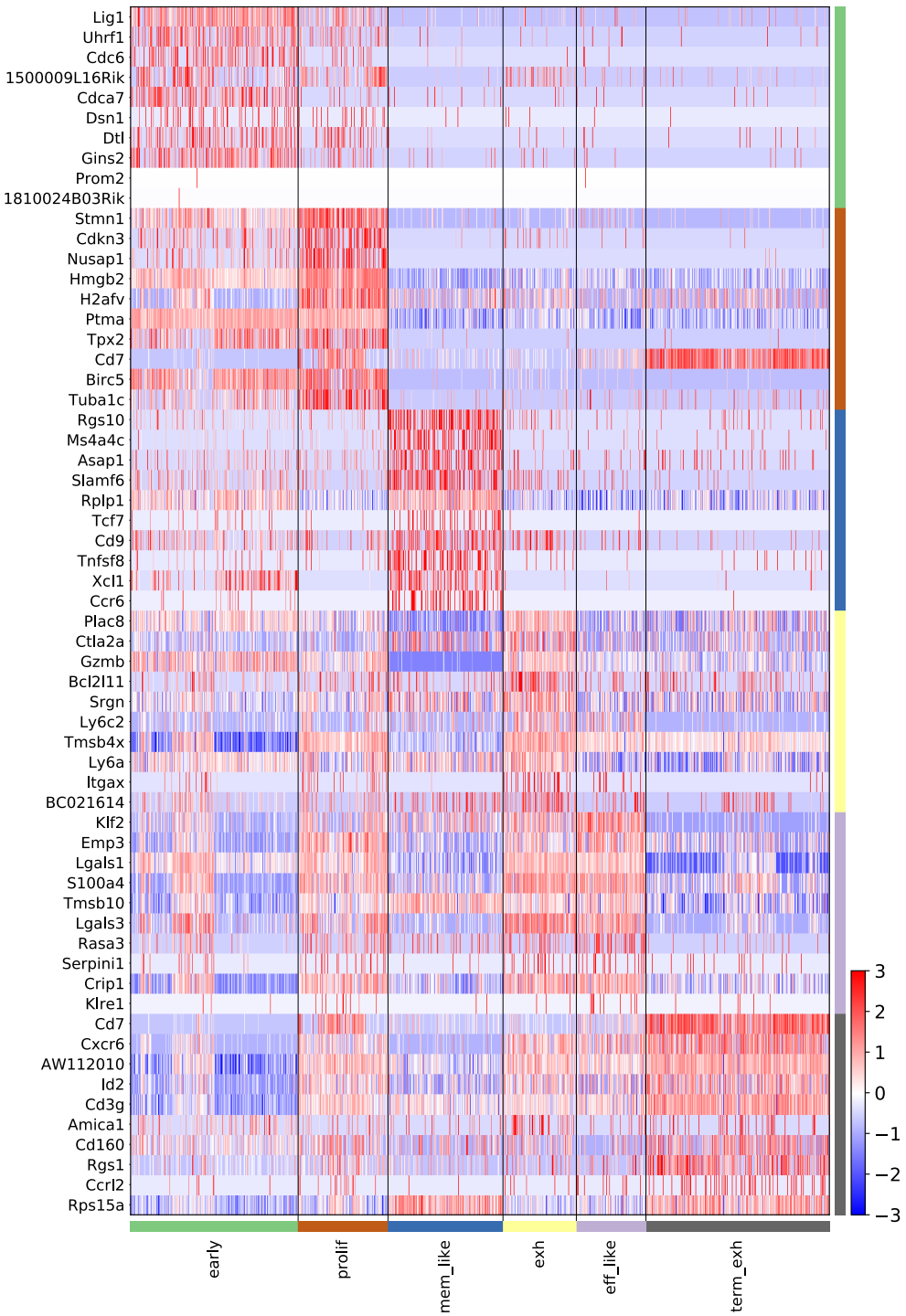


Figure S2: Top 10 differentially expressed genes between relevant populations

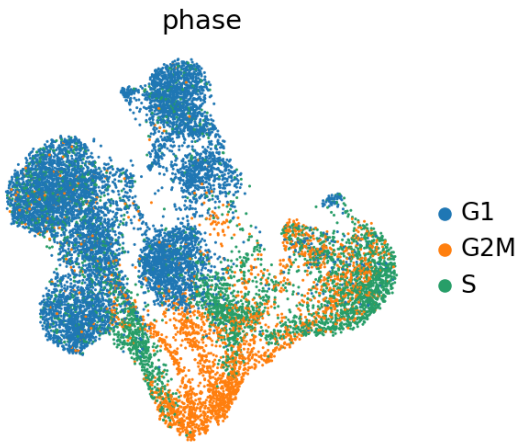


Figure S3: cell cycle scoring confirmed strong G2M activity in the proliferating group

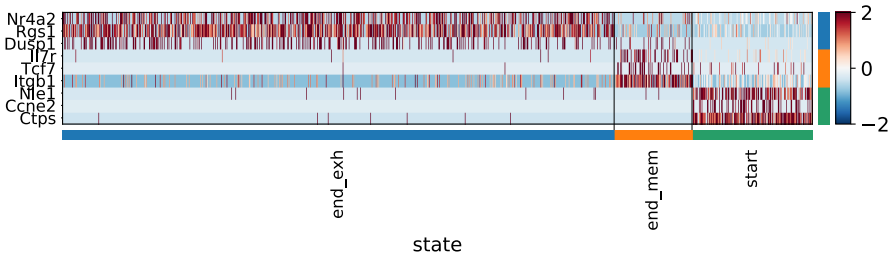


Figure S4: differentially expressed genes in root and end states

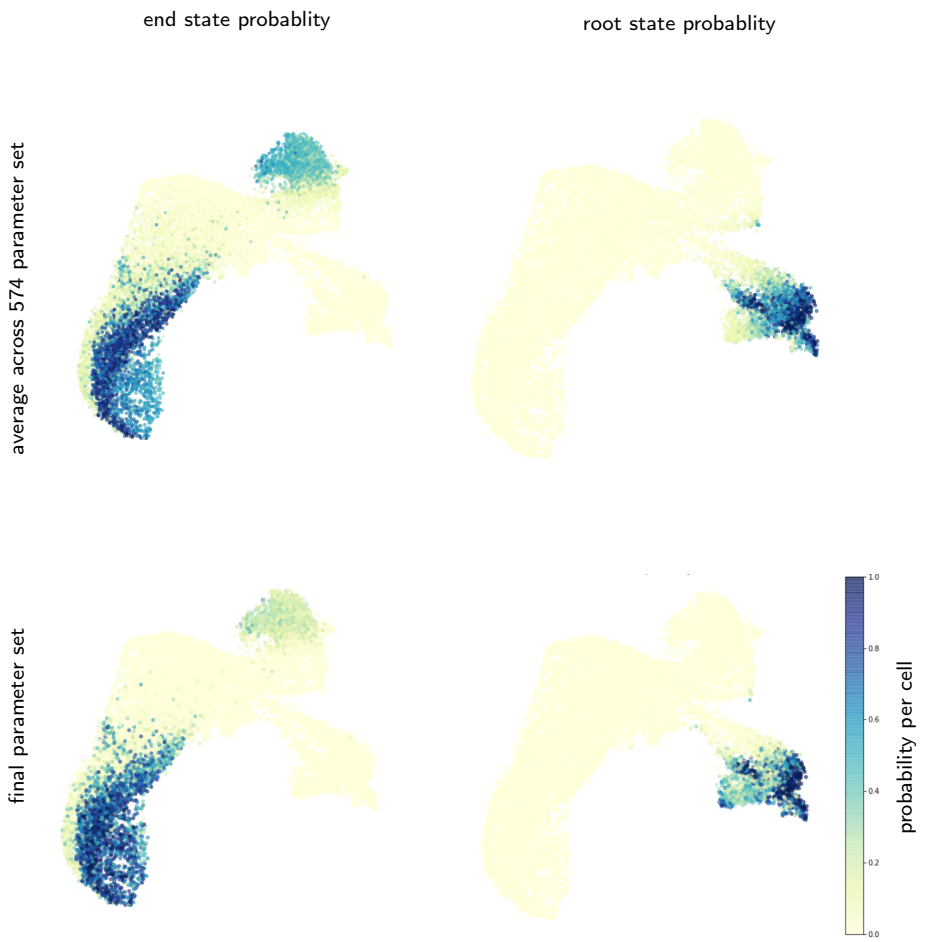


Figure S5: mean end and root point probability across many parameters

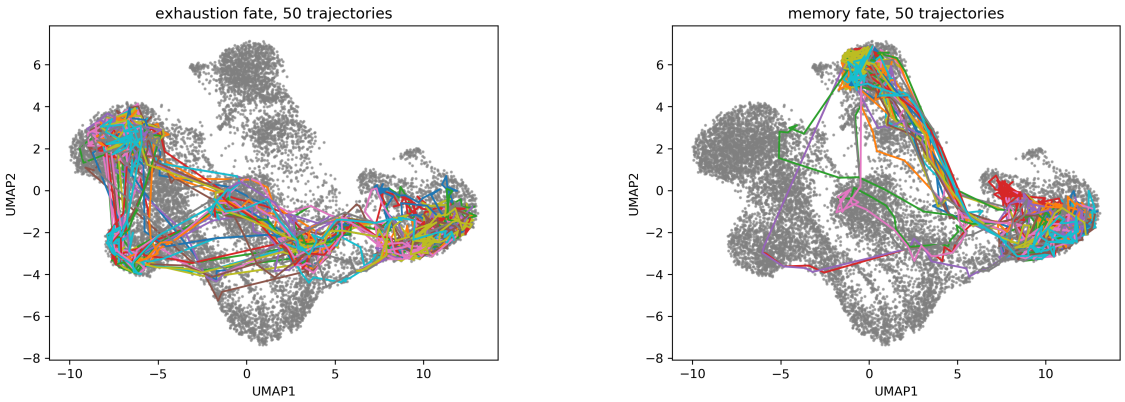


Figure S6: Simulated trajectories from starting region to exhaustion endpoint

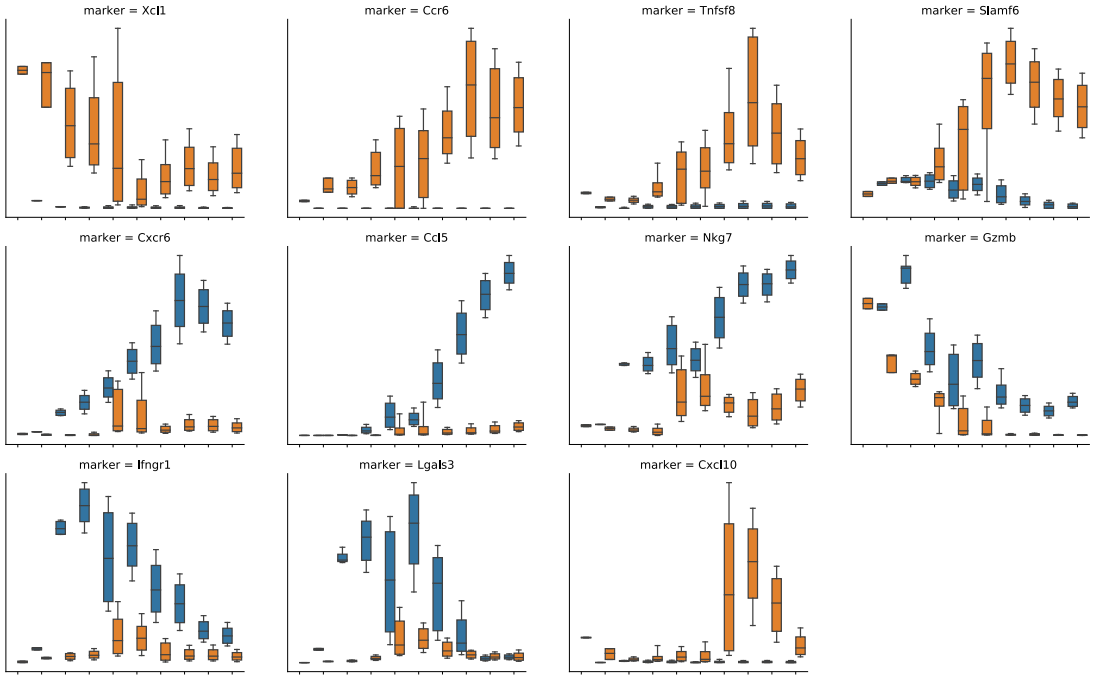


Figure S7: selection of differentially expressed genes between memory-like and exhausted trajectories in order of their significance score.

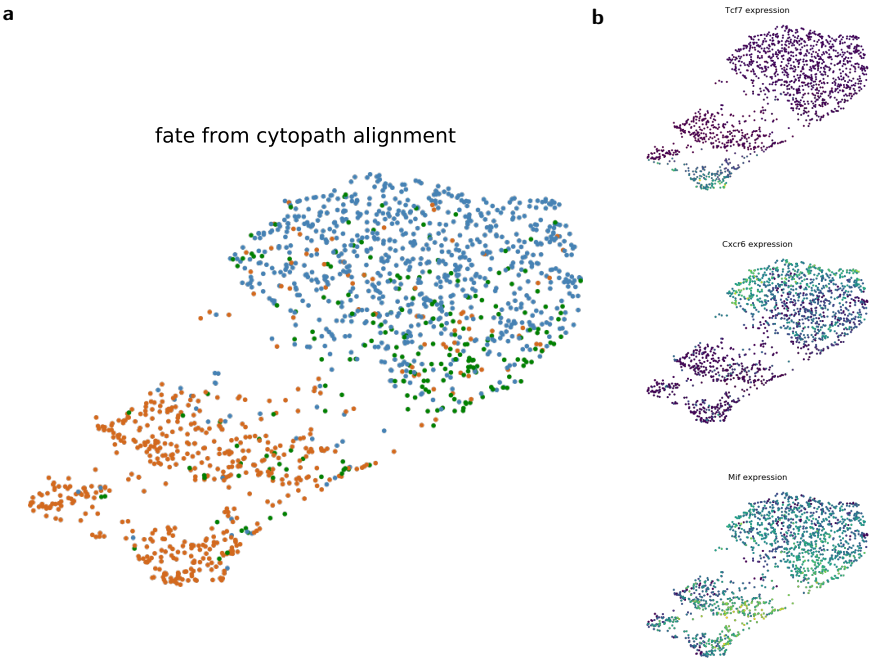


Figure S8: Using only identified marker genes allowed good precition of the assigned branch fate.
(a) Three branches in umap projection.
(b) Branch specific gene expression.

Chapter 5

Opinion Paper: scRNAseq based reconstruction of CD8 T cell exhaustion

Dario Cerletti[1,2], Annette Oxenius[2,*] and Manfred Claassen[1,*,&]

[1]Institute of Molecular Systems Biology, ETH Zurich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland

[2]Institute of Microbiology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland

[&]Current address: Internal Medicine I, University Hospital Tübingen, Faculty of Medicine, University of Tübingen, Otfried-Müller-Straße 10, 72076 Tübingen, Germany

[*]shared contribution

Abstract

CD8 T cells differentiate into multiple functional subsets in the context of chronic infection in an asynchronous and bifurcating fashion. Populations with effector-like, memory-like and exhausted features have been identified and studied using bulk sequencing and flow cytometry. However, the lineage relations between theses populations as well as intermediate states remain elusive. Single-cell RNA sequencing (scRNAseq) allows to identify populations and infer potential lineages between them using computational tools. Here we discuss findings by recently published scRNAseq datasets and our own longitudinal scRNAseq data set that aim at elucidating relevant populations and their differentiation trajectories in chronic infection. We compare suggested differentiation models by trajectory inference algorithms in the context of the performed validation experiments.

5.1 Studying T cell exhaustion in the murine chronic Lymphocytic choriomeningitis model

During chronic infections, such as Lymphocytic choriomeningitis virus (LCMV), CD8 T cells acquire an exhausted phenotype through continued antigen exposure and T cell receptor (TCR) stimulation [Utzschneider et al., 2016b]. Terminally exhausted T cells express high levels of co-inhibitory receptors such as PD-1, LAG-3 and CD39 on their surface and have little proliferative capacity [Wherry and Kurachi, 2015]. The activity of co-inhibitory receptors dampens TCR signaling and results in less cytotoxicity and cytokine production of these cells *in vivo* [Sandu et al., 2020a]. This inhibition by co-inhibitory receptors and corresponding ligands on infected cells serves to avoid detrimental immuno-pathology and protects the host but also hinders viral clearance [Frebel and Oxenius, 2013b, Alfei et al., 2019].

Recently, it has been shown that T cell differentiation during chronic LCMV infection does not exclusively result in differentiation of a uniform population of exhausted CD8 T cells, but that the population of LCMV-specific CD8 T cells comprises different subsets of cells, including a population with more effector cell traits [Hudson et al., 2019], and a population exhibiting a memory-like state. Memory-like cells expressing the transcription factor TCF1 show stem-like properties and respond to check-point blockade by expansion [Siddiqui et al., 2019]. Effector-like cells expressing the surface receptor CX3CR1 show features of effector cells from acute infection

and higher proliferation than terminally exhausted cells, but do not share the stem-like property of TCF1$^+$ cells [Hudson et al., 2019].

Acquisition of the exhausted phenotype is preceded by a spectrum of cellular differentiation states over the course of the infection, characterized by changing transcriptional states and ultimately functional capacity [Wherry and Kurachi, 2015]. Early during LCMV infection, naïve T cells are activated and later during the infection branch off to start to differentiate into these different states. This process is highly asynchronous and dependent on the antigenic exposure of the cells and possibly their local environment. Therefore, identification of these subsets is non-trivial and very difficult using techniques that only capture static phenotypes as in flow cytometry, as these cells continue to differentiate into the different phenotypic subsets as the infection progresses. The resulting ensemble of T cells will be heterogeneous at any time-point of the infection due to both asynchrony of the process as well as bifurcation of lineages into subsets.

Adoptive transfer experiments of cells sorted using flow cytometry imply that these subsets have varying capability of differentiating into one another [Utzschneider et al., 2016b, Wu et al., 2016, Hudson et al., 2019]. However, it is not fully understood which population can give rise to another and at which stage of the infection. Single-cell RNA sequencing (scRNAseq) measures the transcriptional state of many cells at high depth and allows to identify multiple T cell populations as well as intermediate states between them. Sampling T cells at specific time-points of infection combined with computational tools to infer cell relations can aid in uncovering differentiation relations between subsets.

5.2 Single-cell RNA seq enables reconstruction of T cell subset development during chronic infection

Multiple studies have recently made use of the high detail and resolution of single-cell RNA sequencing to shed light on the origin and development of T cell subsets in chronic LCMV infection. In the following we summarize their conclusions based on scRNAseq experiments and validation experiments.

We considered the following studies:

- Chen et al. 2019: TCF-1-Centered Transcriptional Network Drives an Effector versus Exhausted CD8 T Cell-Fate Decision
- Zander et al. 2019: CD4$^+$ T Cell Help Is Required for the Formation of a Cytolytic CD8$^+$ T Cell Subset that Protects against Chronic Infection and Cancer
- Yao et al. 2019: Single-cell RNA-seq reveals TOX as a key regulator of CD8$^+$ T cell persistence in chronic infection
- Raju & Xia 2020: Latent Plasticity of Effector-like Exhausted CD8 T cells contributes to memory responses
- Cerletti et al. 2020 (this work): Fate trajectories of CD8$^+$ T cells in chronic LCMV infection

Comparison and scope of studies

Chen2019 Chen et al. generated samples using virus specific CD8$^+$ T cells from three different infection types using LCMV Armstrong, Clone-13 and Clone-13 with CD4 T cell depletion from 8 days after infection and also included naïve CD8$^+$ T cells as reference. Sequencing data was aligned, filtered and normalized using the CellRanger (10x Genomics) and scran pipeline. After graph-based clustering on 50 principle components and visualizing in a tSNE projection, they used differential gene expression (Hurd model) to identify cluster specific genes. Cluster characterization was done using a mixture of previously known marker genes [Im et al., 2016,

Work	Time-points	Infection setting	Trajectory inference	Conclusion
[Chen et al., 2019]	d8	Arm/C13/C13ΔCD4	tSNE Monocle-2 TSCAN	Early bifurcation between Tex-precursors and Teff-like cells
[Zander et al., 2019]	d8/d30	C13 infection	tSNE, Seurat, Monocle-2	Tcf1^{high} population gives rise to effector-like and terminally exhausted
[Yao et al., 2019]	d4.5/d7	Arm/C13, TOX$^{-/-}$	tSNE, PC-analysis hierarchical clustering	TOX is higher in Tcf1^{high} progenitor cells in chronic compared to memory-precursors in acute infection
[Raju et al., 2020]	d21	C13, BCL-6$^{-/-}$	umap	CX3CR1^{+} cells arise independent from TCF1^{+} cells
Cerletti 2020	d1-4/7/14/21	C13/C13^{mut}	umap, RNA velocity, cytopath	CX3CR1^{+} cells compose the same lineage as terminally exhausted cells, differentiation of TCF1^{+} cells into effector-like or terminally exhausted is antigen driven

Table 5.1: Overview of scRNAseq data generation and computational analysis of discussed work. Time-points are days post infection (d). Infection settings used Armstrong (Arm), Clone-13 (C13), Clone-13 with CD4 depletion (C13ΔCD4) or Clone-13 P14 escape mutant (C13^{mut}).

Utzschneider et al., 2016c, Wherry et al., 2007] as well as the differentially expressed genes (Supp. Fig S1).

The different infection types revealed varying degrees of exhaustion. The authors assumed that there is a gradual path from less to more exhausted cells. By applying the lineage inference method Monocle-2 they built a tree structure into the data from all the samples (Supp. Fig S1). The cells from the naïve sample were put at the base of the tree since they were the obvious starting point of the differentiation process. The two other branches, inferred by monocle, considered the two different phenotypes effector-like (Teff-like) and exhaustion precursor cells (Tex-precursors), based on the expression of previously known markers. Using a combination of previously known relationships between cells expressing PD-1, KLRG1, CD39 and TCF1, the authors propose a terminal effector-like (Teff-like), an exhausted precursor (Tex-precursors) and a terminally exhausted (Tex) population, which they went on to further investigate. Based on expression of apoptotic and survival markers, they proposed these populations to have different survival capacity.

The authors tried to confirm the survival of the previously defined three populations experimentally in an infection setting. They therefore isolated Teff (KLRG1^{+}CD39^{+}), pre-Tex (PD-1^{+}Ly108^{+}) and Tex (PD-1^{+}Ly108^{-}) cells using flow cytometry and adoptively transferred them into infection-matched hosts and indeed found better survival for the Ly108^{+} PD-1^{+} phenotypes, compared to the Ly108^{-} and CD39^{+} populations.

Zander2019 To identify developmental pathways in chronic infection, the authors sampled GP33-specific CD8 T cells at day 8 and day 30 after chronic LCMV infection with Clone-13. Using a tSNE projection of both time points and subsequent clustering (Fig. S2), they identified cell populations that were shared, or exclusive to the time points. They identified a memory-like *Tcf7* expressing population that was found in both d8 and d30 time points. Additionally they discovered a more exhausted Pdcd1^{+} population and a more effector-like Cx3cr1^{+} population on d30.

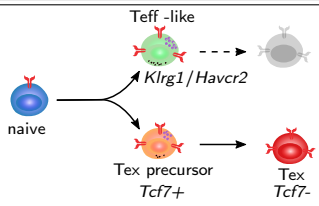
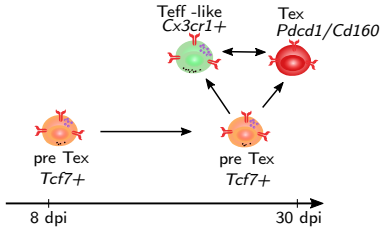
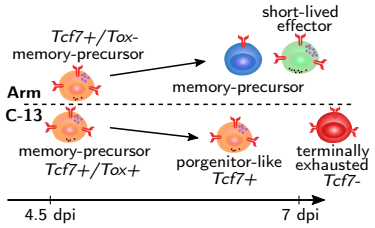
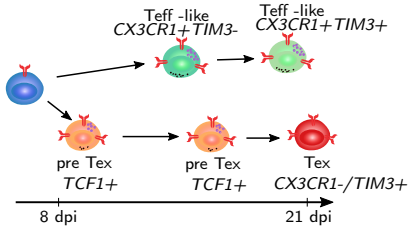
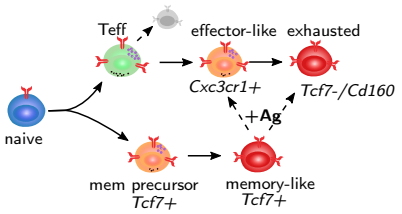
Work	Validation	Markers	Inferred Lineages
[Chen et al., 2019]	adoptive transfer	Teff KLRG1^{+}CD39^{+} pre-Tex PD-1^{+}Ly108^{+} Tex PD-1^{+}Ly108^{-}	
[Zander et al., 2019]	adoptive transfer	Teff CX3CR1^{+} pre-Tex Ly108^{+} Tex CX3CR1^{-}Ly108^{-}	
[Yao et al., 2019]	Tox$^{-/-}$ K.O.	preTex TCF1^{+} termTex TIM3^{+}	
[Raju et al., 2020]	adoptive transfer, Bcl-6$^{-/-}$	Teff CX3CR1^{+}TIM3^{-} Teff CX3CR1^{+}TIM3^{+} pre-Tex TCF1^{+}TIM3^{-} Tex TCF1^{-}TIM3^{+}	
Cerletti 2020	adoptive transfer, C13^{mut}	pre CXCR6^{-}TCF1/Ly108^{-} Tmem CXCR6^{-}TCF1/Ly108^{+} Tex CXCR6,$^{+}$TCF1/Ly108^{-}	

Table 5.2: Overview of the validation experiments in the discussed studies. The type of experiments as well as the most relevant markers are indicted. The identified transitions and sub-populations are illustrated.

To disentangle the lineage relations of the three populations, they applied Monocle-2 to the cells from these clusters including the connecting clusters (Supp. Fig. S2). They determined the memory-like cells giving rise to both the exhausted and the effector-like population.

Validating this direction of development by adoptively transferring memory-like, effector-like and exhausted cells into naive host with subsequent Clone-13 infection, they confirmed their proposed lineage relation. They observed highest proliferative capacity in the memory-like cells, which were also most plastic and formed both an exhausted and an effector-like population and maintained a TCF1^{+} population. Surprisingly, they also observed plasticity of the exhausted cells, giving rise to a substantial pool of effector-like cells. The authors explain this unexpected plasticity by the presence of intermediate states in the exhausted pool with remaining potential to generate effector-like cells. However, effector-like cells proliferated more than exhausted cells but did not show plasticity.

Yao2019 This study investigates the formation of memory-like TCF1$^+$ cells in chronic infection in contrast to memory cells in acute infection and performed scRNAseq on P14 CD8 T cells at day 4.5 and 7 after infection with either Clone-13 (leading to chronic infection) or Armstrong (leading to acute infection). Visualization of the data using tSNE revealed overlapping phenotypes between acute and chronic during the early time point (d4.5), but pronounced differences at the later time point (d7) (Supp. Fig. S3).

Using known markers for memory cells and memory-like cells the authors identified the respective clusters and used differential gene expression analysis between the clusters of interest to identify differences between the two infection settings. Among the top up-regulated genes in the exhaustion cluster was the transcription factor TOX.

The authors demonstrate a crucial role of TOX in the formation and maintenance of memory-like T cells in chronic infection using over-expression and knock-out experiments. Overexpression of *Tox* increased the number and fraction of memory-like cells in the spleen at day 14 and 21 post infection. Conversely the conditional deletion of *Tox* lead to a drastic reduction of these cells, leading to the suggestion that *Tox* is necessary and required to retain memory-like cells in the context of chronic viral infection.

Raju2020 Aiming at understanding the emergence and plasticity of memory-like TCF1$^+$ cells and effector-like CX3CR1$^+$ cells in chronic infection, the authors performed scRNA sequencing on gp33-specific CD8 T cells 21 days after infection with a high dose of LCMV Clone-13. Using dimensionality reduction with umap and clustering using shared nearest neighbor (SNN), the three clusters of exhausted (TIM3$^+$CX3CR1$^-$), effector-like (TIM3$^-$CX3CR1$^+$) and memory-like (TCF-1$^+$) were identified (Supp. Fig. S4). They additionally identified an intermediate phenotype of TIM3$^+$CX3CR1$^+$ cells, that show strong signatures of proliferation, which they confirmed using bulk sequencing of all four subsets. Expression gradients of marker genes pointed towards this population being intermediate of the effector-like CX3CR1$^+$ and the terminally exhausted TIM3$^+$ population.

Through CD4 T cell depletion experiments they linked the maintenance of CX3CR1$^+$ cells to CD4 help but not their initial generation. They make use of *Bcl-6* conditional knock-out mice, that have drastically reduced numbers of TCF1$^+$ cells, to delete specifically the memory-like population. This reduction did however not affect the number of effector-like cells, leading them to the conclusion, that CX3CR1$^+$ cells comprise an independent lineage. Tamoxifen induced expression of tdTomato in CX3CR1 expressing cells at day 8 post infection labeled exclusively CX3CR1$^+$ cells at that time-point. At day 21 post infection, CX3CR1$^+$ cells had maintained high levels of tdTomato, indicating, that they were not replenished from the memory-like pool, but rather were self-maintaining.

Cerletti2020 Our own study focused on the development of memory-like and terminally exhausted cells, in particular investigating the time-point of commitment to either lineage. To have higher time resolution, we sampled P14 CD8 T cells from the beginning of the infection between day 1 and 4, as well as at day 7, 14 and 21 after infection with a high-dose of Clone-13. This resulted in dense coverage of the whole differentiation from start of the infection to the two endpoint populations.

We inferred likely cell transitions using RNA velocity from the estimated start and end-points in an unbiased fashion. The estimated start and end-points were confirmed using marker genes and their robustness to parameter variation was assessed. Interested in the possible differentiation paths towards the endpoints, we applied cytopath (Chapter 3) to infer trajectories to the endpoints. This method uses the velocity transition matrix to sample individual trajectories from start to end-point regions. These samples are aligned and averaged to infer prototypical average trajectories (Supp. Fig. S5, Chapter 4).

The trajectories allowed us to identify of a pre-committed region in the cellular differentiation space, in which cells can still acquire either an exhausted or a memory-like phenotype. However, after this identified branching point, cells were committed to only one of the two endpoints. Using a classification algorithm, we identified a small set of cell surface markers that allowed us to isolate cells from the three branches. Adoptive transfer of these cell populations corroborated plasticity of the pre-committed branch and lineage restriction of the other two branches.

Additionally, we observed that development of the branches was driven by TCR stimulation. Transfer of cells belonging to these three branches into hosts that had been infected with a Clone-13 mutant not recognized by P14 cells, largely arrested the cells' differentiation. This indicated a crucial role of continued TCR stimulation in lineage development.

5.3 Towards a unified differentiation model of T cell exhaustion in chronic LCMV infection

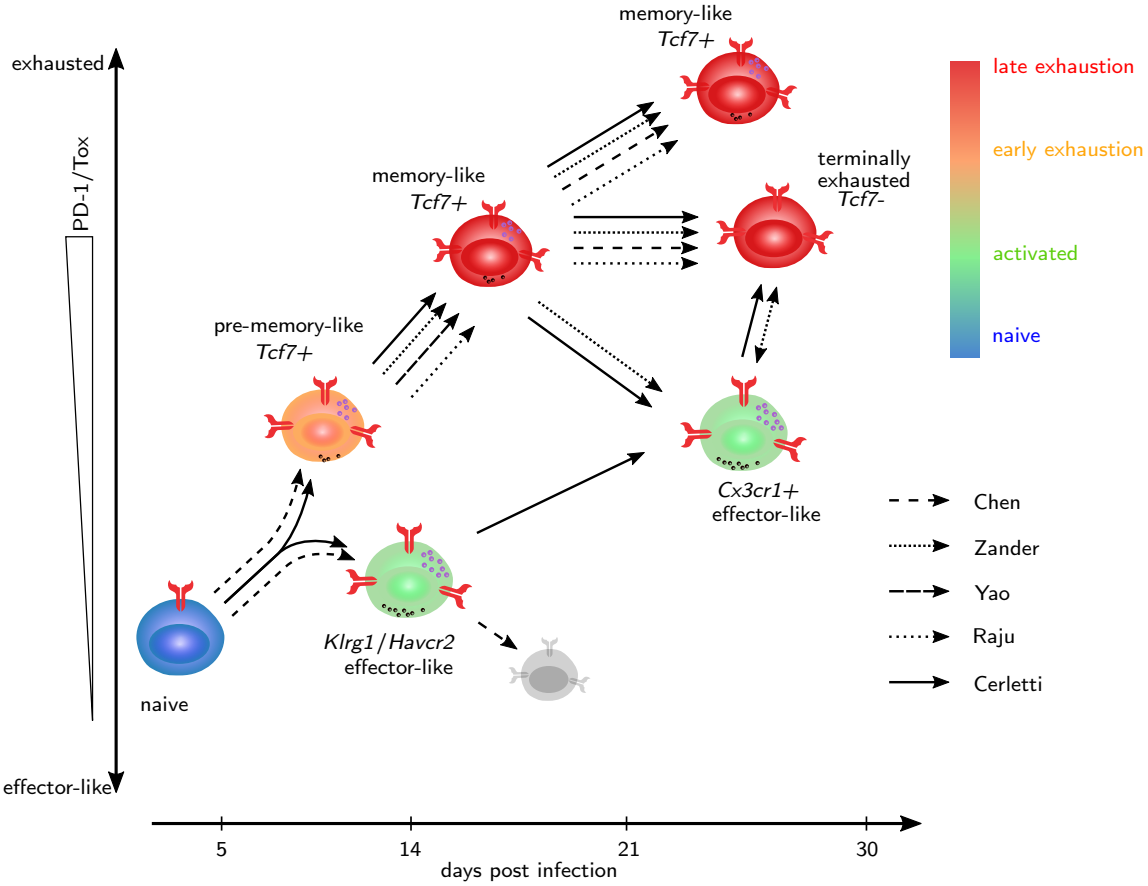


Figure 5.1: Differentiation of relevant population in T cell exhaustion across time with supported transitions by discussed studies.

The discussed studies aim at identification of the major cell states and development trajectories in Tex, and address the challenges of asynchronicity and complexity of this process by resorting to monitoring by scRNAseq. While the resulting data yields expression profiles at the single-cell level, achieving the afore goal requires addressing difficult data analysis challenges.

The result of a scRNAseq measurement is transcript abundance per gene per cell. This places every cell at a point in a high-dimensional state space that is occupied in certain regions by cells.

As cells change their phenotype and differentiate during the course of the infection, they move through different regions and occupy other points in this state space. Dimensionality reduction techniques like tSNE and UMAP try to capture the structure of this high dimensional space in two dimension to help interpret the data. Clustering techniques try to group similar populations based on their similarity.

Since scRNAseq is a destructive method, cells cannot be tracked over time but each sample is a snapshot of cell states at a particular time-point. This makes inference of cell transitions challenging. Firstly, we can only observe states that are actually present in the sample, limiting the phenotypes we can study. Often combination of multiple samples are needed to capture all populations of interest. Additionally, cells are grouped by similarity of their transcriptome and transitions or relationships are inferred therefrom. Furthermore, these analyses ignore other aspects that can define a cell state such as protein abundance or epigenetic modification.

Pseudotime inference methods try to interpolate the snapshot data in scRNAseq by ordering the cellular states along a proposed developmental axis. Current methods use non-linear distance measures based on graph connectivity (diffusion maps, PAGA) to capture cell development more accurately.

Once some time axis has been defined, lineage inference methods can help in understanding differentiation of cell populations. Usually distance measures are used to estimate the most likely cell transitions based on similarity of their transcriptomes. It is difficult to estimate the directionality of these transitions as similarity is symmetric. Previously available domain knowledge can allow to infer start and end of the process to orient directionality.

Recently developed approaches make use of the ratio of un-spliced to spliced mRNA transcripts to estimate whether a specific gene is on the verge of being turned on or off [Manno et al., 2018]. Aggregating this information across all genes and extrapolating future expression allows to infer likely future cell states. These methods can alleviate some of the ambiguities by attaching a likely direction of development to each cell, yielding a transition matrix of all states. Inferring trajectories from these transition matrices is still challenging though, as the signals are noisy and the underlying assumptions not always met, since some genes will not follow the proposed expression dynamic model.

5.3.1 Exhausted, memory- and effector-like states constitute major hallmarks of T cell exhaustion states in chronic LCMV infection

Visualization of the high-dimensional scRNAseq data in two dimensions is both an important exploratory tool for biologists and challenging. It helps in understanding structure and relations of cell state composition of heterogeneous cell populations and cellular processes. t-SNE and UMAP have prevailed as popular visualization tools, each to this end trading off the tasks of preserving local relations in the original high dimensional data, as well as maintaining global distances. However, the inevitable approximations made to this end can result in misleading visualizations, suggesting spurious orientation and relationships across cell state clusters.

All the studies discussed above were relying on such visualization techniques coupled with cluster analysis to identify relevant cell populations. Manual cluster annotation was then achieved through previously known hallmark genes or proteins expressed in these populations or interpretation of differentially expressed genes between clusters. There is a broad consensus on two main populations detected abundantly in chronic infection. All studies discover a terminally exhausted population demarcated by high expression of co-inhibitory receptors (e.g. PD-1, TIM3, CD160) and a memory-like population marked by TCF1 and Ly108.

The effector-like CX3CR1$^+$ population, however, was only reported in three of above studies and predominantly at later time-points between day 14 and 30 post infection. Yao2019 and Chen2019 do not report this population, however only included samples until day 7 and 8 post infection,

respectively, when the effector population either might not be present yet or not present a clear transcriptional signature yet.

5.3.2 Trajectory inference and intervention experiments identify novel regulators & drivers of T cell exhaustion in chronic LMCV infection

To understand the order of phenotypes during lineage development and transitions between populations, it is crucial to order the cells on a time axis. This time axis can be constructed within one sample from one given time-point to honor asynchronous activation of CD8 T cells and therefore different developmental stages; this axis is termed pseudotime. Zander2019 applied Monocle-2 to their day 30 sample to construct connections between exhausted, memory-like and effector-like populations at this time-point. The constructed branches however did not respect the previously identified populations and merely maintained an enrichment of those cells in the three constructed branches. There was a priori no information about the source population of these differentiation stages; however, the authors proposed the memory-like cells based in previous findings [Utzschneider et al., 2016c, He et al., 2016] indicating progenitor-like properties. Adoptive transfer of all three populations supported the memory-like cells having the highest proliferative potential and giving rise to both other phenotypes, placing them at the root of development. However, they also observed exhausted cells giving rise to effector-like cells, indicating unexpected plasticity in the TCF1$^-$ subsets. The proposed bifurcating structure by Monocle-2 might be too simple to capture transitions from exhausted to effector-like end-point, which is a common issue with trajectory inference [Saelens et al., 2018]. The lack of directional information makes it difficult to infer differentiation relationships from scRNAseq data alone, which is where RNA velocity could help to orient direction of differentiation.

There is additional support for plasticity in TCF1$^-$ cells by findings of Raju2020. In their study CX3CR1 expressing cells were genetically labeled with dtTomato early during infection and shown to be maintained three weeks after infection, indicating that they also have stem-like properties. Additional transfer experiments in this study revealed that the CX3CR1$^+$ lineage developed independently of the memory-like population. However, also their scRNAseq analysis did not reveal the direction of development, but was only identified through complementing flow cytometry experiments.

Chen2019 constructed lineages from one time-point but across multiple samples from different infection settings. Since all cells are from the same time-point, the cells are not ordered across time but rather along a development axis towards exhaustion. They too used Monocle-2 to fit lineages into their data but imposed the naïve cells at the start of differentiation, which seems justified as the development starts with naïve CD8 T cells. Their discovered trajectory end-points are consistent with Zander2019, finding that naïve cells give rise to an exhausted as well as an effector-like population. Without additional information though, the differentiation relationship between these two subsets could not be reliably determined.

Our own analysis is so far the only one attempting to construct lineages across a number of different experimental time-points. Dense sampling at many different time-points allowed us to follow populations along the progress of the infection. Additionally, we applied RNA velocity to get informed cell transitions. The estimation of root and end-points in the process was estimated using the equilibrium distribution of the calculated transition matrix. This data driven approach recovered terminally exhausted and memory-like cells as endpoints of the process and put cells from the earliest time-point at the starting point. We estimated lineages from start to end-points using cytopath (Chapter 3) that uses the transition probability matrix to perform simulation based trajectory inference. This reconstructed, in a data driven manner, the phenotypic states a cell would likely traverse through during development towards one of the endpoints. Additionally, this analysis detected a branching point that suggested commitment of cells to memory-like and exhausted lineages around d5. Isolation and transfer of pre-committed, pre-exhausted and pre-memory-like cells into infection matched hosts confirmed commitment to these lineages.

Pre-exhausted cells would give rise to terminally exhausted cells and pre-committed cells would produce both exhausted as well as memory cells. Although, pre-memory-like cells differentiated into a substantial pool of memory-like cells, they also gave rise to a large terminally exhausted population. This was contradictory to our lineage inference, since our trajectories suggested that there is no transition from memory-like to the exhausted population.

One reason why we did not identify memory-like to terminally exhausted transitions in our lineage inference, could be the absence or very low abundance of these transition states in the spleen at a given time-point. In addition, the differentiation of terminally exhausted cells could happen in other lymphoid tissues or peripheral organs [Sandu et al., 2020b]. Absence of detecting such transition states would prohibit us from finding these transition trajectories originating at the population of memory-like cells and terminating at either the effector-like or terminally exhausted state. The same is true, if this process is very fast and therefore unlikely to be captured during sampling.

Yao2019 did not attempt to construct lineages but used the single cell resolution in their data to identify the two corresponding populations in acute and chronic infection. They used differential expression analysis to find drivers of the exhausted phenotype population specific and genome wide. Focusing on the memory-precursor cells in both acute and chronic infection allowed them to identify the specific influence of the different infection settings in these populations specifically. Among other genes they identified *Tox* as being differentially expressed. This factor was already implicated as relevant in CD4 T cell development [Aliahmad and Kaye, 2008], and they validated its relevance in maintenance of the TCF1$^+$ CD8 T cell population in chronic infection.

Other studies focusing on TOX have shown that it drives up-regulation of co-inhibitory receptors [Alfei et al., 2019]. Therefore, TOX will also reduce antigen stimulation of these cells. Our transfer experiments into Clone-13 mutant infection illustrated a crucial role of antigen stimulation of memory-like cells to drive differentiation into terminally exhausted cells. The expression of TOX could allow these cells to shield themselves from strong stimulation and thereby maintain their stem-like properties. The deletion of TOX would lead to depletion of the memory-like pool through stimulation induced differentiation. This highlights the intricate connection between antigen-stimulation and differentiation.

This indicates also a potential shortcoming of isolation and transfer experiments. Adoptive transfer experiments disturb the spatial niche of the cells by removing them from their original residence and exposing them to a new environment after intravenous injection. Transfer of memory-like cells to an environment which is more abundant in antigen or lacks co-inhibitory ligands could result in TCR stimulation that triggers a differentiation program towards a terminally exhausted phenotype, which might be less apparent in their original niche environment.

5.4 Outlook: Temporal, spatial and network information complements transcriptome data

Despite the high-detailed information and single-cell resolution of scRNAseq, there are still gaps of information that need to be filled with complementary technologies.

The destructive nature of the technology allows only for snapshot analyses and studying temporal processes involves connecting multiple samples or error prone pseudotime inference. Recently developed approaches make use of enzymes that modify the sequence of specifically tagged RNA transcripts in a time dependent manner [Rodriques et al., 2020]. This adds a "timestamp" to transcripts during a process and allows retrospective temporal sorting of the cells in scRNAseq experiments with hour-scale accuracy. Including this information as a prior in lineage inferences can crucially improve pseudotime estimation since it does not only rely on transcriptome similarity but additionally a cell intrinsic temporal measure. It could potentially

aid in understanding the time-scale of the transitions between memory-like and exhausted cell populations in chronic infection.

The loss of spatial information is an additional drawback of current single-cell RNA sequencing technology. The cellular context is crucial in many biological processes, such as interaction between B-cells and follicular T helper cells in germinal centers [Seth and Craft, 2019]. Current systems in droplet based scRNAseq destroy spatial relation and cell-cell interactions can not be assessed. The recent development of spatial transcriptomics approaches like 10x Visium [Bergenstråhle et al., 2020], or fluorescence based MERFISH [Chen et al., 2015], as well as spatial proteomic approaches like imaging mass cytometry [Palma and Bodenmiller, 2015] constitute a perspective to include the tissue arrangement and context to develop a more complete model for development of T cell exhaustion.

Additionally, in scRNAseq only transcripts are considered to infer cellular phenotype, despite the fact that it is proteins exerting function and defining phenotypes. This additional layer of regulation might be of relevance. Antibody based protein quantification with CiteSeq allows to include selected protein information in an scRNAseq experiment by including DNA-tagged antibodies against surface markers that can be identified using standard droplet based platforms. However, this method is targeted and limited to validated and available antibodies against relevant epitopes. In CD8 T cell development though, there is an abundance of relevant antibodies available that could complement transcriptional information.

On the computational side we still lack universally applicable and robust lineage inference methods for scRNAseq data. A major difficulty is the structural diversity in datasets. Assumptions about cluster density, transition state occupancy and branching structure by the different methods can lead to vastly different results in cell ordering or lineage inference. There is still no "killer" algorithm, that is able to fully interpret one given data-set, but usually only a combination of methods can reveal the real underlying biological structure. Combining read counts with RNA velocity to infer directionality of gene expression can aid lineage inference like in the here presented cytopath (Chapter 3) or recently published CellRank [Lange et al., 2020]. Other solutions could involve the inclusion of previous knowledge about gene modules or protein-protein interactions [Klimm et al., 2019] to focus on the activity of gene sets with known biological relevance. The activity of gene-set further characterizes cellular states and might allow causal transition inference [Qiu et al., 2020].

5.5 Concluding remarks and future challenges

Differentiation of CD8 T cells during infection remains an attractive field of study. Ease of isolation of these cells combined with extensive domain knowledge, make this field a prime candidate to combine and extend novel single-cell technologies as well as improve computational methods to analyze this data. Integrating data from complementary experimental techniques, such as spatial transcriptomics or genetic lineage tracing, will be needed to fully understand spatially and temporally dependent differentiation processes. Only a comprehensive understanding of the CD8 differentiation process in infection might allow for interventions that shape the composition of the T cell pool and positively affect disease progression and outcome.

Supplementary Figures

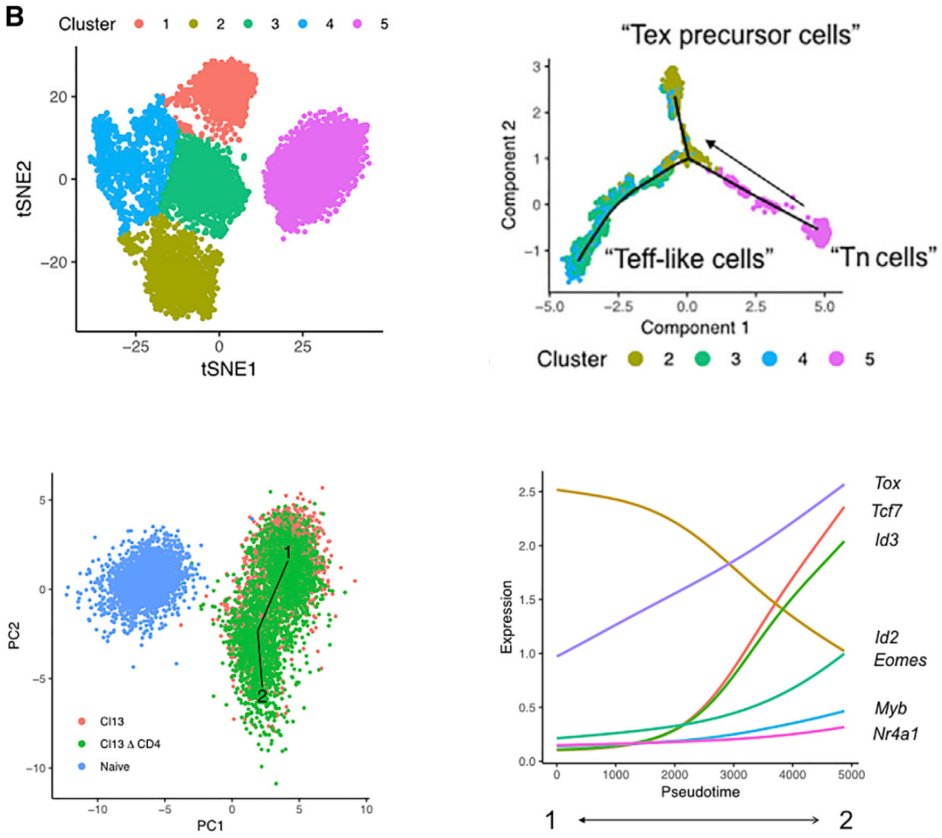


Figure S1: clustering and lineage inference from Chen 2019

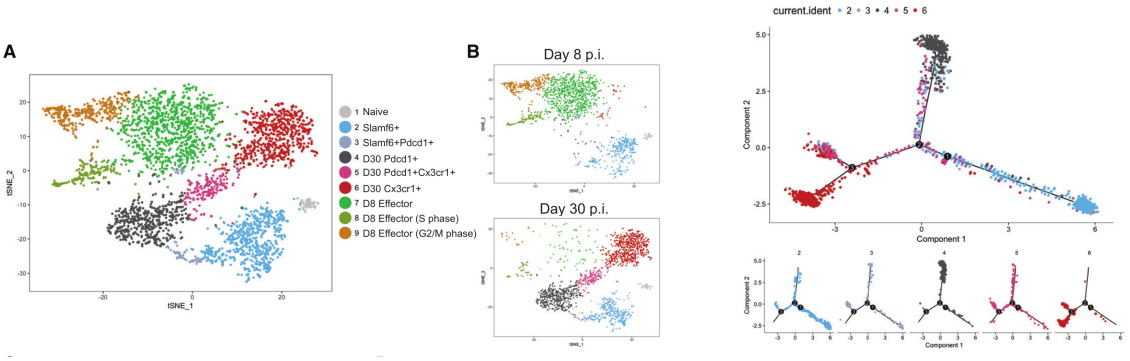


Figure S2: clustering and lineage inference from Zander 2019

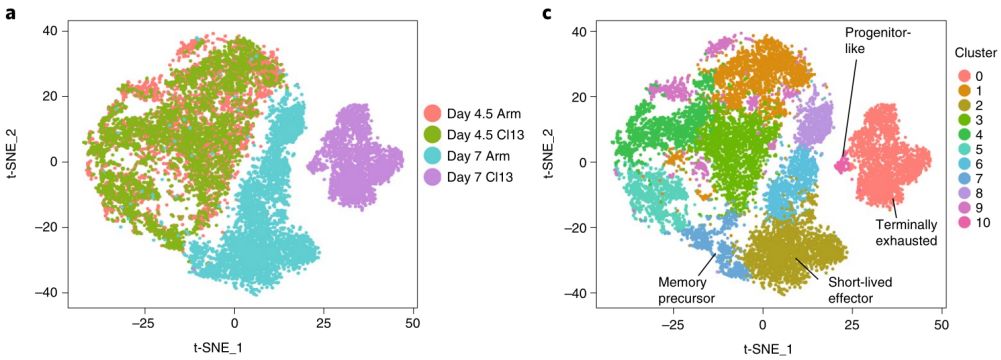


Figure S3: clustering and population annotation from Yao 2019

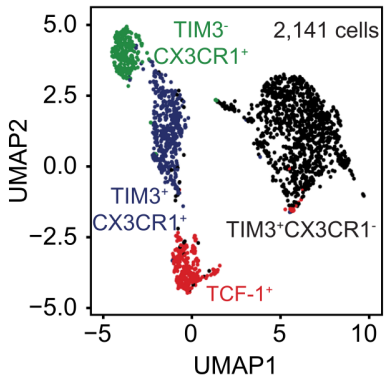


Figure S4: clustering and population annotation from Raju 2020

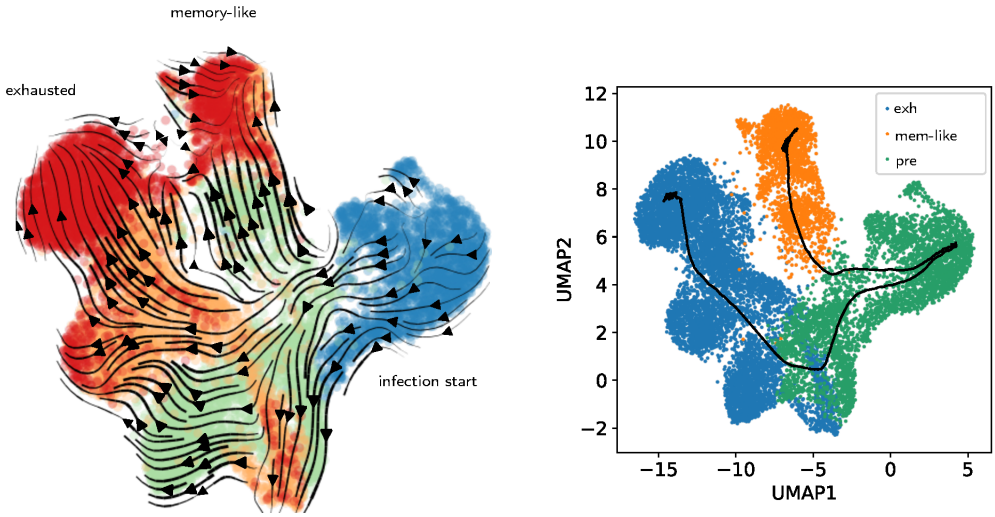
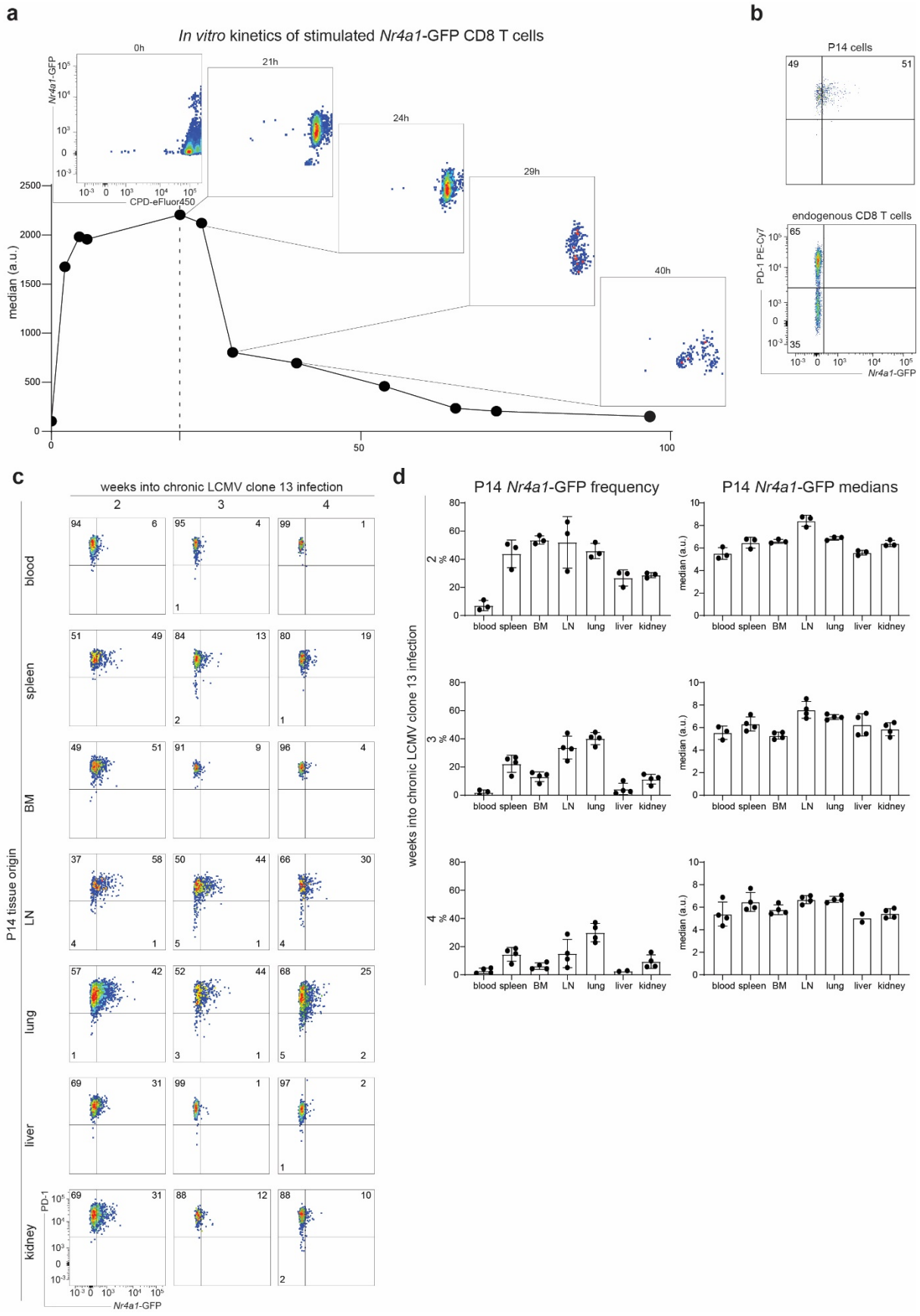


Figure S5: RNA velocity and lineage inference from Cerletti 2020

Supplementary Figures - Chapter 2

Exhausted CD8 T cells exhibit low and strongly inhibited TCR signaling during chronic LCMV infection

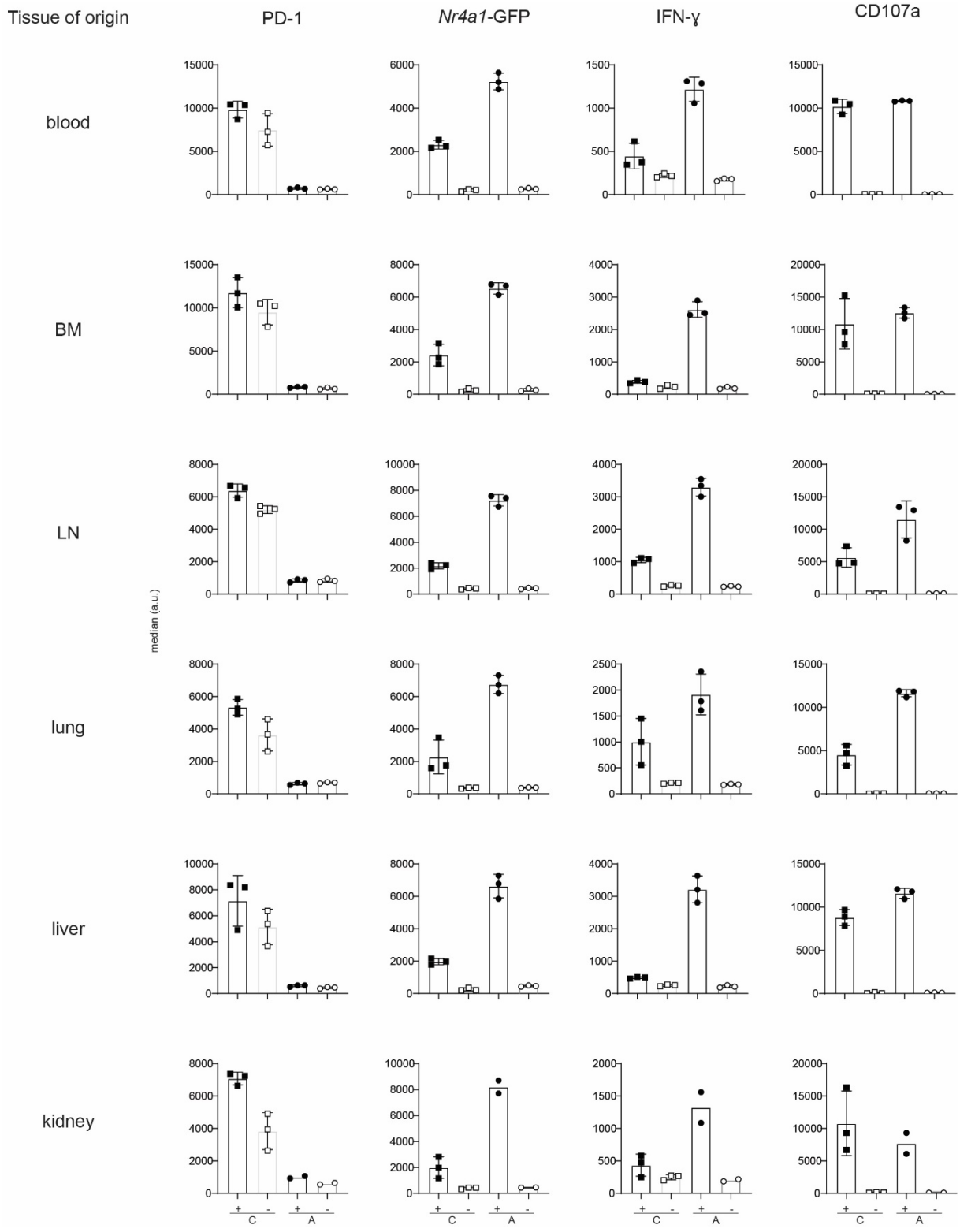
Sandu et al., 2020



Supplementary Fig. 1: *In vitro* kinetics of *Nr4a1*-GFP after stimulation and *ex vivo* expression in P14 cells.

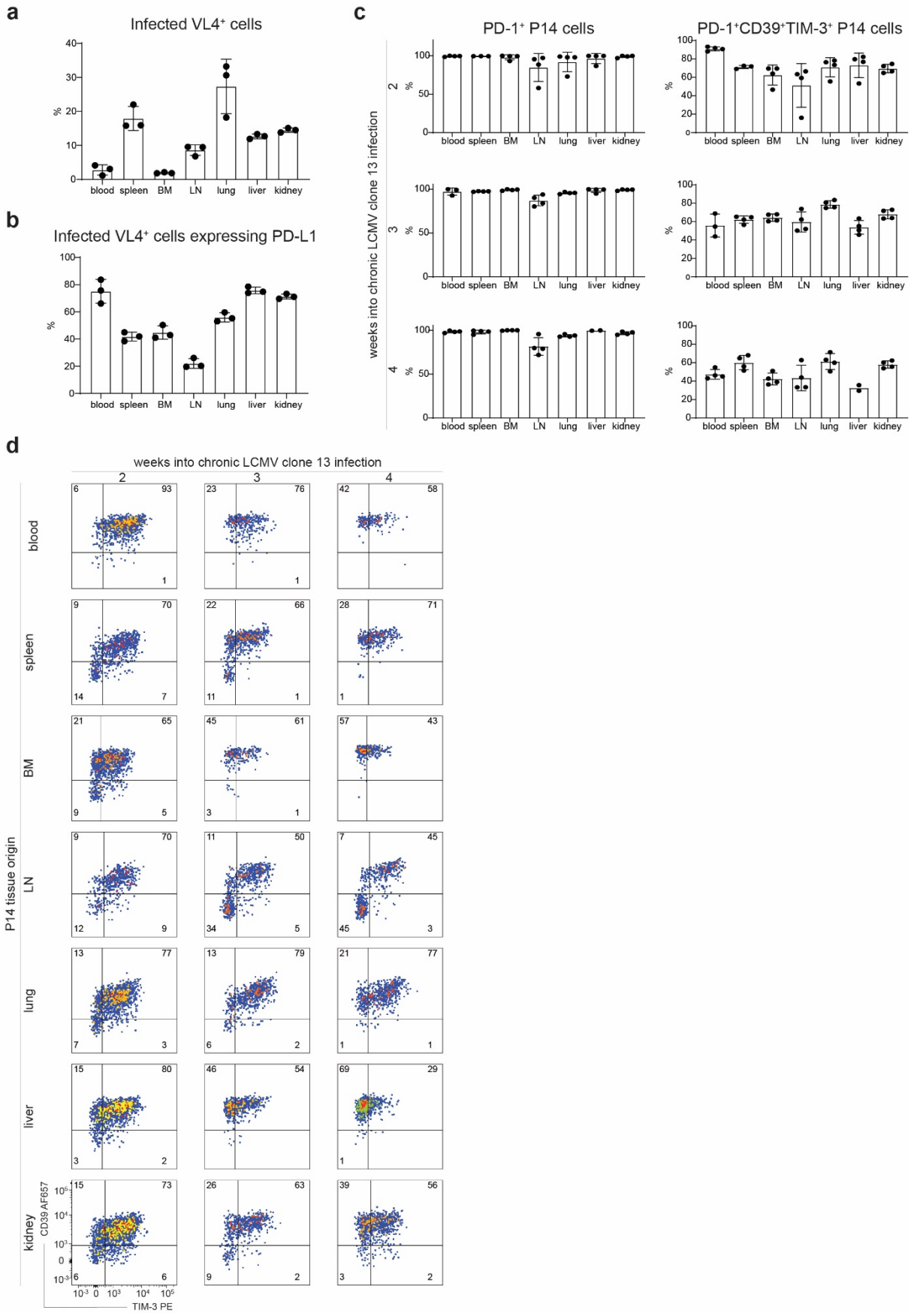
a *Nr4a1*-GFP CD8 T cells labeled with CPD (cell proliferation dye) were in vitro stimulated with plate-bound anti-CD3ε and anti-CD28 for 20 hours. Dotted line indicates the time when the stimulus was removed. Boxes

show the profile of CDP and *Nr4a1*-GFP levels at the specified time-points after activation. Dots represent mean of normalized *Nr4a1*-GFP expression (n = 2 mice). **b-d** CD45.1$^+$ P14-*Nr4a1*-GFP cells were transferred into CD45.2$^+$ hosts one day prior infection with a high dose of LCMV clone 13. **b** Non-transgenic endogenous CD8 T cells were used to set the gates: example showing *Nr4a1*-GFP signal in non-transgenic endogenous CD8 cells (gated on alive single CD8$^+$ cells) or transferred transgenic P14 *Nr4a1*-GFP cells (gated on alive single CD8$^+$ CD45.1$^+$ cells) isolated from the spleen of a chronically infected mouse 14 dpi. **c** Example of flow cytometry plots showing PD-1 and *Nr4a1*-GFP expression in P14 cells isolated from various tissues at the indicated time-points into chronic infection. **d** Bar plots show mean ± SD of the frequency or medians of GFP$^+$ P14-*Nr4a1*-GFP cells isolated at the indicated time points after infection. Each dot represents an individual mouse. One of two experiments is shown (n=3-4 mice). Source data are provided as a Source Data File.



Supplementary Fig. 2: *In vitro* restimulation of sorted P14 cells harvested from different organs. CD45.1$^+$ P14-*Nr4a1*-GFP cells were adoptively transferred into CD45.2$^+$ mice one day prior infection with a high (2x10^6 ffu for chronic (C) infection) or low (200 ffu to induce an acute (A) infection) dose of LCMV clone 13. Three weeks into chronic infection, P14 cells were sorted from different tissues (lymph nodes (LN,) bone marrow (BM), lungs, liver, blood, and kidney) and stimulated in vitro for six hours with plate-bound antibodies. Various

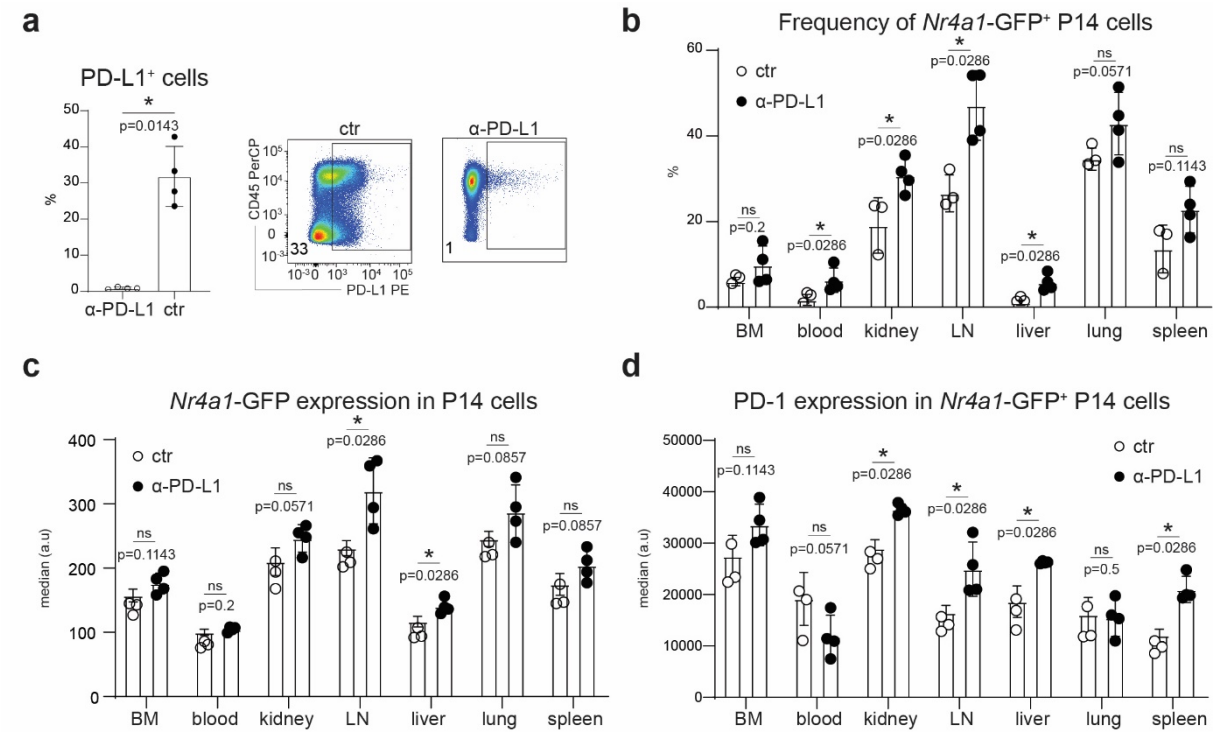
markers were quantified (average of medians ± SD are depicted) in cells that responded ($CD107a^+$), denoted with plus signs, and cells that did not ($CD107a^-$), denoted with minus signs. One of two experiments is shown (n=3 mice). Each dot represents an individual mouse. Source data are provided as a Source Data File.



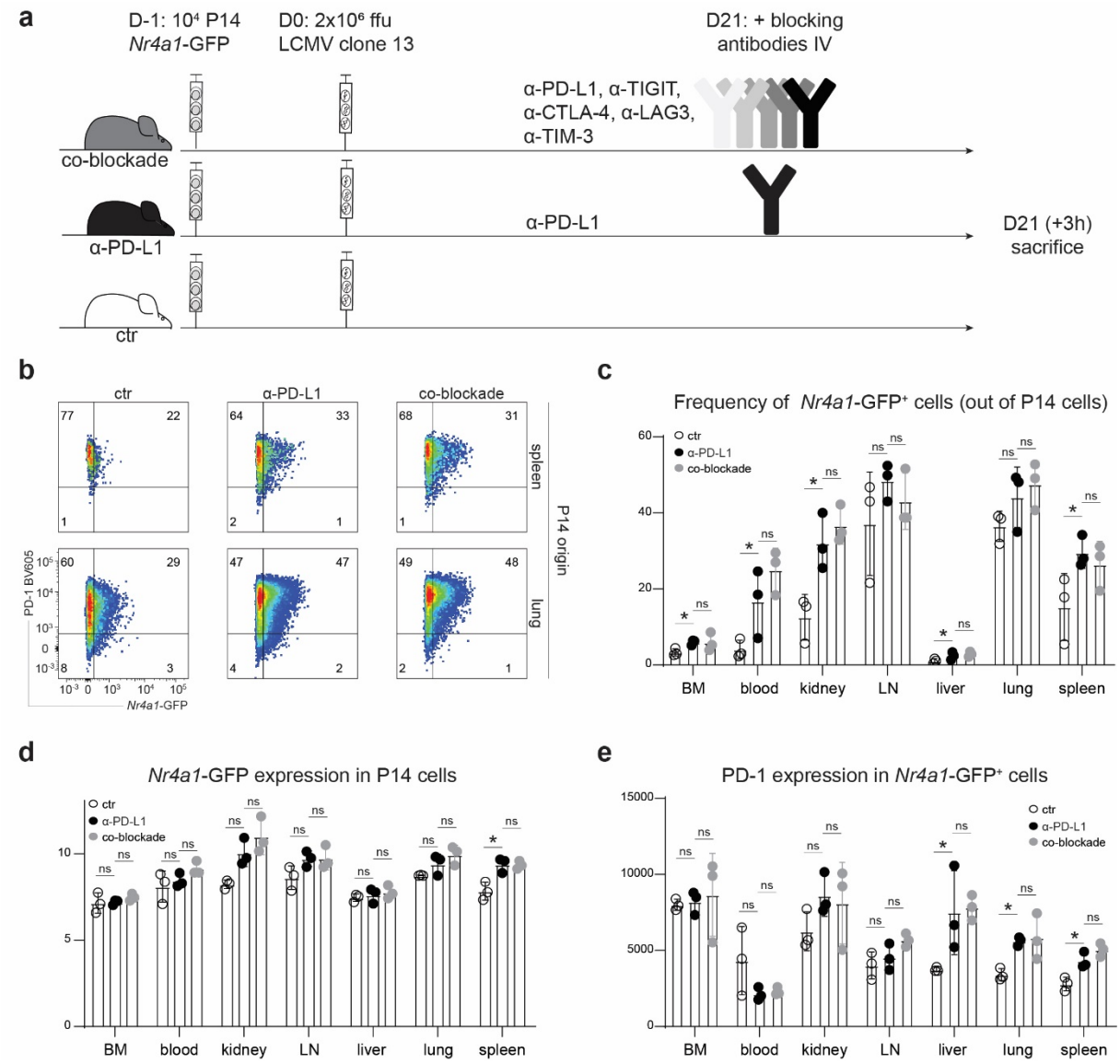
Supplementary Fig. 3: PD-L1 expression on infected cells and inhibitory receptor expression on P14 cells.

Percentage of infected cells (gated on alive cells (**a**) or gated on alive VL4^{+} cells (**b**) expressing PD-L1 are

shown for seven tissues harvested from chronically LCMV infected mice 21 days into chronic infection (n = 3 mice). **c-d** P14-*Nr4a1*-GFP cells were adoptively transferred into hosts one day prior high dose LCMV infection. Example of flow cytometry plots show TIM-3 and CD39 expression in P14 cells (gated on alive single $CD8^+$ $CD45.1^+$ cells) isolated from different tissues at indicated time-points into chronic infection (**c**). Graphs show the frequency of $PD\text{-}1^+$ or $PD\text{-}1^+CD39^+TIM\text{-}3^+$ of P14 cells isolated from various tissues at different time points after infection. Bar plots show mean ± SD. Each dot represents an individual mouse. One of two experiments is shown (n=2-4 mice). Source data are provided as a Source Data File.

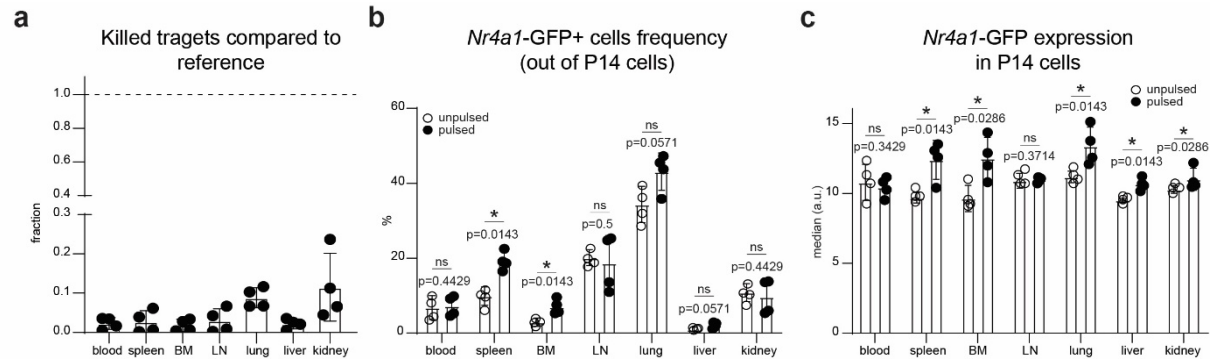


Supplementary Fig. 4: PD-L1 blockade increases TCR signaling in various tissues. CD45.1$^+$ P14-*Nr4a1*-GFP were adoptively transferred into naïve CD45.2$^+$ hosts one day prior high dose LCMV clone 13 infection. 21 days into chronic infection, one group received one dose of 200 µg α-PD-L1 blocking antibody intravenously. The mice were sacrificed six hours after treatment. **a** Representative flow cytometry plots (cells isolated from the spleen, gated on alive single cells). PD-L1 was stained to check blockade efficiency with the same clone used for blocking. Fraction of *Nr4a1*-GFP$^+$ cells (gated on single alive CD8$^+$ CD45.1$^+$ P14 cells) (**b**), mean of GFP (**c**) and PD-1 (**d**) (gated on single alive CD8$^+$ CD45.1$^+$ P14 GFP$^+$ cells) are shown for cells isolated from various tissues. Bar plots represent mean ± SD. * denotes a significant p-value ≤ 0.05 and ns non-significance (one-tailed Mann-Whitney test without correction for multiple comparisons). One of three experiments is shown (n=4 mice). Each dot represents an individual mouse. Source data are provided as a Source Data File.

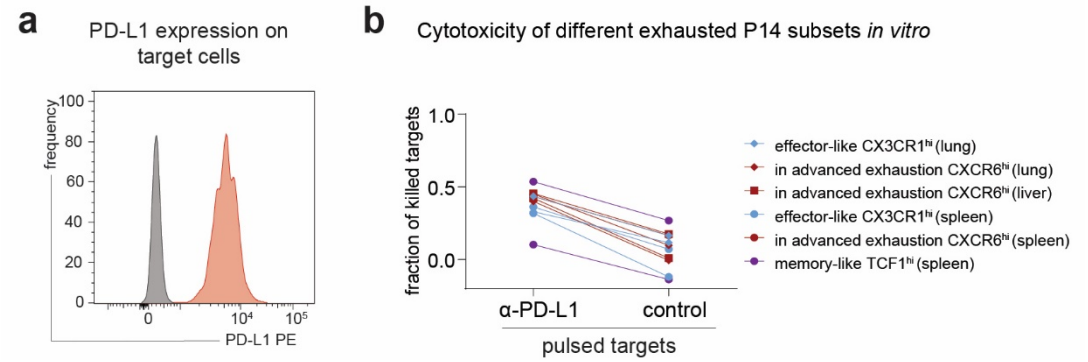


Supplementary Fig. 5: Short-term inhibitory receptors co-blockade does not increase TCR signaling compared to PD-L1 blockade alone. a CD45.1$^+$ P14-*Nr4a1*-GFP were adoptively transferred into naïve CD45.2$^+$ hosts one day prior high dose LCMV clone 13 infection. 21 days post-infection, mice received one dose of α-PD-L1, α-TIGIT, α-CTLA-4, α-LAG3, α-TIM-3 (co-blockade), one dose of α-PD-L1 blocking antibody alone intravenously or no antibody (ctr). The mice were sacrificed three hours after treatment. **b** Representative flow cytometry plots showing PD-1 and *Nr4a1*-GFP expression in P14 cells (gated on alive single CD8$^+$ CD45.1$^+$ cells) isolated from spleen or lung. **b** Representative plots showing *Nr4a1*-GFP signal in the three experimental groups, control (ctr) and treated (α-PD-L1 and co-blockade) are shown. Fraction of *Nr4a1*-GFP+ cells (gated on alive single CD8$^+$ CD45.1$^+$ cells) (**c**), median of GFP (**d**) and PD-1 (**e**) (gated on alive single CD8$^+$ CD45.1$^+$ GFP$^+$ cells) are shown for cells isolated from different tissues. Bar plots represent mean ± SD. * denotes p-value = 0.05 and ns non-significance (Mann-Whitney one tailed-test without correction for multiple

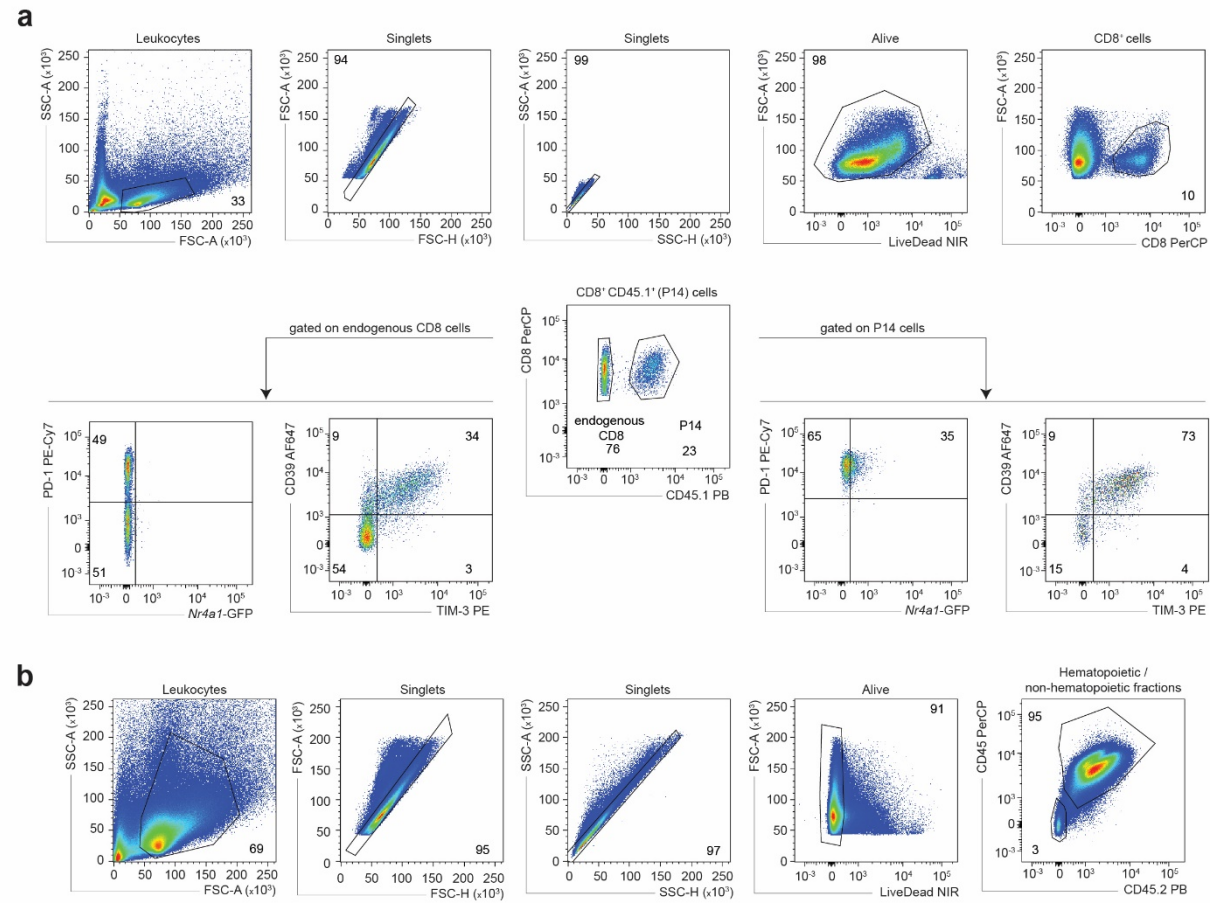
comparisons). One of two experiments is shown (n=3 mice). Each dot represents an individual mouse. Source data are provided as a Source Data File.



Supplementary Fig. 6: *In vivo* killing assay three weeks into chronic infection. CD45.1$^+$ P14-*Nr4a1*-GFP cells were adoptively transferred into naïve CD45.2$^+$ mice one day prior high dose infection. 21 days into chronic infection, the hosts received GP_{33-41}-pulsed or unpulsed targets together with a reference population. Three hours later the mice were sacrificed. **a** Fraction of recovered targets (calculated based on the reference population) in different tissues isolated from unpulsed or pulsed groups is shown. Frequency of GFP$^+$ cells (gated on alive single CD8$^+$ CD45.1$^+$ P14 cells) (**b**) and medians of GFP (gated on alive single CD8$^+$ CD45.1$^+$ GFP$^+$ cells) (**c**) are shown. Bar plots represent mean ± SD. * denotes a significant p-value ≤ 0.05 and ns non-significance (Mann-Whitney one tailed-test without correction for multiple comparisons). Each dot represents an individual mouse. One of three experiments is shown (n=4 mice). Source data are provided as a Source Data File.



Supplementary Fig. 7: ***In vitro*** **killing assay three weeks into chronic infection.** CD45.1$^+$ P14-*Tcf7* cells were adoptively transferred into naïve CD45.2$^+$ mice one day prior high dose infection. P14 cells were sorted from chronically infected mice 21 dpi based on CX3CR1, CXCR6, and TCF1-GFP expression into advanced exhausted CXCR6^{hi}CX3CR1^{neg}, effector-like CX3CR1^{hi}CXCR6^{lo}, and memory-like GFP^{hi} cells. Sorted populations were incubated with GP_{33-41} pulsed EL4 target cells at 5:1 E/T ratio. **a** EL4 cells were incubated with IFN-γ at 100 U/mL for 6 hours prior pulsing to induce PD-L1 upregulation (red histogram). For one condition, PD-L1 expression was blocked with 30 µg/mL α-PD-L1 for one hour at 37ºC (gray histogram). **b** The graph shows percentage of killing; lines denote paired samples. One of two experiments is shown. Each data point represents a sorted population from 2 or 3 pooled mice. n=2 groups of pooled mice. 2 and 3 mice were pooled per group to increase recovered cell numbers. Source data are provided as a Source Data File.



Supplementary Fig. 8: Flow cytometry gating strategy. a For all samples the following gating strategy was used: leukocytes (SSC-A/FSC-A), exclusion of doublets (FSC-H/FSC-A and SSC-H/SSC-A), alive cells (Near IR/FSC-A). For P14 staining, cells were gated on CD8 (FSC-A/CD8 gate), then CD45.1 (CD8/CD45.1 gate). *Nr4a1*-GFP gates were set using endogenous CD8 cells (gated on alive single CD8+ cells) not expressing the transgene from the same mouse. Inhibitory receptor gates (PD-1, CD39, TIM-3) were set on endogenous CD8 cell population where distinct populations could be clearly identified. **b** For hematopoietic or non-hematopoietic populations, a wider FSC-A/SSC-A gate was used (including all cells). In order to distinguish the hematopoietic from the non-hematopoietic fraction, CD45 and CD45.2 staining was used.

Bibliography

[Agnellini, 2007] Agnellini, P. (2007). Impaired nfat nuclear translocation results in split exhaustion of virus-specific cd8+ t cell functions during chronic viral infection. *Proc. Natl Acad. Sci. USA*, 104.

[Agnellini et al., 2007] Agnellini, P., Wolint, P., Rehr, M., Cahenzli, J., Karrer, U., and Oxenius, A. (2007). Impaired NFAT nuclear translocation results in split exhaustion of virus-specific CD8+ t cell functions during chronic viral infection. *Proceedings of the National Academy of Sciences*, 104(11):4565–4570.

[Ahmed et al., 1984] Ahmed, R., Salmi, A., Butler, L. D., Chiller, J. M., and Oldstone, M. B. (1984). Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice. role in suppression of cytotoxic t lymphocyte response and viral persistence. *Journal of Experimental Medicine*, 160(2):521–540.

[Alfei et al., 2019] Alfei, F., Kanev, K., Hofmann, M., Wu, M., Ghoneim, H. E., Roelli, P., Utzschneider, D. T., von Hösslin, M., Cullen, J. G., Fan, Y., Eisenberg, V., Wohlleber, D., Steiger, K., Merkler, D., Delorenzi, M., Knolle, P. A., Cohen, C. J., Thimme, R., Youngblood, B., and Zehn, D. (2019). TOX reinforces the phenotype and longevity of exhausted t cells in chronic viral infection. *Nature*.

[Aliahmad and Kaye, 2008] Aliahmad, P. and Kaye, J. (2008). Development of all CD4 t lineages requires nuclear factor TOX. *Journal of Experimental Medicine*, 205(1):245–256.

[Anderson et al., 2016] Anderson, A. C., Joller, N., and Kuchroo, V. K. (2016). Lag-3, tim-3, and TIGIT: Co-inhibitory receptors with specialized functions in immune regulation. *Immunity*, 44(5):989–1004.

[Angerer et al., 2015] Angerer, P., Haghverdi, L., Büttner, M., Theis, F. J., Marr, C., and Buettner, F. (2015). destiny - diffusion maps for large-scale single-cell data in r. *Bioinformatics*, 32(December 2015):btv715—btv715–.

[Are et al., 2016] Are, P., Driver, E., Exhaustion, T. C., Bengsch, B., Johnson, A. L., Kurachi, M., Paley, M. A., Delgoffe, G. M., Wherry, E. J., Bengsch, B., Johnson, A. L., Kurachi, M., Odorizzi, P. M., and Pauken, K. E. (2016). Bioenergetic insufficiencies due to metabolic alterations regulated by the inhibitory receptor article bioenergetic insufficiencies due to metabolic alterations regulated by the inhibitory receptor pd-1 are an early driver of cd8 + t cell exhaustion. *Immunity*, 45(2):1–16.

[Au-Yeung et al., 2014] Au-Yeung, B. B., Zikherman, J., Mueller, J. L., Ashouri, J. F., Matloubian, M., Cheng, D. a., Chen, Y., Shokat, K. M., and Weiss, a. (2014). A sharp t-cell antigen receptor signaling threshold for t-cell proliferation. *Proceedings of the National Academy of Sciences*, 111(25):1413726111—1413726111–.

[Barber et al., 2006] Barber, D. L., Wherry, E. J., Masopust, D., Zhu, B., Allison, J. P., Sharpe, A. H., Freeman, G. J., and Ahmed, R. (2006). Restoring function in exhausted cd8 t cells during chronic viral infection. *Nature*, 439(7077):682–687.

[Battegay et al., 1991] Battegay, M., Cooper, S., Althage, A., Bänziger, J., Hengartner, H., and Zinkernagel, R. M. (1991). Quantification of lymphocytic choriomeningitis virus with an immunological focus assay in 24- or 96-well plates. *Journal of Virological Methods*, 33(1-2):191–198.

[Becht et al., 2018] Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I. W. H., Ng, L. G., Ginhoux, F., and Newell, E. W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology*, 37(1):38–44.

[Beltra et al., 2020] Beltra, J.-C., Manne, S., Abdel-Hakeem, M. S., Kurachi, M., Giles, J. R., Chen, Z., Casella, V., Ngiow, S. F., Khan, O., Huang, Y. J., Yan, P., Nzingha, K., Xu, W., Amaravadi, R. K., Xu, X., Karakousis, G. C., Mitchell, T. C., Schuchter, L. M., Huang, A. C., and Wherry, E. J. (2020). Developmental relationships of four exhausted CD8+ t cell subsets reveals underlying transcriptional and epigenetic landscape control mechanisms. *Immunity*, 52(5):825–841.e8.

[Bengsch et al., 2018] Bengsch, B., Ohtani, T., Khan, O., Setty, M., Manne, S., O'Brien, S., Gherardini, P. F., Herati, R. S., Huang, A. C., Chang, K.-M., Newell, E. W., Bovenschen, N., Pe'er, D., Albelda, S. M., and Wherry, E. J. (2018). Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 t cells. *Immunity*, 48(5):1029–1045.e5.

[Bergen et al., 2020] Bergen, V., Lange, M., Peidli, S., Wolf, F. A., and Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*.

[Bergenstråhle et al., 2020] Bergenstråhle, J., Larsson, L., and Lundeberg, J. (2020). Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics*, 21(1).

[Blackburn, 2009] Blackburn, S. D. (2009). Coregulation of cd8+ t cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat. Immunol.*, 10.

[Blackburn, 2010] Blackburn, S. D. (2010). Tissue-specific differences in pd-1 and pd-l1 expression during chronic viral infection: implications for cd8 t-cell exhaustion. *J. Virol.*, 84.

[Blackburn et al., 2010] Blackburn, S. D., Crawford, A., Shin, H., Polley, A., Freeman, G. J., and Wherry, E. J. (2010). Tissue-specific differences in pd-1 and pd-l1 expression during chronic viral infection: implications for cd8 t-cell exhaustion. *Journal of virology*, 84(4):2078–2089.

[Blackburn et al., 2008a] Blackburn, S. D., Shin, H., Freeman, G. J., and Wherry, E. J. (2008a). Selective expansion of a subset of exhausted cd8 t cells by alphapd-l1 blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39):15016–15021.

[Blackburn et al., 2008b] Blackburn, S. D., Shin, H., Haining, W. N., Zou, T., Workman, C. J., Polley, A., Betts, M. R., Freeman, G. J., Vignali, D. A. A., and Wherry, E. J. (2008b). Coregulation of CD8+ t cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nature Immunology*, 10(1):29–37.

[Blondel et al., 2008] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

[Bowie, 2007] Bowie, A. G. (2007). Recent advances in understanding the role of toll-like receptors in anti-viral immunity. *Clinical & Experimental Immunology*, 147(2):217–226.

[Byrnes et al., 2018] Byrnes, L. E., Wong, D. M., Subramaniam, M., Meyer, N. P., Gilchrist, C. L., Knox, S. M., Tward, A. D., Ye, C. J., and Sneddon, J. B. (2018). Lineage dynamics of murine pancreatic development at single-cell resolution. *Nature Communications*, 9(1).

[Böttcher and Knolle, 2015] Böttcher, J. and Knolle, P. A. (2015). Global transcriptional characterization of CD8+ t cell memory. *Seminars in Immunology*, 27(1):4–9.

[Carty et al., 2020] Carty, M., Guy, C., and Bowie, A. G. (2020). Detection of viral infections by innate immunity. *Biochemical Pharmacology*, page 114316.

[Chen et al., 2015] Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., and Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, 348(6233):aaa6090–aaa6090.

[Chen et al., 2019] Chen, Z., Ji, Z., Ngiow, S. F., Manne, S., Cai, Z., Huang, A. C., Johnson, J., Staupe, R. P., Bengsch, B., Xu, C., Yu, S., Kurachi, M., Herati, R. S., Vella, L. A., Baxter, A. E., Wu, J. E., Khan, O., Beltra, J.-C., Giles, J. R., Stelekati, E., McLane, L. M., Lau, C. W., Yang, X., Berger, S. L., Vahedi, G., Ji, H., and Wherry, E. J. (2019). TCF-1-centered transcriptional network drives an effector versus exhausted CD8 t cell-fate decision. *Immunity*.

[Cook and Whitmire, 2016] Cook, K. D. and Whitmire, J. K. (2016). Lag-3 confers a competitive disadvantage upon antiviral cd8+ t cell responses. *J. Immunol.*, 197.

[core team, 2019] core team, R. (2019). R core development team. r: A language and environment for statistical computing. (r core development team, 2019).

[Dixon, 2018] Dixon, K. O. (2018). Functional anti-tigit antibodies regulate development of autoimmunity and antitumor immunity. *J. Immunol.*, 200.

[Frebel, 2012] Frebel, H. (2012). Programmed death 1 protects from fatal circulatory failure during systemic virus infection of mice. *J. Exp. Med.*, 209.

[Frebel and Oxenius, 2013a] Frebel, H. and Oxenius, A. (2013a). The risks of targeting co-inhibitory pathways to modulate pathogen-directed t cell responses. *Trends in Immunology*, 34(5):193–199.

[Frebel and Oxenius, 2013b] Frebel, H. and Oxenius, A. (2013b). The risks of targeting co-inhibitory pathways to modulate pathogen-directed t cell responses. *Trends Immunol.*, 34.

[Garcia et al., 2015] Garcia, V., Richter, K., Graw, F., Oxenius, A., and Regoes, R. R. (2015). Estimating the in vivo killing efficacy of cytotoxic t lymphocytes across different peptide-mhc complex densities. *PLoS Comput. Biol.*, 11.

[Graw et al., 2011] Graw, F., Richter, K., Oxenius, A., and Regoes, R. R. (2011). Comparison of cytotoxic t lymphocyte efficacy in acute and persistent lymphocytic choriomeningitis virus infection. *Proceedings of the Royal Society B: Biological Sciences*, 278(1723):3395–3402.

[Gupta, 2015] Gupta, P. K. (2015). Cd39 expression identifies terminally exhausted cd8+ t cells. *PLoS Pathog.*, 11.

[Gupta et al., 2015] Gupta, P. K., Godec, J., Wolski, D., Adland, E., Yates, K., Pauken, K. E., Cosgrove, C., Ledderose, C., Junger, W. G., Robson, S. C., Wherry, E. J., Alter, G., Goulder, P. J. R., Klenerman, P., Sharpe, A. H., Lauer, G. M., and Haining, W. N. (2015). CD39 expression identifies terminally exhausted CD8+ t cells. *PLOS Pathogens*, 11(10):e1005177.

[Haghverdi et al., 2016] Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F., and Theis, F. J. (2016). Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods*, 13(10):845–848.

[He et al., 2016] He, R., Hou, S., Liu, C., Zhang, A., Bai, Q., Han, M., Yang, Y., Wei, G., Shen, T., Yang, X., Xu, L., Chen, X., Hao, Y., Wang, P., Zhu, C., Ou, J., Liang, H., Ni, T., Zhang, X., Zhou, X., Deng, K., Chen, Y., Luo, Y., Xu, J., Qi, H., Wu, Y., and Ye, L. (2016). Erratum: Follicular CXCR5-expressing CD8+ t cells curtail chronic viral infection. *Nature*, 540(7633):470–470.

[Heinemann and Zenobi, 2011] Heinemann, M. and Zenobi, R. (2011). Single cell metabolomics. *Current Opinion in Biotechnology*, 22(1):26–31.

[Huang et al., 2014] Huang, Y.-H., Zhu, C., Kondo, Y., Anderson, A. C., Gandhi, A., Russell, A., Dougan, S. K., Petersen, B.-S., Melum, E., Pertel, T., Clayton, K. L., Raab, M., Chen, Q., Beauchemin, N., Yazaki, P. J., Pyzik, M., Ostrowski, M. A., Glickman, J. N., Rudd, C. E., Ploegh, H. L., Franke, A., Petsko, G. A., Kuchroo, V. K., and Blumberg, R. S. (2014). CEACAM1 regulates TIM-3-mediated tolerance and exhaustion. *Nature*, 517(7534):386–390.

[Hudson et al., 2019] Hudson, W. H., Gensheimer, J., Hashimoto, M., Wieland, A., Valanparambil, R. M., Li, P., Lin, J.-X., Konieczny, B. T., Im, S. J., Freeman, G. J., Leonard, W. J., Kissick, H. T., and Ahmed, R. (2019). Proliferating transitory t cells with an effector-like transcriptional signature emerge from PD-1+ stem-like CD8+ t cells during chronic infection. *Immunity*, 51(6):1043–1058.e4.

[Im et al., 2016] Im, S. J., Hashimoto, M., Gerner, M. Y., Lee, J., Kissick, H. T., Burger, M. C., Shan, Q., Hale, J. S., Lee, J., Nasti, T. H., Sharpe, A. H., Freeman, G. J., Germain, R. N., Nakaya, H. I., Xue, H.-H., and Ahmed, R. (2016). Defining cd8+ t cells that provide the proliferative burst after pd-1 therapy. *Nature*, 537(7620):417–421.

[Jennings et al., 2019] Jennings, E., Kanabar, S., Ono, M., Wraith, D. C., and Bending, D. (2019). Distinct regulation of nr4a receptors by NFAT and ERK signalling during t cell activation.

[Jin et al., 2010] Jin, H.-T., Anderson, A. C., Tan, W. G., West, E. E., Ha, S.-J., Araki, K., Freeman, G. J., Kuchroo, V. K., and Ahmed, R. (2010). Cooperation of tim-3 and PD-1 in CD8 t-cell exhaustion during chronic viral infection. *Proceedings of the National Academy of Sciences*, 107(33):14733–14738.

[Johnston et al., 2014] Johnston, R. J., Comps-Agrar, L., Hackney, J., Yu, X., Huseni, M., Yang, Y., Park, S., Javinal, V., Chiu, H., Irving, B., Eaton, D. L., and Grogan, J. L. (2014). The immunoreceptor TIGIT regulates antitumor and antiviral CD8 + t cell effector function. *Cancer Cell*, 26(6):923–937.

[Joller, 2011] Joller, N. (2011). Cutting edge: Tigit has t cell-intrinsic inhibitory functions. *J. Immunol.*, 186.

[Kakaradov et al., 2017] Kakaradov, B., Arsenio, J., Widjaja, C. E., He, Z., Aigner, S., Metz, P. J., Yu, B., Wehrens, E. J., Lopez, J., Kim, S. H., Zuniga, E. I., Goldrath, A. W., Chang, J. T., and Yeo, G. W. (2017). Early transcriptional and epigenetic regulation of cd8+ t cell differentiation revealed by single-cell rna sequencing. *Nature Immunology*, 18(4):422–432.

[Khan et al., 2019] Khan, O., Giles, J. R., McDonald, S., Manne, S., Ngiow, S. F., Patel, K. P., Werner, M. T., Huang, A. C., Alexander, K. A., Wu, J. E., Attanasio, J., Yan, P., George, S. M., Bengsch, B., Staupe, R. P., Donahue, G., Xu, W., Amaravadi, R. K., Xu, X., Karakousis, G. C., Mitchell, T. C., Schuchter, L. M., Kaye, J., Berger, S. L., and Wherry, E. J. (2019). TOX transcriptionally and epigenetically programs CD8+ t cell exhaustion. *Nature*.

[Klimm et al., 2019] Klimm, F., Toledo, E. M., Monfeuga, T., Zhang, F., Deane, C. M., and Reinert, G. (2019). Functional module detection through integration of single-cell RNA sequencing data with protein–protein interaction networks.

[Kulkarni et al., 2019] Kulkarni, A., Anderson, A. G., Merullo, D. P., and Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology*, 58:129–136.

[Lange et al., 2020] Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H. B., Pe'er, D., and Theis, F. J. (2020). CellRank for directed single-cell fate mapping.

[Love et al., 2014] Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol.*, 15.

[Mackerness et al., 2010] Mackerness, K. J., Cox, M. A., Lilly, L. M., Weaver, C. T., Harrington, L. E., and Zajac, A. J. (2010). Pronounced virus-dependent activation drives exhaustion but sustains IFN-γ transcript levels. *The Journal of Immunology*, 185(6):3643–3651.

[Macnair et al., 2019] Macnair, W., Roditi, L. D. V., Ganscha, S., and Claassen, M. (2019). Tree-ensemble analysis assesses presence of multifurcations in single cell data. *Molecular Systems Biology*, 15(3).

[Manno et al., 2018] Manno, G. L., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lönnerberg, P., Furlan, A., Fan, J., Borm, L. E., Liu, Z., van Bruggen, D., Guo, J., He, X., Barker, R., Sundström, E., Castelo-Branco, G., Cramer, P., Adameyko, I., Linnarsson, S., and Kharchenko, P. V. (2018). RNA velocity of single cells. *Nature*, 560(7719):494–498.

[Martinez et al., 2015] Martinez, G. J., Pereira, R. M., Äijö, T., Kim, E. Y., Marangoni, F., Pipkin, M. E., Togher, S., Heissmeyer, V., Zhang, Y. C., Crotty, S., Lamperti, E. D., Ansel, K. M., Mempel, T. R., Lähdesmäki, H., Hogan, P. G., and Rao, A. (2015). The transcription factor NFAT promotes exhaustion of activated CD8 + t cells. *Immunity*, 42(2):265–278.

[Matloubian et al., 1994] Matloubian, M., Concepcion, R. J., and Ahmed, R. (1994). Cd4+ t cells are required to sustain cd8+ cytotoxic t-cell responses during chronic viral infection. *J. Virol.*, 68.

[Matloubian et al., 1993] Matloubian, M., Kolhekar, S. R., Somasundaram, T., and Ahmed, R. (1993). Molecular determinants of macrophage tropism and viral persistence: importance of single amino acid changes in the polymerase and glycoprotein of lymphocytic choriomeningitis virus. *J. Virol.*, 67.

[McLane et al., 2019] McLane, L. M., Abdel-Hakeem, M. S., and Wherry, E. J. (2019). CD8 t cell exhaustion during chronic viral infection and cancer. *Annual Review of Immunology*, 37(1):457–495.

[Menner, 2015] Menner, A. J. (2015). Id3 controls cell death of 2b4+ virus-specific cd8+ t cells in chronic viral infection. *J. Immunol.*, 195.

[Miller et al., 2019] Miller, B. C., Sen, D. R., Abosy, R. A., Bi, K., Virkud, Y. V., LaFleur, M. W., Yates, K. B., Lako, A., Felt, K., Naik, G. S., Manos, M., Gjini, E., Kuchroo, J. R., Ishizuka, J. J., Collier, J. L., Griffin, G. K., Maleri, S., Comstock, D. E., Weiss, S. A., Brown, F. D., Panda, A., Zimmer, M. D., Manguso, R. T., Hodi, F. S., Rodig, S. J., Sharpe, A. H., and Haining, W. N. (2019). Subsets of exhausted CD8+ t cells differentially mediate tumor control and respond to checkpoint blockade. *Nature Immunology*, 20(3):326–336.

[Moran et al., 2011] Moran, A. E., Holzapfel, K. L., Xing, Y., Cunningham, N. R., Maltzman, J. S., Punt, J., and Hogquist, K. A. (2011). T cell receptor signal strength in tregand iNKT cell development demonstrated by a novel fluorescent reporter mouse. *The Journal of Experimental Medicine*, 208(6):1279–1289.

[Moskophidis et al., 1993] Moskophidis, D., Lechner, F., Pircher, H., and Zinkernagel, R. M. (1993). Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector t cells. *Nature*, 362(6422):758–761.

[Palma and Bodenmiller, 2015] Palma, S. D. and Bodenmiller, B. (2015). Unraveling cell populations in tumors by single-cell mass cytometry. *Current Opinion in Biotechnology*, 31:122–129.

[Parry et al., 2005] Parry, R. V., Chemnitz, J. M., Frauwirth, K. A., Lanfranco, A. R., Braunstein, I., Kobayashi, S. V., Linsley, P. S., Thompson, C. B., and Riley, J. L. (2005). CTLA-4 and PD-1 receptors inhibit t-cell activation by distinct mechanisms. *Molecular and Cellular Biology*, 25(21):9543–9553.

[Patsoukis, 2015] Patsoukis, N. (2015). Pd-1 alters t-cell metabolic reprogramming by inhibiting glycolysis and promoting lipolysis and fatty acid oxidation. *Nat. Commun.*, 6.

[Pircher, 1990] Pircher, H. (1990). Viral escape by selection of cytotoxic t cell-resistant virus variants in vivo. *Nature*, 346.

[Pircher et al., 1989] Pircher, H., Bürki, K., Lang, R., Hengartner, H., and Zinkernagel, R. M. (1989). Tolerance induction in double specific t-cell receptor transgenic mice varies with antigen. *Nature*, 342(6249):559–561.

[Povinelli et al., 2018] Povinelli, B. J., Rodriguez-Meira, A., and Mead, A. J. (2018). Single cell analysis of normal and leukemic hematopoiesis. *Molecular Aspects of Medicine*, 59:85–94.

[Qiu et al., 2020] Qiu, X., Rahimzamani, A., Wang, L., Ren, B., Mao, Q., Durham, T., McFaline-Figueroa, J. L., Saunders, L., Trapnell, C., and Kannan, S. (2020). Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Systems*.

[Quigley, 2010] Quigley, M. (2010). Transcriptional analysis of hiv-specific cd8+ t cells shows that pd-1 inhibits t cell function by upregulating batf. *Nat. Med.*, 16.

[Raju et al., 2020] Raju, S., Xia, Y., Daniel, B., Yost, K. E., Bradshaw, E., Tonc, E., Verbaro, D. J., Satpathy, A. T., and Egawa, T. (2020). Latent plasticity of effector-like exhausted CD8 t cells contributes to memory responses. *BiorXive*.

[Ren et al., 2018] Ren, X., Kang, B., and Zhang, Z. (2018). Understanding tumor ecosystems by single-cell sequencing: promises and limitations. *Genome Biology*, 19(1).

[Richter et al., 2009] Richter, K., Agnellini, P., and Oxenius, A. (2009). On the role of the inhibitory receptor LAG-3 in acute and chronic LCMV infection. *International Immunology*, 22(1):13–23.

[Richter et al., 2010] Richter, K., Agnellini, P., and Oxenius, A. (2010). On the role of the inhibitory receptor lag-3 in acute and chronic lcmv infection. *Int. Immunol.*, 22.

[Rodriques et al., 2020] Rodriques, S. G., Chen, L. M., Liu, S., Zhong, E. D., Scherrer, J. R., Boyden, E. S., and Chen, F. (2020). RNA timestamps identify the age of single molecules in RNA sequencing. *Nature Biotechnology*.

[Ryseck et al., 1989] Ryseck, R. P., Macdonald-Bravo, H., Mattei, M. G., Ruppert, S., and Bravo, R. (1989). Structure, mapping and expression of a growth factor inducible gene encoding a putative nuclear hormonal binding receptor. *EMBO J.*, 8.

[Saelens et al., 2018] Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2018). A comparison of single-cell trajectory inference methods: towards more accurate and robust tools.

[Saelens et al., 2019] Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature Biotechnology*.

[Salvador and Chan, 2007] Salvador, S. and Chan, P. (2007). Fastdtw: Toward accurate dynamic time warping in linear time and space. Armament and Technical Products.

[Sandu et al., 2020a] Sandu, I., Cerletti, D., Claassen, M., and Oxenius, A. (2020a). Exhausted CD8+ t cells exhibit low and strongly inhibited TCR signaling during chronic LCMV infection. *Nature Communications*, 11(1).

[Sandu et al., 2020b] Sandu, I., Cerletti, D., Oetiker, N., Borsa, M., Wagen, F., Spadafora, I., Welten, S. P., Stolz, U., Oxenius, A., and Claassen, M. (2020b). Landscape of exhausted virus-specific CD8 t cells in chronic LCMV infection. *Cell Reports*, 32(8):108078.

[Sen et al., 2016] Sen, D. R., Kaminski, J., Barnitz, R. A., Kurachi, M., Gerdemann, U., Yates, K. B., Tsao, H.-W., Godec, J., LaFleur, M. W., Brown, F. D., Tonnerre, P., Chung, R. T., Tully, D. C., Allen, T. M., Frahm, N., Lauer, G. M., Wherry, E. J., Yosef, N., and Haining, W. N. (2016). The epigenetic landscape of t cell exhaustion. *Science*, 354(6316):1165–1169.

[Seth and Craft, 2019] Seth, A. and Craft, J. (2019). Spatial and functional heterogeneity of follicular helper t cells in autoimmunity. *Current Opinion in Immunology*, 61:1–9.

[Siddiqui et al., 2019] Siddiqui, I., Schaeuble, K., Chennupati, V., Marraco, S. A. F., Calderon-Copete, S., Ferreira, D. P., Carmona, S. J., Scarpellino, L., Gfeller, D., Pradervand, S., Luther, S. A., Speiser, D. E., and Held, W. (2019). Intratumoral tcf1+PD-1+CD8+ t cells with stem-like properties promote tumor control in response to vaccination and checkpoint blockade immunotherapy. *Immunity*, 50(1):195–211.e10.

[Stocco et al., 2002] Stocco, C. O., Lau, L. F., and Gibori, G. (2002). A calcium/calmodulin-dependent activation of erk1/2 mediates jund phosphorylation and induction of nur77 and 20alpha-hsd genes by prostaglandin f2alpha in ovarian cells. *J. Biol. Chem.*, 277.

[Street et al., 2018] Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, 19(1).

[Stubbington et al., 2017] Stubbington, M. J. T., Rozenblatt-Rosen, O., Regev, A., and Teichmann, S. A. (2017). Single-cell transcriptomics to explore the immune system in health and disease. *Science*, 358(6359):58–63.

[Svensson, 2019] Svensson, V. (2019). Droplet scRNA-seq is not zero-inflated.

[Trapnell et al., 2014] Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32(4):381–386.

[Tritschler et al., 2019] Tritschler, S., Büttner, M., Fischer, D. S., Lange, M., Bergen, V., Lickert, H., and Theis, F. J. (2019). Concepts and limitations for learning developmental trajectories from single cell genomics. *Development*, 146(12):dev170506.

[Utzschneider et al., 2016a] Utzschneider, D., Charmoy, M., Chennupati, V., Pousse, L., Ferreira, D., Calderon-Copete, S., Danilo, M., Alfei, F., Hofmann, M., Wieland, D., Pradervand, S., Thimme, R., Zehn, D., and Held, W. (2016a). T cell factor 1-expressing memory-like cd8+ t cells sustain the immune response to chronic viral infections. *Immunity*, 45(2):415–427.

[Utzschneider et al., 2016b] Utzschneider, D. T., Alfei, F., Roelli, P., Barras, D., Chennupati, V., Darbre, S., Delorenzi, M., Pinschewer, D. D., and Zehn, D. (2016b). High antigen levels induce an exhausted phenotype in a chronic infection without impairing t cell expansion and survival. *The Journal of Experimental Medicine*, 213(9):jem.20150598–jem.20150598.

[Utzschneider et al., 2016c] Utzschneider, D. T., Charmoy, M., Chennupati, V., Pousse, L., Ferreira, D. P., Calderon-Copete, S., Danilo, M., Alfei, F., Hofmann, M., Wieland, D., Pradervand, S., Thimme, R., Zehn, D., and Held, W. (2016c). T cell factor 1-expressing memory-like CD8 + t cells sustain the immune response to chronic viral infections. *Immunity*, 45(2):415–427.

[Van Der Maaten and Hinton, 2008] Van Der Maaten, L. J. P. and Hinton, G. E. (2008). Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605.

[Wherry et al., 2003] Wherry, E. J., Blattman, J. N., Murali-Krishna, K., Most, R., and Ahmed, R. (2003). Viral persistence alters cd8 t-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *J. Virol.*, 77.

[Wherry et al., 2007] Wherry, E. J., Ha, S.-J., Kaech, S. M., Haining, W. N., Sarkar, S., Kalia, V., Subramaniam, S., Blattman, J. N., Barber, D. L., and Ahmed, R. (2007). Molecular signature of cd8+ t cell exhaustion during chronic viral infection. *Immunity*, 27(4):670–684.

[Wherry and Kurachi, 2015] Wherry, E. J. and Kurachi, M. (2015). Molecular and cellular insights into t cell exhaustion. *Nature Reviews Immunology*, 15(8):486–499.

[Wolf et al., 2018] Wolf, F. A., Angerer, P., and Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1).

[Wolf et al., 2019] Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, 20(1).

[Wu et al., 2016] Wu, T., Ji, Y., Moseman, E. A., Xu, H. C., Manglani, M., Kirby, M., Anderson, S. M., Handon, R., Kenyon, E., Elkahloun, A., Wu, W., Lang, P. A., Gattinoni, L., McGavern, D. B., and Schwartzberg, P. L. (2016). The tcf1-bcl6 axis counteracts type i interferon to repress exhaustion and maintain t cell stemness. *Science Immunology*, 1(6):54–60.

[Yang et al., 2019] Yang, L., George, J., and Wang, J. (2019). Deep profiling of cellular heterogeneity by emerging single-cell proteomic technologies. *PROTEOMICS*, 20(13):1900226.

[Yao et al., 2019] Yao, C., Sun, H.-W., Lacey, N. E., Ji, Y., Moseman, E. A., Shih, H.-Y., Heuston, E. F., Kirby, M., Anderson, S., Cheng, J., Khan, O., Handon, R., Reilley, J., Fioravanti, J., Hu, J., Gossa, S., Wherry, E. J., Gattinoni, L., McGavern, D. B., O'Shea, J. J., Schwartzberg, P. L., and Wu, T. (2019). Single-cell RNA-seq reveals TOX as a key regulator of CD8+ t cell persistence in chronic infection. *Nature Immunology*, 20(7):890–901.

[Zajac, 1998] Zajac, A. J. (1998). Viral immune evasion due to persistence of activated t cells without effector function. *J. Exp. Med.*, 188.

[Zander et al., 2019] Zander, R., Schauder, D., Xin, G., Nguyen, C., Wu, X., Zajac, A., and Cui, W. (2019). CD4+ t cell help is required for the formation of a cytolytic CD8+ t cell subset that protects against chronic infection and cancer. *Immunity*, 51(6):1028–1042.e4.

[Zhang and Zhang, 2019] Zhang, L. and Zhang, S. (2019). Comparison of computational methods for imputing single-cell RNA-sequencing data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pages 1–1.

[Zheng et al., 2017] Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen,

T. S., Hindson, B. J., and Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1).

[Ziegenhain et al., 2017] Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Molecular Cell*, 65(4):631–643.e4.