Molecular mechanisms of $CD8^+$ T cell differentiation

A dissertation presented

by

Jernej Godec

to

The Division of Medical Sciences

in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Immunology

Harvard University
Cambridge, Massachusetts

March 2016

Dissertation Advisors: Dr. Arlene H. Sharpe Jernej Godec
Dr. W. Nicholas Haining

Molecular mechanisms of CD8$^+$ T cell differentiation

Abstract

CD8$^+$ T cells are a crucial component of the adaptive immune system and are required for optimal protection from most pathogens and cancer. They function by secreting pro-inflammatory cytokines and by directly eliminating infected and malignant cells. In order to be effective, CD8$^+$ T cells must be activated through a complex sequence of steps involving engagement of the antigen-specific T cell receptor (TCR) and other receptors, which orchestrate transcriptional, epigenetic, and metabolic changes that direct the differentiation of the responding cells. Following optimal activation, naive CD8$^+$ T cells rapidly undergo clonal expansion and effector differentiation that enables prompt resolution of infection. Following pathogen clearance, a fraction of effector CD8$^+$ T cells differentiate into long-lived memory CD8$^+$ T cells that provide robust protection from re-challenge with the same microbe. However, in the context of persistent abundance of antigen and inflammation, such as in chronic infections and in cancer, the T cells instead become gradually more dysfunctional – a state known as T cell exhaustion.

The overarching goal of this thesis is to identify the cardinal features and molecular mechanisms associated with three main states in which CD8$^+$ T cells exist: T cell memory, T cell exhaustion, and T cell effector differentiation. I used two complementary approaches to examine CD8$^+$ T cells at the different states *in vivo*. First, I used classical immunology techniques including knockout mice and cellular phenotypic analyses to examine the role of cell surface molecules PD-1 and CD39 on CD8$^+$ T cells in the context of memory and exhaustion, respectively. Secondly, I developed a novel experimental platform that enables

gene perturbation in naive $CD8^+$ T cells *in vivo* during their differentiation. I used this approach to systematically interrogate the transcriptional programming of activated $CD8^+$ T cells and to identify novel regulators of effector differentiation. In a proof of concept study, I used this system to further define how the transcription factor BATF regulates $CD8^+$ T cell activation. Additionally, I used this experimental platform to systematically interrogate the functional role of a set of ~80 transcription factors in $CD8^+$ T cell differentiation, and identified TGIF1 as a novel regulator of this process.

The role of the co-inhibitory receptor PD-1 on $CD8^+$ T was examined in mice using an acute respiratory infection model. PD-1 is a co-inhibitory receptor that is up-regulated on T cells following activation and recruits SHP1/2 phosphatases to directly antagonize signals through the TCR and this way inhibit the activation of T cells. It is down-regulated following the resolution of an acute infection but remains persistently expressed on $CD8^+$ T cells in chronic infections and cancer. As such, PD-1 has been exhaustively studied for its contribution to the functional exhaustion of T cells. However, its role in acute infections remains less defined. We found that this receptor prevents over-activation and over-expansion of $CD8^+$ T cells following initial differentiation, and is crucial for optimal differentiation of effector $CD8^+$ T cells into functional memory cells.

Exhausted $CD8^+$ T cells express several markers distinctive of the state. Some, like PD-1, Tim-3, and Lag-3 are well known. However, genome-wide transcriptional studies identified numerous additional genes that are differentially expressed in the exhausted state. Thus, we hypothesized that additional markers may provide characteristic features of the exhausted cell state and may function in chronic infections. We investigated one such gene – *ENTPD1* – that encodes for CD39. This cell surface molecule is an ectonucleotidase that hydrolyzes extracellular ATP into ADP and AMP, which can be further broken down to immunosuppressive adenosine by CD73. In the context of the immune system, CD39 has largely been studied on $CD4^+$ regulatory T cells, which use CD39 as a mechanism to

suppress immune responses. However, surprisingly, we found that $CD8^+$ T cells can also express CD39, but its expression is largely restricted to terminally exhausted $CD8^+$ T cells. These cells are most dysfunctional as measured by the most limited proliferative capacity and ability to produce pro-inflammatory cytokines. We have observed this biology in both human and mouse chronic viral infections. Additional studies further demonstrated the importance of CD39 and the purinergic pathway in regulating lethal immunopathology associated with chronic LCMV infection in mice.

In addition to interrogating memory and exhaustion fates of $CD8^+$ T cells, we also examined the initial regulatory programs involved in $CD8^+$ T cell differentiation *in vivo* through gene silencing. Gene perturbation in naive T cells without prior cellular stimulation has been a continuous challenge in the field. To circumvent this limitation, we engineered a novel experimental platform that enables inducible gene knock-down in any immune cell in mice *in vivo* without prior manipulation of these cells. Initially, I validated this system by knocking down BATF and confirmed its essential role in $CD8^+$ T cell responses to acute LCMV infection. Additionally, leveraging the inducible nature of the platform, I showed that BATF functions in the early stages of T cell activation but becomes dispensable once its transcriptional program is initiated.

Several other transcription factors such as T-bet, Eomes, Bcl6, and Blimp-1 have been described to regulate $CD8^+$ T cell differentiation. However, numerous additional transcription factors may function in this process based on their rapid up-regulation following T cell activation. I used the novel platform to systematically test the functional relevance of ~80 additional transcription factors in a pooled setting. These experiments identified several novel candidate regulators of this process. We validated one such gene – *Tgif1* – to confirm its role in the effector $CD8^+$ T cell differentiation following acute LCMV infection and provide clues to the potential mechanism in which it may function.

The above projects have benefited significantly from genome-wide transcriptional datasets of cells at various states or of different genotypes that we generated or that originate from published studies. One particularly powerful approach to examine differences between different groups is gene set enrichment analysis (GSEA) that examines coordinate up- or down-regulation of sets of genes rather than assessing differential expression of specific genes. This is particularly important because changes in biological processes are often guided by relative small changes of groups of genes that act in concert rather than by a robust expression change of a single gene. This approach, however, is only informative if a relevant gene-set collection is used to analyze the data. Existing collections are largely centered around cancer biology and general biological processes but no extensive gene-set collection existed that contained information describing immune processes. Thus, we created ImmuneSigDB – the largest collection of immunology-focused gene sets to date by identifying, annotating, and reanalyzing ~400 published immunology studies. To show its broad use, we used it to examine the cross-species conservation of transcriptional responses in the immune system. We focused on analyzing transcriptional data from systemic responses to sepsis using GSEA and a novel approach, called leading edge metagene analysis. Using these approaches, we uncovered shared and species-specific biology in mouse and human transcriptional responses to sepsis.

Deciphering $CD8^+$ T cell biology is key for conceptualizing new medical interventions that may boost their activation, memory development, and rejuvenation from functional exhaustion. We have determined that PD-1 is essential for optimal $CD8^+$ T cell memory responses, and that BATF is a key transcription factor initiating effector T cell transcriptional programming. We also identified CD39 as a new marker of terminally exhausted $CD8^+$ T cells and uncovered a key role for purinergic signaling in regulating lethal immunopathology in LCMV Clone 13 infection in mice. Furthermore, we developed a new experimental platform that enables systematic interrogation of gene function in any hematopoietic cell type by

inducible knock-down of genes and identified TGIF1 as a novel negative regulator of $CD8^+$ T cell responses. We have also developed a new computational resource to improve analyses of transcriptional profiles in the immune system. Together, the body of work presented in this thesis advances our knowledge of major states of $CD8^+$ T cell biology, uncovering both novel mechanisms underlying $CD8^+$ T cell function, as well as highlighting potential novel therapeutic targets that may be transformative in creating better vaccines, treating infections, or fighting cancer.

One school is finished, and the time has come for another to begin.

– Richard Bach, *Jonathan Livingston Seagull*

Acknowledgements

I am deeply appreciative to the many individuals who have supported me through my PhD with their encouragement, thoughtful feedback, and patience. Without them, I could not have succeeded in coming this far.

Above all, I would like to thank my PhD advisors – Dr. W. Nicholas Haining and Dr. Arlene H. Sharpe – for their continuous support and inspiration. Their enthusiasm, patience, immense knowledge, and devotion to mentorship were essential for my success. Despite their busy schedules, Nick and Arlene would always make themselves available to discuss not only the scientific direction, but also my personal growth. Their mentoring stroke the perfect balance in providing direction while enabling scientific freedom to pursue my (sometimes wild and crazy) ideas, which truly enabled my development as a scientist. The environment they have created in their laboratories is truly unique as it fosters pursuit of collaborative science, thus resulting in a simultaneously productive and interesting culture. I cannot imagine having better mentors or culture during my PhD studies.

I would also like to thank the extraordinary friends and colleagues in both the Haining and Sharpe laboratories for their continuous assistance; whether thinking critically about my projects and providing advice, assisting with the experiments in the most selfless manner, or helping enjoy the down time to make these past years truly memorable.

I would also like to thank my dissertation advisory committee consisting of Dr. Ulrich von Andrian, Dr. Nir Hacohen, Dr. Shiv Pilai, and Dr. Glenn Dranoff for their critical insights and advice that provided additional guidance on my research projects.

My parents, Valerija and Matjaž Godec, have instilled in me the genuine curiosity about the natural world surrounding us, and the utmost value of education. My current scholastic achievement was only attainable because of their dedication, whether by driving early morning to swimming practice and school, or just standing by my side for all my life's decisions. My sister, Špela Godec, has always been my biggest fan that coached me through any adversities and celebrated any accomplishments that arose. Without her, my journey would not have been the same. I will forever be indebted to my family for their continuous unconditional love, support, and encouragement.

Lastly, I would like to thank my best friend and my life partner – my wife, Dr. Oluwatobi Ogbechie-Godec – for standing by me both through late nights in lab and enjoying life to the fullest outside of research. I am grateful her for support throughout this path – for celebrating the successes and for providing comfort in the failures.

Table of Contents

Chapter 1. Introduction

Optimal protection of an organism from infections relies on the adaptive immune system, comprised of T and B lymphocytes. T lymphocytes are divided into two main subsets based on their function and expression of surface molecules CD4 and CD8. $CD4^+$ T helper cells produce cytokines that can enhance and regulate appropriate immune responses. $CD8^+$ T cells, also known as cytotoxic T lymphocytes (CTL) directly kill aberrant cells to eliminate danger and can produce cytokines that stimulate other arms of adaptive as well as innate immune system (1-3). Both subsets are generated in the thymus where the developing thymocytes undergo genetic rearrangement of T cell receptors (TCR) to create a unique TCR on each mature naive cell as it enters the periphery.

T cells are activated by a series of signals that together induce appropriate transcriptional reprogramming of naive T cells as well as inducing metabolic and epigenetic changes that together differentiate the responding cells into effector T cells (4-7). Binding of the TCR to the cognate peptide presented on appropriate major histocompatibility complex (MHC) molecules is the key signal that initiates the activation. Simultaneous engagement of the co-stimulatory surface receptor CD28 by B7-1/B7-2 (CD80/CD86) is the key checkpoint for T cell activation to proceed (8, 9). Concurrently, the T cells can be modulated by signaling through additional co-inhibitory and co-stimulating receptors that, together with cytokine receptor signaling, further direct the fate of the cell (9-20).

Most commonly, the life cycle of a T cell is divided into three main stages: clonal expansion, contraction of the T cell pool, and the formation of long-lived T cell memory (6, 21-25). Immediately following stimulation in the secondary lymphoid organs by dendritic cells, T cells proliferate, differentiate into functional effector cells, and migrate to the affected tissues where they can eliminate the pathogen (1, 2, 26-29). Following pathogen clearance, the majority of the responding T cell clones is no longer needed and is thus eliminated. A minor subset of the effector cells, however, differentiates into a long-lived self-renewing pool

of memory T cells (6, 30-32). Even though a brief encounter with an antigen is sufficient for inducing proliferation and even limited memory differentiation (33-35), it is thought that the optimal differentiation of T cells into effector and memory $CD8^+$ T cells is much more complex and includes signals through co-stimulatory receptors, cytokines, and other factors (6, 25, 36-38). The responding pool of effector cells, however, is heterogeneous, consisting of at least two distinct sub-populations that have been described to have unique characteristics and fates (39-42). Both of these are functional effector cells that degranulate to eliminate infected cells, but they have distinct developmental outcome (39-44). These subsets are referred to as short-lived memory effector cells (SLEC), which proliferate extensively in response to a stimulus but largely contract shortly following resolution of infection, and memory progenitor effector cells (MPEC) that expand less initially but have the capacity to generate the pool of long-lived memory cells (39, 40). These cells can be distinguished by their expression of cell surface molecules KLRG1, CD127 (IL-7Rα), and CD25 (IL-2Rα): the SLEC cells are $KLRG1^{high}$ and $CD25^{high}$ while the MPECs express high levels of CD127 (39-42, 45-47). Memory T cells can be maintained in the body for the life span of an organism and rapidly respond to clear the same pathogen upon re-challenge.

The fate of responding cells is altered in the context of persistent antigen exposure and inflammation, such as in the case of chronic viral infection and cancer. Persistent antigen exposure alters T cell contraction and memory formation, leading cells instead into a state of dysfunction known as exhaustion (21). This dysfunction is associated with loss of proliferative capacity, ability to produce effector cytokines such as IFN-γ and TNF-α, as well as other effector functions. Exhausted T cells are also characterized by persistent expression of inhibitory receptors such as PD-1, TIM-3, LAG-3 and others, and it is thought that continuous engagement of these receptors largely contributes to the formation and maintenance of exhaustion state (21, 48). During the course of chronic infection, T cells become progressively more exhausted as they increasingly lose their ability to proliferate and

acquire the expression of more co-inhibitory receptors. Two separate subsets of exhausted $CD8^+$ T cells have been described and can be distinguished based on their level of expression of PD-1 and transcription factors T-bet and Eomesodermin. While T-bet^{high} PD-1^{int} progenitor exhausted T cells maintain some ability to proliferate and produce cytokines, the $Eomes^{high}$ PD-1^{high} terminally exhausted cells express highest levels of co-inhibitory function and while they maintain some cytotoxic effect, they lose the ability to proliferate and produce pro-inflammatory cytokines (49, 50). While exhaustion is likely to be a product of evolution to protect the infected host from overt immunopathology, the rejuvenation of exhausted T cells through PD-1 pathway blockade shows to be effective in combating certain chronic infections and cancer (51-54).

The PD-1:PD-L co-inhibitory pathway regulates dysfunctional T cells during chronic viral infection and cancer, but the role of this pathway in effector and memory responses following acute infection or vaccination remains less clear (55). PD-1 pathway is described in more detail in Chapter 2. We examined the role of the PD-1 pathway in regulating responses of $CD8^+$ T cells during a respiratory infection and demonstrated that inhibitory signals from the PD-1:PD-L pathway are needed for optimal generation and function of memory $CD8^+$ T cells following an acute viral infection (Chapter 2). Mice deficient in the PD-1 pathway (PD-1 KO or PD-L1/L2 DKO) exhibit impaired $CD8^+$ T cell memory following influenza infection, including reduced virus-specific memory $CD8^+$ T cell numbers and compromised recall responses. PD-1 deficiency in $CD8^+$ T cells results in cell intrinsic alterations during initial $CD8^+$ T cell priming that leads to excessive early $CD8^+$ T cell expansion, but increased $CD8^+$ T cell contraction and aberrant effector to memory $CD8^+$ T cell transition. Overall, our studies reveal a critical and previously unappreciated role for PD-1 as an integrator of early $CD8^+$ T cell activation signals that promote optimal $CD8^+$ T cell memory formation and durability. This novel PD-1 function has therapeutic implications for the generation of T cell memory during PD-1 cancer immunotherapy and modulation of the PD-1 pathway to enhance immune

memory following acute infection or prophylactic vaccination.

PD-1 has been extensively characterized in T cell exhaustion, but there is an increasing number of molecules that are becoming appreciated to function on T cells during persistent infections in cancer. While cancer therapies targeting PD-1 have had transformative effects on disease management, many patients remain who do not respond or respond poorly. Thus, defining novel markers of exhaustion is important both for identifying and potentially reversing T cell exhaustion, by targeting other molecules in addition to or in lieu of PD-1. Herein, we show a new cell surface molecule, CD39, to be a marker of exhausted $CD8^+$ T cells in chronic viral infection (Chapter 3). $CD8^+$ T cells specific for hepatitis C virus (HCV) or human immunodeficiency virus (HIV) in infected patients express high levels of CD39, but those specific for latent Epstein-Barr Virus (EBV) and cytomegalovirus (CMV) do not. CD39 is an ectonucleotidase that hydrolyzes extracellular ATP into ADP and AMP. A secondary enzyme, CD73, further breaks down AMP to free adenosine that has been shown to suppress the immune system (56-60). As such, CD39 has largely been studied on regulatory T cells in the context of the immune system, since production of free adenosine is one of the immunoregulatory mechanisms they employ (61, 62). Thus, it is perhaps surprising that CD39 is also expressed by $CD8^+$ T cells, as previous studies have shown that it is not expressed on $CD8^+$ T cells or other subsets of T cells in healthy patients such as memory and naive cells (63-66). CD39 expressed by $CD8^+$ T cells in chronic infection is enzymatically active, co-expressed with PD-1, marks cells with a transcriptional signature of T cell exhaustion and correlates with viral load in both HIV and HCV. We also examined this phenomenon in the mouse model of Lymphocytic Choriomeningitis Virus (LCMV) using the acute Armstrong strain that is cleared within 8-10 days, and the chronic Clone 13 strain that can persist for several months (67-69). In chronic LCMV infection, virus-specific $CD8^+$ T cells contain a population of $CD39^{high}$ $CD8^+$ T cells that is absent in functional memory cells elicited by acute infection. This $CD39^{high}$ $CD8^+$ T cell

population is enriched for cells with the phenotypic and functional profile of terminal exhaustion in which T cells further lose proliferative capacity, the ability to produce IFN-γ and TNF-α and are marked by highest expression of PD-1 and the transcription factor Eomes (49, 50). Furthermore, CD39-deficient animals succumb to Clone 13 infection due to overt immunopathology, similar to mice lacking other immunosuppressive pathways such as PD-1 or PD-L1 (48). These findings provide a new marker of T cell exhaustion, and implicate the purinergic pathway in the regulation of chronic infections.

While $CD8^+$ T cells memory and exhaustion are a long-term outcomes to acute or chronic infection, respectively, immediately following initial stimulation the $CD8^+$ T cells undergo effector $CD8^+$ T cell stimulation accompanied by proliferation with the goal of prompt resolution of infection. This differentiation of effector $CD8^+$ T cells is a critical step for the development of protective responses to pathogens and for effective vaccines. In the first few hours after activation, naive $CD8^+$ T cells initiate a transcriptional program that leads to the formation of effector and memory T cells, but the regulation of this process is poorly understood. Investigating the role of specific transcription factors (TFs) in determining $CD8^+$ effector T cell fate by gene knockdown with RNAi is challenging because T cells are refractory to transduction with viral vectors without extensive *ex vivo* stimulation, obscuring the earliest events in effector differentiation. To overcome this obstacle, we developed a novel strategy to test the function of genes in naive $CD8^+$ T cells *in vivo* by creating bone marrow chimera from hematopoietic progenitors transduced with an inducible shRNA construct (Chapter 4). Following hematopoietic reconstitution, this strategy allowed inducible *in vivo* gene knockdown in any cell type that developed from this transduced progenitor pool. We showed that lentivirus-transduced progenitor cells reconstituted normal hematopoiesis and allowed the development of naive $CD8^+$ T cells that were indistinguishable from wild-type naive T cells. This approach allowed efficient gene knock-down to be induced *in vivo* without subsequent manipulation. We applied this strategy to study the effect of the transcription

factor Basic Leucine Zipper Transcription Factor, ATF-Like (BATF) in $CD8^+$ T effector T cell differentiation. We show that BATF is essential for initial commitment of naive $CD8^+$ T cells to effector differentiation but becomes dispensable by 72h. This approach allows the study of gene function *in vivo* in unperturbed cells of hematopoietic origin that are refractory to viral transduction.

Additionally, we used this system to systematically test the functional role of ~80 transcription factors in $CD8^+$ effector T cell activation and proliferation through a targeted *in vivo* RNAi screen. We identified many known and several potentially novel genes functioning in this process. Through validation studies, we focused on a potentially novel negative regulator of T cell activation, *Tgif1* (Chapter 5). While this gene has not yet been studied in immune cells, previous work in other cells suggest it may have a role downstream of TGF-β and retinoic acid (70-72). Knock down of TGIF1 in $CD8^+$ T cells responding to LCMV infection *in vivo* results in significantly increased accumulation of these cells and skews the effector differentiation toward memory progenitor effector cells (MPEC). Similar results were observed in mice that selectively lack TGIF1 in peripheral $CD8^+$ T cells. Our data hint at a potential mechanism for this effect, suggesting that TGIF1 may be able to directly regulate Blimp-1, a transcription factor well appreciated for driving terminal effector $CD8^+$ T cells differentiation (73, 74). Additionally, these TGIF1 deficient mice exhibit enhanced anti-tumoral immunity, consistent with the hypothesis that TGIF1 dampens $CD8^+$ T cell responses.

Transcription factors included in the targeted *in vivo* RNAi screen in $CD8^+$ T cells were prioritized by analyzing previously published publicly accessible gene expression profiling datasets from mouse $CD8^+$ T cells responding to infection (39, 40, 75, 76). Gene expression profiling has become a mainstay in immunology, but subtle changes in gene networks related to biological processes are hard to discern when comparing various

datasets (77, 78). For instance, conservation of the transcriptional response to sepsis in mouse models and human disease remains controversial (79, 80). To improve transcriptional analysis in immunology, we created ImmuneSigDB: a manually annotated compendium of ~5,000 gene-sets from diverse cell states, experimental manipulations and genetic perturbations in immunology (Chapter 6). Analysis using ImmuneSigDB identified signatures induced in activated myeloid cells and differentiating lymphocytes that were highly conserved between humans and mice. Sepsis triggered conserved patterns of gene expression in humans and mouse models. However, we also identified species-specific biological processes in the sepsis transcriptional response: while both species up-regulated phagocytosis-related genes, a mitosis signature was specific to humans. ImmuneSigDB enables granular analysis of transcriptome data to improve biological understanding of immune processes in humans and mice.

Together, the data presented in this thesis provide new insights in the biology of $CD8^+$ T cells, establish a novel experimental platform and a new resource for analyzing immune transcriptome datasets, and uncover conserved and distinct transcriptional patterns in mouse and human inflammatory responses.

Chapter 2: PD-1 pathway regulates development and function of memory $CD8^+$ T cells following respiratory viral infection

Parts of this chapter are currently under revision publication as:

Jernej Godec*, Pamela M. Odorizzi*, Keturah E. Brown, Kristen E. Pauken, Kathleen Yates, Shannon M. Grande, Loise M. Francisco, Mohammed-Alkhatim Ali, Sabrina Imam, W. Nicholas Haining, E. John Wherry, Arlene H. Sharpe. The PD-1 pathway regulates development and function of memory $CD8^+$ T cells following respiratory viral infection.

*Co-first author.

Introduction

The development of effector and memory $CD8^+$ T cells requires coordinated signals from antigen-TCR (signal 1), co-stimulation (signal 2), and inflammation (signal 3) (10, 11). The quantity and quality of signals 1, 2, and 3 can impact $CD8^+$ T cell activation, but exactly how such signals regulate memory $CD8^+$ T cell differentiation remains poorly understood. Signal 2 encompasses a large group of co-stimulatory and co-inhibitory pathways. In general, co-stimulatory signals such as CD28 and ICOS can augment survival and function, enhance metabolic activity of effector T cells and promote sustained responses (9, 81). Conversely, co-inhibitory receptors such as CTLA-4 and PD-1 dampen these positive signals. Signal 2 has come under intense focus given the recent application of antibodies blocking co-inhibitory receptors for the treatment of cancer and chronic infections (48, 51, 53, 54). Disrupting the PD-1:PD-L pathway during chronic infections and cancer can enhance T cell activity and improve outcomes (48, 53, 54), and has shown 30–50% response rates in melanoma and other cancers (82, 83). Given these clinical successes and increasing use of PD-1 pathway blockade, a better understanding of how disruption of this pathway affects

immune responses is warranted. A major unanswered question is how this pathway regulates the formation and/or differentiation of $CD8^+$ memory T cells, particularly important considering susceptibility of cancer patients to acute infections.

There are currently conflicting data about the role of the PD-1 pathway in effector and memory $CD8^+$ T cell differentiation during acute infection. Lack of PD-1:PD-L signals during primary infection with some strains of LCMV results in more robust effector T cell responses (48, 84), and can enhance $CD8^+$ T cell memory and/or skew T cells toward central memory (85). In addition, the secondary expansion of unhelped memory $CD8^+$ T cells is increased by PD-1 blockade (86). Other studies, in contrast, reported that loss of the PD-1 pathway during acute infection can diminish T cell responses (87-90). Consequently, the role of the PD-1 pathway in the development of effector and memory $CD8^+$ T cells following acute infections remains incompletely understood.

The PD-1:PD-L pathway is particularly important is in the respiratory tract. While human anti-PD-1 cancer immunotherapy has been generally well tolerated, several reports of pneumonitis have raised concerns about respiratory complications following PD-1 pathway blockade (53, 54). During acute infection in the lung, PD-1:PD-L interactions may inhibit $CD8^+$ T cell function, which can limit tissue damage but also impair optimal pathogen control (91). Indeed, antibody blockade of PD-L1 during HMPV and/or genetic deletion of PD-1 during influenza infection enhanced $CD8^+$ T cell functionality early in the immune response (91). Moreover, blockade of PD-L1 during HMPV re-challenge improved secondary $CD8^+$ T cell response and viral clearance (91). However, a better understanding of the consequences of PD-1 signals during acute respiratory infections is needed to determine how PD-1 controls the balance between immunopathology and generation of long-term $CD8^+$ T cell memory.

To address these issues we interrogated the role of the PD-1 pathway in effector and memory $CD8^+$ T cell differentiation during influenza virus infection in mice lacking PD-1 (PD-1 KO) or both ligands, PD-L1 (B7-H1) and PD-L2 (B7-DC) (PD-L1/L2 DKO). Lack of PD-1:PD-

L signals led to compromised $CD8^+$ T cell memory, including reduced memory $CD8^+$ T cell numbers and impaired secondary responses. While some effects may be due to altered inflammation and/or viral control in the complete absence of PD-1 pathway, there were major cell-intrinsic alterations in $CD8^+$ T cell memory, as PD-1 KO $CD8^+$ T cells adoptively transferred into WT mice had similar defects. These data demonstrate a crucial and previously unappreciated role for PD-1 in tempering the strength of initial activation to promote optimal $CD8^+$ T cell memory formation, and have implications for clinical use of PD-1/PD-L1 blockade.

Results

CD8$^+$ T cell memory is impaired in mice lacking PD-1 signaling

To investigate the impact of PD-1 pathway deficiency on CD8$^+$ T cell memory development following acute respiratory viral infection, WT, PD-1 KO or PD-L1/L2 DKO mice were infected with H3N2 influenza virus X31-GP33 (92) and virus-specific CD8$^+$ T cells were examined 60-85 days later. The frequency and number of virus-specific memory CD8$^+$ T cells were reduced in the lungs of PD-L1/L2 DKO and PD-1 KO mice (Figure 2.1, A – D). While true for all responses examined, the magnitude of the difference was greater for subdominant CD8$^+$ T cell responses ($PA_{224-233}$-specific) than for immunodominant ($NP_{366-374}$-specific) responses. A similar decrease in CD8$^+$ T cell responses was observed in the spleen, mediastinal (lung-draining) lymph node (dLN), and bone marrow (Figure 2.2, A – C). In addition, fewer memory CD8$^+$ T cells from the PD-L1/L2 DKO mice produced IFN-γ and TNF-α (Figure 2.2D).

We also examined the effects of PD-1 deficiency on a systemic infection using the acute lymphocytic choriomeningitis virus (LCMV) Armstrong infection. We observed a similar defect in the relative abundance of memory CD8$^+$ T cells in the absence of PD-1 signaling suggesting that this effect was not restricted to the respiratory infection with influenza but may have broader implications to other infections (Figure 2.3, A and B).

Decreased CD8$^+$ T cell memory in mice lacking PD-1 signaling results in poor resolution of secondary infection

To test whether PD-1:PD-L interactions during primary viral infection affected secondary responses, we re-challenged X31-GP33 immune WT, PD-1 KO and PD-L1/L2 DKO mice with the heterotypic H1N1 strain of influenza virus PR8-GP33 (93). Since neutralizing antibodies do not cross-react between these viruses, memory CD8$^+$ T cells play a central role in recall responses and protective immunity in this setting (93-95). PD-1 KO

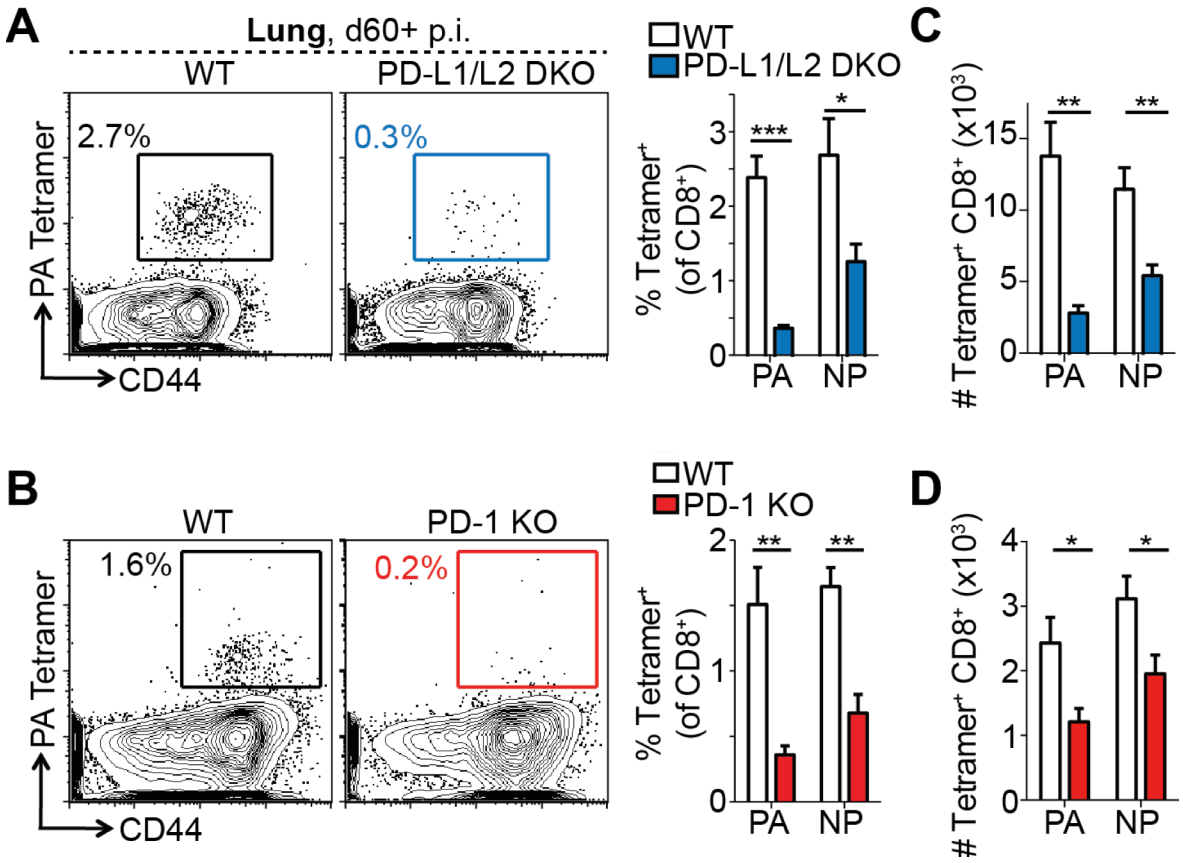

Figure 2.1. CD8$^+$ T cell memory is impaired in PD-1 pathway-deficient mice. (A, B) Representative plots (left) show the frequency of PA_{224}^+ CD8$^+$ T cells in the lungs of X31-GP33 infected WT or PD-L1/L2 DKO **(A)** and WT or PD-1 KO mice **(B)** on d60+ p.i. Numbers indicate percent of PA_{224} (PA) and NP_{366} (NP) tetramer$^+$ CD8$^+$ T cells. **(C, D)** Summary of numbers of memory CD8$^+$ T cells in the lung gated as in **(**A, B**)**. Data are representative of 3-5 independent experiments with 4-5 mice per experiment. Significance was assessed using Student's t-test; **P* < 0.05, ***P* < 0.01, *** *P* < 0.001.

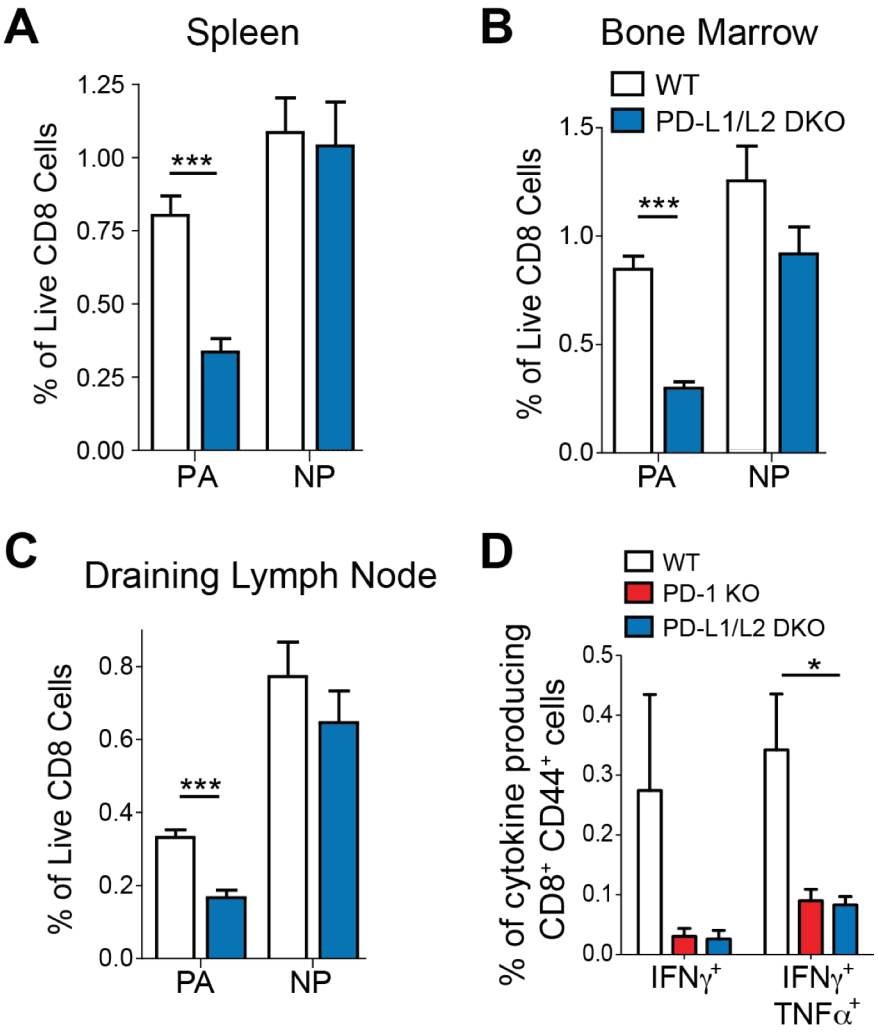

Figure 2.2. CD8$^+$ T cell memory is impaired in PD-1 pathway-deficient mice. (A-C) Summary of frequencies of WT and PD-L1/2 DKO memory CD8$^+$ T cells specific for influenza epitopes in the spleen **(A)**, bone marrow **(B)** and lung draining LN **(C)** at d35+ p.i. with X31-GP33. **(D)** Summary of frequencies of cytokine-producing CD8$^+$ CD44$^+$ T cells from the lung at d60+ p.i. stimulated *ex vivo* with NP_{366}. All data are representative of 3-5 independent experiments with 4-5 mice per experiment. Significance was assessed using Student's t-test; *$P < 0.05$, *** $P < 0.001$.
Parts of this figure were generated by Keturah Brown.

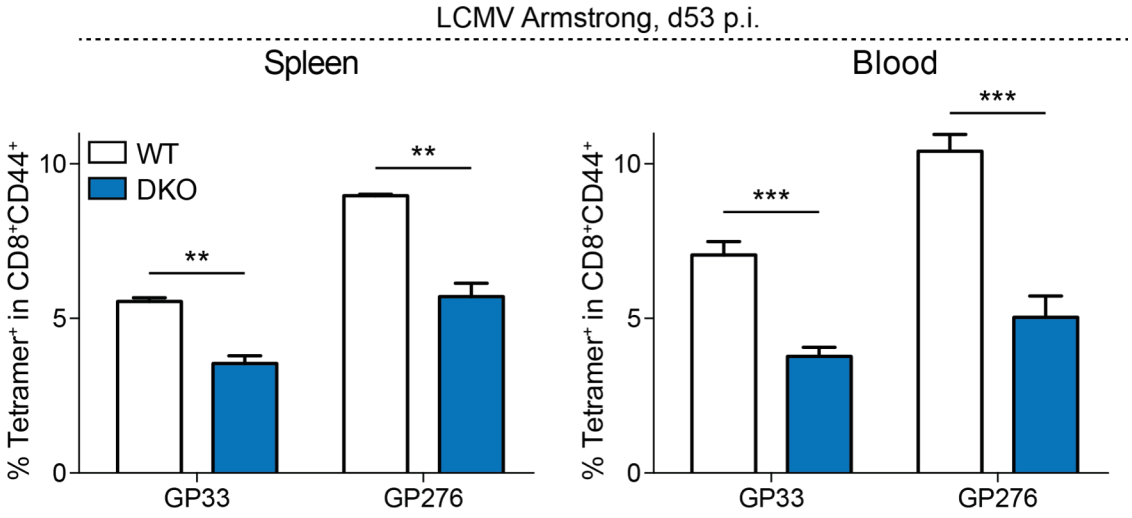

Figure 2.3. **PD-L1/L2-DKO mice have decreased relative abundance of virus-specific memory CD8$^+$ T cells following systemic LCMV-Armstrong infection.** Fraction of GP_{33-41} (GP33) or $GP_{276-286}$ (GP276) tetramer positive CD8$^+$ T cells in the spleen (left) and peripheral blood (right) at d53 following primary intraperitoneal LCMV-Armstrong infection.

and PD-L1/L2 DKO mice had significantly increased viral titers in the lung compared to WT mice 3.5 days (d3.5) following re-challenge (Figure 2.4A) and lost more weight (Figure 2.4B). Additionally, lower frequencies and total numbers of tetramer$^+$ as well as cytokine-producing CD8$^+$ T cells were observed in the lung at d3.5 post-rechallenge in PD-L1/L2 DKO mice (Figure 2.4, C – E). By d7 post re-challenge the frequencies, numbers, and function of virus-specific CD8$^+$ T cells in the lungs were comparable in WT and PD-L1/L2 DKO mice (Figure 2.5, A – C). Thus, PD-1 pathway deficiency results in compromised memory CD8$^+$ T cells that failed to mount optimal recall responses. This defect in memory CD8$^+$ T cell responses in the absence of PD-1 signals manifests as a delay in the kinetics of viral clearance (Figure 2.5D). Taken together, these data indicate a key role for the PD-1 pathway in regulating the development of optimally functional CD8$^+$ T cell memory.

CD8$^+$ memory T cell defect in absence of PD-1 signaling is not due to aberrant mTOR signaling

While PD-1 is largely appreciated for directly antagonizing TCR-proximal signaling, it has also been shown to increase the expression of phosphatase PTEN, which can facilitate PIP_3 degradation to antagonize PI3K/mTOR function (96, 97). Importantly, signaling through mTOR complex 1 (mTORC1) has been previously established to antagonize CD8$^+$ T cell memory differentiation in a cell-intrinsic way (98). Additionally, mTOR inhibition using the small molecule antagonist Rapamycin reduced this antagonism and resulted in increased CD8$^+$ T cell memory (98). To examine whether the memory CD8$^+$ T cell defect observed in the absence of PD-1 signaling was due to aberrant mTORC1/2 signaling, we analyzed whether mTOR inhibition with Rapamycin would rescue the memory defect in PD-L1/L2-DKO mice. Consistent with previous reports, daily administration of Rapamycin for 35 days resulted in increased relative abundance of virus-specific memory CD8$^+$ T cells in WT mice (Figure 2.6). However, Rapamycin did not significantly affect the virus-specific memory CD8$^+$

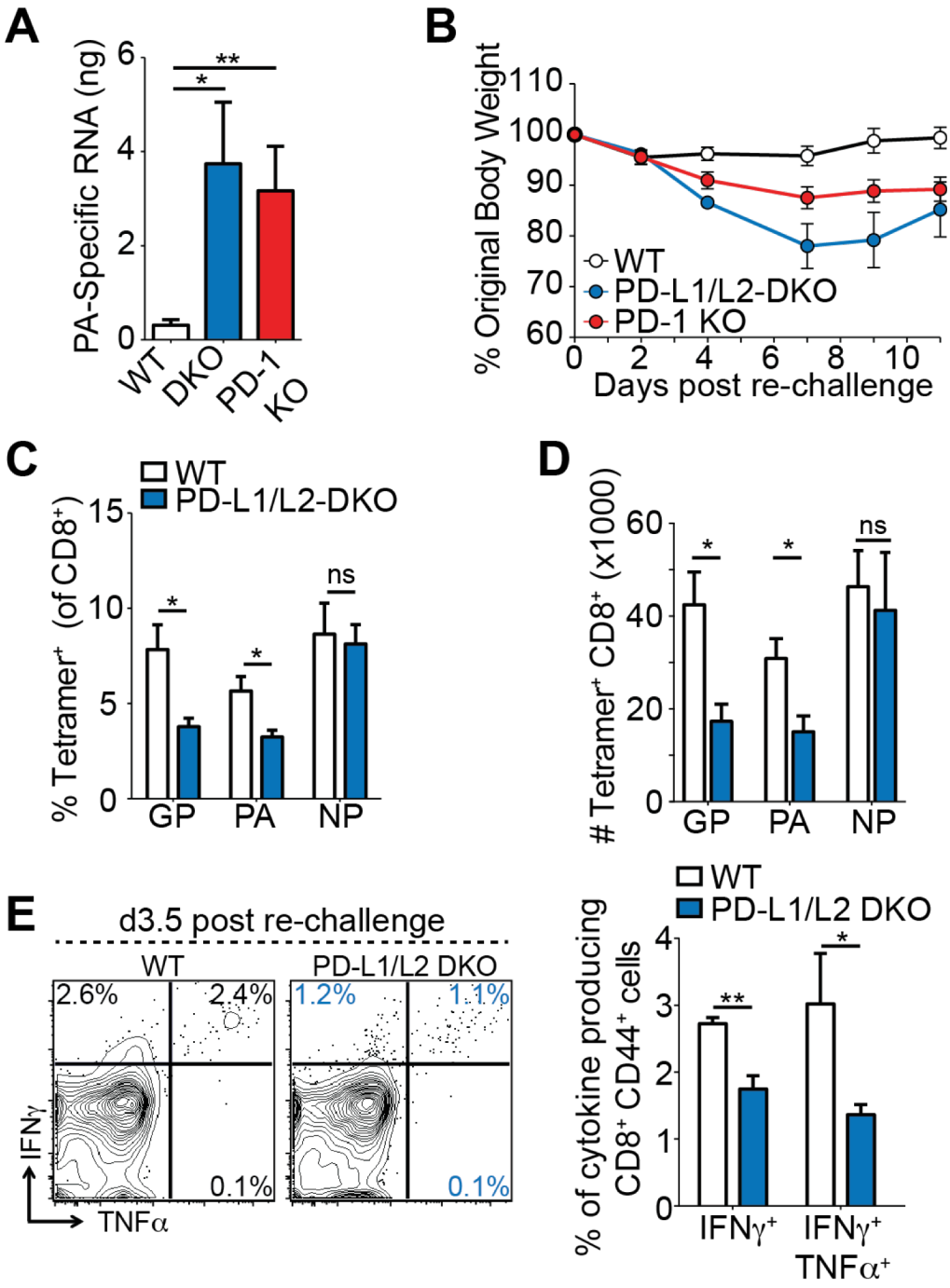

Figure 2.4. CD8$^+$ T cell memory is impaired in PD-1 pathway-deficient mice. (A) Influenza viral titers in the lung at d3.5 p.i. following re-challenge of X31-GP33 immune WT and PD-L1/L2 DKO mice with PR8-GP33. (**B**) Weight loss in WT and PD-L1/L2 DKO immune mice following re-challenge with PR8-GP33. **(C**, **D)** Frequencies (C) and number (D) of tetramer$^+$ CD8$^+$ T cells in lungs of X31-GP33 immune WT and PD-L1/L2 DKO mice at d3.5 following re-challenge. **(E)** Representative plots of intracellular cytokine staining for IFN-γ and TNF-α production in cells from **(C)** stimulated *ex vivo* with NP_{366} (E, left) and summary of the percentages of responding cells (E, right). Data are representative of 3-5 independent experiments with 4-5 mice per experiment. Significance was assessed using Student's t-test; **P* < 0.05, ***P* < 0.01.

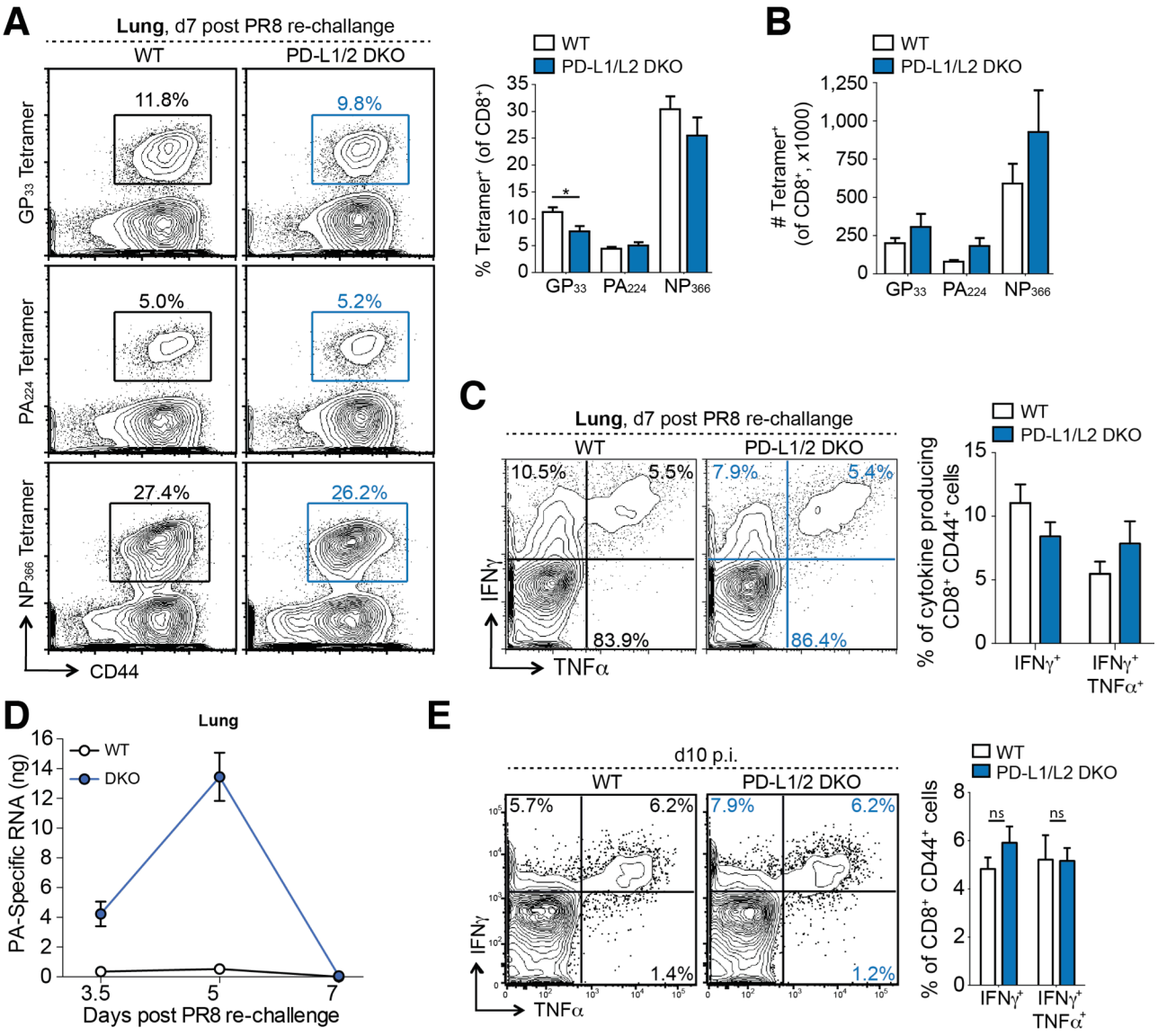

Figure 2.5. Influenza-specific CD8$^+$ T cell numbers normalize at the height of secondary infection, but functionality remains impaired in PD-1 pathway-deficient mice. (A) Representative plots (left) show the frequency of GP_{33}^+ (upper), PA_{224}^+ (middle) and NP_{366}^+ (lower) CD8$^+$ T cells in the lungs of X31-GP33 immune WT and PD-L1/L2 DKO mice on d7 post-rechallenge with PR8-GP33. Summary of frequencies of CD8$^+$ T cells specific for influenza epitopes is shown to the right. **(B)** Summary of numbers of WT and PD-L1/L2 DKO CD8$^+$ T cells specific for influenza epitopes in the lung on d7 post-rechallenge with PR8-GP33. **(C)** Representative plots of ICS for IFN-γ and TNF-α production in lung CD8$^+$ T cells from **(A)** stimulated *ex vivo* with NP_{366}. Fraction of responding cells summarized to the right. (**D**) Kinetics of viral clearance in WT and PD-L1/L2 DKO mice following secondary infection with PR8-GP33 at d35 following primary X31-GP33 infection. **(E)** Representative plots of intracellular cytokine staining for IFN-γ and TNF-α production in lung CD8$^+$ T cells on d10 p.i. following *ex vivo* stimulation with NP_{366}. Representative plots gated on CD8$^+$ CD44$^+$ T cells (left) and summary of percent responding cells (right) from 5 mice per group. Data are representative of 2-4 independent experiments with 4-5 mice per experiment. Significance was assessed using Student's t-test; $*P < 0.05$.

Parts of this figure were generated by Keturah Brown.

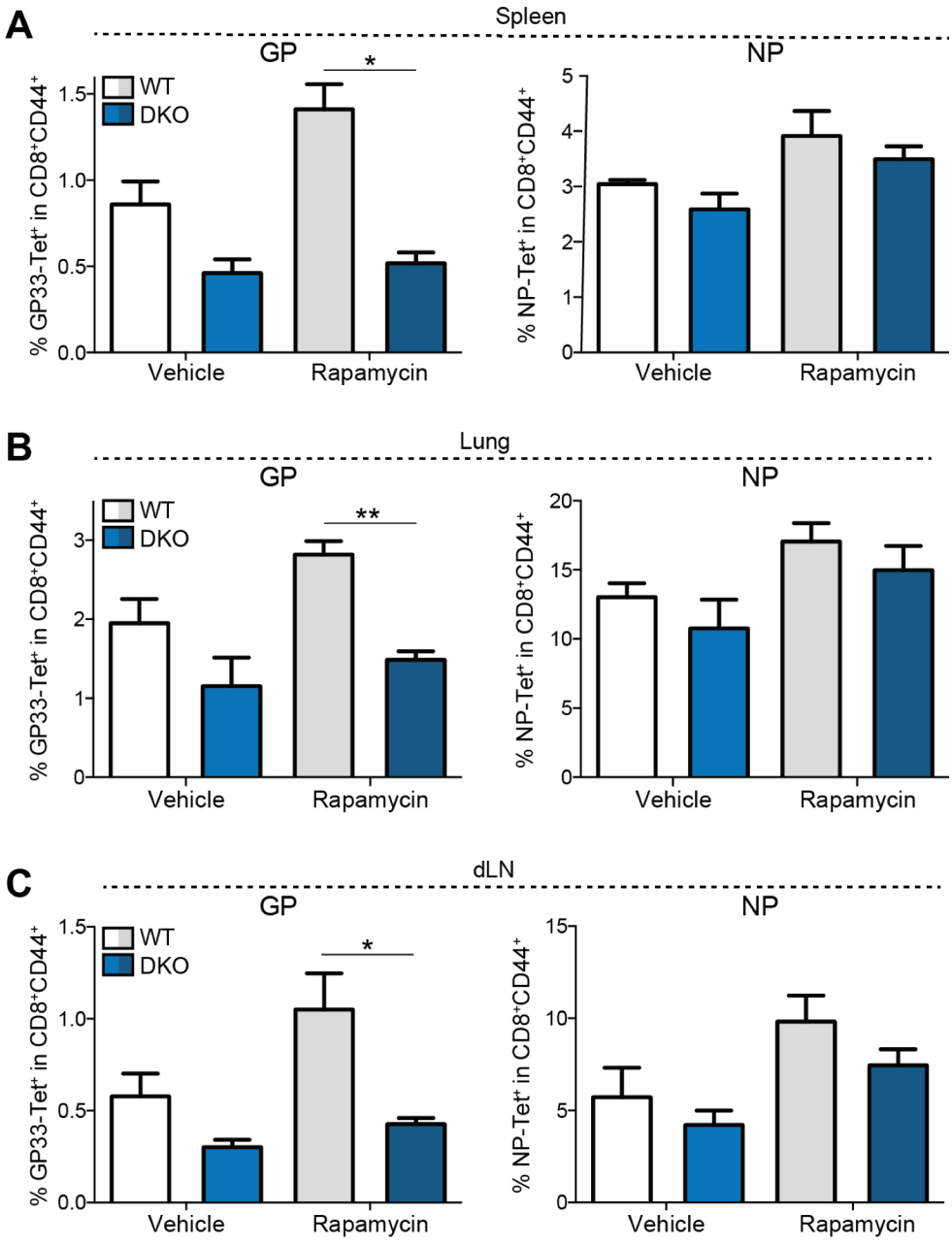

Figure 2.6. **mTOR inhibition with Rapamycin does not rescue memory CD8$^+$ T cell defect observed in the absence of PD-1 signaling.** Fraction of GP33 (left) or NP366 (right) tetramer positive CD8$^+$ T cells in the spleen (A), lung (B), and dLN (C) at d35 following X31-GP33 influenza infection. Mice were either treated with Rapamycin daily from d-1 to d35 i.p. or treated with the vehicle control.

T cell abundance in mice lacking PD-1 signaling like it did in WT mice (Figure 2.6). These data suggest that the functional T cell memory defect associated with PD-1 signaling deficiency is associated with a mechanism other than aberrant mTORC1/2 signaling.

The PD-1 pathway regulates initial CD8$^+$ T cell expansion and contraction

Given the ability of PD-1 to modulate T cell activation, we hypothesized that alterations in CD8$^+$ T cell memory stem from changes early in T cell priming. To test this, we compared influenza-specific CD8$^+$ T cell responses at the height of the primary response, d8-10 post-infection (p.i.) with X31-GP33. While all mice had similar frequencies of influenza-specific CD8$^+$ T cells in the lung, there was a significant increase in the total number of CD8$^+$ T cells in PD-L1/L2 DKO mice (Figure 2.7A). The enhanced early expansion of CD8$^+$ T cells in PD-L1/L2 DKO mice was also associated with a subsequent decrease in the frequency of influenza-specific CD8$^+$ T cells during the contraction phase, d15-20 p.i. (Figure 2.7B). While there was no significant difference in total numbers at d15-20, the ratio of DKO:WT virus-specific CD8$^+$ T cells began to dramatically decline after d10 p.i. (Figure 2.7C). Of note, the immunodominant (NP_{366}-specific) CD8$^+$ T cell responses displayed slower kinetics in this deterioration (Figure 2.7C). Thus, while PD-1 pathway deficiency led to an initial increase in virus-specific CD8$^+$ T cells, there was an erosion of virus-specific CD8$^+$ T cell responses in PD-L1/L2 DKO mice during the establishment of immunological memory. Analysis of CD8$^+$ T cell functionality at d10 p.i. revealed no significant differences in the frequency or quality of cytokine-producing cells in the lungs of WT or PD-L1/L2 DKO mice (Figure 2.4E). In addition, KLRG1, CD127, and CD122 were not dramatically different between WT, PD-1 KO or PD-L1/L2 DKO at effector or memory time points (Figure 2.8A). We further examined the potential for regulation through the PD-1 pathway and found that PD-L1 and PD-L2 expression increased substantially on both hematopoietic and non-hematopoietic APCs including DCs, macrophages, B cells, epithelial and endothelial cells in the lung during

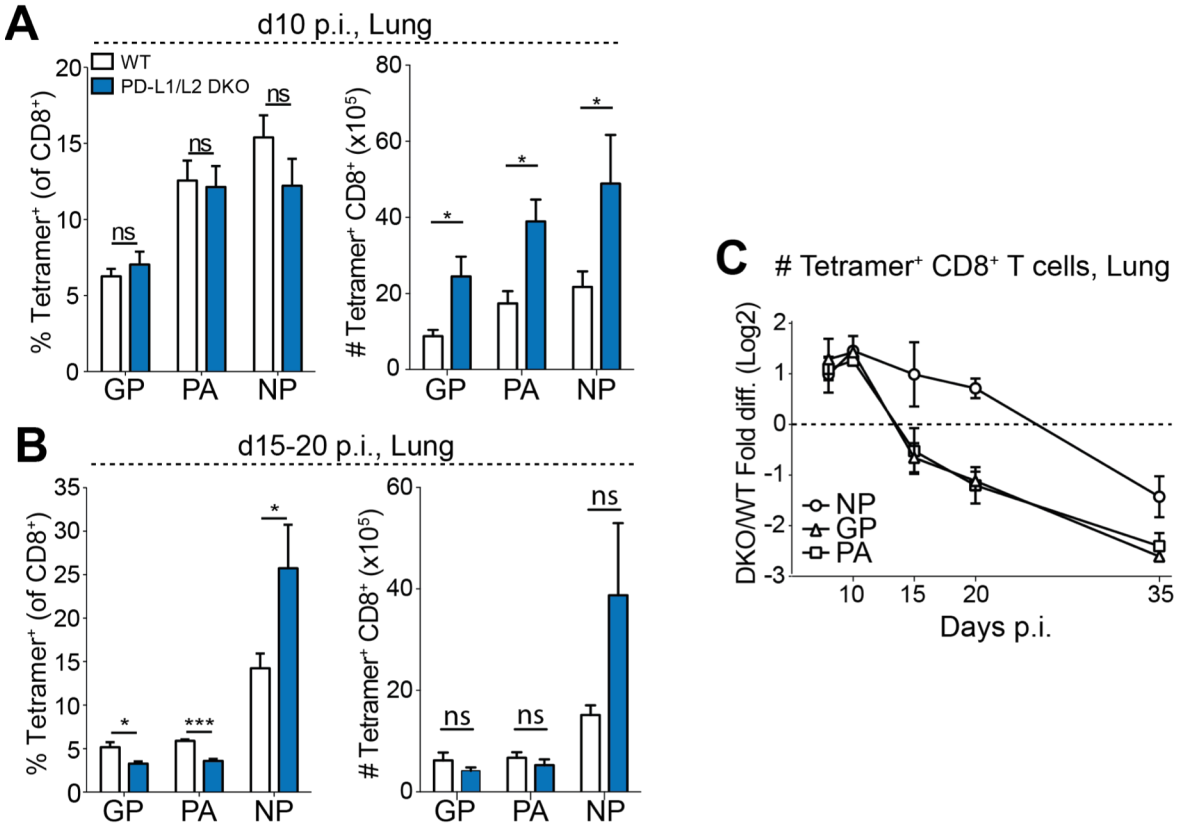

Figure 2.7. Altered effector CD8$^+$ T cell expansion and contraction in the absence of PD-1. (A, B) Frequencies (left) and numbers (right) of indicated tetramer$^+$ CD8$^+$ T cells in the lungs of WT and PD-L1/L2 DKO mice at d10 **(A)** and d15-20 **(B)** after primary X31-GP33 infection. (**C**) Ratios of numbers of influenza-specific CD8$^+$ T cells in PD-L1/L2 DKO and WT mice over time after primary X31-GP33 infection. Significance was assessed using Student's t-test; *P < 0.05, *** P < 0.001.

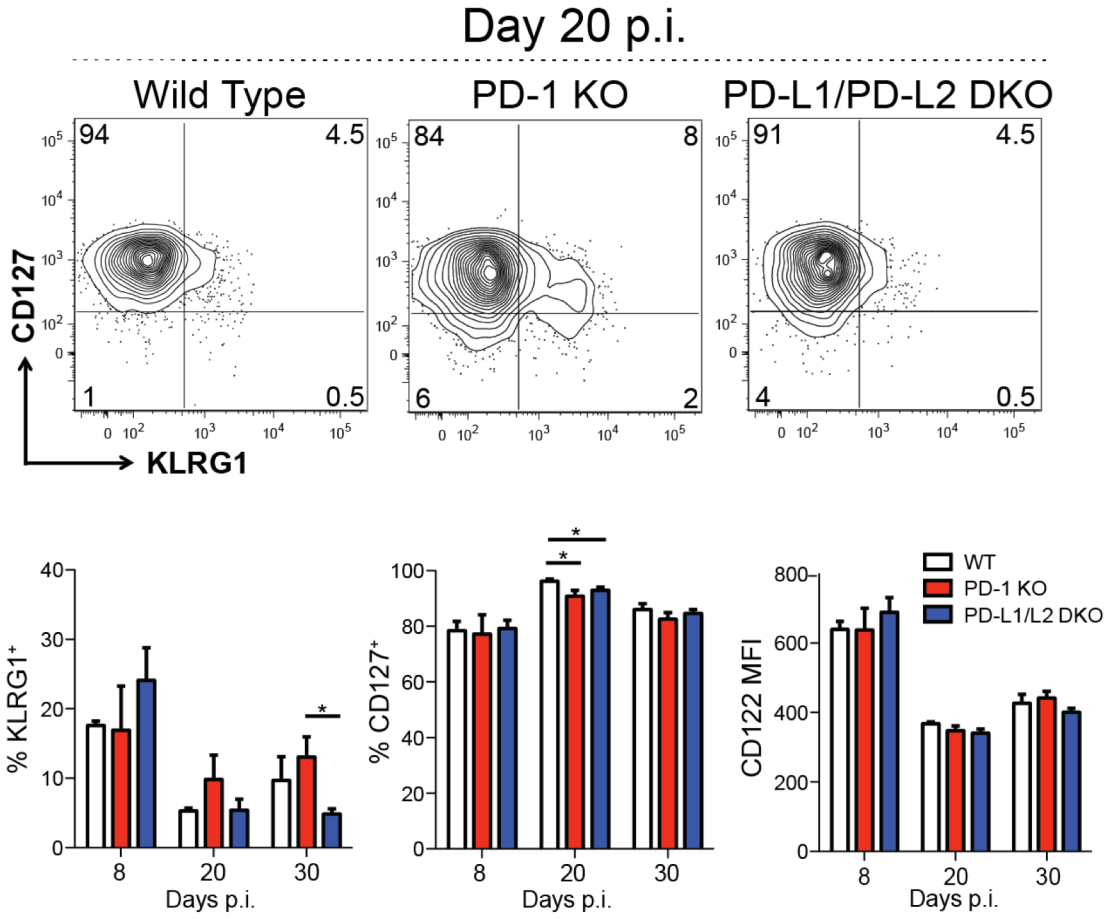

Figure 2.8. Phenotypic analysis of memory markers and PD-1, PD-L1, and PD-L2 expression on cell subsets following influenza infection. (**A**) Longitudinal analysis of KLRG1, CD127, and CD122 expression in WT, PD-1 KO, and PD-L1/PD-L2 DKO mice at 8, 20, and 30 days p.i. in the lung. Representative plots of KLRG1 and CD127 expression are shown in upper panels. Summary of KLRG1 and CD127 frequencies and CD122 MFI at the indicated time points are shown in upper panels.

influenza infection (Figure 2.9, A – C). Moreover, upon influenza infection PD-1 expression is rapidly up-regulated by $CD4^+$ and $CD8^+$ T cells (Figure 2.9, D and E). These findings indicate substantial and dynamic interactions between PD-1 and PD-L1/L2 can occur in the respiratory tract during influenza infection. Together, these data suggest a key regulatory role for the PD-1 pathway in the effector phase that promotes an optimal transition to $CD8^+$ T cell memory.

Increased proliferation and death of $CD8^+$ T cells without PD-1 signaling

To directly investigate the mechanism responsible for the quantitative differences between WT and PD-1 pathway deficient effector $CD8^+$ T cell responses, we performed transcriptional profiling of influenza-specific $CD8^+$ T cells from the lungs of WT and PD-L1/L2 DKO mice at d8 p.i. We focused on subdominant epitope (PA_{244} and GP_{33}) responses that exhibited more profound phenotypic and functional differences. Gene set enrichment analysis (GSEA) using Gene Ontology (GO) (78, 99) revealed that most of the significantly enriched pathways ($p<0.001$, $FDR<0.001$) were related to cell division (Figure 2.10A). We further analyzed three representative pathways enriched in the WT cells – mitosis, spindle assembly, and DNA replication (Figure 2.11A). Among those genes highly over-expressed in WT relative to PD-L1/L2 DKO $CD8^+$ T cells, the majority were members of these pathways (Figure 2.10B). These data suggest a direct link between PD-1:PD-L signals and control of cell cycle during early activation and proliferation of virus-specific $CD8^+$ T cells in the lung during influenza infection.

To assess the relationship between PD-1 signals and cell division, we compared the proliferation and cell death of influenza-specific $CD8^+$ T cells on d6-8 p.i. in WT and PD-L1/L2 DKO mice. At d8 p.i., more influenza-specific $CD8^+$ T cells in the spleens of PD-L1/L2 DKO mice expressed the cell cycle protein Ki-67 compared to WT mice (Figure 2.11D), consistent with the higher cell numbers observed in the lung of PD-L1/L2 DKO mice at this time point

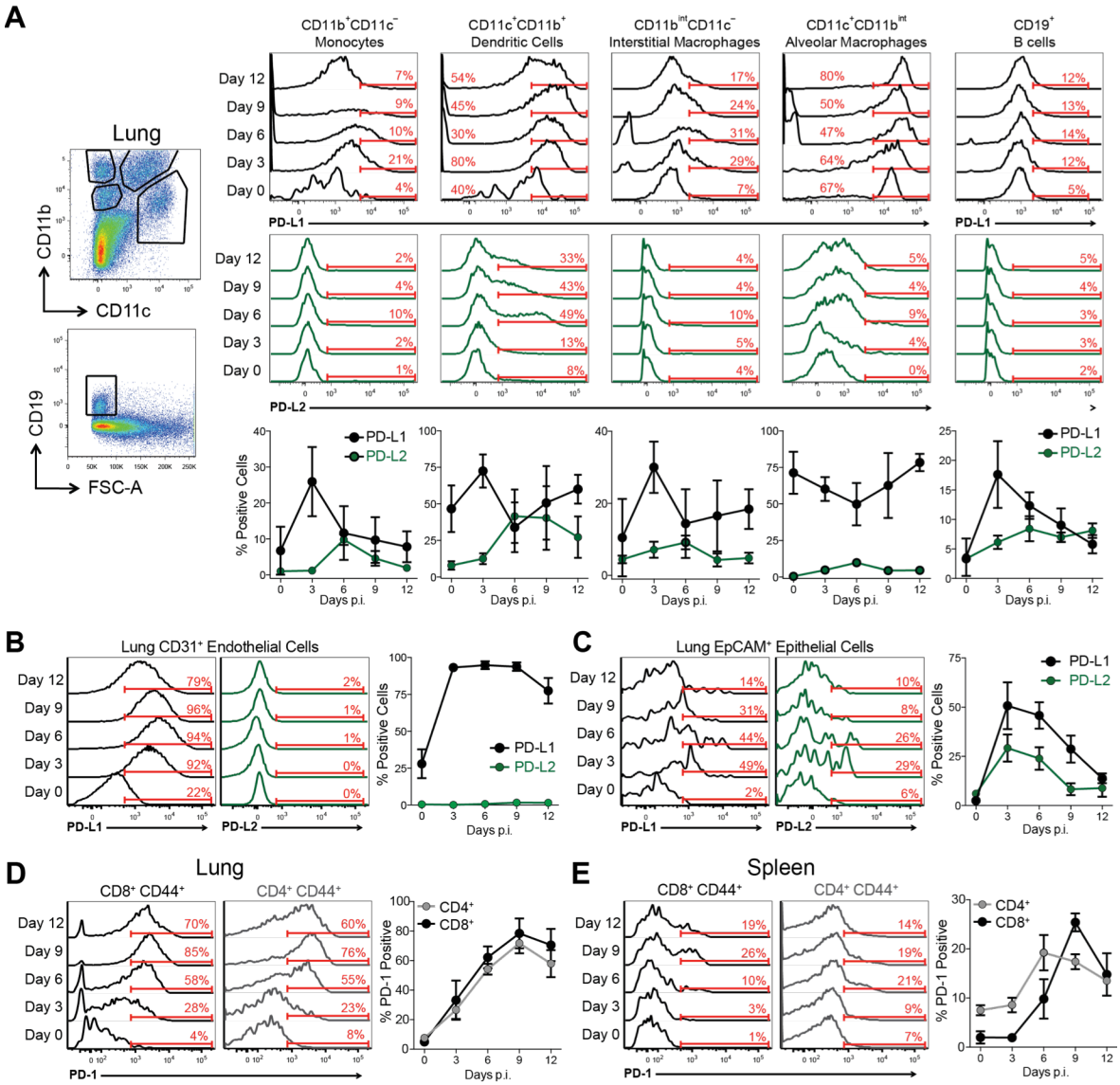

Figure 2.9. Phenotypic analysis of memory markers and PD-1, PD-L1, and PD-L2 expression on cell subsets following influenza infection. (A – C) Expression of PD-L1 (black) and PD-L2 (green) in lung myeloid subsets and B cells **(A)**, endothelial cells **(B)** and epithelial cells **(C)** following primary X31-GP33 influenza infection. Gating for lung myeloid subsets and B cells shown to the left and summarized data to the right. (**D, E**) Wild type mice were infected with X31-GP33 influenza and antigen-experienced $CD44^+$ $CD8^+$ and $CD4^+$ T cells were analyzed for their expression of PD-1 before (d0) and after (d3-d12) infection in lung **(D)** and spleen **(E)**. Representative flow cytometry histograms and summary of data of 5 mice per group shown for each population. Significance was assessed using Student's t-test; $*P < 0.05$, $**P < 0.01$, $*** P < 0.001$.

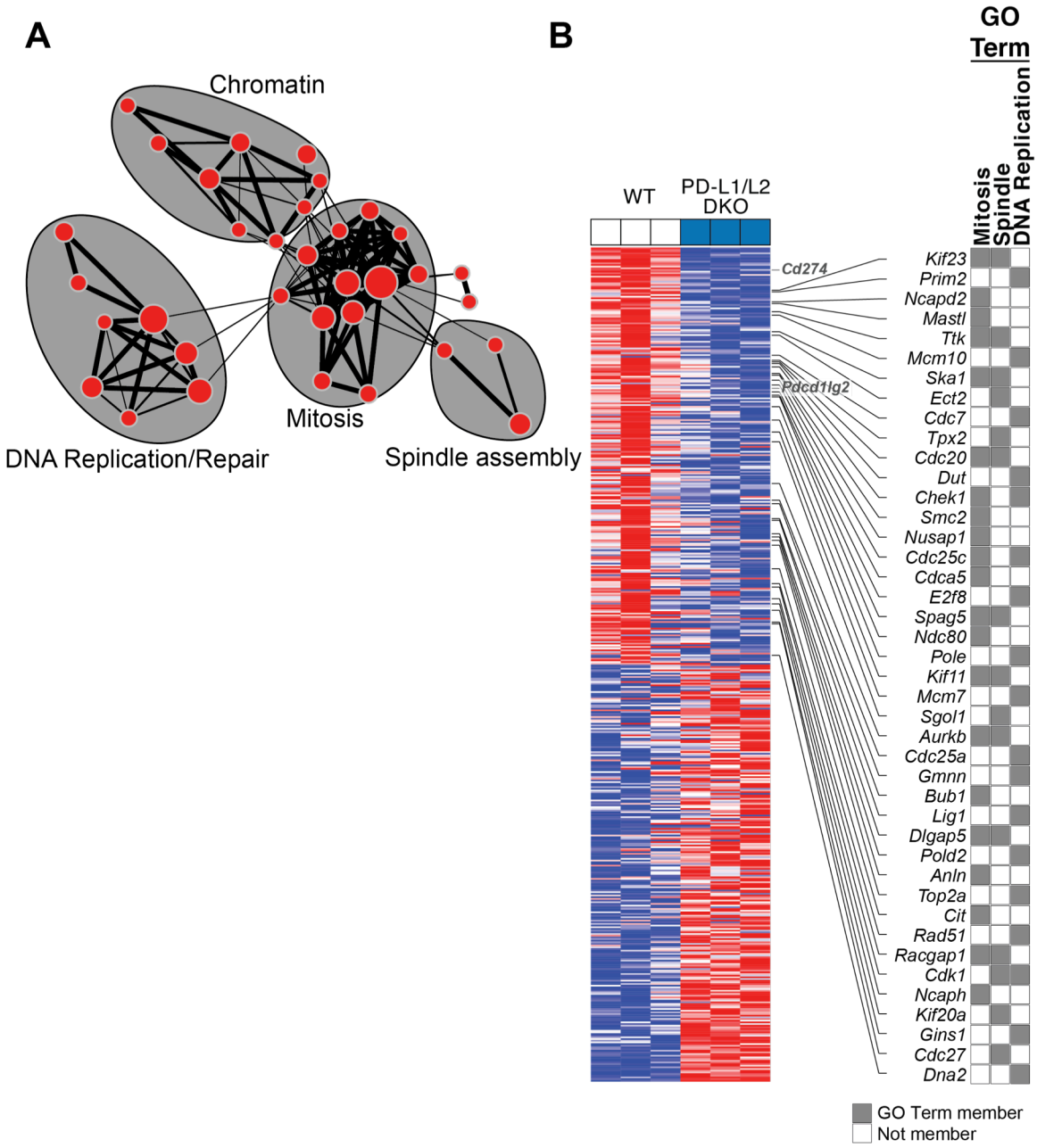

Figure 2.10. Altered expression of genes representing cell division pathways in WT versus PD-L1/L2 DKO virus-specific CD8$^+$ T cells and reduced cell survival in the absence of PD-1 signaling. (A) Significantly enriched (FDR<0.01) gene sets in WT CD8$^+$ T cells clustered based on the extent of gene member overlap and annotated for the biological states/processes they represent. (**B**) Top and bottom 250 differentially expressed genes in subdominant-epitope (GP_{33} and PA_{224}) specific WT and DKO CD8$^+$ T cells ranked by signal-to-noise values following filtration to top 10% of genes with highest mean absolute deviation across samples. Membership of genes in the three representative GO terms is shown on the right.

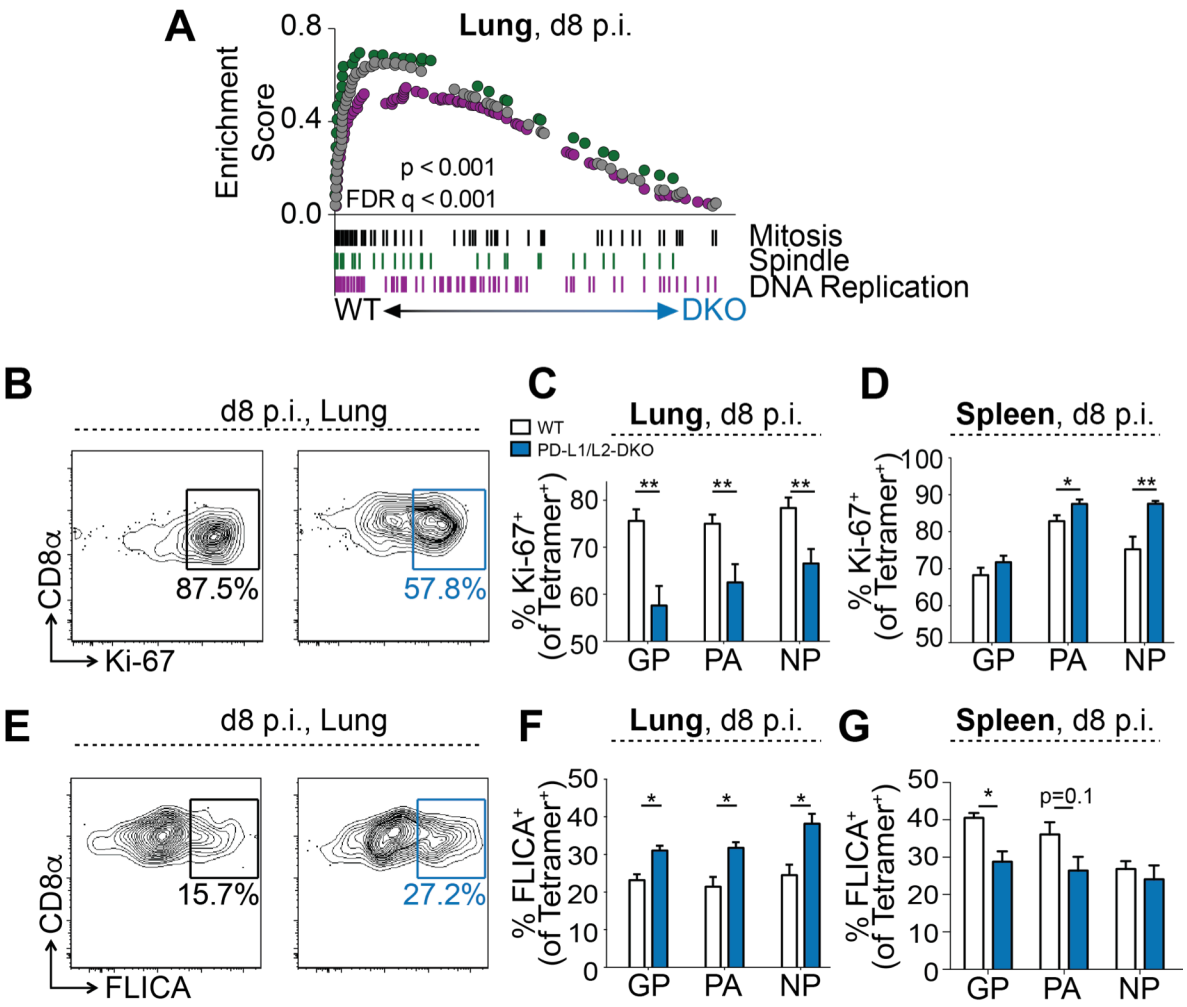

Figure 2.11. CD8$^+$ T cells undergo robust proliferation but have reduced survival in the absence of PD-1 signaling in influenza-infected mice. **(A)** Representative GO gene sets enriched in whole genome transcriptional profiles from GP_{33} and PA_{224} specific CD8$^+$ T cells from lungs of WT versus PD-L1/L2 DKO mice d8 p.i. (**B**) Representative flow cytometric analysis of Ki-67 expression by NP_{366}-specific CD8$^+$ T cells from the lung of WT and PD-L1/L2 DKO mice at d8 p.i. (**C**, **D**) Numbers indicate fraction of Ki-67$^+$ $NP_{366}{}^+$ CD8$^+$ T cells in lung (C) and spleen (D). (**E**) Flow cytometric analysis of active caspases by FLICA staining of NP_{366}-specific lung CD8$^+$ T cells from WT and PD-L1/L2 DKO mice at d8 p.i. (**F**, **G**) Numbers indicate fraction of FLICA$^+$ $NP_{366}{}^+$ CD8$^+$ T cells in lung (F) and spleen (G) based on unstained controls. Significance was assessed using Student's t-test; **P* < 0.05, ***P* < 0.01. Data are representative of 2 independent experiments with 4-5 mice per experiment. Significance was assessed using Student's t-test; **P* < 0.05.

(Figure 2.7A). However, significantly fewer virus-specific $CD8^+$ T cells were Ki-67^+ in the lungs of PD-L1/L2 DKO mice (Figure 2.11, B and C). These findings suggest that the increased number of $CD8^+$ T cells in the lungs of PD-1 KO and PD-L1/L2 DKO mice is largely due to robust proliferation in lymphoid organs and homing/trafficking of these $CD8^+$ T cells to the lung. Interestingly, PD-L1/L2 DKO mice had a higher frequency of influenza-specific $CD8^+$ T cells that were positive for active caspase staining (FLICA) in the lung, but not in the spleen, on day 8 p.i. (Figure 2.11, E – G), consistent with the increased contraction of the DKO populations. Thus, the PD-1 pathway tempers early $CD8^+$ T cell proliferation in lymphoid tissues, which is critical for $CD8^+$ T cell survival at the site of infection and subsequent memory formation.

Absence of PD-1 signaling enhances primary influenza infection clearance

One reason for altered proliferation/survival dynamics of $CD8^+$ T cells in PD-1 pathway deficient mice could be enhanced clearance of primary influenza infection, resulting in truncated exposure to the optimal levels of antigen and inflammation needed for proper memory development. Indeed, PD-1 pathway deficient mice cleared influenza virus more rapidly than WT mice (Figure 2.12A). Enhanced control of infection was also associated with reduced weight loss and faster recovery (Figure 2.12B). We also noted slight yet statistically significant differences in initial viral clearance between PD-1 KO and PD-L1/L2 DKO mice (Figure 2.12A). We hypothesize that these minor differences may arise from additional binding partners for these molecules, such as PD-L1 interacting with B7-1 and PD-L2 with RGMb (100, 101). Altered viral control and weight loss were associated with a trend toward more severe lymphocytic infiltration and injury in the lungs in the absence of PD-1 pathway signals (Figure 2.12, C and D). Thus, in PD-1 KO and PD-L1/L2 DKO mice, control of infection was accelerated and altered memory $CD8^+$ T cell development could be due to changes in antigen stimulation or inflammation and/or differences in T cell differentiation due

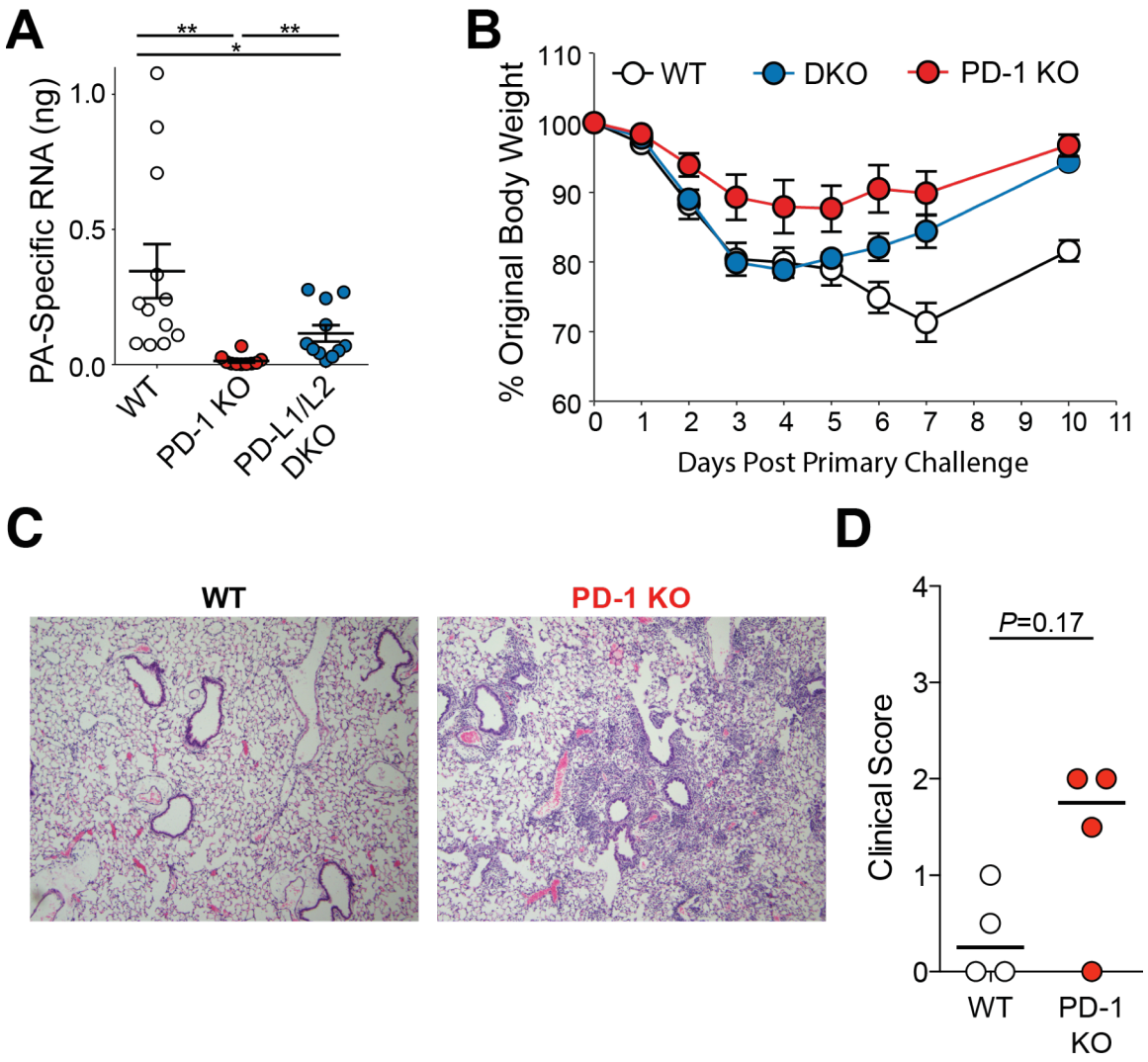

Figure 2.12. PD-1 pathway deficient mice clear primary influenza infection faster than WT mice and show a modestly enhanced immune infiltration. (A) Influenza viral titers in the lung at d7 p.i. and (**B**) weight loss in WT, PD-1 KO, and PD-L1/L2 DKO mice following primary infection with X31-GP33. **(C, D)** Representative H&E stains **(C)** and quantification **(D)** of the histopathological clinical score of WT and PD-1 KO mice at d7 p.i. with X31-GP33. Significance was assessed using Student's t-test (A) or Mann Whitney test (D); **P* < 0.05, ***P* < 0.01.

to cell-intrinsic PD-1 signals.

Cell-intrinsic regulation of CD8$^+$ T cell memory differentiation by PD-1

To test whether PD-1 signals had a cell-intrinsic role in regulating CD8$^+$ T cell memory differentiation during influenza infection, we used a distinct, but complementary model. We co-transferred equal numbers of WT and PD-1 KO P14 TCR transgenic CD8$^+$ T cells (recognizing GP_{33-41} peptide presented by H-2D^b) into WT mice followed by infection with X31-GP33 virus. This approach allows for direct comparison of WT and PD-1 KO CD8$^+$ T cells in the same WT mice and excludes differences in precursor frequency, TCR repertoire, viral control, effects of PD-L1 and/or PD-L2 on APCs, and/or influenza pathogenesis. Viral clearance was similar in recipient WT mice receiving no P14 cells or the mixture of WT and PD-1 KO P14 cells (Figure 2.14A). Consistent with findings in knockout mice, the frequency and numbers of PD-1 KO P14 cells were significantly higher than WT P14 cells in the lung and spleen at d7 p.i. (Figure 2.13, A and B), with similar trends in the blood and dLN (Figure 2.14B). Similar trends were observed with PR8-GP33 (Figure 2.13, C and D). The increased frequency of PD-1 KO P14 cells was associated with increased BrdU incorporation in the spleen and dLN, indicating that PD-1 regulates CD8$^+$ T cell proliferation in a cell-intrinsic manner (Figure 2.13*C* and 2.14E), similar to virus-specific CD8$^+$ T cells in PD-L1/L2 DKO mice. However, unlike the global KO mice, when PD-1 deficiency was restricted to P14 cells, proliferation in the lungs was not reduced. We attribute these differences to the altered viral clearance in the KO animals. Also, PD-1 KO P14 cells had moderately reduced KLRG1 but increased CD127 and CD122 at memory time points compared to WT P14 cells, highlighting potential cell-intrinsic effects of PD-1 signals not observed in the global KO mice (Figure 2.15, Figure 2.8). However, similar to CD8$^+$ T cells from the global KO animals, the PD-1 KO P14 cells in WT hosts were more prone to cell death in the lung, as demonstrated by increased FLICA staining at d10 p.i. (Figure 2.13D). Thus, absence of cell-intrinsic PD-1

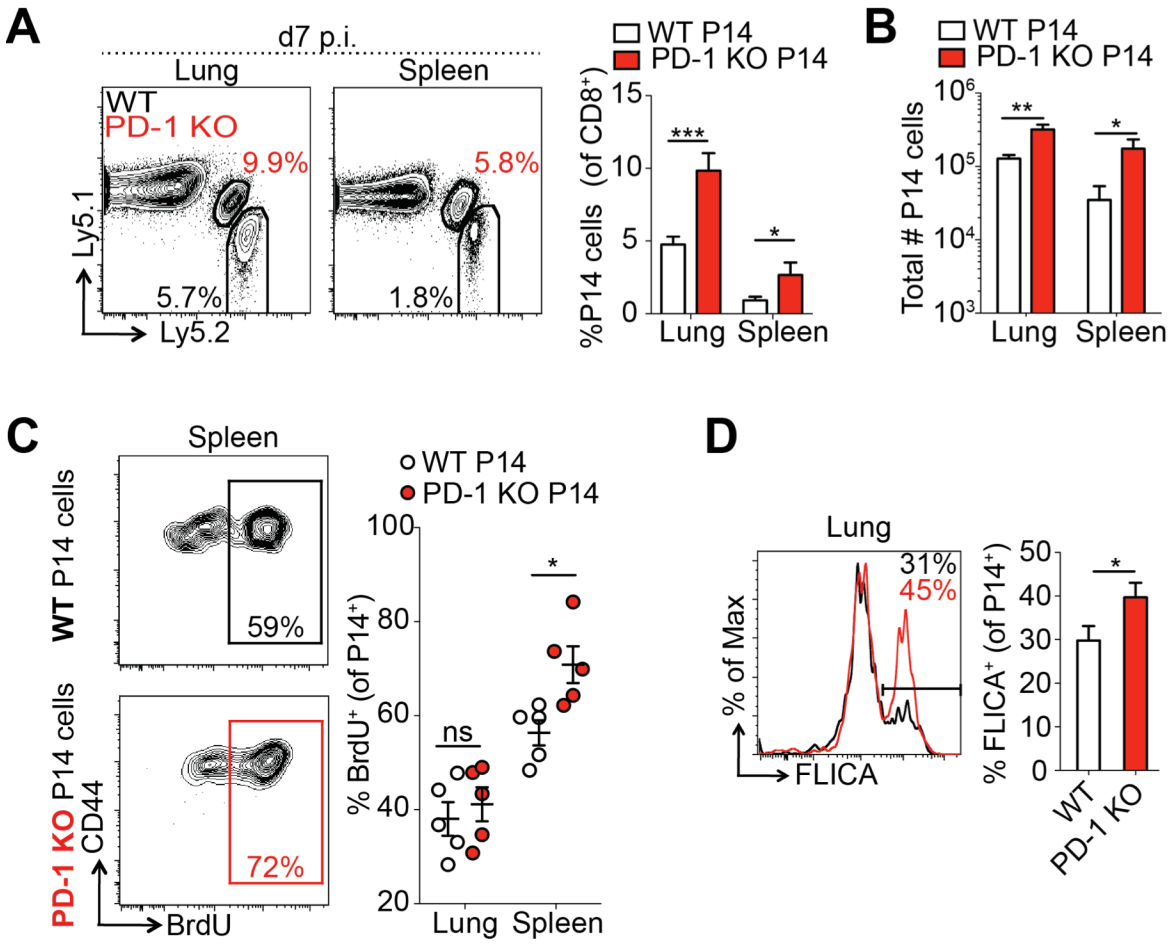

Figure 2.13. PD-1 controls CD8$^+$ T cell proliferation and cell death in a cell-intrinsic manner. **(A, B)** Representative plots (left) show the frequency of WT and PD-1 KO P14 cells in the lung and spleen at d7 after X31-GP33 infection. Numbers indicate frequency of P14 cells as a percent of CD8$^+$ T cells. Summaries of frequencies **(A)** and numbers **(B)** of P14 cells are shown to the right. **(C)** Flow cytometric analysis of BrdU incorporation by WT and PD-1 KO P14 cells in the spleen at d7 p.i. Numbers indicate fraction of BrdU$^+$ P14 cells. Representative plots shown to the left. Summary of BrdU$^+$ P14 cells in different organs shown to the right. **(D)** Flow cytometric analysis of active caspase by FLICA staining of lung WT and PD-1 KO P14 T cells on d8 p.i. Numbers indicate fraction of P14 cells positive for FLICA staining based on unstained controls. Representative plot shown (left) and summary of FLICA$^+$ cells shown (right). Significance was assessed using Student's t-test; *P < 0.05, **P < 0.01, *** P < 0.001.
The data in this figure were generated by Pamela Odorizzi and Kristen Pauken.

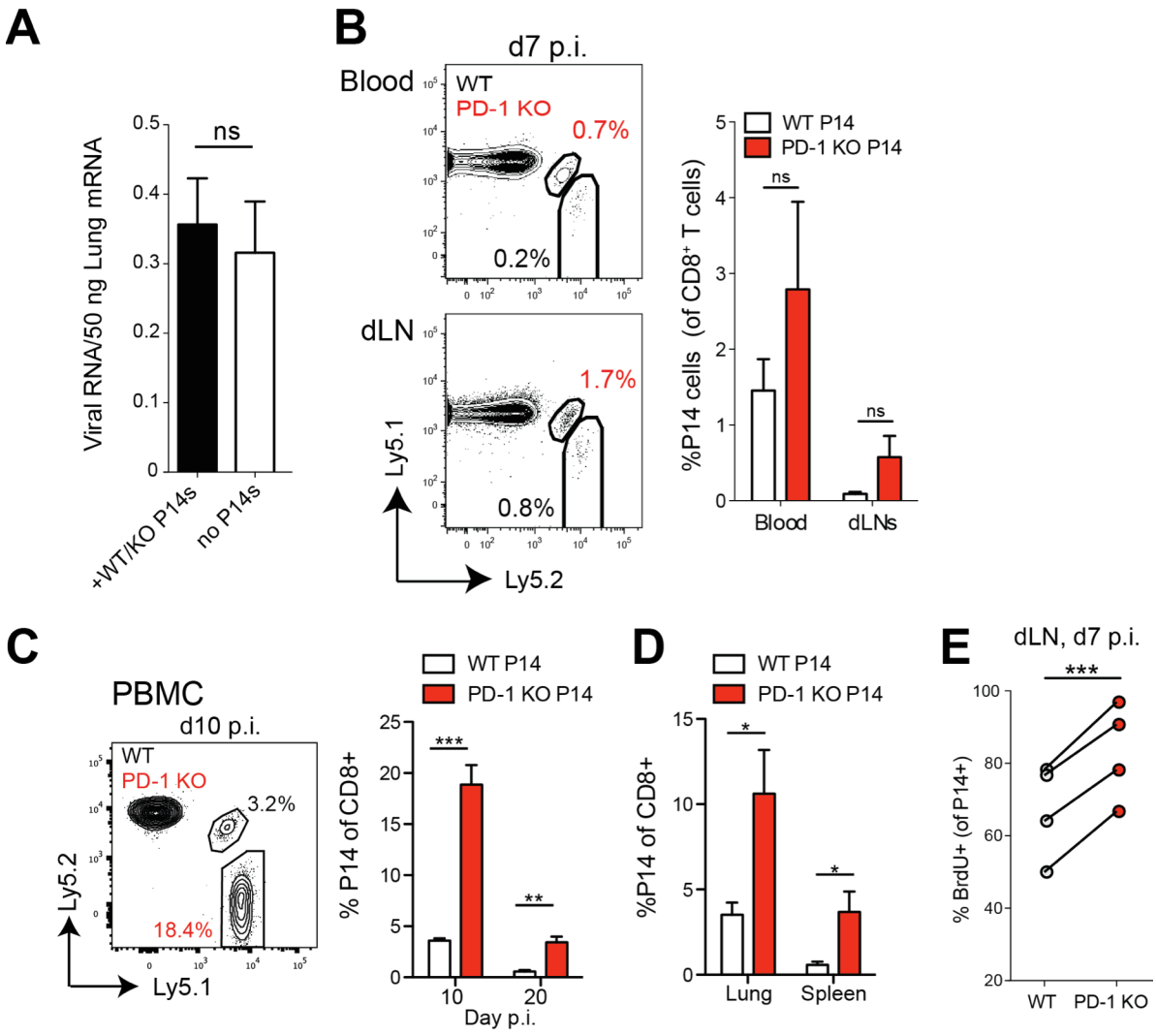

Figure 2.14. PD-1 controls CD8$^+$ T cell proliferation in a cell-intrinsic manner. (A) Quantification of viral load d8 after X31-GP33 infection in WT mice compared to WT mice that received WT/PD-1 KO mixed P14 chimeras. (**B**) Representative plots (left) show the frequency of WT and PD-1 KO P14 cells in indicated organs at d7 after X31-GP33 infection. Numbers indicate frequency of P14 cells as a percent of CD8$^+$ T cells. Summary of frequencies (right). (**C**, **D**) Frequency of WT and PD-1 KO P14 cells in chimeras following primary 0.3 LD_{50} PR8-GP33 infection on d10 and d20 in blood **(C)** and at d30 in the lung and spleen **(D). (E)** Summary of BrdU$^+$ P14 cells in dLN at d7 p.i. Data are representative of 3 independent experiments with 4-5 mice per group. Significance was assessed using Student's t-test; *P < 0.05, **P < 0.01, *** P < 0.001.
The data in this figure were generated by Pamela Odorizzi and Kristen Pauken.

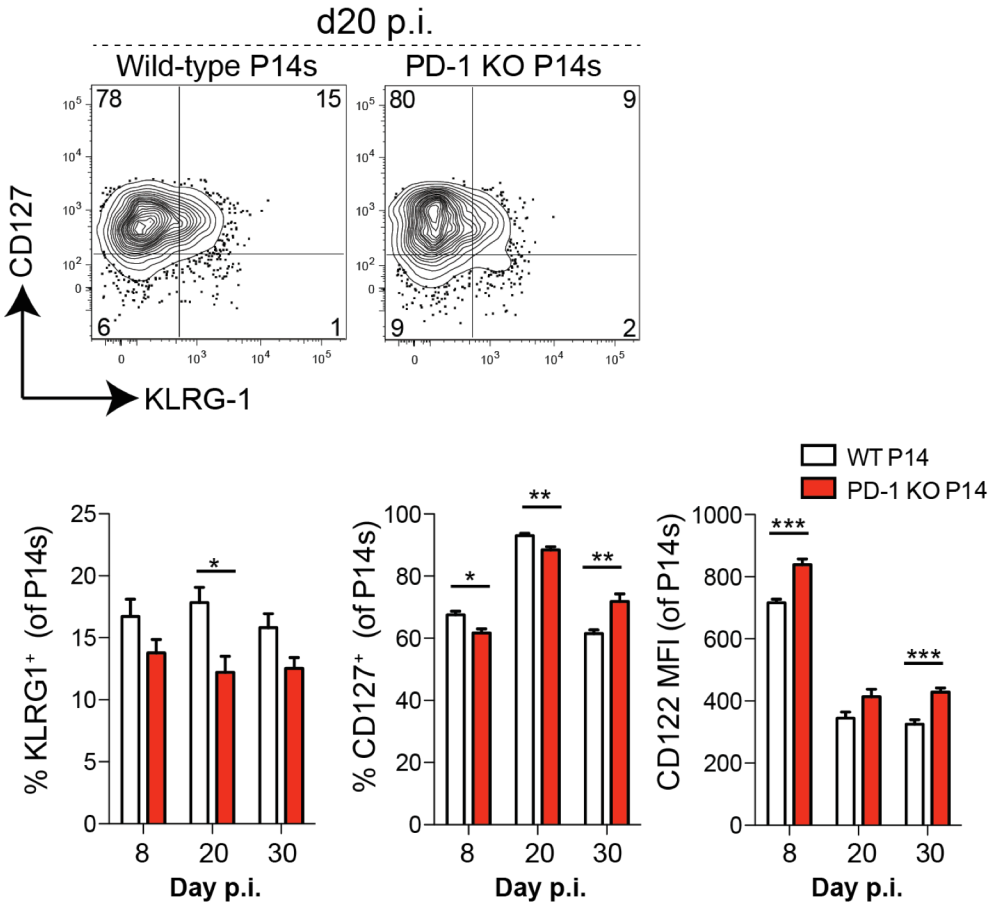

Figure 2.15. Longitudinal analysis of KLRG1, CD127, and CD122 expression on transferred PD-1 KO and WT P14 cells 8, 20, and 30 days p.i. in the lung. Representative plots of KLRG1 and CD127 expression are shown in upper panels. Summary of KLRG1 and CD127 frequencies and CD122 MFI at the indicated time points are shown in the lower panels. Significance was assessed using Student's t-test; *$P < 0.05$, **$P < 0.01$, *** $P < 0.001$.
The data in this figure were generated by Pamela Odorizzi and Kristen Pauken.

signals largely, though not completely, phenocopies $CD8^+$ T cells from the global KO mice and leads to increased proliferation of virus-specific $CD8^+$ T cells in lymphoid tissues and subsequent excessive cell death at the site of infection.

We next used the P14 co-transfer system to interrogate whether these early cell-intrinsic alterations in proliferation and cell death also resulted in aberrant memory development. While WT P14 cells formed stable memory, PD-1 KO P14 cells underwent progressive contraction and were less abundant than WT P14 cells in the lung and spleen after 3 months (Figure 2.16, A and B). While the kinetics differed slightly compared to the whole animal KO model, these results are consistent with the memory attrition observed in the global KO mice described above. Significantly fewer PD-1 KO memory P14 cells from the lung or spleen produced IFN-γ and TNF-α, upon re-stimulation compared to WT cells and PD-1 KO P14 cell function was diminished on a per cell basis (Figure 2.16C). Most importantly, following re-challenge with PR8-GP33, PD-1 KO P14 memory cells displayed dramatically impaired secondary expansion compared to WT P14 cells in the same animals (Figure 2.16D). These results point to a key function for PD-1 signals in regulating T cell proliferation and survival during the acute phase of infection and subsequent $CD8^+$ T cell memory.

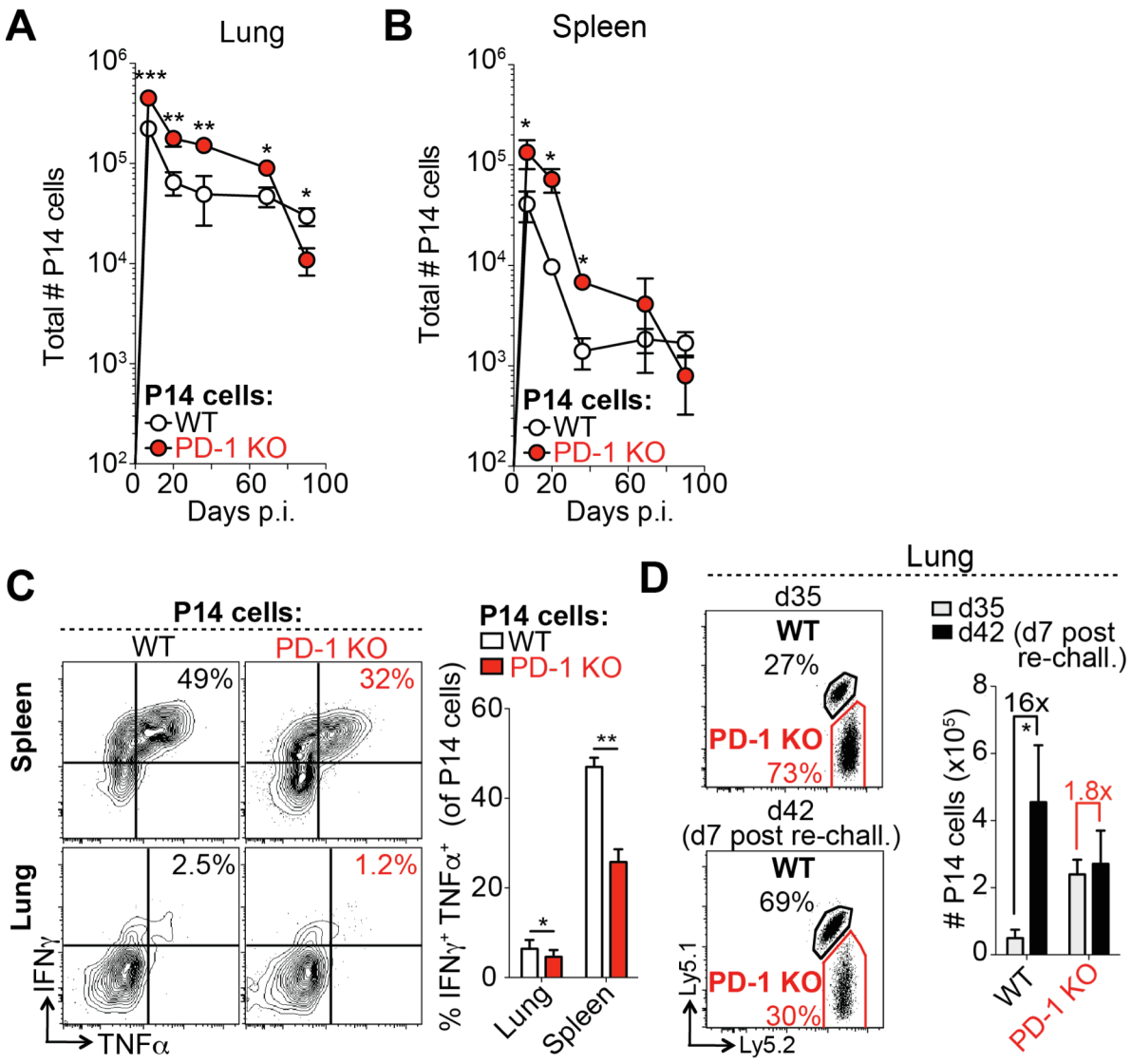

Figure 2.16. Defect in CD8$^+$ T cell memory in the absence of PD-1 is cell-intrinsic. (A) Longitudinal analysis of WT and PD-1 KO P14 cell numbers in the lung during primary X31-GP33 infection. **(B)** Longitudinal analysis of WT and PD-1 KO P14 absolute cell numbers in the spleen during primary X31-GP33 infection. **(C)** Representative plots of intracellular cytokine staining for IFN-g and TNF-a production (left) in WT and PD-1 KO P14 cells from spleen (upper) and lung (lower) at d47+ p.i. and summary quantification (right). **(D)** Representative plots of WT and PD-1 KO P14 cell percentages d35 after infection with X31-GP33 (d35) and 7 days post re-challenge with PR8-GP33 (d42 after primary infection). Numbers indicate frequency of P14 cells as a percent of CD8$^+$ T cells. Summary of fold change of numbers of P14 cells pre- and post-re-challenge shown to the right. Data are representative of 3-4 independent experiments with 4-6 mice per group. Significance was assessed using Student's t-test; *P < 0.05, **P < 0.01, *** P < 0.001.
The data in this figure were generated by Pamela Odorizzi and Kristen Pauken.

Discussion

The PD-1 pathway plays a key role in regulating T cell responses and is an important immunotherapeutic target in cancer and chronic infection. With the recent increase in clinical administration of long-term PD-1 pathway blockade, patients with sustained blockade of PD-1 signals are likely to be exposed to subsequent infections and vaccinations. PD-1 is rapidly up-regulated during acute infection or vaccination, but the role of PD-1 signals in this context remains poorly understood. Here we found that PD-1 KO and PD-L1/L2 DKO mice developed sub-optimal memory $CD8^+$ T cells that lack long-term stability and respond poorly to re-challenge, key changes also observed with cell-intrinsic PD-1 deficiency. Without PD-1 signals, virus-specific $CD8^+$ T cells lose or never develop the robust recall capacity that is characteristic of memory T cells. This defect was associated with early excessive proliferation of effector $CD8^+$ T cells, coupled with increased cell death. The PD-1 pathway regulates this process, at least in part, by controlling the expression of key cell cycle genes and pathways in virus-specific $CD8^+$ T cells. Taken together, our work suggests a model (Figure 2.17) where the PD-1 pathway plays an important role in the optimal development of long-lived memory responses by tempering early proliferation of $CD8^+$ T cells.

PD-1 can inhibit TCR and co-stimulatory signals (101-105), implying a role in regulation of T cell differentiation. However, there are conflicting data about how the PD-1:PD-L pathway impacts effector and memory $CD8^+$ T cell responses (87-91), making the precise role of PD-1:PD-L signals in the formation of long-term T cell memory unclear. Here, we show that loss of PD-1 pathway signals impairs the development of optimal $CD8^+$ T cell memory following influenza infection. Previous work showed that antibody blockade or genetic deletion of the PD-1 pathway enhances early virus-specific effector $CD8^+$ T cell responses to respiratory infection (91), consistent with our data. Our current study adds to these findings by demonstrating that the early benefit of loss of PD-1 signals ultimately results in long-term deficiencies in $CD8^+$ T cell memory. PD-1 regulates key cell cycle genes

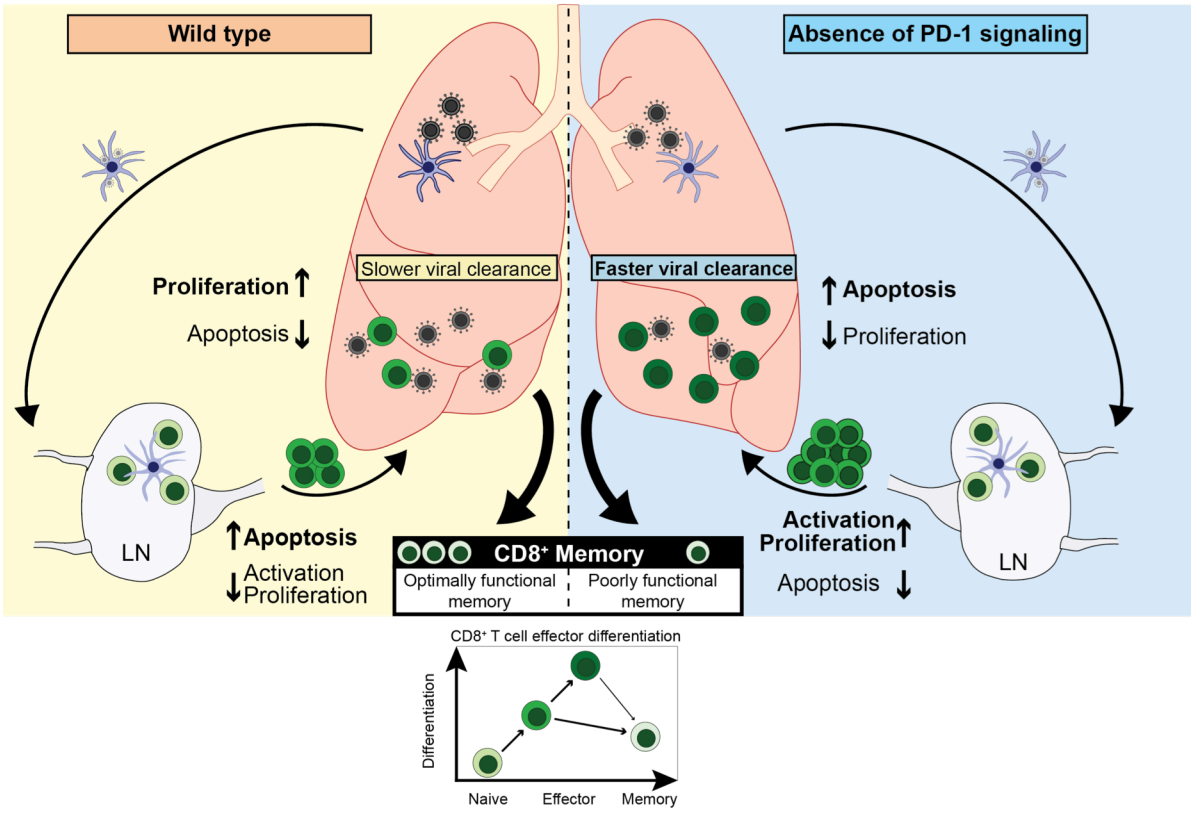

Figure 2.17. Regulation of CD8$^+$ T cell responses to influenza virus infection by the PD-1 pathway. In wild-type mice (left), PD-1 pathway signals during influenza virus infection restrict early activation and proliferation of virus-specific CD8$^+$ T cells in dLNs. These T cells then migrate into the infected lung, where inhibition by the PD-1 pathway promotes their survival and continued proliferation, allowing for clearance of influenza virus and the development of optimally functional memory CD8$^+$ T cells. In settings of PD-1 pathway deficiency (right), virus-specific CD8$^+$ T cells undergo excessive proliferation in dLNs and migrate to the lungs in high numbers. While this leads to faster clearance of primary influenza infection, virus-specific CD8$^+$ T cells undergo increased apoptosis in the lung and ultimately develop into poorly functional and unstable memory cells.

controlling proliferation and survival of virus-specific $CD8^+$ T cells into the memory phase. We show that PD-1 signals regulate $CD8^+$ T cell proliferation in a cell-intrinsic manner. Our transcriptional profiling studies suggest a broad and pervasive effect on coordinating cell cycle progression. Our observations are consistent with *in vitro* work showing that PD-1 inhibits cell cycle progression through the G_1 phase by modulating key cell-cycle regulators in $CD4^+$ T cells (106, 107). PD-1 signaling also can inhibit PI3K/Akt signaling (102, 105). Therefore, PD-1 may target multiple key pathways to inhibit cell cycle progression in T cells. Our data suggest a model whereby PD-1 serves as a critical integrator of early $CD8^+$ T cell activation signals, operating at least partly through control of cell cycle progression and thereby promoting optimal $CD8^+$ T cell memory formation.

Interestingly, there were some notable differences between mice globally deficient in PD-1 signaling and the P14 chimera model, including altered viral control, different kinetics of changes in proliferation/cell death and KLRG1, CD127, and CD122 expression, suggesting a role for $CD8^+$ T cell extrinsic effects. PD-1, PD-L1 and PD-L2 are expressed on non-$CD8^+$ T cells, including myeloid cells, B cells, $CD4^+$ T cells, as well as endothelial and epithelial cells, and their expression is dynamic during respiratory influenza virus infection. Global loss of PD-1:PD-L interactions accelerated viral clearance and may have other systemic effects on inflammation and/or pathology that could contribute to differences between cell-intrinsic versus systemic PD-1 signaling deficiencies. While the P14 chimera system provides a means to control for these potential cell-extrinsic effects and define mechanisms, the systemic loss of PD-1 signals may have important implications for humans receiving long-term treatment with PD-1 pathway blocking reagents.

The ability of the PD-1 pathway to modulate effector and memory $CD8^+$ T cell differentiation makes it an attractive therapeutic target for infectious diseases and cancer. However, in order to most effectively target the PD-1 pathway to improve immunity, several questions must be addressed. First, it is not yet clear when during the course of infection PD-

1 pathway blockade is most beneficial. Our data suggest that very early removal of PD-1 signals during primary infection can prevent optimal memory differentiation, while other data demonstrate that blockade of PD-1 just prior to re-challenge improves recall responses (86, 91). Targeting the PD-1 pathway during different stages of infection may result in distinct effects on $CD8^+$ T cell differentiation, recall responses, and/or memory maintenance. Second, the PD-1:PD-L pathway seems to have a particularly important role in controlling the balance between immunopathology and effective viral immunity, particularly in the respiratory tract (48, 84, 91). Therefore, PD-1 signals may help prevent excessive immune-mediated damage by restricting T cell responses during acute infection in the lung. Our findings reveal that another key function of PD-1 is to temper excessive T cell stimulation and facilitate the development of optimal $CD8^+$ T cell memory. How PD-1 integrates signals to balance effective pathogen control and memory development, while limiting immunopathology, will be an important area for future investigation. The answers to these questions will be critical for determining how to best modulate PD-1 to enhance responses to infectious agents and cancer.

In summary, we have identified a key role for the PD-1 pathway in regulating effector and memory $CD8^+$ T cell differentiation during influenza virus infection. PD-1 signaling is needed to promote optimal formation of long-lived, protective memory $CD8^+$ T cells. We have identified control of cell cycle and apoptosis as one mechanism by which PD-1:PD-L signals influence $CD8^+$ T cell memory. Future studies are needed to identify additional molecular pathways by which the PD-1:PD-L pathway regulates $CD8^+$ T cell differentiation in different disease settings.

Materials and Methods

Mouse Strains

Wild-type C57BL/6 mice were purchased from The Jackson Laboratory. PD-1 KO and PD-L1/L2 DKO have been described previously (108, 109). PD-1 KO P14 mice were generated by crossing germline PD-1 KO mice to P14 TCR transgenic mice (110). $Ly5.2^+$ or $Ly5.2^+Ly5.1^+$ PD-1 KO and WT P14 cells were isolated from peripheral blood and transferred i.v. into C57BL/6 recipient mice at a 1:1 ratio (500 cells each). All mice were maintained in a specific pathogen-free facility and used according to Institutional Animal Care and Use Committee and National Institutes of Health guidelines.

Viral Infections

For primary infection, 6 to 12 week old mice were infected with recombinant influenza virus X31-GP33 ($1.6x10^5$ $TCID_{50}$) intranasally (i.n.). For re-challenge, mice were infected with PR8-GP33 (10 LD_{50} i.n.) at least 35 days after primary infection. Recombinant influenza strains containing the LCMV GP_{33-41} epitope inserted into the neuraminidase protein stalk region were provided by Dr. Richard Webby (St. Jude Children's Research Hospital, Memphis, TN) (92, 93, 111). Prior to infections, mice were anesthetized by intraperitoneal (i.p.) injection of 2.5% Avertin (Sigma-Aldrich). Viral titers in lungs were determined by quantitative real-time PCR (qRT-PCR) as described (111). For Rapamycin experiments, the drug was administered as a suspension in carboxymethylcellulose (CMC) i.p. at 75 µg/kg daily from d-1 to d35 as described previously (98, 112). For LCMV infections, 2×10^5 p.f.u. of LCMV Armstrong (gift of E. John Wherry, University of Pennsylvania Perelman School of Medicine) was administered i.p. at d0.

Flow cytometry

Lymphocytes were isolated from spleen, lung and draining lymph node as described

(111). Single cell suspensions were stained with antibodies specific to specific to CD8a (53-6.7), CD44 (IM7), IFN-γ (XMG1.2), TNF-α (MP6-XT22), CD127 (SB/199), CD122 (TM-b1), CD62L (MEL-14) CD45.1 (A20), CD45.2 (104), PD-1 (RMP1-30), PD-L1 (10F.9G2), PD-L2 (TY25), CD11b (M1/70), CD11c (N418), CD19 (6D5), CD4 (GK1.5), Ep-CAM (G8.8), and CD31 (390) purchased from BioLegend, Ki-67 from BD Biosciences, and KLRG1 (2F1) from Abcam. Poly-caspase analysis was performed with the FLICA Vybrant FAM Assay Kit (Life Technologies). Dead cell exclusion was performed by live/dead fluorescent reactive dye (Life Technologies) staining. $H2D^b$ GP_{33-41}, $NP_{366-374}$, and $PA_{224-233}$ tetramers were prepared and used as described (36). For intracellular cytokine staining (ICS), single cell suspensions from spleens or lungs, which included T cells as well as APCs from the same animal, were incubated with 0.5 μg/ml influenza $NP_{366-374}$ or GP_{33-41} peptide (Genscript) or no peptide for 5 hours at 37°C in the presence of GolgiPlug (BD Biosciences), surface stained, fixed/permeabilized, and intracellularly stained using the Cytofix/Cytoperm kit (BD Biosciences), as directed by the manufacturer. For BrdU detection, animals were treated with 2mg of BrdU (Sigma-Aldrich) i.p. 12-24 hours prior to analysis. BrdU incorporation was assessed by the BrdU Flow Kit per manufacturer's instructions (BD Biosciences). Single cell suspensions were stained using fluorescently labeled antibodies, and data acquired on BD LSR II flow cytometer and analyzed using FlowJo software (Tree Star).

Gene Expression

$CD8^+$ T cells specific for the immunodominant ($NP_{366-374}$) and subdominant (GP_{33-41}, $PA_{224-233}$) influenza epitopes were sorted from the lungs of wild-type and PD-L1/L2 DKO mice on d8 following X31-GP33 infection. For microarray, RNA was extracted using RNAdvance tissue isolation kit (Agencourt) and amplified using the WT-Ovation One Direct System (NuGEN). Fragmented and labeled cDNA was hybridized to Affymetrix Mouse430_2 microarray. Microarray data were processed and analyzed as described previously (24, 37,

38). For qRT-PCR, WT and PD-1 KO P14 cells were sorted on d8 following infection, lysed, and RNA was isolated using RNeasy Mini Kit (Qiagen) followed by reverse transcription (Applied Biosystems). qRT-PCR was performed on the resulting cDNA using the Taqman Universal PCR Master Mix (Roche) and pre-designed primers and hydrolysis probes (IDT) for *Pdcd1, Rab4a, Prim2, CD200, Tcf4,* and *Dsp.* Data were normalized against the reference gene *Hprt1* and expressed as fold change in PD-1 KO P14 compared to WT P14.

Histopathology

Lungs were fixed in buffered 10% formalin (VWR), sectioned, and stained with hematoxylin and eosin (H&E). Sections were scored by a pathologist in a blinded fashion based on a standard 0-4 scale (0 = no inflammation, 1 = perivascular/peribronchial infiltration without involvement of alveolar space, 2 = focal infiltration of alveolar space with organization; 3 = bridging between foci, 4 = confluent inflammation with parenchymal destruction).

Statistical Analysis

Data were analyzed using Student's t-test for normally distributed data and *P* values <0.05 were considered significant.

Acknowledgments

I would like to acknowledge Keturah Brown (Sharpe Lab, Harvard Medical School) who has performed initial work on this project and established the phenotypic difference in memory defect with PD-1 deficiency. Pamela Odorizzi and Kristen Pauken (Wherry Lab, University of Pennsylvania Perelman Medical School) performed the P14 $CD8^+$ T cell transfer experiments. Kathleen Yates and Sabrina Imam (Haining Lab, Dana-Farber Cancer Institute) assisted with the microarray and RNA-sequencing experiments. Clinical scoring of H&E slides was done by Roderick T. Bronson (Rodent Histopathology Core, Harvard Medical School). I would also like to thank Justin Trombley (Sharpe Lab, Harvard Medical School) for technical support with experiment as well as Arlene H. Sharpe (Harvard Medical School) and E. John Wherry (University of Pennsylvania Perelman Medical School) for invaluable discussions and being closely involved with all parts of this study.

Chapter 3: CD39 expression identifies terminally exhausted CD8$^+$ T cells and limits immunopathology in chronic infection

Parts of this chapter have previously been published in:

Prakash K. Gupta*, **Jernej Godec***, David Wolski*, Emily Adland, Kathleen Yates, Cormac Cosgrove, Carola Ledderose, Wolfgang G. Junger, Simon C. Robson, E. John Wherry, Galit Alter, Philip J. R. Goulder, Paul Klenerman, Arlene H. Sharpe, Georg M. Lauer, W. Nicholas Haining. CD39 expression identifies terminally exhausted CD8$^+$ T cells. *PLOS Pathogens*. 2015 Oct 20;11(10):e1005177. (PMID: 26485519). *Co-first author.

Introduction

In acute infections, antigen-specific T cells differentiate into activated effector cells and then into memory T cells which rapidly gain effector functions and re-expand on subsequent encounter with the same pathogen (6). In contrast, during chronic infections, pathogen-specific T cells gradually lose effector functions, fail to expand, and can eventually become physically deleted (21). These traits are collectively termed T cell exhaustion, and have been described both in animal models of chronic viral infection as well as in human infections with hepatitis C virus (HCV) and human immunodeficiency virus (HIV) (21, 51, 113). Identifying reversible mechanisms of T cell exhaustion is therefore a major goal in medicine.

Prolonged or high-level expression of multiple inhibitory receptors such as PD-1, Lag3, and CD244 (2B4) is a cardinal feature of exhausted T cells in both animal models and human disease (23, 48, 114). Expression of PD-1 appears to be a particularly important feature of exhausted CD8$^+$ T cells, as the majority of exhausted cells in mouse models of chronic infection express this receptor, and blockade of the PD-1:PD-L1 axis can restore the

function of exhausted $CD8^+$ T cells in humans and mouse models (21, 48). However, in humans, many inhibitory receptors also can be expressed by a large fraction of fully functional memory $CD8^+$ T cells. PD-1, for instance, can be expressed by up to 60% of memory $CD8^+$ T cells in healthy individuals, making it challenging to use PD-1 to identify exhausted $CD8^+$ T cells in humans, particularly when the antigen-specificity of potentially exhausted $CD8^+$ T cells is not known (115).

Studies in mice and humans suggest that exhausted $CD8^+$ T cells are not a homogeneous population, but instead include at least two subpopulations of T cells that differentially express the transcription factors T-bet and Eomesodermin (Eomes) (49, 116, 117). T-bet^{high} $CD8^+$ T cells represent a progenitor subset with proliferative potential that give rise to $Eomes^{high}$ $CD8^+$ T cells, which are terminally differentiated and can no longer proliferate in response to antigen or be rescued by PD-1 blockade (49, 118). Both populations express PD-1, but $Eomes^{high}$ exhausted cells express the highest levels of PD-1. However, no specific cell-surface markers of this terminally differentiated population of exhausted cells have thus far been identified.

CD39 (*ENTPD1*) is an ectonucleotidase originally identified as an activation marker on human lymphocytes and as the vascular ecto-ADPase (119), but has subsequently been shown to be a hallmark feature of regulatory T cells (61, 63, 120). CD39 hydrolyzes extracellular ATP and ADP into adenosine monophosphate, which is then processed into adenosine by CD73, an ecto-5'-nucleotidase (121). Adenosine is a potent immunoregulator that binds to A2A receptors expressed by lymphocytes causing accumulation of intracellular cAMP, preventing T cell activation and NK cytotoxicity (122-124). Loss of CD39 in Tregs markedly impairs their ability to suppress T cell activation, suggesting that the juxtacrine activity of CD39 serves to negatively regulate T cell function (61). However, blood $CD8^+$ T cells have generally been reported to be $CD39^-$ (63-66), and the expression of this marker on exhausted T cells has not been examined.

In this study, we demonstrate that, in contrast to $CD8^+$ T cells from healthy donors, antigen-specific $CD8^+$ T cells responding to chronic viral infection in humans and a mouse model express high levels of biochemically active CD39. $CD39^+$ $CD8^+$ T cells co-express PD-1 and are enriched for a gene signature of T cell exhaustion. In the mouse model of chronic LCMV infection, high levels of CD39 expression demarcate terminally differentiated virus-specific $CD8^+$ T cells within the pool of exhausted $CD8^+$ T cells. Thus, CD39 provides a specific, pathological marker of exhausted $CD8^+$ T cells in chronic viral infection in humans and mouse models of chronic viral infection.

Results

CD39 is expressed by CD8$^+$ T cells responding to chronic infection

We surveyed the expression of CD39 by CD8$^+$ T cells from healthy adult subjects without chronic viral infection. Consistent with previous reports we found that only a small fraction (mean 6%) of CD8$^+$ T cells in healthy individuals expressed CD39 (Figure 3.1, A and B) (63-66). This small population of CD39$^+$ CD8$^+$ T cells in healthy donors was primarily found in the central and effector memory compartments while virtually no naive CD8$^+$ T cells expressed CD39 (Figure 3.2). We next focused on CD39 expression by antigen-specific CD8$^+$ T cells specific for latent viruses in healthy subjects and found that only a very small fraction of CMV- or EBV-specific CD8$^+$ T cells expressed CD39 (Figure 3.1, A and B) (mean 3% and 7% respectively).

We next measured CD39 expression by T cells specific for the chronic viral pathogens HCV and HIV. We measured CD39 expression in 57 subjects with acute HCV infections (23 with acute resolving infection and 34 with chronically evolving infection), and in 40 subjects with HIV infection (28 chronic progressors and 12 controllers; clinical characteristics of the subjects are summarized in Supplemental Tables 3.1 and 3.2). We found a mean of 51% of HCV-specific CD8$^+$ T cells and 31% of HIV-specific CD8$^+$ T cells expressed CD39, a number significantly higher than CD8$^+$ T cells specific for EBV or CMV, or in total CD8$^+$ T cell populations from healthy individuals (Figure 3.1, A and B). A slightly greater fraction of virus-specific CD8$^+$ T cells from HCV-infected subjects expressed CD39 than did those from HIV-infected subjects.

In subjects with chronic infection, the frequency of CD39-expressing cells in the virus-specific (tetramer$^+$) CD8$^+$ T cell population was significantly higher than in the total CD8$^+$ T cell population (Figure 3.1, C and D). However the fraction of total CD8$^+$ T cells expressing CD39 in the CD8$^+$ T cell compartment of individuals with HCV or HIV infection was slightly increased compared to healthy controls (Figure 3.1E), consistent with the presence of other,

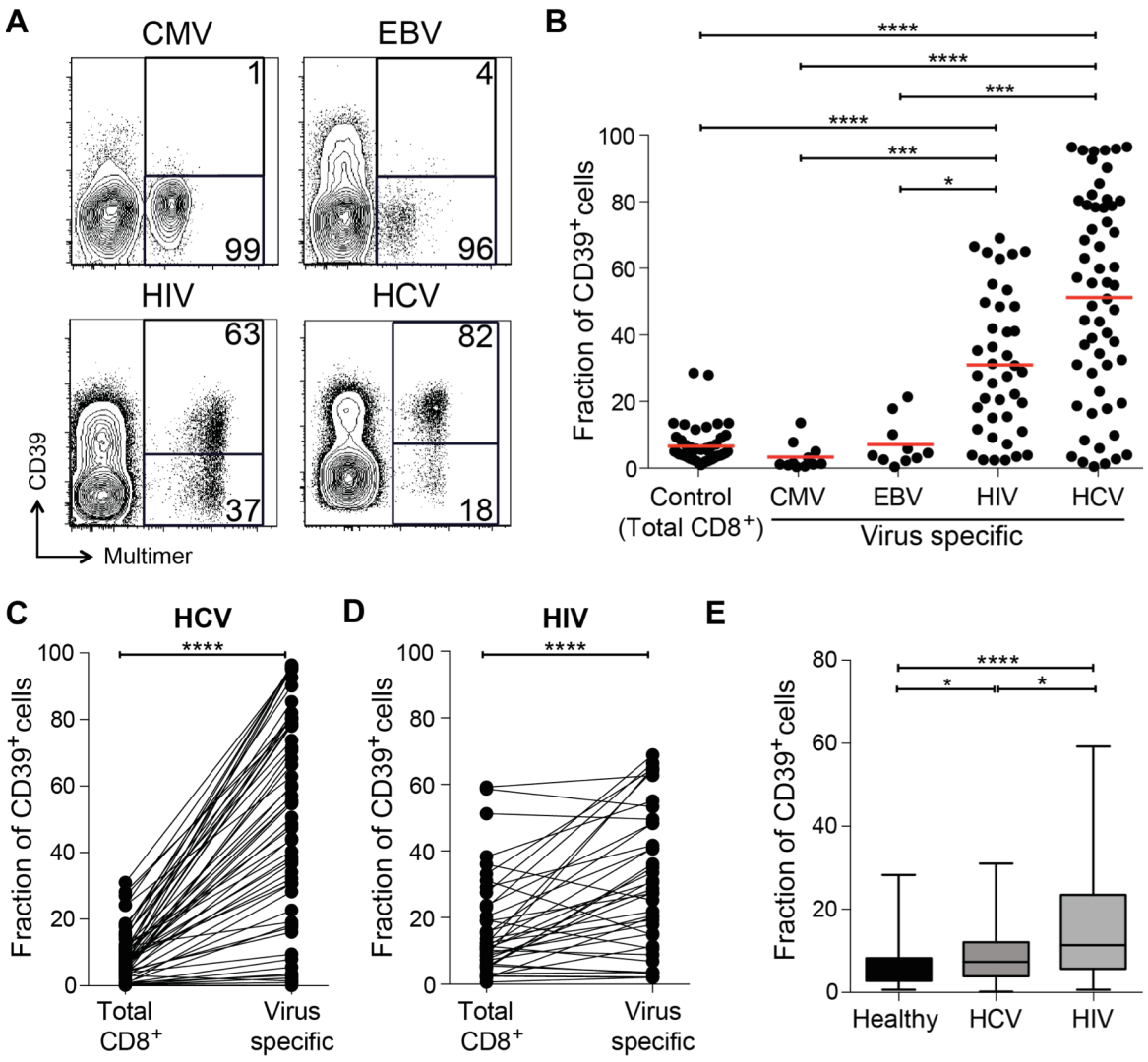

Figure 3.1. CD39 is highly expressed by virus-specific $CD8^+$ T cells in chronic viral infection. **(A)** Expression of CD39 by virus-specific $CD8^+$ T cells. Plots are gated on $CD8^+$. **(B)** Fraction of total or antigen-specific $CD8^+$ T cells expressing CD39. **(C, D)** Comparison of CD39 expression by total $CD8^+$ T cells with virus-specific $CD8^+$ T cells from patients with HCV (C) and HIV (D) infections. **(E)** Fraction of total $CD8^+$ T cells expressing CD39 in healthy, HIV or HCV infected donors. Error bars represent SEM. Statistical significance was assessed by Kruskal-Wallis test (B, E), or Wilcoxon test (C, D). **P* <0.05, ****P* <0.001, *****P* <0.0001.
The data in this figure were generated with the help of Prakash Gupta, David Wolski, and Kathleen Yates.

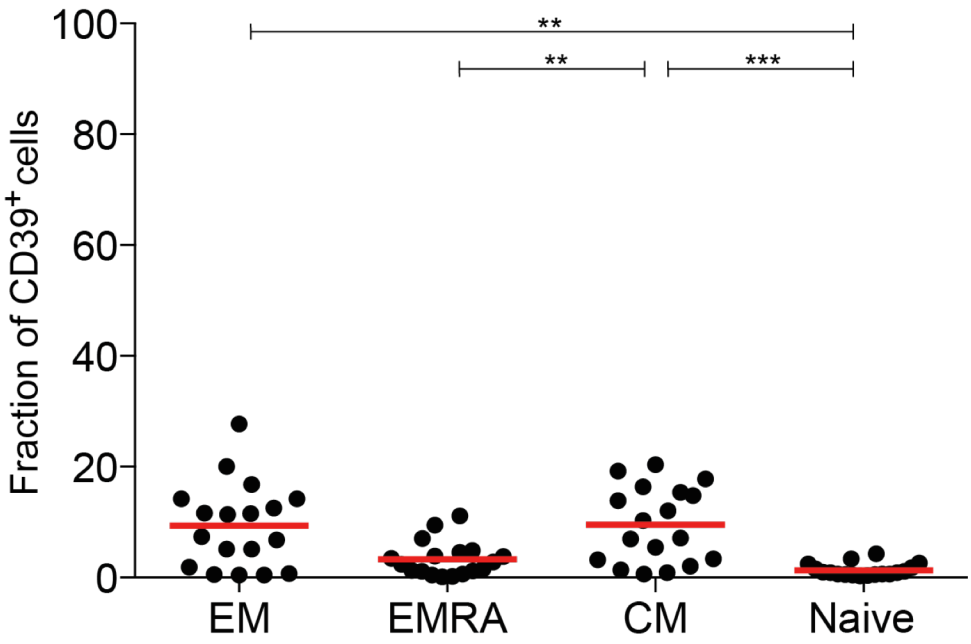

Figure 3.2. CD39 is expressed by few CD8$^+$ T cells in health donors. Fraction of CD39$^+$ cells in naive CD8$^+$ T and central memory (CM), effector memory (EM) and effector memory RA$^+$ (EMRA) subpopulations of CD8$^+$ T cells based on CD45RA and CCR7 staining from 18 healthy human donors. Error bars represent SEM. Statistical significance was assessed by Friedman test. ***P* <0.01, ****P* <0.001.

The data in this figure were generated with the help of Prakash Gupta, David Wolski, and Kathleen Yates

unmeasured virus-specific $CD8^+$ T cells that were also $CD39^+$ in the $tetramer^-$ fraction of $CD8^+$ T cells. Thus CD39 is expressed infrequently by $CD8^+$ T cells in healthy donors, but marks a large fraction of pathogen-specific cells $CD8^+$ T cells in patients with chronic infection.

CD39 expressed by $CD8^+$ T cells hydrolyzes ATP

CD39 expressed by regulatory T cells catalyzes the hydrolysis of ADP to 5'-AMP (61, 63, 120) but its enzymatic activity can be regulated by a range of post-transcriptional mechanisms [PMID. We therefore tested CD39 expressed by $CD8^+$ T cells from patients infected with chronic HCV was functional using ATP hydrolysis as a surrogate marker of CD39 activity (125-127). We sorted $CD39^-$ and $CD39^+$ $CD8^+$ T cells from six HCV-infected individuals (four with chronic infection and two with resolved infection) and incubated equal numbers of cells in the presence of extracellular ATP (eATP)**.** Remaining levels of eATP were measured in the supernatant by HPLC. As a control, we assessed ATP hydrolysis by $CD4^+$ $CD25^+$ $CD39^+$ regulatory T cells (Tregs) sorted from the same individuals (Figure 3.3A).

Within the $CD39^+$ $CD8^+$ T cell population the level of CD39 expression was lower than in Tregs (Figure 3.3B). Consistent with reduced CD39 expression relative to Tregs, ATP hydrolysis by $CD39^+$ $CD8^+$ T cells was less than that by Tregs (Figure 3.3C). However ATP hydrolysis by $CD39^+$ $CD8^+$ T cells was significantly greater than that of $CD39^-$ cells (Figure 3.3C). Thus CD39 expressed by $CD8^+$ T cells in HCV infection is enzymatically active and capable of hydrolyzing ATP.

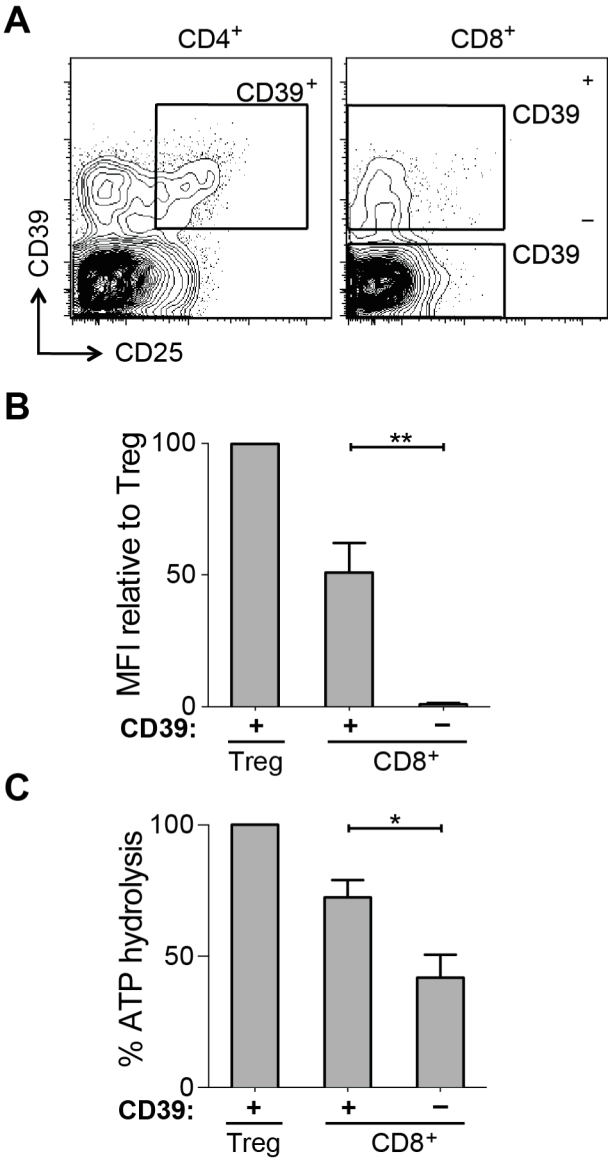

Figure 3.3. CD39 expressed by CD8$^+$ T cells in HCV infection is enzymatically active. (A) Flow cytometry sorting gates of CD39$^+$ and CD39$^-$ CD8$^+$ T cells and CD39$^+$ CD25$^+$ CD4$^+$ Tregs used for rpHPLC analysis of CD39 activity**. (B)** Summary of CD39 expression level relative to Tregs in the same subjects. **(C)** ATP hydrolysis by CD8$^+$ T cell populations relative to Tregs. Data represent 6 patients with chronic HCV infection. Error bars represent SEM. Statistical significance was assessed by paired Student's t-test (B, C). *P <0.05, **P <0.01.
The data in this figure were generated by Prakash Gupta and Carola Ledderose.

CD39 is co-expressed with PD-1 on virus-specific CD8$^+$ T cells and correlates with viral load in both HCV and HIV infection

CD8$^+$ T cells specific for chronic viruses such as HCV and HIV express increased levels of PD-1 (51, 128). We therefore examined the relationship between CD39 and PD-1 expression by virus-specific CD8$^+$ T cells in 54 patients with HCV (23 chronically infected and 31 resolvers) and 40 patients infected with HIV (28 chronic progressors, 7 viraemic controllers and 5 elite controllers). In both diseases we found a significant association between the level of expression (mean fluorescence intensity, MFI) of CD39 and PD-1 on antigen-specific CD8$^+$ T cells in subjects with HCV and with HIV ($r=0.70$, $P<0.0001$ and $r=0.54$, $P<0.05$, respectively) (Figure 3.4, A and B).

We next examined the relationship between CD39 and PD-1 expression and viral load in HCV and HIV infection. We found that in both the HCV and HIV infection there was a modest but significant correlation between viral load and the level of CD39 expression on virus-specific CD8$^+$ T cells measured by MFI (Figure 3.4C). The fraction of CD39$^+$, virus-specific CD8$^+$ T cells was significantly higher in HIV progressors compared with those from HIV controllers (Figure 3.5). A similar, but non-significant, trend was seen comparing CD39 expression in HCV-specific CD8$^+$ T cells in patients with chronic versus resolved disease. However, in HCV, a significantly higher fraction of virus-specific CD8$^+$ T cells co-expressed both CD39 and PD-1 in patients with chronic versus resolved disease (Figure 3.5). Consistent with these findings, there was a significant correlation between viral load and the fraction of virus-specific CD8$^+$ T cells that were CD39$^+$ PD-1$^+$ double positive in both HCV and HIV infection (Figure 3.5). PD-1 expression was also modestly correlated with the viral load in HCV and in HIV-infected patients (Figure 3.4D) (51, 128). Thus CD39 expression by virus-specific CD8$^+$ T cells is greatest in setting of high antigen burden.

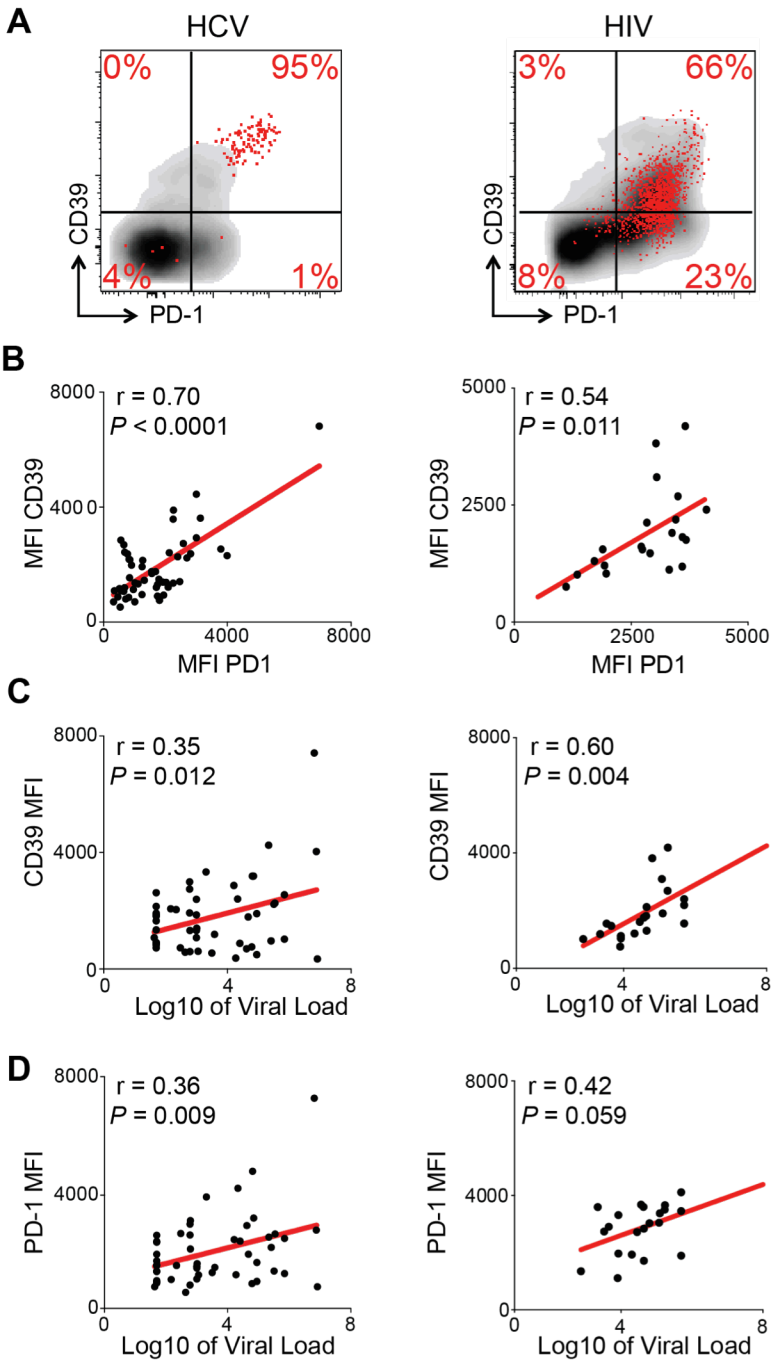

Figure 3.4. CD39 expression correlates with PD-1 expression and viral load in chronic viral infection. (A) CD39 and PD-1 expression in chronic HCV (left) or HIV infection (right). Representative plots demonstrate total (gray) and virus-specific (red) $CD8^+$ T cells. **(B)** Correlation between CD39 and PD-1 expression by HCV- (left) and HIV-specific (right) $CD8^+$ T cells. **(C)** Correlation between CD39 expression by virus-specific $CD8^+$ T cells and viral load count in HCV (left) or HIV (right) infection. **(D)** Correlation between PD-1 expression by virus-specific $CD8^+$ T cells and viral load in HCV (left) or HIV (right) infection. Correlation was assessed by Pearson correlation coefficient (B, C, D). MFI; mean fluorescence intensity.
The data in this figure were generated with the help of Prakash Gupta, David Wolski, and Kathleen Yates.

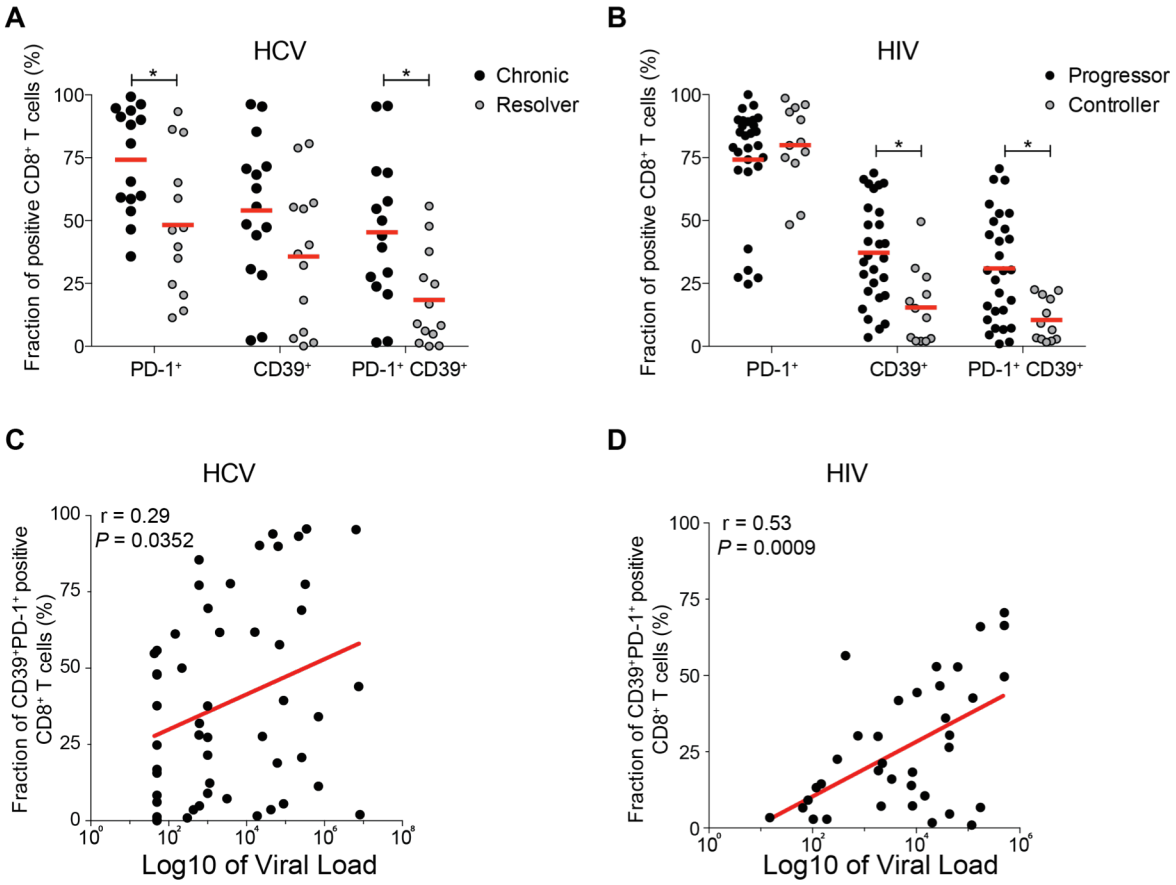

Figure 3.5. CD39 and PD-1 co-expression in HCV and HIV. (A, B) Fraction of HCV-specific (A) and HIV-specific (B) $CD8^+$ T cells expressing PD-1, CD39, or both in patients with persistent high viral load (black) or patients controlling the disease (grey). Correlation of the fraction of PD-1 and CD39 double positive virus specific $CD8^+$ T cells with the viral load in the blood in HCV (C) and HIV (D) infected patients. Statistical significance was assessed by Mann-Whitney test with Bonferroni correction (A, B). *P <0.05. Correlation was assessed by Pearson correlation coefficient (C, D). MFI; mean fluorescence intensity.
The data in this figure were generated with the help of Prakash Gupta, David Wolski, and Kathleen Yates

Transcriptional analysis of CD39$^+$ CD8$^+$ T cells in HCV infection

In order to characterize more broadly the phenotype of CD39$^+$ CD8$^+$ T cells from individuals with chronic infection, we compared the global gene expression profiles of sorted CD39$^+$ and CD39$^-$ CD8$^+$ T cells from 8 HCV-infected subjects (3 with acute resolving infection and 5 with chronically evolving infection. Limited numbers of cells precluded the comparison of CD39$^+$ and CD39$^-$ CD8$^+$ T cells within HCV-specific cells, leading us to focus on the total CD8$^+$ population of antigen-experienced CD8$^+$ T cells. Because naive CD8$^+$ T cells express little CD39 (Figure 3.2), we excluded this population from the sorted cells (Figure 3.6) to enable direct comparison of antigen-experienced CD39$^+$ and CD39$^-$ CD8$^+$ T cells.

We first used unbiased clustering approaches to identify whether CD39$^+$ and CD39$^-$ CD8$^+$ T cells showed distinct patterns of gene expression. Analysis of gene expression profiles using consensus hierarchical clustering (Figure 3.7A) showed two distinct clusters of samples that corresponded almost exactly to CD39$^+$ and CD39$^-$ populations, suggesting that that in both acute and chronic infection, CD39 expression demarcates two types of CD8$^+$ T cells with markedly different patterns of gene expression. Supervised analysis of differential gene expression identified 619 genes differentially expressed (FDR<0.15) between CD39$^+$ and CD39$^-$ CD8$^+$ T cells. Inspection of the list of differentially expressed genes revealed many with known roles in CD8$^+$ T cell biology including increased expression of the inhibitory receptors PD-1 and CTLA-4 in CD39$^+$ CD8$^+$ T cells.

To identify biological processes that were differentially active in CD39$^+$ vs. CD39$^-$ cells, we performed gene set enrichment analysis using the Gene Ontology collection of gene sets (99). We found no significant enrichment of GO terms in the CD39$^-$ CD8$^+$ subset. In contrast, 21 gene sets significantly enriched (FDR<0.1) in CD39$^+$ population, almost all of which were related to mitosis and cell-cycle associated genes or cytoskeleton organization

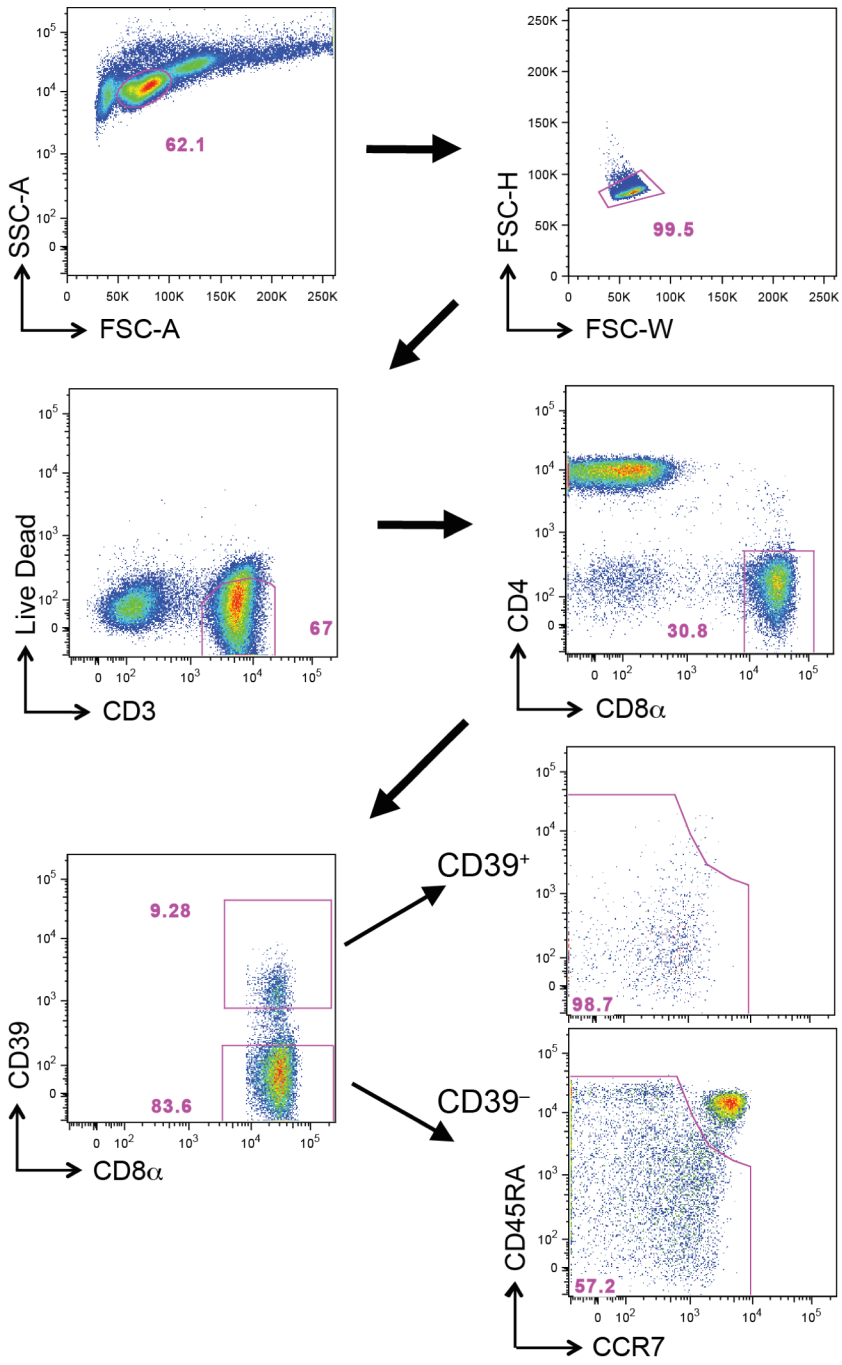

Figure 3.6. Cell sorting strategy for microarray analysis. Gating strategy for $CD39^+$ and $CD39^-$ live non-naive $CD8^+$ T cells from HCV-infected patients.
The data in this figure were generated with the help of Prakash Gupta.

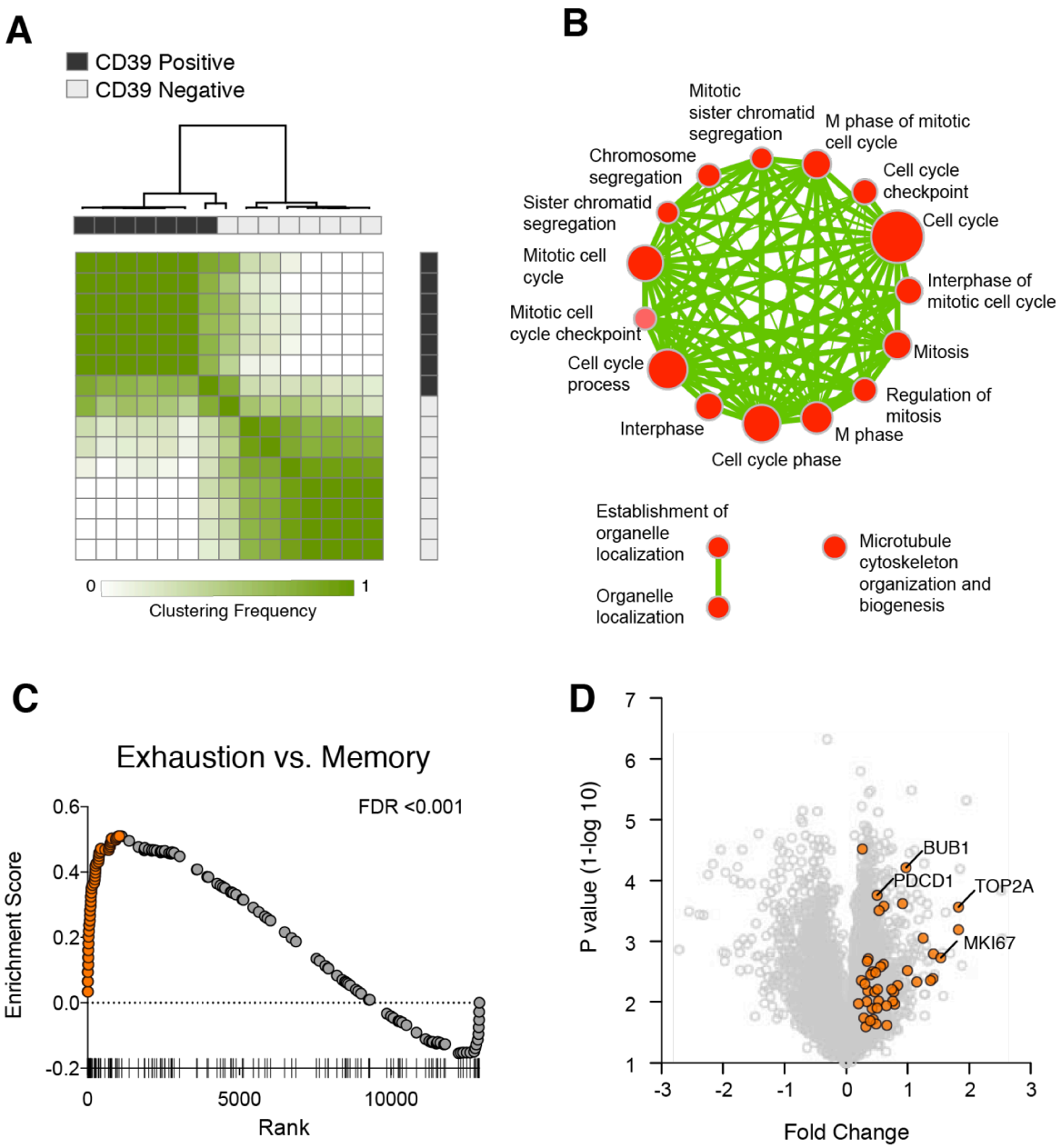

Figure 3.7. Transcriptional analysis of CD39$^+$ and CD39$^-$ CD8$^+$ T cells in HCV infection. (A) Consensus hierarchical clustering of expression profiles from CD39$^+$ (black) and CD39$^-$ (grey) CD8$^+$ T cells from 8 HCV infected patients. Clustering is based on the top 10% of genes by variance across the dataset. Sample similarity (1-Pearson correlation coefficient) is annotated with color from low (white) to high (green). **(B)** Gene set enrichment map displaying Gene Ontology gene sets enriched (FDR < 0.1) in CD39$^+$ CD8$^+$ T cells from (A). Nodes (in red) are sized in proportion to gene set size; connecting line thickness represents extent of gene member overlap between gene sets. **(C)** Gene set enrichment analysis of a signature of 200 genes up-regulated in exhausted CD8$^+$ T cells from the mouse model of chronic viral infection versus acute infection (day 30 post infection) in the ranked list of genes differentially expressed by CD39$^+$ vs. CD39$^-$ CD8$^+$ T cells. Leading edge genes are indicated by orange symbols. **(D)** Volcano plot of all genes (grey) or exhausted leading edge genes (orange) from (C).

The data in this figure were generated with the help of Prakash Gupta, Kathleen Yates, and W. Nicholas Haining.

(Figure 3.7B). This suggests that $CD39^+$ $CD8^+$ T cells in chronic viral infection show coordinate up-regulation of genes related to proliferation.

The expression of CD39 by $CD8^+$ T cells in chronic but not acute/latent infection, suggests that it may be a marker of T cell exhaustion. We therefore tested whether the profile of $CD39^+$ $CD8^+$ T cells was enriched for genes expressed by exhausted $CD8^+$ cells. Previous studies of gene expression in $CD8^+$ T cells in the mouse model of chronic viral infection with the Clone 13 strain of LCMV have identified signatures of T cell exhaustion that are also enriched in exhausted $CD8^+$ T cells in humans (75, 129, 130). We therefore curated a signature of 200 genes up-regulated by exhausted $CD8^+$ T cells responding to chronic infection relative to functional memory $CD8^+$ T cells generated by acute infection (LCMV Armstrong strain). We found that the exhausted $CD8^+$ T cell signature from LCMV model was significantly enriched in $CD39^+$ vs. $CD39^-$ $CD8^+$ T cells in subjects with HCV infection (Figure 3.7C). We focused on the "leading edge" genes contributing most to the enrichment (78), which correspond to genes up-regulated both in the mouse exhausted signature and in the human $CD39^+$ profile. As expected, the leading edge genes included PD-1 (*PDCD1*), a feature of both human $CD39^+$ $CD8^+$ T cells and of exhausted $CD8^+$ T cells in the mouse model (Figure 3.7D). In addition we found that up-regulation of many genes associated with proliferation including *BUB1*, *TOP2A* and *MKI67* was common to mouse exhausted $CD8^+$ T cells and human $CD39^+$ $CD8^+$ T cells. Thus $CD39^+$ $CD8^+$ T cells in HCV infection and exhausted $CD8^+$ T cells in a mouse model of chronic infection share transcriptional features that include genes related to proliferation.

CD39 is increased in exhausted $CD8^+$ T cells in the mouse model of chronic LCMV infection.

Because the mouse signature of $CD8^+$ T cell exhaustion was significantly enriched in the transcriptional profile of $CD39^+$ $CD8^+$ T cells in HCV-infected patients, we next asked if

CD39 was up-regulated by $CD8^+$ T cells in the mouse model of chronic viral infection. To address this question we compared two well-described mouse models of viral infection using two strains of LCMV: LCMV Armstrong that causes an acute infection that is resolved in up to 8 days; and LCMV Clone 13 that persists in mice for up to 3 months and leads to T cell exhaustion (23, 48).

We measured CD39 expression and compared it to PD-1 expression in $CD8^+$ T cells responding to each infection. While naive $CD8^+$ T cells expressed neither CD39 nor PD-1 (Figure 3.8A), both were rapidly and coordinately up-regulated by antigen-experienced cells following either infection at d7 p.i. (Figure 3.8B). However, in acute infection, the fraction of CD39 bright PD-1^+ population decreased with time. In contrast, high expression of CD39 and PD-1 was maintained in Clone 13 infection. The accumulation of CD39 bright PD-1^+ cells among the total $CD8^+$ population was most apparent in the H-$2D^b$ $GP_{276-286}$ tetramer-specific $CD8^+$ T cells (Figure 3.8B).

Thus after chronic viral infection, antigen-specific $CD8^+$ T cells can be identified by high expression of both CD39 and PD-1. This difference in expression of both markers between chronic and acute infection is noticeable as early as d7 p.i. but becomes more pronounced with time after infection.

CD39 expression correlates with a terminally exhausted phenotype in virus-specific $CD8^+$ T cells in chronic infection.

Having determined that high, persistent expression of CD39 is a feature of LCMV-specific $CD8^+$ T cells during chronic LCMV infection, we next sought to further characterize the phenotype of $CD39^+$ $CD8^+$ T cells during Clone 13 infection. We analyzed CD39 expression in antigen-experienced, $CD44^+$ $CD8^+$ T cells and found that mice infected with Clone 13 developed a population of cells with particularly high expression of CD39 ($CD39^{high}$). This population was entirely absent in mice infected with the acute LCMV

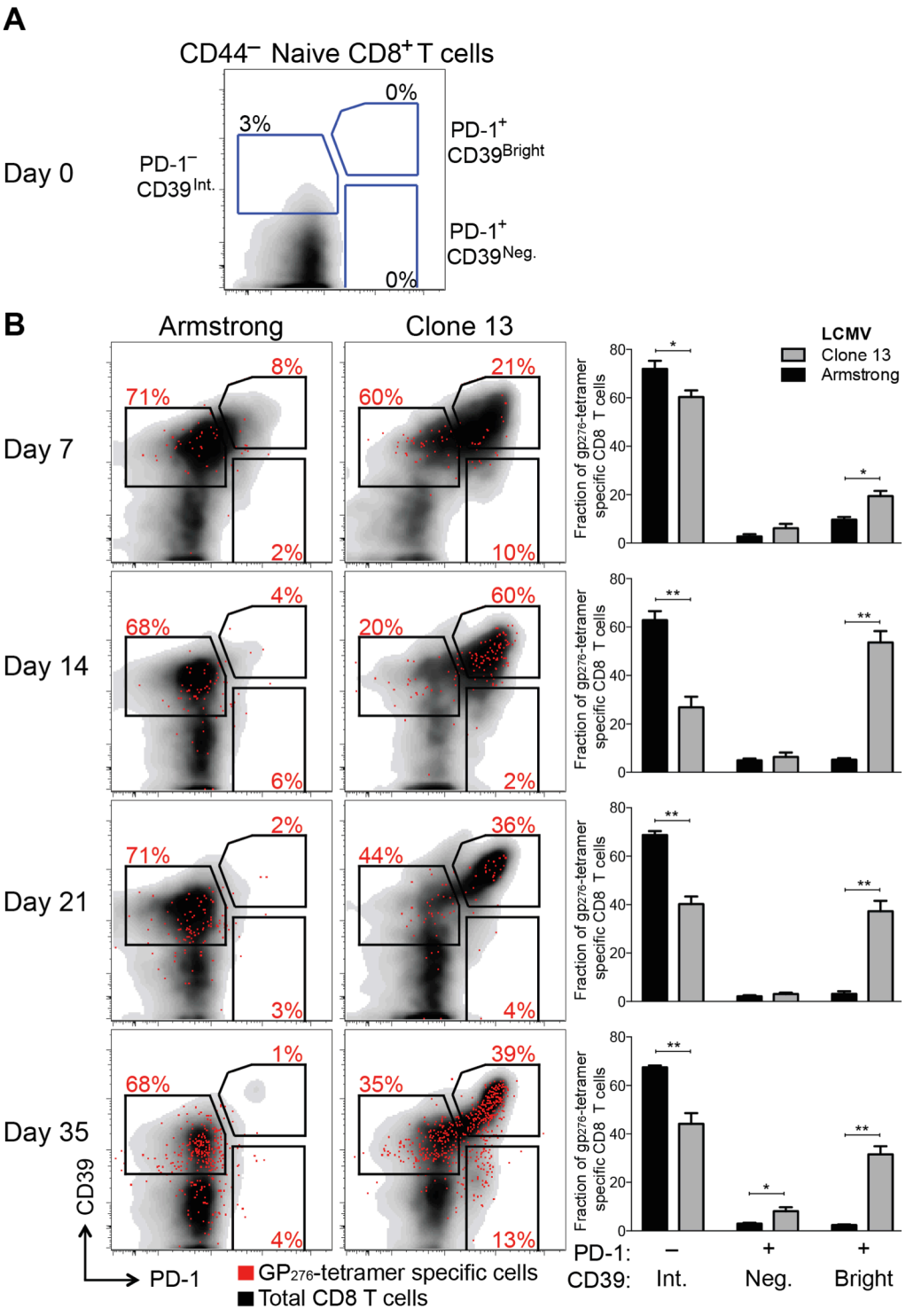

Figure 3.8. CD39 is highly up-regulated by exhausted CD8$^+$ T cells in a mouse model of chronic infection. (**A, B**) Expression of CD39 and PD-1 by CD44$^-$ naive mouse CD8$^+$ T cells (A) and in CD8$^+$ T cells at indicated times following LCMV Armstrong (acute) or Clone 13 (chronic) infection (B). Representative plots show total (black) and H-2Db $GP_{276-286}$ tetramer-specific CD8$^+$ T cells (red). Summary of results in 5 mice per group is shown in bar graphs on the right. Statistical significance was assessed with Mann-Whitney test. *P < 0.5, **P < 0.01.

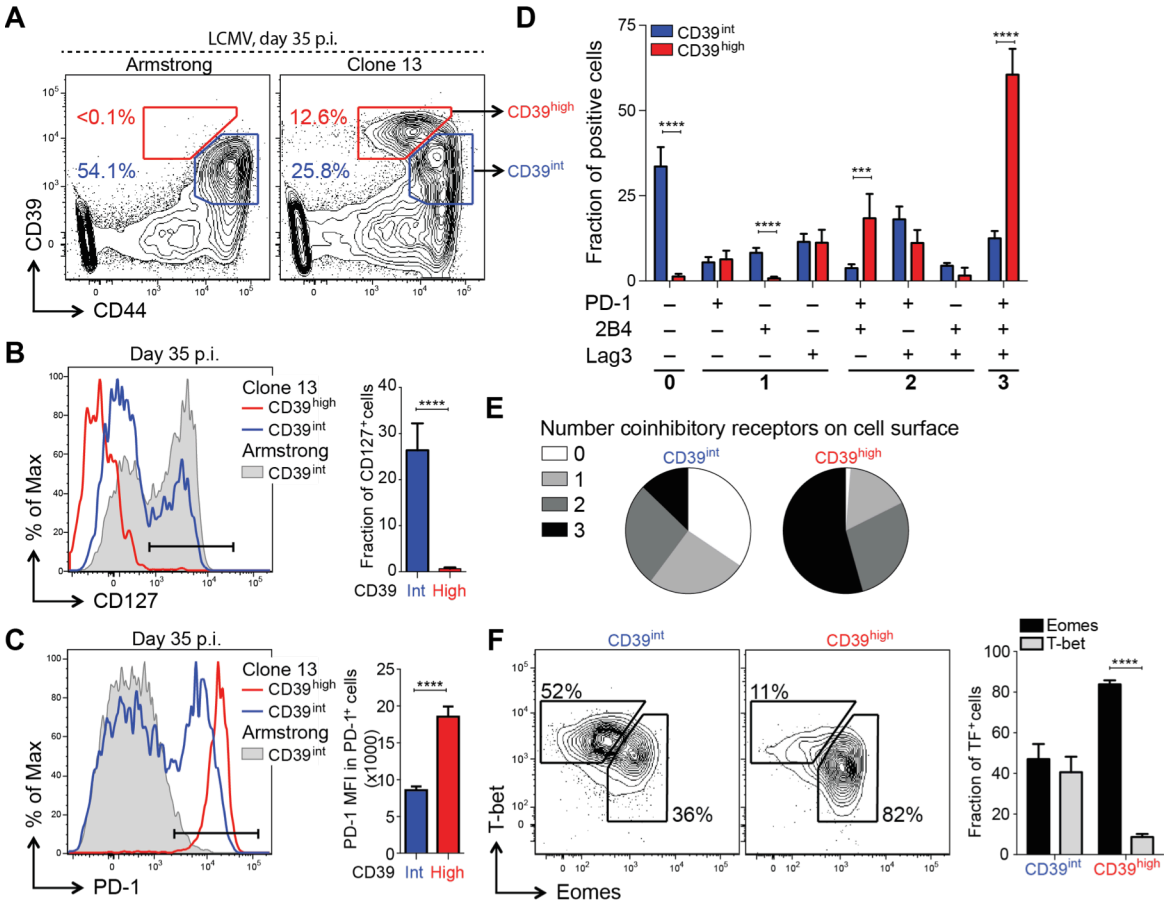

Figure 3.9. CD39 identifies terminally exhausted CD8$^+$ **T cells in mice with chronic LCMV infection.** (**A**) Expression of CD39 and CD44$^+$ by mouse CD8$^+$ T cells 30-35 days following LCMV Armstrong (left) or Clone 13 (right) infection. (**B**, **C**) Representative histograms (left) of CD127 (B) and PD-1 (C) expression by CD39^{high} and CD39^{int} CD8$^+$ T cells from Clone 13 (red and blue, respectively) and CD39^{int} from Armstrong (filled gray) infected mice on d35 p.i. (left). Fraction of CD127$^+$ (B) and MFI of PD-1 in PD-1$^+$ cells (C) is shown on the right. Results are from 5 mice. (**D**) Fraction of CD39^{high} and CD39^{int} CD44$^+$ CD8$^+$ T cells expressing different combinations of co-inhibitory receptors PD-1, 2B4, and Lag3. (**E**) Average number of co-inhibitory receptors expressed by CD39^{int} (left) or CD39^{high} (right) CD8$^+$ T cells at d35 following LCMV Clone 13 infection. (**F**) Representative plots of T-bet and Eomes expression in CD39^{int} (left) and CD39^{high} (right) cells as in (A). Summary of results is shown on the right. Data are representative of three experiments of 5 mice per group. Statistical significance was assessed with Student's t-test (B, C, F) with Holm-Sidak multiple comparison correction (D). $^{**}P < 0.01$, $^{****}P < 0.0001$.

Armstrong strain at d35 p.i., which only exhibited the presence of intermediate levels of CD39 staining ($CD39^{int}$) (Figure 3.9A). Further characterization of the two sub-populations in Clone 13 infected mice revealed that the $CD39^{high}$ cells showed more down-regulation of CD127 (Figure 3.9B) and higher expression of PD-1 (Figure 3.9C) than did the $CD39^{int}$ population.

Because the highest levels of PD-1 are characteristic of terminally exhausted $CD8^+$ T cells in chronic infection (118, 131), we tested whether $CD39^{high}$ T cells in chronic infection showed other phenotypic characteristics of terminal exhaustion. Analysis of expression of two additional co-inhibitory receptors, CD244 (2B4) and Lag3, showed that a significantly higher fraction of $CD39^{high}$ cells co-expressed multiple receptors, consistent with terminal exhaustion. In contrast, $CD39^{int}$ $CD8^+$ T cells were generally negative for all three receptors analyzed (Figure 3.9, D and E). We next examined the expression of the transcription factors T-bet and Eomes. We found that the $CD39^{high}$ subset of $CD8^+$ T cells was comprised primarily of $Eomes^{high}$ $T\text{-}bet^{low}$ terminally exhausted phenotype, while the $CD39^{int}$ $CD8^+$ T cells showed a comparable distribution of both (Figure 3.9F). Similarly, we found that in $CD8^+$ T cells from subjects with either HCV or HIV infection, the $CD39^+$ $CD8^+$ T cell compartment contained a significantly higher ratio of $Eomes^{high}$ $T\text{-}bet^{low}$: $Eomes^{low}$ $T\text{-}bet^{high}$ relative to $CD39^-$ $CD8^+$ T cells (Figure 3.10). Thus in both humans and mice with chronic viral infection, $CD39^+$ $CD8^+$ T cells show a phenotype consistent with previous descriptions of terminal exhaustion (49).

CD39 correlates with reduced functionality in virus-specific $CD8^+$ T cells in chronic infection.

We next examined the functional properties of $CD39^{high}$ and $CD39^{int}$ $CD8^+$ T cells from mice with chronic LCMV infection. Co-production of cytokines IFN-γ and TNF-α is a feature of virus-specific T cells responding to acute infection and in the early stages of chronic

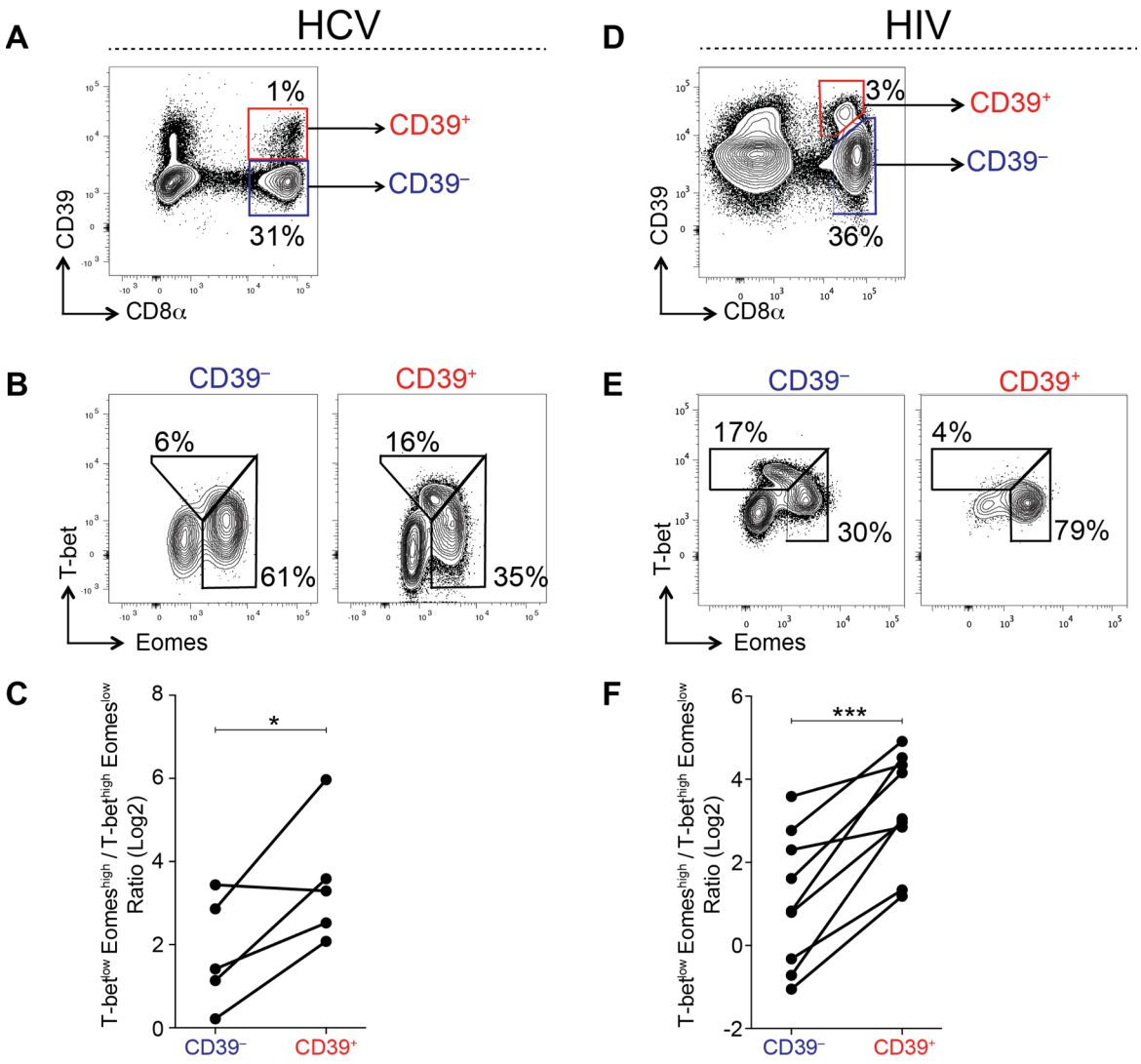

Figure 3.10. Comparison of T-bet and Eomes expression by CD39$^+$ and CD39$^-$ CD8$^+$ T cells in patients with chronic viral infection. (**A, D**) Expression of CD39 in CD8$^+$ T cells in patients infected with HCV (A) and HIV (D). (**B, E**) Expression of transcription factors T-bet and Eomes on CD39$^-$ and CD39$^+$ populations identified in (A) and (D). (**C, F**) Summary of the ratio of terminally exhausted Eomes^{high}/T-bet^{low} CD8$^+$ T cells in CD39$^-$ and CD39$^+$ subsets in HCV (C) and HIV (F) infection. Statistical significance was assessed with paired Student's t-test. *P < 0.05, ***P < 0.001. The data in this figure were generated by Kathleen Yates and Cormac Cosgrove.

infection but is progressively lost as exhaustion evolves (21). To compare the functionality of $CD39^{high}$ and $CD39^{int}$ virus-specific $CD8^+$ T cells, we isolated $CD8^+$ T cells from mice with chronic infection at d35 post-infection and stained for IFN-γ and TNF-α following in vitro stimulation with GP_{33-41} peptide. We found a significantly smaller fraction of antigen-specific coproduced IFN-γ and TNF-α in $CD39^{high}$ $CD8^+$ T cells compared to $CD39^{int}$ $CD8^+$ T cells (Figure 3.11, A and B).

To confirm this finding, we analyzed the function of transferred P14 $CD8^+$ T cells in chronic infection. The P14 TCR transgene recognizes the GP_{33-41} peptide of LCMV presented on $H\text{-}2D^b$. We found that both the frequency of IFN-γ-producing and IFN-γ/TNF-α co-producing P14 T cells was significantly lower in $CD39^{high}$ $CD8^+$ T cells compared to $CD39^{int}$ $CD8^+$ T cells (Figure 3.11, C and D). The defect in cytokine secretion was not only observed in terms of the frequency of cytokine-secreting cells, but also in the amount of cytokine detected per cell. Even among cells that did secrete IFN-γ, we found the MFI of expression to be significantly lower in $CD39^{high}$ $CD8^+$ T cells compared to $CD39^{int}$ $CD8^+$ T cells (Figure 3.11, E and F). Thus high levels of CD39 expression demarcate a population of exhausted cells with the poorest function in chronic infection.

CD39 limits lethal immunopathology associated with Clone 13 infection in mice.

Our data has shown that CD39 expression on $CD8^+$ T cells tightly correlates with the exhausted state of $CD8^+$ T cells in chronic infection in both humans and mice. To examine the functional role for CD39 during chronic viral infection we infected CD39 deficient mice with LCMV Clone 13 (132). It has been reported previously that in the absence of other immunoregulatory mechanisms, such as PD-1:PD-L1 pathway, mice succumb to LCMV Clone 13 several days following infection (48, 84). Similarly, in the absence of CD39, the majority of animals died within 10 days of infection (Figure 3.12A). This mortality could be attributed to enhanced immune infiltration to target organs such as liver and the lung

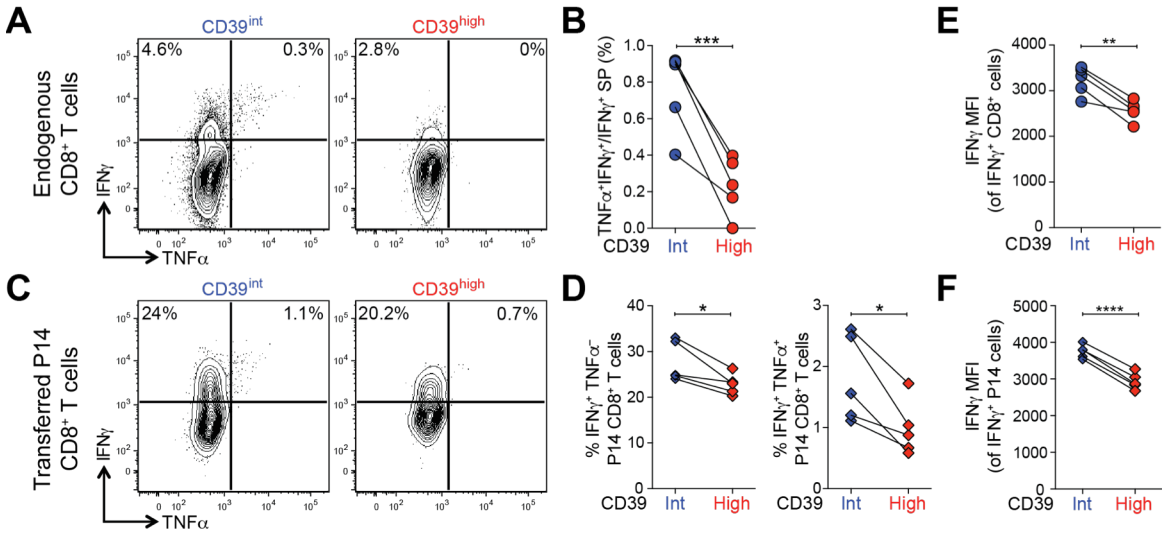

Figure 3.11. Terminally exhausted $CD8^+$ **T cells marked by high levels of CD39 are most impaired in their effector function.** (**A**) Representative plots showing the production of IFN-γ and TNF-α in endogenous $CD44^+$ $CD39^{int}$ or $CD39^{high}$ $CD8^+$ T cells d36 following LCMV Clone 13 infection. (**B**) Quantification of cells in (A) that produce both TNF-α and IFN-γ relative to IFN-γ only. (C, D) 500 congenic P14 cells were transferred i.v. at d-1 and their ability to produce cytokines at d36 p.i. with LCMV Clone 13 is shown in representative flow plots (C) gated as in (A) and the percentages of IFN-γ and TNF-α is summarized in (D). (E, F) Mean fluorescence intensity (MFI) of IFN-γ in IFN-γ positive endogenous (E) and transferred P14 cells (F). Statistical significance was assessed with paired Student's t-test. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.
The data in this figure were generated with the help of Kristen Pauken.

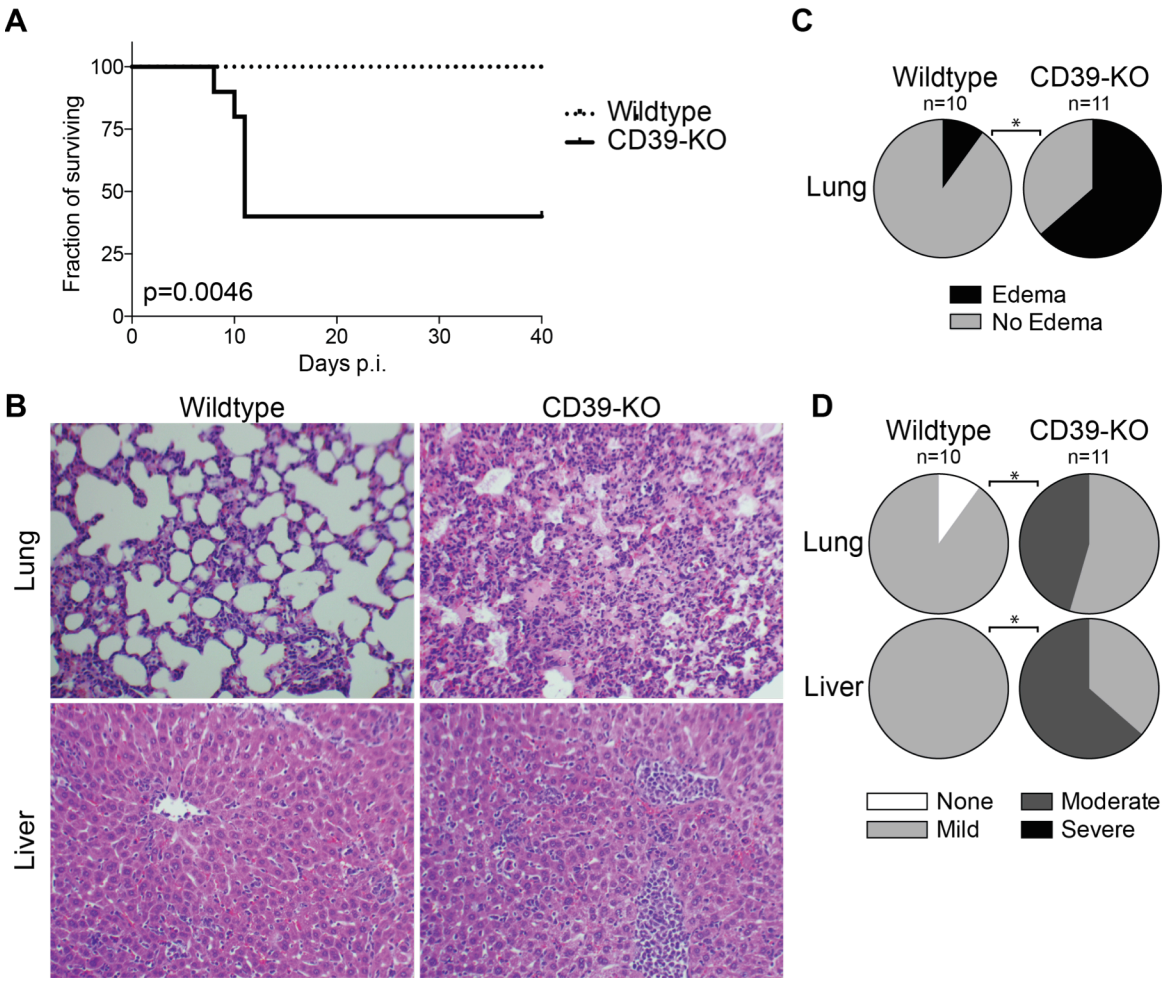

Figure 3.12. CD39 deficient mice exhibit increased mortality to LCMV Clone 13 due to exacerbated immune response in target organs. (**A**) Survival of CD39 knockout (CD39-KO) and wild-type mice following LCMV Clone 13 infection. (**B-D**) Histopathological analyses (H&E) of target organs at d8 following LCMV Clone 13 infection. (B), representative lung (top) and liver (bottom) images with quantification of the severity of leukocyte infiltration in lung (D, top) and liver (D, bottom) and the severity of pulmonary edema (C). Statistical analysis was done using Mantel-Cox test (A), Chi-square test (C), or Mann Whitney test (D), * $P<0.05$. Data are representative of two independent experiments with 4-6 mice per group.

examined at d8 following infection (Figure 3.12, B and C). Additionally, LCMV Clone 13 infected CD39-KO animals suffered from pulmonary edema (Figure 3.12D), similar to what has been described in mice lacking PD-1 (84). Our preliminary data suggest that the viral clearance is not altered in absence of CD39 (data not shown). These data highlight the essential role of the purinergic pathway in protecting the organism from lethal complications following a virulent infection.

CD39 expression on CD8$^+$ T cells is regulated by TGF-β.

The fate of activated effector CD8$^+$ T cells is tightly regulated by the microenvironment in which they are activated, which may include other cells and a variety of cytokines (27). We, therefore, examined how cytokines influenced CD39 expression on CD8$^+$ T cells. We first noted that activation of the mouse CD8$^+$ T cells through TCR and CD28 alone was not sufficient to induce expression of CD39 (Figure 3.13). However, if the CD8$^+$ T cells were stimulated in the presence of heterogeneous population of splenocytes, there was some up-regulation of CD39 on CD8$^+$ T cells, suggesting an additional role for the cytokines or intercellular contacts that contribute to CD39 induction (Figure 3.13, A and B). To test whether a specific cytokine was inducing CD39 expression, we stimulated purified CD8$^+$ T cells with anti-CD3 and anti-CD28 antibodies in the presence of IL-2 and a variety of cytokines including IL-1β, IL-6, IL-7, IL-12, IL-15, TGF-β1, and GM-CSF. TGF-β1 turned out to strongly induce the expression of CD39 on CD8$^+$ T cells soon after the activation (Figure 3.13, C and D). While IL-12 also showed a limited capacity to induce CD39 on CD8$^+$ T cells, no other cytokine was able to promote CD39 expression (Figure 3.13, C and D). Interestingly, Interestingly, TGF-β is critical for the differentiation of exhausted CD8$^+$ T cells during chronic viral infection (133), consistent with our results that it may be inducing exhaustion-specific high expression of CD39 in CD8$^+$ T cells.

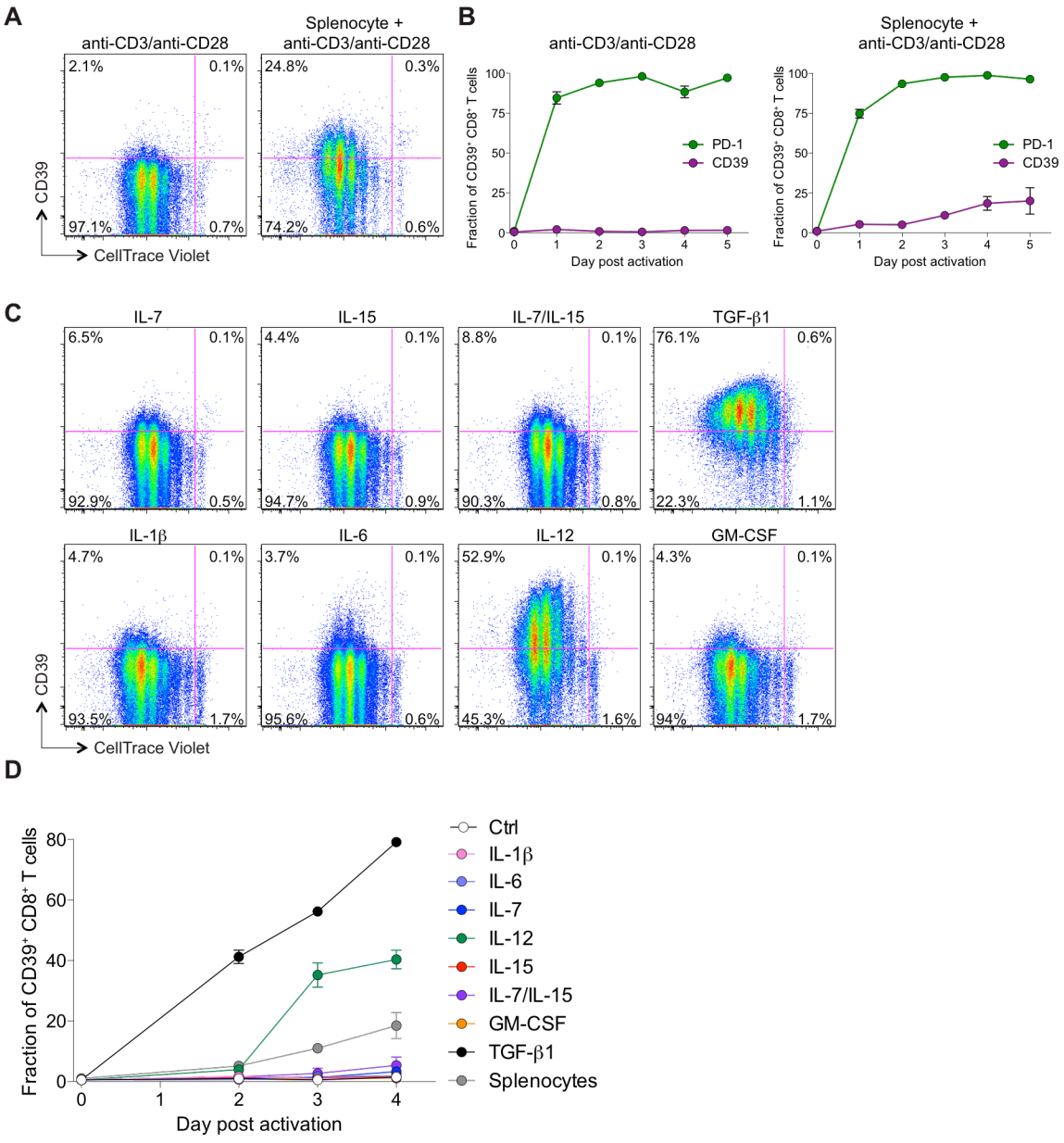

Figure 3.13. TGFb1 drives CD39 expression in activated CD8$^+$ T cells. (**A**) Expression of CD39 and CellTrace Violet dilution in CD8$^+$ T cells 3 days following activation with plate-bound anti-CD3/anti-CD28 (left) and with antigen presenting cells with soluble anti-CD3 (right). (**B**) Quantification of CD39 (purple) and PD-1 (green) expression on CD8$^+$ T cells at indicated days following stimulation as in (A). (**C**) Expression of CD39 and CellTrace Violet dilution in CD8$^+$ T cells 3 day following activation with plate-bound anti-CD3/anti-CD28 incubated with IL-2 and indicated cytokines. (**D**) Summary of the kinetics of CD39 expression in the conditions described in (C).

Discussion

The state of $CD8^+$ T cell exhaustion is characterized by widespread changes in gene expression relative to functional memory $CD8^+$ T cells (23). However, in humans, identification of specific T cell exhaustion markers that are not shared by more functional $CD8^+$ T cell populations has been challenging (115). We show that high-level expression of the ectonucleotidase CD39 is characteristic of $CD8^+$ T cells specific for chronic viral infections in humans and mice, but is otherwise rare in the $CD8^+$ T cell compartment of healthy donors. Persistent, high-level expression is also seen in the LCMV mouse model of chronic viral infection, suggesting that CD39 expression is a phenotypic marker of $CD8^+$ T cell exhaustion. Moreover, within the exhausted population in the mouse model, $CD39^{high}$ $CD8^+$ T cells express the highest levels of PD-1, co-express multiple inhibitory receptors and have profoundly impaired function. We found that in both mice and humans, CD39 is expressed preferentially by $CD8^+$ T cells that are T-bet^{low}/$Eomes^{high}$. These data suggest that CD39 expression by $CD8^+$ T cells is a pathological finding and demarcates the population of $CD8^+$ T cells previously identify as being terminally exhausted (49).

The fact that peripheral blood $CD8^+$ T cells in humans can express CD39 is surprising. Previous data have shown that CD39 expression is restricted to $CD4^+$ regulatory T cells, Th17 cells, and small populations of regulatory-like $CD8^+$ T cells (63-66). Indeed, we find that in the bulk population of $CD8^+$ T cells in healthy donors only a small minority of $CD8^+$ T cells expresses CD39. However, CD39 is abundantly expressed by virus-specific $CD8^+$ T cells in two human chronic infections (HIV and HCV). This helps explain why $CD39^+$ $CD8^+$ T cells have not been appreciated in earlier studies that have focused on healthy individuals, and suggests that, in steady-state conditions, the expression of CD39 by $CD8^+$ T cells is a pathological occurrence that is related to the development of T cell exhaustion. Whether the small fraction of $CD8^+$ T cells expressing CD39 in healthy donors represents acutely

activated $CD8^+$ T cells, or those exhausted by asymptomatic chronic pathogens or inflammatory signals is an important question for future studies.

Several features of CD39-expressing $CD8^+$ T cells suggest that CD39 is a diagnostically valuable marker of T cell exhaustion. First, in both human and mouse $CD8^+$ T cells responding to chronic infection, CD39 is co-expressed with PD-1, an inhibitory receptor expressed by the majority of exhausted T cells (23, 48). Second, CD39 expression correlates with viral load in subjects with HIV and HCV infection suggesting that the conditions of high levels of inflammation and antigen load that lead to exhaustion also increase CD39 expression in the virus-specific pool of $CD8^+$ T cells, as has been observed for PD-1 (51, 134). Third, gene signatures characteristic of exhausted mouse $CD8^+$ T cells are enriched in $CD39^+$ cells relative to $CD39^-$ $CD8^+$ T cells in subjects with HCV infection, underscoring the association between CD39 expression and T cell exhaustion. Finally, chronic LCMV infection in the mouse model increases CD39 expression by exhausted virus-specific $CD8^+$ T cells, and elicits a population of $CD39^{high}$ cells that are absent in functional memory cells. Previous studies show that CD39, like PD-1, is transiently up-regulated by acute T cell activation (63, 135). Additional studies will therefore be required to determine the extent to which T cell activation (rather than exhaustion *per se*) contributes to the up-regulation of CD39 and PD-1 in chronic infection. However, the strong association between CD39 expression and the hallmark phenotypic features of T cell exhaustion in humans and a mouse model suggests that it can serve as a valuable marker of the exhausted T cells state.

The expression of molecules, such as PD-1, that inhibit T cell function has been used to identify exhausted $CD8^+$ T cells in several studies of human chronic infection and cancer (21). However, there are important distinctions between the pattern of CD39 expression and that of inhibitory receptors. Many inhibitory receptors, such as PD-1 (51, 115, 136) and CD244 (137, 138) are also expressed by a substantial fraction of $CD8^+$ T cells in healthy donors that are not exhausted. In contrast, CD39 expression is found in only a very small

minority of $CD8^+$ T cells from healthy donors. This expression pattern suggests that CD39 expression, particularly in combination with PD-1, may be useful as a more specific phenotype of exhausted $CD8^+$ T cells, at least in HCV and HIV infection. In addition, CD39 may provide a useful marker to isolate exhausted $CD8^+$ T cells in settings such as tumor-specific responses where very few reagents are available to identify antigen-specific T cells. Importantly, while CD39 is rare in the $CD8^+$ compartment in healthy donors, it is expressed by $CD4^+$ Tregs – as is PD-1 – making it difficult to distinguish between exhausted $CD4^+$ T cells and Tregs by CD39 expression alone.

Analysis of global expression profiles of $CD39^+$ *versus* $CD39^-$ $CD8^+$ T cells in HCV-infected subjects showed that the $CD39^+$ fraction was strongly enriched for genes related to proliferation. This may at first seem counterintuitive, given the functional defects that have been described in exhausted $CD8^+$ T cells (21, 23). However, data from the mouse model of chronic infection suggest that, unlike memory $CD8^+$ T cells, exhausted $CD8^+$ T cells are dependent on continuous exposure to viral antigen to ensure their survival and undergo extensive cell division at a rate higher than that seen in physiological homeostatic proliferation of the memory $CD8^+$ T cell pool (139). Exhausted $CD8^+$ T cells therefore have a paradoxical increase in their proliferation *in vivo* despite reduced proliferative potential *in vitro* (140), explaining the increased expression of proliferation-associated genes in $CD39^+$ $CD8^+$ T cells in HCV infection and in mouse exhausted $CD8^+$ T cells (49, 141).

Recent studies of exhausted $CD8^+$ T cells have revealed that two distinct states of virus-specific $CD8^+$ T cells exist in chronically infected mice and humans (49, 116). Differential expression of the T-box transcription factors T-bet and Eomes characterize two populations, which form a progenitor-progeny relationship. $T\text{-bet}^{high}$ cells display low intrinsic turnover but are capable of proliferation in response to persisting antigen, giving rise to $Eomes^{high}$ terminal progeny. In contrast, $Eomes^{high}$ $CD8^+$ T cells responding to chronic infection had reduced capacity to undergo additional proliferation *in vivo.* The $T\text{-bet}^{low}$

/Eomes^{high} exhausted subset of CD8^{+} T cells correspond to the PD-1 bright population that has also been shown to be unresponsive to PD-1:PD-L1 blockade. These data suggest that the differential expression of these transcription factors identifies subpopulations of exhausted CD8^{+} T cells with fundamentally different fates and functional profiles. Our data show that in the LCMV mouse model of chronic infection and in HIV infection, the CD39^{high} subset of CD8^{+} T cells demarcates T-bet^{low} /Eomes^{high} cells. Consistent with this, CD39^{+} CD8^{+} T cells in the mouse model express the highest levels of PD-1, co-express multiple inhibitory receptors and show marked functional defects. These findings suggest that CD39 may be a marker not only of the exhausted state, but specifically of the most terminally exhausted cells, at least in the mouse model. Additional studies of the fate of transferred CD39^{+} vs. CD39^{-} exhausted CD8^{+} T cells in the mouse model, and broader surveys of CD39 expression in human chronic infections will be required to determine whether this marker can be used as a surrogate for terminal exhaustion. However, the strong association between CD39 expression and the key features of terminal exhaustion suggests that it may prove a useful marker to help distinguish between "reversible" and "irreversible" T cell exhaustion. Moreover, the fact that isolating CD39^{+} cells does not require intracellular staining (as is required for T-bet and Eomes) makes this marker useful for studying the function of this terminally exhausted cells *ex vivo.*

The fact that CD39 is expressed by a slightly larger fraction of HCV-specific CD8^{+} T cells than HIV-specific CD8^{+} T cells may be related to differences in the timing of blood sampling during the course of infection, or may be due to differences in the extent of antigen-load and inflammation in the two infections. Alternatively, it may be consistent with a model in which HCV-specific CD8^{+} T cells are in a more “terminal” state of exhaustion than CD8^{+} T cells specific for HIV. This possibility is supported by profound loss of HCV-specific CD8^{+} T cells over the course of chronic infection (142) that is not seen in the HIV-specific CD8^{+} T cell

pool, consistent with the clonal deletion seen in mouse models of extreme $CD8^+$ T cell exhaustion (143, 144)

It is tempting to speculate that expression of CD39 contributes to the dysfunction of exhausted T cells (145). For instance, the expression of CD39 might enable $CD8^+$ T cells to provide negative regulation in an autocrine or juxtacrine fashion via adenosine (122-124) in the same manner as Tregs (61, 135). The fact that CD39 requires both a substrate (ATP) and a downstream enzyme (CD73) to generate adenosine could provide a mechanism to ensure that this negative signaling occurred only in certain contexts such as in inflamed, damaged tissues where the extracellular concentrations of ATP are high and CD73-expressing cells are present (146). Moreover, CD39-expressing $CD8^+$ T cells may contribute to the general inhibitory milieu by contributing to the inhibition of activated T cells that express the adenosine receptor but are not yet exhausted. It will therefore be important to determine whether inhibition of CD39 activity could provide an additional therapeutic strategy to rescue the function of exhausted T cells.

Materials and Methods

Human Subjects

Healthy human donors were recruited at the Kraft family Blood Donor Center, Dana-Farber Cancer Institute. All human subjects with HCV infection were recruited at the Gastrointestinal Unit and the Department of Surgery of the Massachusetts General Hospital (Boston, MA) (Supplemental Table 3.1).

Individuals with chronic HCV infection (n = 82) were defined by positive anti-HCV antibody and detectable viral load. Patients with spontaneous clearance of HCV, termed resolvers (n = 30), were defined by positive anti-HCV antibody but an undetectable viral load for at least 6 months. The estimated time of infection was calculated either using the exposure date or the time of onset of symptoms and peak ALT (which are assumed to be 7 weeks post infection). All HCV patients were treatment naive and studied at 5.9 and 219.7 weeks post infection. HCV RNA levels were determined using the VERSANT HCV RNA 3.0 (bDNA 3.0) assay (Bayer Diagnostics).

All HIV infected subjects (n = 40) were recruited at the Ragon Institute at the Massachusetts General Hospital (Boston, USA) or the Peter Medawar Building for Pathogen Research (Oxford, UK) (Supplemental Table 3.2). HIV controllers included elite controllers (n = 5) defined as having HIV RNA below the level of detection (<75 viral copies per ml) and viraemic controllers (n = 7) with HIV RNA levels < 2,000 viral copies per ml. HIV chronic progressors (n = 28) were defined as having > 2,000 viral copies per ml. All subjects were off therapy. Viral load during chronic infection was measured using the Roche Amplicor version 1.5 assay.

Mice and infections

Wild-type C57BL/6J mice were purchased from The Jackson Laboratory. CD39-KO mice have been described previously (132) and were obtained from Dr. Simon Robson (Beth

Israel Deaconess Medical Center). Female mice (6-8 weeks old) were infected with 2 x 10^5 plaque forming units (p.f.u.) of LCMV-Armstrong intraperitoneally or 4 x 10^6 p.f.u. of LCMV-Clone 13 intravenously and analyzed at indicated time points by homogenizing the spleen into a single-cell suspension, Ammonium-Chloride-Potassium lysis of red blood cells, followed by antibody staining. For experiments involving P14 cell transfers, Ly5.1$^+$ P14s were isolated from peripheral blood, and 500 P14 cells were transferred i.v. into 5-6 week old wild-type female mice one day prior to infection. Viruses were propagated as described previously (69, 147, 148). The mouse work was performed under a protocol 01214 approved by the HMA Institutional Animal Care and Use Committee (IACUC), in strict accordance with the recommendations in the Guide for the care and use of Laboratory Animals of the National Institutes of Health. The Harvard Medical School animal management program is accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International (AAALAC).

MHC Class I Tetramers

Major histocompatibility complex (MHC) class I HIV Gag-specific tetramers were produced as previously described (149) or obtained from Proimmune. CMV- and EBV-specific MHC class I dextramers conjugated with FITC and APC were purchased from Immudex. Mouse MHC class I tetramers of H-2D^b complexed with LCMV $GP_{276-286}$ were produced as previously described (69, 147). Biotinylated complexes were tetramerized using allophycocyanin-conjugated streptavidin (Molecular Probes). The complete list of multimers can be found in supplemental materials (Supplemental Table 3.3).

Antibodies and flow cytometry

The following anti-human (hu) and anti-mouse (m) fluorochrome-conjugated antibodies were used for flow cytometry: huCD8α (RPA-T8), huCD4 (OKT4), huCD3 (OKT3),

huCD39 (A1), huPD-1 (EG12.2H7), huCD25 (BC96), huCCR7 (G043H7), huCD45RA (HI100), huT-bet (4B10), mCD8a (53-6.7), mCD4 (GK1.5), mCD3 (145-2C11), mCD244.2 (m2B4 (B6)458.1), mPD-1 (RMP1-30), mLag3 (C9B7W), mCD44 (IM7), mCD127 (A7R34), mTNF-α (MP6XT22) (all from Biolegend), mT-bet (04-46; BD Pharmingen), mCD39 (24DMS1), mIFN-γ (XMG1.2), huEomes (WD1928) and mEomes (Dan11mag) (eBioscience). Intracellular staining was performed following surface staining and fixed and permeabilized using the FoxP3/Transcription Factor Staining Buffer Set (eBioscience). Cells were sorted by BD FACS ARIA II and all other analyses were performed on BD LSR II and BD LSR Fortessa flow cytometers equipped with FACSDiva v6.1. Gates were set using Full Minus One (FMO) controls. Data were analyzed using FlowJo software v9.8 (Treestar).

For intracellular cytokine analysis of mouse T cells, $2x10^6$ splenocytes were cultured in the presence of GP_{33-41} peptide (0.2 mg/ml) (sequence KAVYNFATM), brefeldin A (BD), and monensin (BD) for 4.5 hours at 37°C. Following staining for surface antigens, cells were permeabilized and stained for intracellular cytokines with the Cytofix/Cytoperm kit according to manufacturer's instructions (BD Biosciences).

In vitro T cell stimulations

Naive $CD8^+$ T cells were extracted from wild-type mouse spleens using magnetic negative selection kit (Miltenyi). Cells were labeled with CellTrace Violet (Thermo Fischer scientific) according to manufacturer's instructions and stimulated in two different ways. First method was stimulation through incubation of T cells with 30 Gy-irradiated antigen-presenting cells (splenocytes from TCRα-KO mice) in the presence of 4 µg/ml anti-CD3 (clone 145-2C11). In the second method, the T cells were plated on tissue culture treated plate pre-coated overnight at 4°C with 4 µg/ml anti-CD3 (clone 145-2C11) and anti-CD28 (clone 37.51) (BioXcell) and incubated with 100 U/ml recombinant human IL-2 (Peprotech) in complete RPMI (RPMI 1640 supplemented with 10% FBS, 2 mM L-glutamine, 10 mM

HEPES, 100 U/ml penicillin/ streptomycin (all Gibco), and 50 μM 2-mercaptoethanol (Sigma)). For some experiments, one of the following was added to the medium: 10 ng/ml mIL-1β, 200ng/ml mIL-6, 20 ng/ml mIL-7, 10 ng/ml mIL-12, 100 ng/ml mIL-15, 20 ng/ml mGM-CSF, 25 ng/ml mTGF-β1, or both 20 ng/ml mIL-7 and 100 ng/ml mIL-15 (all Peprotech).

HPLC analysis of ATP levels

The concentration of ATP hydrolyzed by $CD8^+$ T cells from subjects with HCV infection (n = 6) was assessed by high performance liquid chromatography (HPLC) as previously described (150). Briefly, 10,000 $CD39^+$ $CD8^+$ T cells were sorted and placed on ice to minimize ATP production by cells. 20 mM of ATP was added and incubated for 1 h at 37°C in 5% CO_2 to allow for cellular activity to increase and CD39-mediated ATP hydrolysis to occur. Samples were then placed in an ice bath for 10 min to halt enzymatic activity, collected, and centrifuged for 10 min at 380 x g and 0°C. Cells were discarded and supernatant centrifuged again to remove remaining cells (2350 x g, 5 min, 0°C). The resulting RPMI samples (160 ml) were treated with 10 ml of an 8 M perchloric acid solution (Sigma-Aldrich) and centrifuged at 15,900 x g for 10 min at 0°C to precipitate proteins. In order to neutralize the pH of the resulting solutions and to remove lipids, supernatants (80 ml) were treated with 4 M K_2HPO_4 (8 ml) and tri-N-octylamine (50 ml). These samples were mixed with 50 ml of 1,1,2-trichloro-trifluoroethane and centrifuged (15,900 x g, 10 min, 0°C) and this last lipid extraction step was repeated once. The resulting supernatants were subjected to the following procedure to generate fluorescent etheno-adenine products: 150 ml supernatant (or nucleotide standard solution) was incubated at 72°C for 30 min with 250 mM Na_2HPO_4 (20 ml) and 1 M chloroacetaldehyde (30 ml; Sigma-Aldrich) in a final reaction volume of 200 ml, resulting in the formation of 1,N6-etheno derivatives as previously

described (150). Samples were placed on ice, alkalinized with 0.5 M NH_4HCO_3 (50 ml), filtered with a 1 ml syringe and 0.45 mM filter and analyzed using a Waters HPLC system and Supelcosil 3 μM LC-18T reverse phase column (Sigma), consisting of a gradient system described previously, a Waters autosampler, and a Waters 474 fluorescence detector (151). Empower2 software was used for the analysis of data and all samples were compared with water and ATP standard controls as well as a sample with no cells to determine background degradation of ATP.

Histopathology

Tissues were fixed in 10% formalin, and paraffin-embedded sections were stained with hematoxylin and eosin (H&E). Photomicrographs were acquired on an Olympus BH-2 light microscope at the indicated magnifications using an Olympus DP71 camera and software provided by the manufacturer. Horizontal bars on 400x images represent 50 μm.

Microarray data acquisition

$CD8^+$ T cells from subjects with HCV infection were sorted and pelleted and re-suspended in TRIzol (Invitrogen). RNA extraction was performed using the RNAdvance Tissue Isolation kit (Agencourt). Concentrations of total RNA were determined with a Nanodrop spectrophotometer or Ribogreen RNA quantification kits (Molecular Probes/Invitrogen). RNA purity was determined by Bioanalyzer 2100 traces (Agilent Technologies). Total RNA was amplified with the WT-Ovation Pico RNA Amplification system (NuGEN) according to the manufacturer's instructions. After fragmentation and biotinylation, cDNA was hybridized to HG-U133A 2.0 microarrays (Affymetrix). Data have been deposited in Gene Expression Omnibus with accession code GSE72752.

Statistics

Prior to analysis, microarray data were pre-processed and normalized using robust multichip averaging, as previously described (152). Differentially gene expression and consensus clustering (153) were performed using Gene-E software (www.broadinstitute.org/cancer/software/GENE-E/), and gene set enrichment analysis was performed as described previously using gene sets from MSigDB (154) or published resources (75, 78).

Consensus hierarchical clustering was performed using the top 10% of genes that varied across the dataset, without reference to sample identity. Consensus cluster assesses the "stability" of the clusters discovered using unbiased methods such as hierarchical clustering i.e. the robustness of the putative clusters to sampling variability. The basic assumption is that if the data represent a sample of items drawn from distinct sub-populations, a different sample drawn from the same sub-populations, would result in cluster composition and number should not be radically different. Therefore, the more the attained clusters are robust to sampling variability, the greater the likelihood that the observed clusters represent real structure. The result of consensus clustering is a matrix that shows, for each pair of samples, the proportion of clustering runs on sub-sampled data in which those two items cluster together (shown on a scale of 0 to 1).

Enrichment Map analysis of GSEA results was performed as described (155). The gene signature of exhaustion was generated by identifying the top 200 genes up-regulated in $CD8^+$ T cells responding to chronic vs. acute LCMV infection in microarray data from a previously published study (75).

Acknowledgments

Prakash K. Gupta (Haining and Klenerman Labs, Dana-Farber Cancer Institute and Oxford University, respectively) performed initial work on this project and made the discovery that CD39 expression is associated with $CD8^+$ T cell exhaustion in human patients. Kathleen Yates (Haining Lab, Dana-Farber Cancer Institute), David Wolski (Lauer Lab, Massachusetts General Hospital), Emily Adland (Klenerman Lab, Oxford University), Cormac Cosgrove (Alter Lab, Ragon Institute of MGH, MIT and Harvard) assisted with sourcing, curating, and analyzing human patient samples. Carola Ledderose (Junger Lab, Beth Israel Deaconess Medical Center) assisted with HPLC assays. Kristen Pauken (Wherry Lab, University of Pennsylvania Perelman Medical School) assisted with *ex vivo* stimulation and analysis of exhausted mouse $CD8^+$ T cells. The H&E slides were prepared and scored in a blinded fashion by pathologist Roderick T. Bronson. I would also like to thank E. John Wherry (University of Pennsylvania Perelman Medical School) for providing reagents for mouse experiments including LCMV and tetramers.

Chapter 4. Inducible RNAi *in vivo* reveals that BATF is required to initiate but not maintain CD8$^+$ T cell effector differentiation

Parts of this chapter have previously been published in:

Jernej Godec, Glenn S. Cowley, R. Anthony Barnitz, Ozan Alkan, David E. Root, Arlene H. Sharpe, W. Nicholas Haining. Inducible RNAi *in vivo* reveals that BATF is required to initiate but not to maintain CD8+ T cell effector differentiation. *Proceedings of the National Academy of Sciences USA*. 2015 Jan 13;112(2):512-7. (PMID: 25548173)

Introduction

Optimal effector and memory CD8$^+$ T cell differentiation requires 3 signals: signal 1 from TCR–MHC/peptide; signal 2 from co-stimulatory pathways; and signal 3 from inflammation. Following activation by antigen, co-stimulation and inflammation, naive CD8$^+$ T cells initiate a differentiation program resulting in massive changes in gene expression and cell function, which leads to the formation of effector and memory T cells (75). This differentiation program is critical for the development of effective tumor immunity (156) and the control of pathogens (2). However, the developing effector CD8$^+$ T cell population is strikingly heterogeneous (157), and several phenotypically distinct subpopulations of CD8$^+$ T cells exist within the effector pool that have different lineage potential and function (40, 158). Thus CD8$^+$ effector T cells face complex lineage choices during differentiation, which can have profound effects on the development of immunity to pathogens (6) and the development of effective responses to vaccines (159-162).

Although the development of effector CD8$^+$ T cells occurs over a period of days (163), early events in the life history of CD8$^+$ T cells are critical in determining their fate (33, 35, 164-166). For instance, asymmetric segregation of cell contents during the first cell division

after encounter with antigen can profoundly influence the ultimate differentiation state of effector cells (167), suggesting that investigating the events that occur in the hours following initial antigen encounter will be essential to define the mechanisms that regulate the fate of effector $CD8^+$ T cells.

Effector $CD8^+$ T cell differentiation is regulated by a set of transcription factors (TFs) including T-bet (168), Eomes (169), Blimp1 (170, 171), Id2 (172) and Runx3 (173). We have recently shown that the AP-1 family TF BATF is absolutely required for effector $CD8^+$ T cell differentiation and coordinates the program of gene expression essential for this process (174). Thus, many TFs that are expressed immediately after stimulation may play a role in specifying the fate of developing effector cells from the earliest point in differentiation.

The role of specific TFs in regulating $CD8^+$ T cell effector differentiation has been investigated using germline or conditional KOs. However, these approaches are limited to studying a small number of candidate genes (6). In contrast, perturbing genes with RNAi could permit the study of many more candidate regulators in parallel (175). However, techniques to deliver shRNAs to T cells are limited by the need to stimulate cells to divide using TCR cross-linking (176), infection (158, 177), or cytokine stimulation (178) in order to achieve meaningful transduction frequencies with viral vectors encoding shRNA constructs. The need to activate T cells for delivery of shRNAs raises concerns about whether this activation alters these T cells at a critical phase of time when even subtle perturbations of TFs can profoundly influence T cell fate (167).

To address these limitations, we have developed an experimental system to knockdown gene expression in T cells *in vivo* using shRNA without the need to transduce T cells directly. To do this, we generated bone marrow chimera from HSCs transduced with an inducible shRNA vector. Following hematopoietic reconstitution, this allows inducible gene knockdown in any cell type that developed from this transduced progenitor pool, including resting, naive $CD8^+$ T cells *in vivo*. We have applied this system to show that BATF is

essential for initial commitment of naive $CD8^+$ T cells to effector differentiation but becomes dispensable after 72h.

Results

Lentivirus-transduced stem cells reconstitute blood immune lineages and give rise to effector CD8$^+$ T cells with unaltered functionality.

Resting T cells are refractory to lentiviral transduction, but hematopoietic stem cells (HSC) are readily transduced. We therefore generated bone marrow chimeric animals using lentivirus-transduced hematopoietic progenitor cells in which hematopoietic lineages (including T cells) are reconstituted with transduced cells (Figure 1*A*). We isolated lineage$^-$/c-kit$^+$/sca-1$^+$ (LSK) cells (which include hematopoietic stem cells and multipotent progenitors, hereafter referred to as "HSCs"), from the bone marrow of P14 TCR transgenic mice in which most CD8$^+$ T cells express a TCR specific for LCMV GP_{33-41} peptide presented on H-2D^b (Figure 4.1A), and transduced them with a lentivirus carrying a GFP expression cassette so that the fate of transduced cells could be tracked. We used congenic markers to distinguish transplanted cells from recipient cells in bone marrow chimeras.

In order to test whether lentivirus-transduced LSK cells could be used to generate fully functional CD8$^+$ T cells, we first transduced the LSK cells with lentivirus encoding only GFP under a PGK promoter (PGK-eGFP) and transplanted them into lethally irradiated animals (50,000 cells/animal). Following reconstitution (8 – 12 weeks later) analyses of major lineages in the immune system showed that the frequencies of GFP$^+$ B cells (B220$^+$), CD4$^+$ and CD8$^+$ T cells, dendritic cells (CD11c$^+$), and monocytes (CD11b$^+$) were similar to that of the LSK inoculum (Figure 4.1B), suggesting efficient engraftment of transduced cells. CD4$^+$ and CD8$^+$ T, total T, B and myeloid lineages developed from transplanted GFP$^+$ (transduced) and GFP$^-$ (untransduced) LSK with equal efficiency (Figure 4.1C and Figure 4.2). We next compared effector CD8$^+$ T cell differentiation of naive CD8$^+$ T cells derived from transplanted PGK-eGFP transduced (GFP$^+$) LSK cells with differentiation of naive CD8$^+$ T cells derived from transplanted but untransduced (GFP$^-$) LSK cells. We transferred equal ratios of GFP$^+$ and GFP$^-$ naive P14 CD8$^+$ T cells to naive wild-type recipients (10,000 cells/animal) and

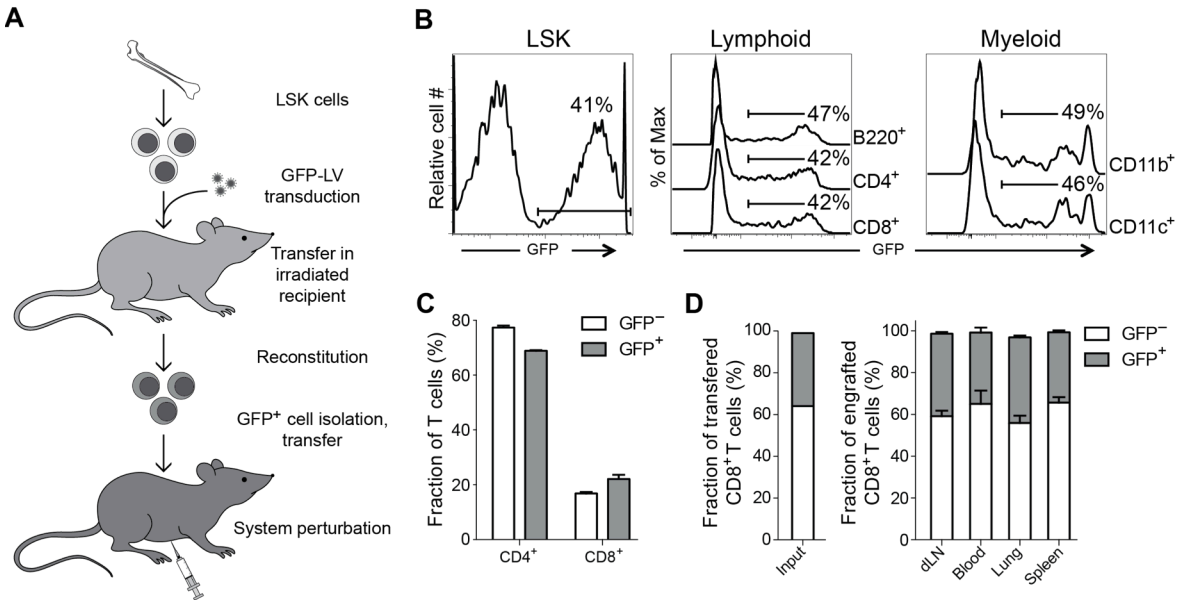

Figure 4.1. Transduced bone marrow progenitor populations efficiently reconstitute myeloid and lymphoid compartments and develop normally into functional CD8$^+$ T cells. (**A**) A schematic diagram of transduction strategy. (**B**) Fraction of LSK cells (left panel) transduced with GFP-expressing lentivirus at the time of transplant, and in lymphoid (middle panel) and myeloid (right panel) cell populations following engraftment. (**C**) Quantitation of fractions of transduced (GFP$^+$) and untransduced (GFP$^-$), donor-derived cells in immune lineages indicated following engraftment as in (B). (**D**) Fraction of transduced (GFP$^+$) and untransduced (GFP$^-$), donor-derived naive P14 CD8$^+$ T cells (left panel) prior to adoptive transfer, and of effector (right panel) P14 CD8$^+$ T cells in tissues indicated 10 days following transfer and subsequent host infection with PR8-GP33 influenza virus. Representative data are shown from 3 experiments with 3-5 mice per group.

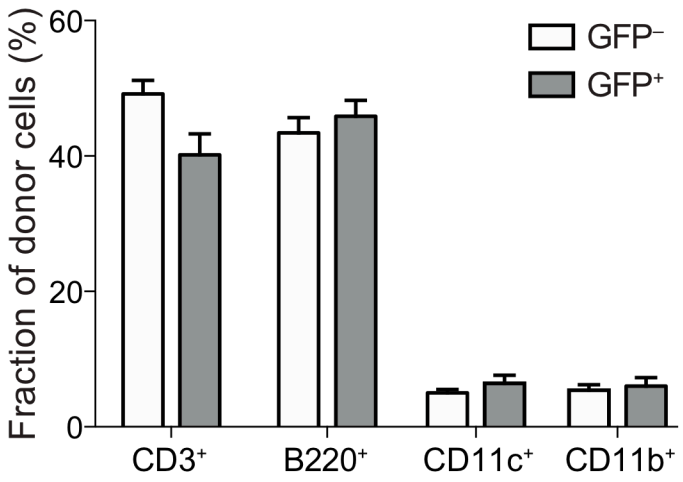

Figure 4.2. Transduced bone marrow progenitor populations efficiently reconstitute myeloid and lymphoid compartments. Fraction of indicated donor-derived immune lineages in blood that differentiated from transduced (GFP^+) and untransduced (GFP^-) congenic WT LSK cells.

infected them with PR8-GP33 influenza (Figure 4.1D). We found equal expansion and persistence of GFP^+ and GFP^- effector $CD8^+$ T cells at 10 days after infection.

We next compared the phenotype and function of effector $CD8^+$ T cells arising from naive $CD8^+$ T cells that developed from transduced LSKs with effector $CD8^+$ T cells differentiating from untransduced naive $CD8^+$ T cells. We analyzed the proliferative capacity, expression of cell surface molecules, key transcription factors, and production of cytokines upon re-stimulation and found no difference between untransduced and transduced effector $CD8^+$ T cells at d8 p.i. (Figure 4.3). Thus, lentiviral transduction of LSK neither impairs the development of lymphoid and myeloid lineages following transplantation nor alters effector $CD8^+$ T cell generation, proliferative capacity or survival following transfer of naive $CD8^+$ T cells.

To compare the persistence and phenotype of effector $CD8^+$ T cells derived from untransduced naive $CD8^+$ T cells, or naive $CD8^+$ T cells generated from bone marrow chimeras, we transferred either unmodified naive P14 $CD8^+$ T cells or naive P14 $CD8^+$ T cells carrying 1xLacO-shLacZ construct into congenically distinct LCMV-infected wild-type recipients. We analyzed the fraction of transferred cells at d28 p.i. and observed no difference in the frequency of unmodified $CD8^+$ T cells and those carrying 1xLacO-shLacZ construct (Figure 4.4, *A* and *B*). Additionally, 1xLacO and WT P14 memory cells were indistinguishable in their expression of cell surface molecules and production of cytokines upon re-stimulation (Figure 4.4, *C* and *D*). This suggests that the persistence of $CD8^+$ T cells following effector differentiation is not altered by the presence of the lentiviral vector.

Naive T cells that develop from transduced LSK are indistinguishable from wild-type naive T cells.

We next examined whether naive $CD8^+$ T cells that developed from transduced LSK cells show any alterations of surface phenotype, proliferative status or gene expression that

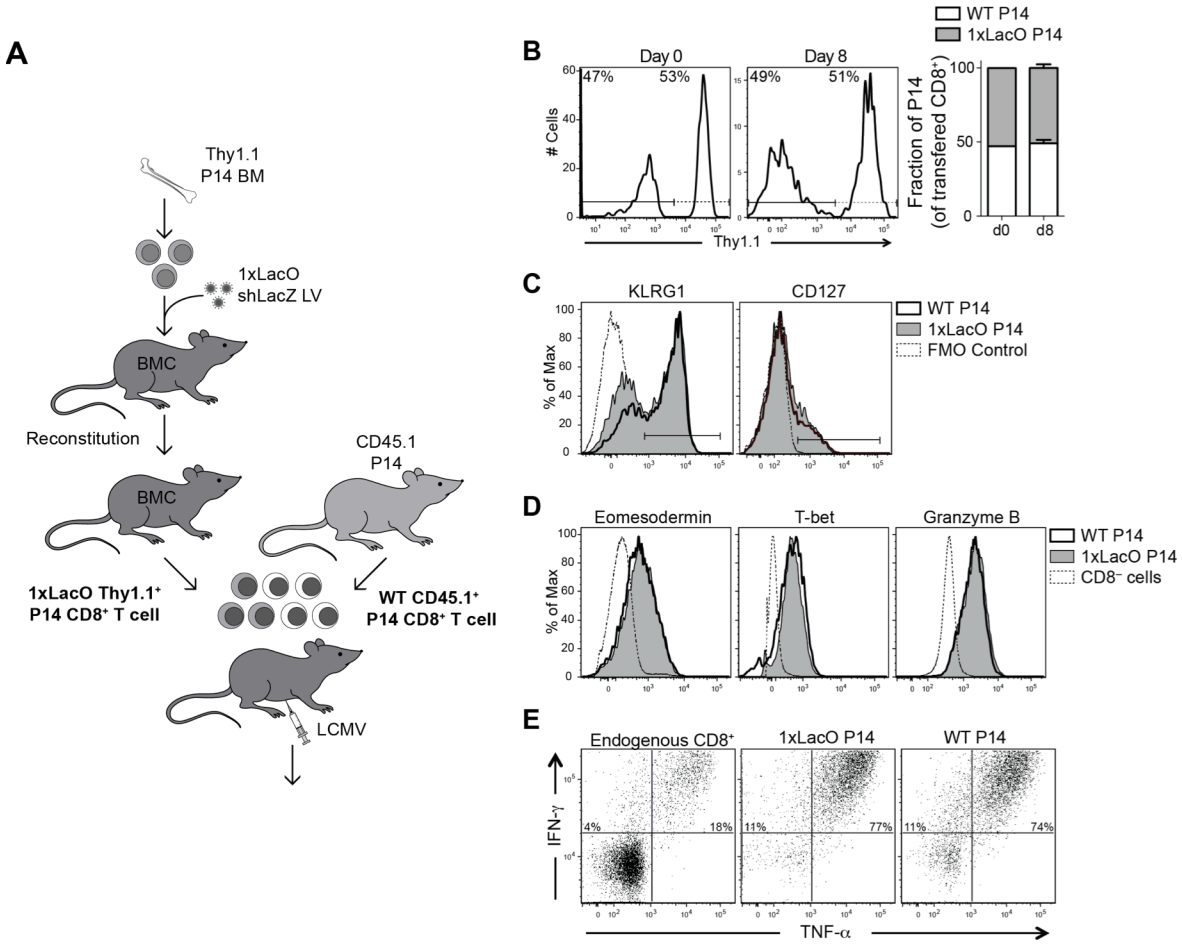

Figure 4.3. Development from 1xLacO transduced LSK cells does not alter the functional capacity and differentiation of effector CD8$^+$ T cells. (**A**) The experimental schema used to analyze the phenotype of 1xLacO-shRFP-carrying cells relative to unperturbed P14 CD8$^+$ T cells responding early to the LCMV infection. (**B**) The ratio of 1xLacO P14 Thy1.1$^+$ cells and CD8$^+$ T cells directly isolated from the Thy1.1$^-$ CD45.1$^+$ P14 TCR transgenic mouse at the time of injection (d0) and at d8 following transfer into mice that were subsequently infected with LCMV Armstrong. Representative plots (left) and summarized data (right). (**C – E**) Analysis of effector CD8$^+$ T cells from (B) for their expression of cell surface markers CD127 and KLRG1 (C), the transcription factors T-bet and Eomes, and Granzyme B (D) as well as their ability to produce cytokines following re-stimulation with GP_{33-41} peptide (E). Data are representative from two independent experiments with 5-10 mice per group.

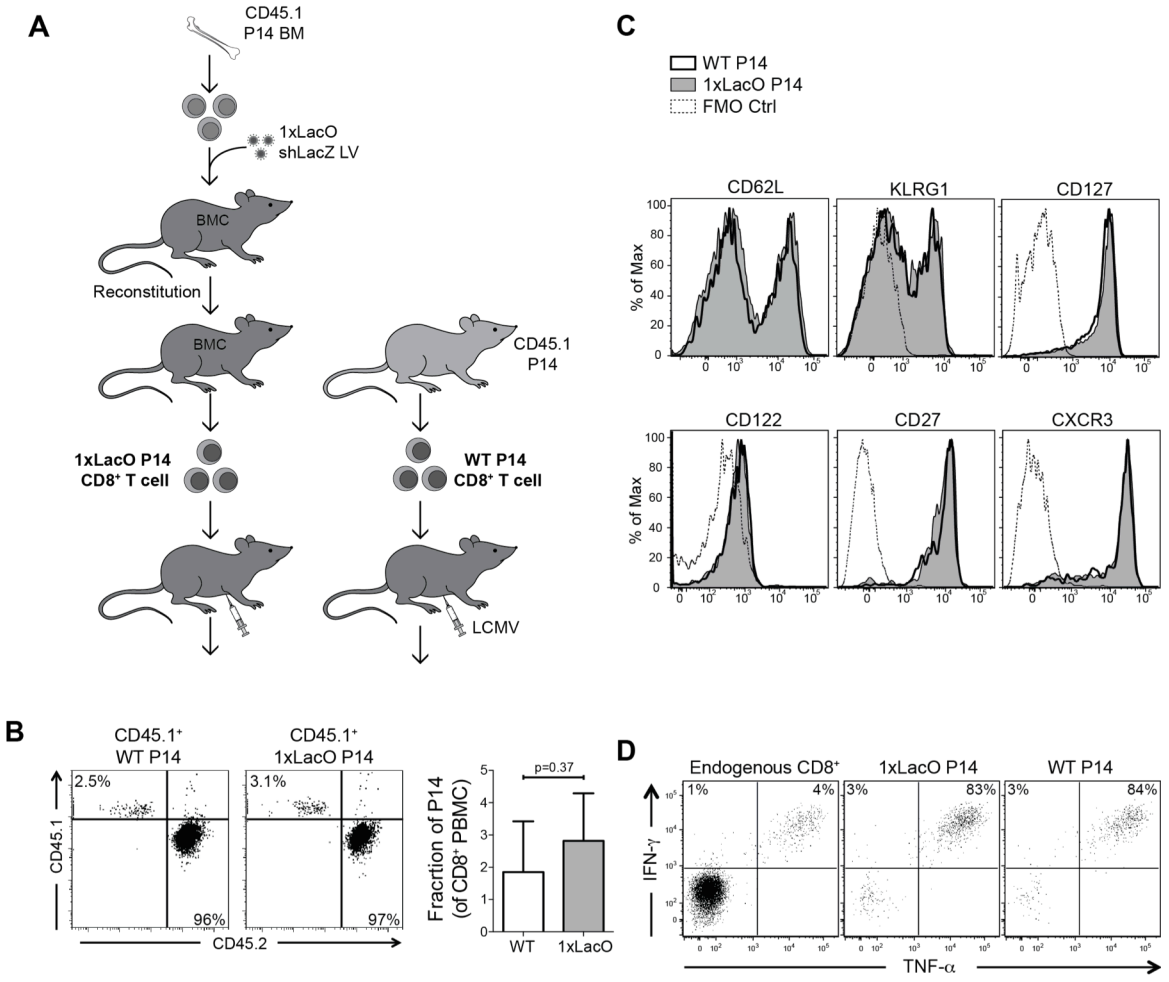

Figure 4.4. Stable integration of 1xLacO in effector CD8$^+$ T cells does not cause preferential rejection or alter the functional capacity, differentiation, or maintenance of memory CD8$^+$ T cells. (**A**) The experimental schema to assess the memory development of 1xLacO-shLacZ-carrying cells and unmodified P14 CD8$^+$ T cells in response to LCMV infection. (**B**) Fraction of transferred 1xLacO P14 cells and unmodified P14 cells at d28 p.i. Shown are representative plots, gated on total CD8$^+$ T cells (left) and summary data from 5 mice (right). Statistical significance was assessed using Student's t-test. (**C**, **D**) Analysis of cells from spleen as in (B) at d60 p.i. for their expression of cell surface markers (C) and ability to produce cytokines upon re-stimulation with GP_{33-41} peptide (D).

might obscure analysis of early differentiation events. We compared naive $CD8^+$ T cells that were derived from transduced LSK cells with control wild-type P14 $CD8^+$ T cells. We also studied P14 $CD8^+$ T cells cultured in conditions used in previous studies to facilitate direct viral transduction of T cells (158, 176-179) to compare $CD8^+$ T cells generated by our approach and previously reported methods: 1) activation of $CD8^+$ T cells *in vivo* by infecting P14 mice with 2 x 10^5 p.f.u. LCMV Armstrong; 2) activation *in vitro* by stimulation with anti-CD3 and anti-CD28; or 3) incubation *in vitro* with a combination of IL-7 and IL-15 cytokines.

The proportions of naive ($CD62L^+$ $CD44^-$), central memory ($CD62L^+$ $CD44^+$) and effector memory ($CD62L^-$ $CD44^+$) cells were similar in the GFP^+ naive $CD8^+$ T cells from the transduced BM chimeras and in naive $CD8^+$ T cells from wild-type mice, but were markedly altered by the other stimulation conditions, particularly with cytokine treatment (Figure 4.5A). The expression of cytokine receptors including CD25, CD127, CD122 was not different in GFP^+ naive and wild-type naive $CD8^+$ T cells, but was altered in naive $CD8^+$ T cells treated with anti-CD3/CD28 or cytokines (Figure 2*B*). The GFP^+ naive $CD8^+$ T cells also showed a low rate of homeostatic turnover that was similar to wild-type naive $CD8^+$ T cells (Figure 4.5C). In contrast, all of the other stimulation conditions induced varying degrees of cell proliferation (Figure 4.5C).

We measured transcript abundance for TFs and effector molecules that change during $CD8^+$ effector T cell differentiation. Important regulators of effector differentiation such as T-bet (*Tbx21*), Eomesodermin (*Eomes*), and Blimp1 (*Prdm1*), as well as effector molecules including granzymes A and B, perforin 1, as well as IFN-γ and TNF-α were unchanged in GFP^+ naive $CD8^+$ T cells relative to wild-type naive $CD8^+$ T cells, but were up-regulated in the other stimulation conditions (Figure 4.5D). Thus, GFP^+ naive $CD8^+$ T cells that had developed from transduced LSK cells were indistinguishable from untransduced naive $CD8^+$ T cells. In contrast, existing protocols used to achieve viral transduction of naive $CD8^+$ T cells were associated with marked perturbation of the T-cell state.

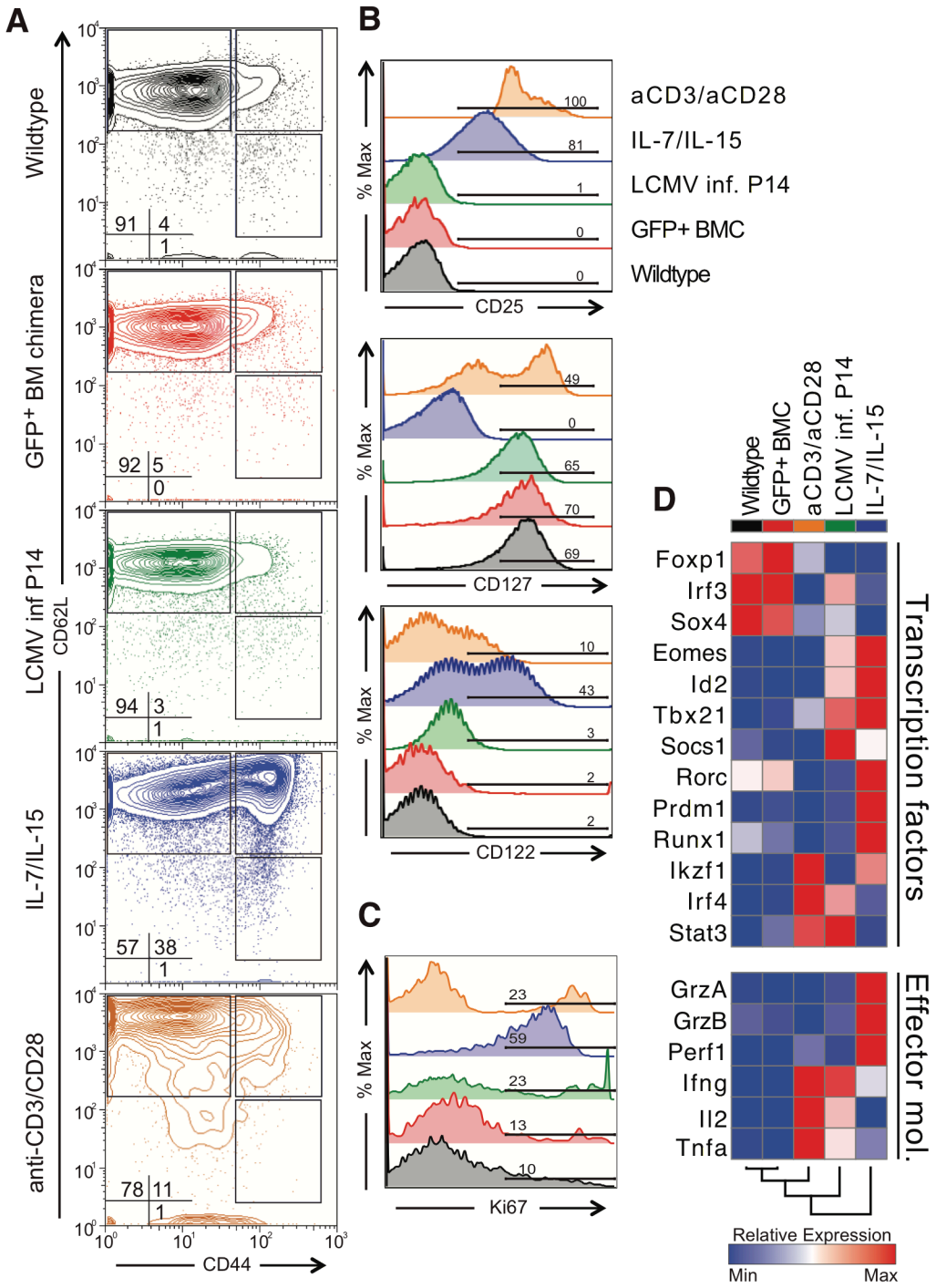

Figure 4.5. CD8$^+$ T cells derived from transduced LSK cells are indistinguishable from untransduced naive CD8$^+$ T cells. (**A** – **C**) Expression of (A) naive surface markers, (B) effector molecules, and (C) Ki-67 by wild-type naive P14 CD8$^+$ T cells (black); naive P14 CD8$^+$ T cells derived from transduced LSKs (red); naive P14 CD8$^+$ T cells stimulated by LCMV infection (green); by the cytokines IL-7 and IL-15 (blue); or by anti-CD3 plus anti-CD28 antibodies (orange). (**D**) Comparison of transcript abundance of transcriptional regulators and effector molecules changes measured by quantitative RT-PCR in wild-type naive P14 CD8$^+$ T cells; naive P14 CD8$^+$ T cells derived from transduced LSKs; naive P14 CD8$^+$ T cells stimulated by anti-CD3 plus anti-CD28 antibodies, LCMV infection, or by the cytokines IL-7 and IL-15.

Lac operon-regulated shRNA allows inducible, efficient, and transient gene knockdown in vivo at low concentrations of IPTG

Hematopoiesis depends on the expression of appropriate genes at the proper developmental stage. Because constitutive gene knockdown in LSK could compromise the development of immune lineages, we used an inducible shRNA expression vector that uses the Lac operon system to regulate the shRNA promoter following addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (Figure 4.6A). We confirmed the inducibility of gene knockdown by targeting a control gene in a Jurkat cell line. Target gene (GFP) expression was only minimally affected in the uninduced state (Figure 4.7A). However, gene knockdown following IPTG induction of shRNA expression was as efficient as that achieved by a constitutive shRNA expressing vector (Figure 4.7A) even at low concentrations of IPTG (Figure 4.6B).

To test knockdown efficiency in primary $CD8^+$ T cells, we generated bone marrow chimeras with an IPTG-inducible vector encoding an shRNA targeting BATF (shBATF) and a GFP expression cassette to create GFP^+ naive T cells that carried the inducible shRNA vector (hereafter "shBATF-naive T cells"). We first tested inducible knockdown *in vitro* by stimulating the cells with anti-CD3/CD28 and assessing the *Batf* transcript levels 3 days following activation. IPTG was administered to the bone marrow chimeras 3 days prior to activation (d-3) or 1 day following activation (d+1). Decreased target gene expression was apparent in both transcript and protein abundance as early as 2 days following IPTG addition *in vitro* (Figure 4.7, B and C) and was comparable to knockdown with the constitutive vector (Mean = 80.4%, SD = 7.8%). To test inducible knockdown *in vivo*, we transferred shBATF-naive P14 $CD8^+$ T cells into mice that were at the same time infected with LCMV-infected mice treated with IPTG, and measured BATF expression after three days. Initiating IPTG induction one day following cell activation resulted in modest (18%) gene knockdown but treating bone marrow chimera three days prior to transfer resulted in a significantly greater

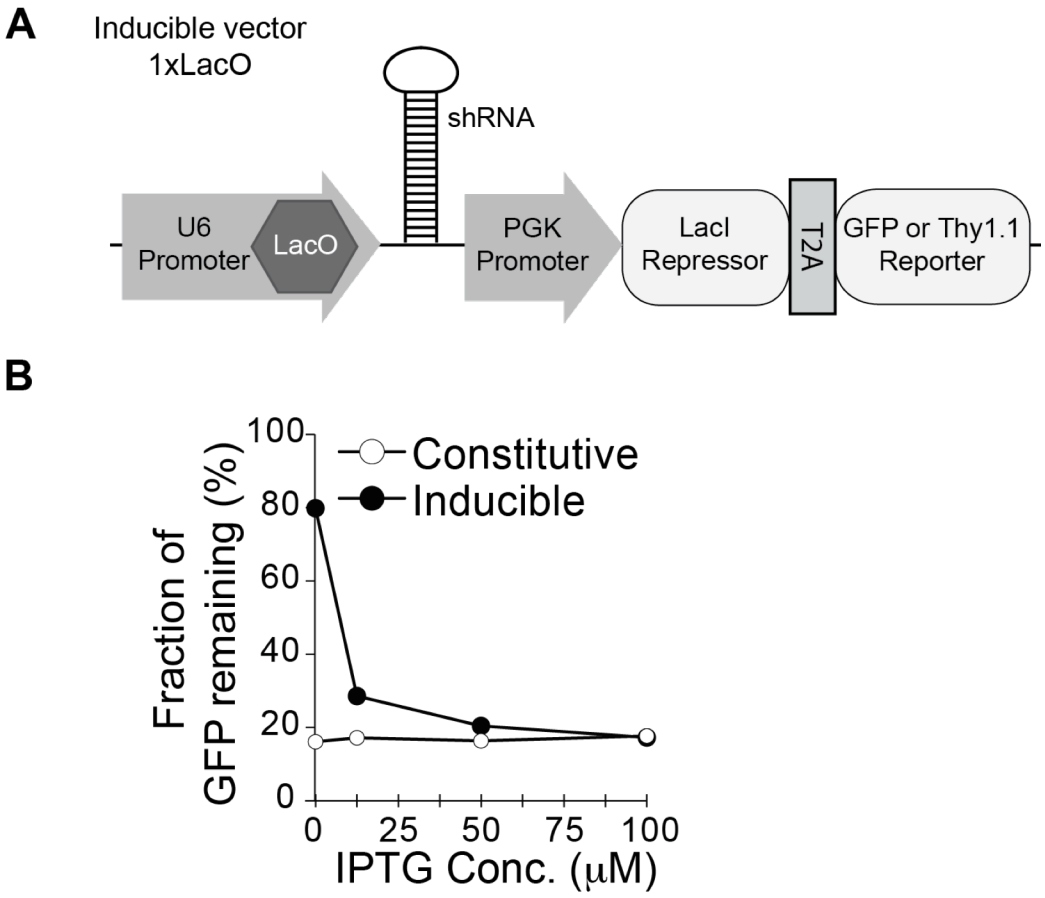

Figure 4.6. Novel shRNA vector enables inducible gene knockdown at low IPTG concentrations *in vitro*. (*A*) Schematic diagram of inducible (1xLacO) vector. (*B*) Fraction of GFP-expressing Jurkat cells transduced with lentivirus expressing an shRNA targeting GFP under constitutive (white symbols) or inducible (black) promoters, cultured with indicated concentrations of IPTG for 7 days.

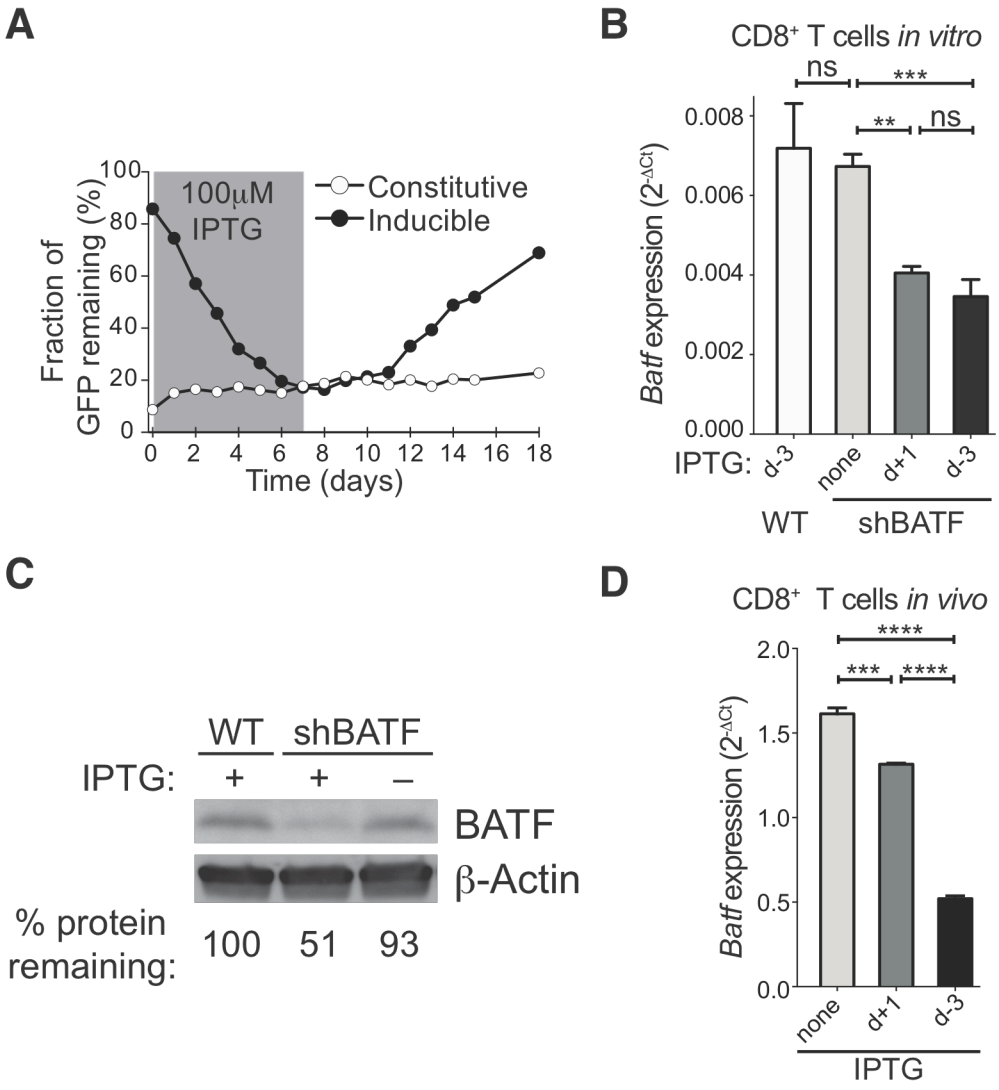

Figure 4.7. Novel shRNA vector enables efficient, inducible, and transient gene knockdown *in vitro* and *in vivo*. (**A**) Fraction of GFP-expressing Jurkat cells transduced with lentivirus expressing an shRNA targeting GFP under constitutive (white symbols) or inducible (black) promoters, cultured with 100 mM IPTG (grey box) for times indicated. (B) BATF expression in anti-CD3/CD28-stimulated shBATF-naive $CD8^+$ T cells cultured in vitro with (grey or black bars) or without (white) IPTG starting at the day indicated. Cells were continuously exposed to IPTG by *in vivo* exposure in bone marrow chimeric mice 3 days prior to T cell sort (d–3) or to the medium 1 day following activation (d+1) and for the remainder of the experiment. (**C**) BATF protein abundance in anti-CD3/CD28-stimulated wild-type and shBATF–naive $CD8^+$ T cells exposed to IPTG d −3 or incubated in medium alone. Numbers represent BATF densiometry values normalized to β-actin in shBATF relative to the wild-type cells. (**D**) shBATF-carrying P14 cells were transferred in LCMV-infected recipient mice and IPTG exposure was maintained by treating mice with 20 mM IPTG in drinking water starting 3 days prior to transfer (in bone marrow chimeras) or 1 day following transfer until the analysis at 3 days following cell transfer. *Batf* mRNA level was normalized to *Hprt* and $2^{-\Delta Ct}$ values reported. Significance was assessed with one-way ANOVA; $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. Representative data are shown from two experiments.

degree of gene silencing (68% knockdown) when it was measured three days following transfer and infection (Figure 4.7C). Thus efficient, inducible gene knockdown can be achieved in following activation of shBATF-naive $CD8^+$ T cells *in vivo*.

BATF knockdown impairs $CD8^+$ effector T cell activation and differentiation following acute viral infection.

We have recently shown that *Batf*$^{-/-}$ $CD8^+$ T cells show profoundly impaired effector $CD8^+$ T cell differentiation (174). To test whether BATF knockdown in wild-type $CD8^+$ T cells also impaired $CD8^+$ effector T cell development, we adoptively transferred naive P14 $CD8^+$ T cells from bone marrow chimeras transduced with either an inducible shBATF vector or a control shRNA vector targeting LacZ in a 1:1 ratio with naive P14 $CD8^+$ T cells from a bone marrow chimera transduced with a second control shRNA (shRFP) into wild-type recipients (Figure 4.8A). Endogenous, shBATF- (or control shLacZ-) and shRFP-naive $CD8^+$ T cells were distinguished by the use of congenic markers. Comparison of the ratios of numbers shBATF- or shLacZ-effector T cells to shRFP-effector T cells was used to determine the effect of BATF knockdown while controlling for any effect of shRNA expression on differentiation. We found markedly reduced numbers of P14 shBATF-effector $CD8^+$ T cells at day 7 – 9 post infection (p.i.) relative to shRFP P14 $CD8^+$ T cells, when the cells were exposed to IPTG from d-3 until the end of the experiment. In contrast, the ratio of shLacZ-effectors to shRFP-effectors remained constant (Figure 4.9, B and C). This reduction in shBATF-effector cell numbers was seen with three different BATF shRNAs designed with different seed regions, making this unlikely to be due to off-target effects (Figure 4.9C) (180).

To identify the reason for the reduced population size of effector $CD8^+$ T cells following BATF knockdown, we measured both cell death and proliferation in shBATF-effector $CD8^+$ T cells at d5 p.i. Analysis of active caspase abundance showed significantly higher apoptosis in shBATF-effector $CD8^+$ T cells (Figure 4.9D). In addition, there was a

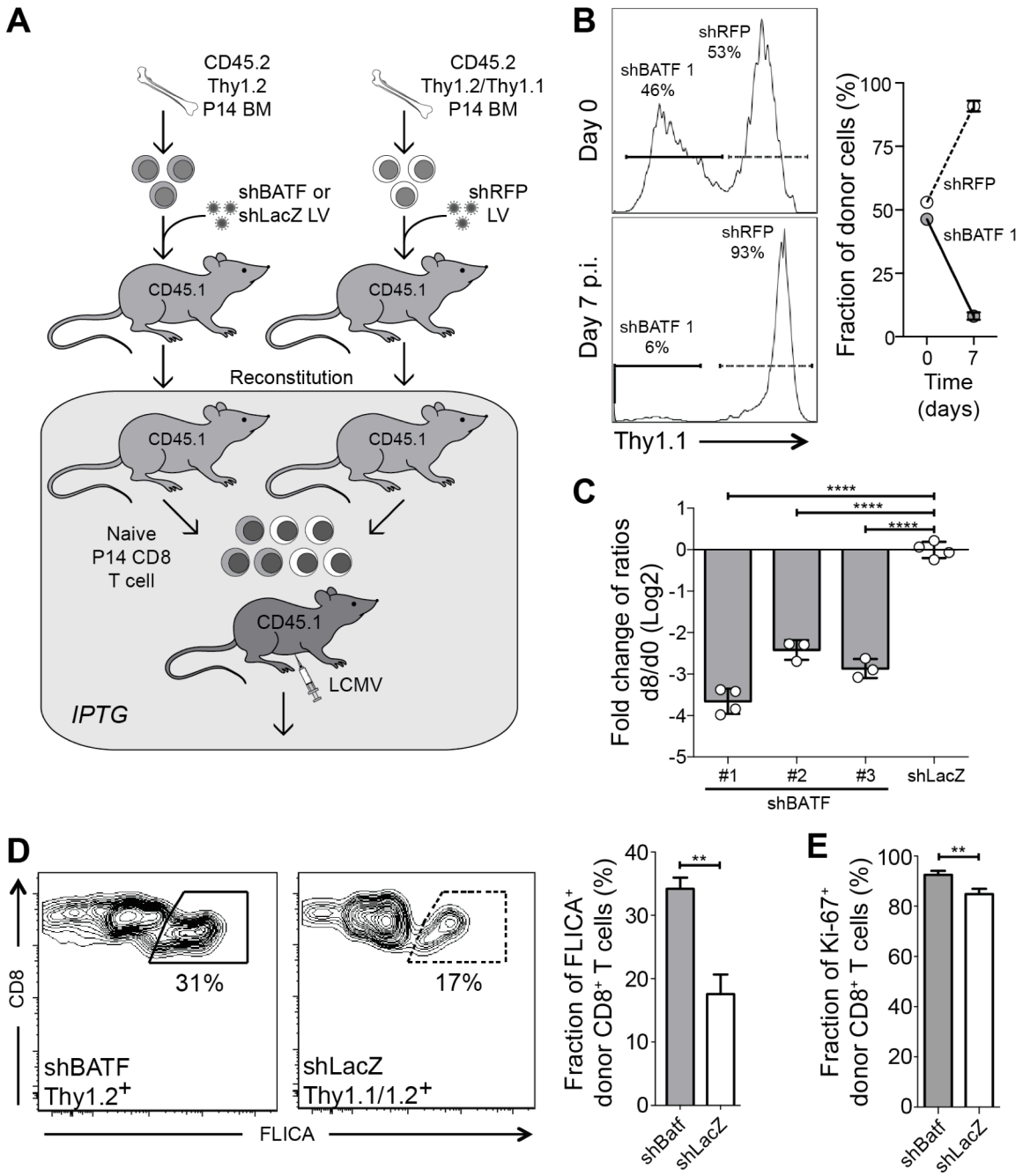

Figure 4.8. BATF knockdown *in* vivo in primary $CD8^+$ T cells impairs effector differentiation. (**A**) Schematic diagram of the experiment. (**B**) Relative fraction of P14 shBATF- (solid gates/lines) or shRFP- (dotted gates/lines) $CD8^+$ T cells at the time of transfer or at d7 p.i. in IPTG-treated animals (from d-3 on). Representative plots (left, middle panels) from a single animal and summary data from 5 mice (right panel). (**C**) Ratio of P14 shBATF- or shLacZ-effector $CD8^+$ T cells to shRFP-effector $CD8^+$ T cells at d8 p.i. with IPTG induction. Ratio at d8 p.i. was normalized to ratio at d0 and shown for 3 different shRNAs targeting BATF. (**D**) Apoptosis of shBATF- or shLacZ-effector $CD8^+$ T cells measured by active Poly-Caspase staining (FLICA) at d5 p.i. (**E**) Ki-67 expression in shBATF and shLacZ effector $CD8^+$ P14 cells on d5 following cell transfer and LCMV infection. Significance was assessed with Student's t-test; $^{**}P < 0.01$, $^{****}P < 0.0001$. Representative data are shown from three (*B*, *C*) or two (*D*, *E*) experiments with 3-5 mice per group.

modest increase in the proliferation of the fraction of remaining shBATF-effector CD8$^+$ T cells compared to shLacZ-effector CD8$^+$ T cells (Figure 4.9E). Thus, knockdown of BATF impairs the development of an effector CD8$^+$ T cell response primarily by increasing cell death during early differentiation. These findings are consistent with previous studies using germline deletion of BATF, which have demonstrated that naive *Batf*$^{-/-}$ T cells undergo massive cell death at 72 – 96h after stimulation (174).

BATF is required to initiate but not maintain effector CD8$^+$ T cell differentiation.

Because previous studies of the role of BATF in effector CD8$^+$ T cell differentiation have been carried out using T cells with constitutive germline deletion, it is not known whether BATF is required only to initiate the development of CD8$^+$ effector T cells (i.e. at the time of initial antigen encounter), or whether BATF is also needed to maintain CD8$^+$ effector T cell development once underway. To address this question, we adoptively transferred 1:1 mixtures of congenically distinguishable P14 shBATF- and shLacZ-CD8$^+$ T cells into recipient wild-type animals, which were then infected with LCMV-Armstrong. IPTG was administered to induce BATF knockdown either before infection, at the time of infection, or 72h p.i. (Figure 4.9A). We assessed BATF knockdown at d8 p.i., and found that BATF transcript abundance was significantly reduced in shBATF compared to shRFP-effector CD8$^+$ T cells regardless of when IPTG was initiated (Figure 4.9B), and was not significantly different between any of the shBATF-effector cell conditions.

We observed profound differences in the ratio of shBATF:shLacZ-CD8$^+$ T cells at d8 p.i., depending on the time at which BATF knockdown had been initiated. BATF knockdown initiated 3 days prior to infection or at the time of infection was associated with a significant reduction in the numbers of d8 p.i. effector CD8$^+$ T cells compared to controls with no IPTG induction. In contrast, inducing BATF knockdown 72h post infection did not significantly change the numbers of effector CD8$^+$ T cells d8 p.i. (Figure 4.9C). These finding show that

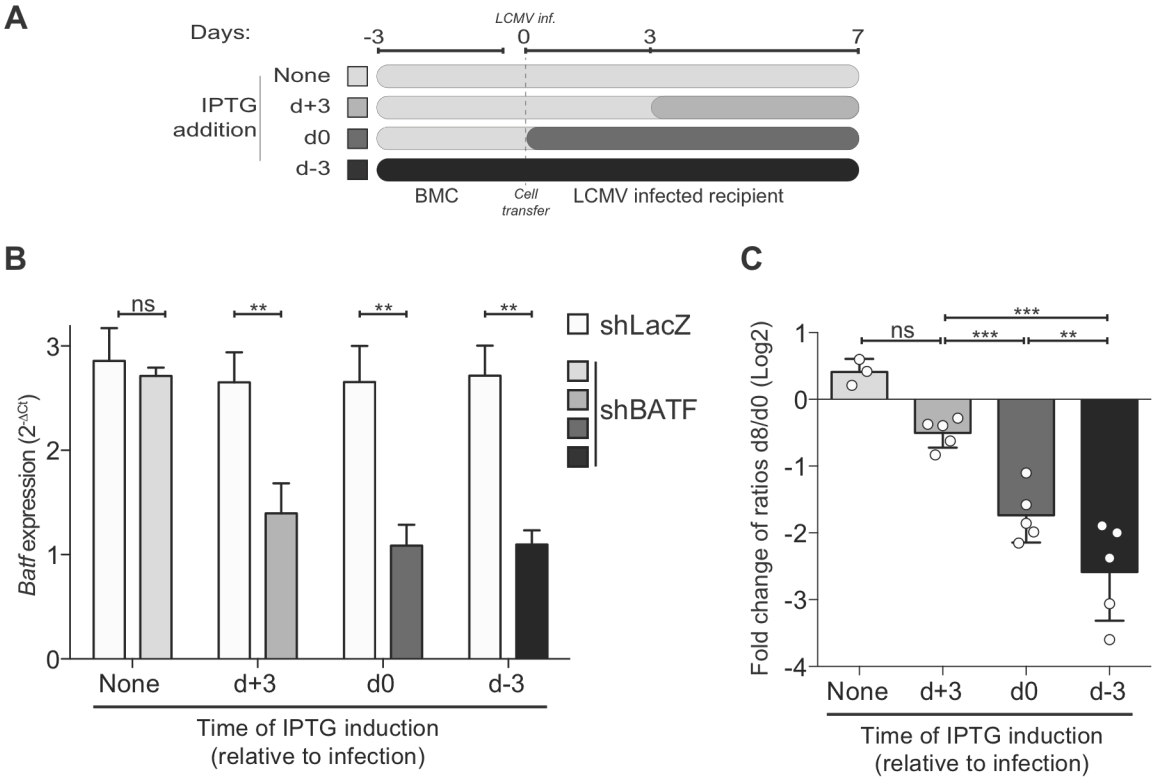

Figure 4.9. BATF is required to initiate but not maintain effector differentiation. (**A**) Schematic diagram of timing of IPTG administration. (**B**) BATF transcript abundance in shBATF– and shRFP–effector CD8$^+$ T cells on d 8 p.i. as described in A. (**C**) Ratio of shBATF– or shLacZ–effector CD8$^+$ T cells to shRFP–effector CD8+ T cells at d 8 p.i. with continuous IPTG exposure initiated at the times indicated. Day 8 p.i. ratios were normalized to the d 0 ratio and Log2 transformed. Significance was assessed with one-way ANOVA; **P < 0.01, ***P < 0.001, ****P < 0.0001. Representative data are shown from three experiments with three to five mice per group.

while BATF is required for effector $CD8^+$ T cell development at the time of initial antigen encounter, by 72h p.i., BATF becomes largely dispensable, at least through d8 of $CD8^+$ effector T cell differentiation.

Transduced LSK cells enable gene perturbation in B cells and highlight the role of BATF in humoral responses.

Bone marrow chimeric mice generated with inducible shRNA-transduced LSK cells had 90-95% of the immune system derived from the donor cells (data not shown). On average, 10-30% of these cells had stably integrated the lentiviral construct as measured by their expression of GFP. Thus, we were able to obtain all major immune cells with the integrated construct, providing the potential for their gene perturbation upon exposure to IPTG. To test if we could achieve inducible gene silencing in a cell type other than the $CD8^+$ T cell, we investigated shBATF-integrated B cells to assess the function of BATF in humoral responses, since BATF has previously been shown to play a vital role in B cell differentiation (181, 182).

To examine if we could show the functional role for BATF in B cells, we set up an *in vivo* competitive transfer system in which both control shLacZ- and shBATF-carrying B cells can be interrogated in the same microenvironment. This was done by transferring NP-OVA-induced follicular T helper (Tfh) cells and a 1:1 mixture of shLacZ:shBATF carrying B cells into Rag1-KO mice (Figure 4.10A), which lack all endogenous T and B cells. Rag1-KO mice were immunized with NP-OVA on the same day as they received the mixture of B and Tfh cells. At d10 following immunization, all B cells were thus derived from the transferred cells at d0. While naive shBATF-carrying CD45.1/CD45.2 double positive B cells were only slightly decreased in their abundance in the draining lymph node relative to the shLacZ-control CD45.1 positive naive B cells, there was a stark reduction of $GL7^+$ $CD38^-$ germinal center B cells with BATF knock-down compared to the control cells (Figure 4.10B). Additionally, there

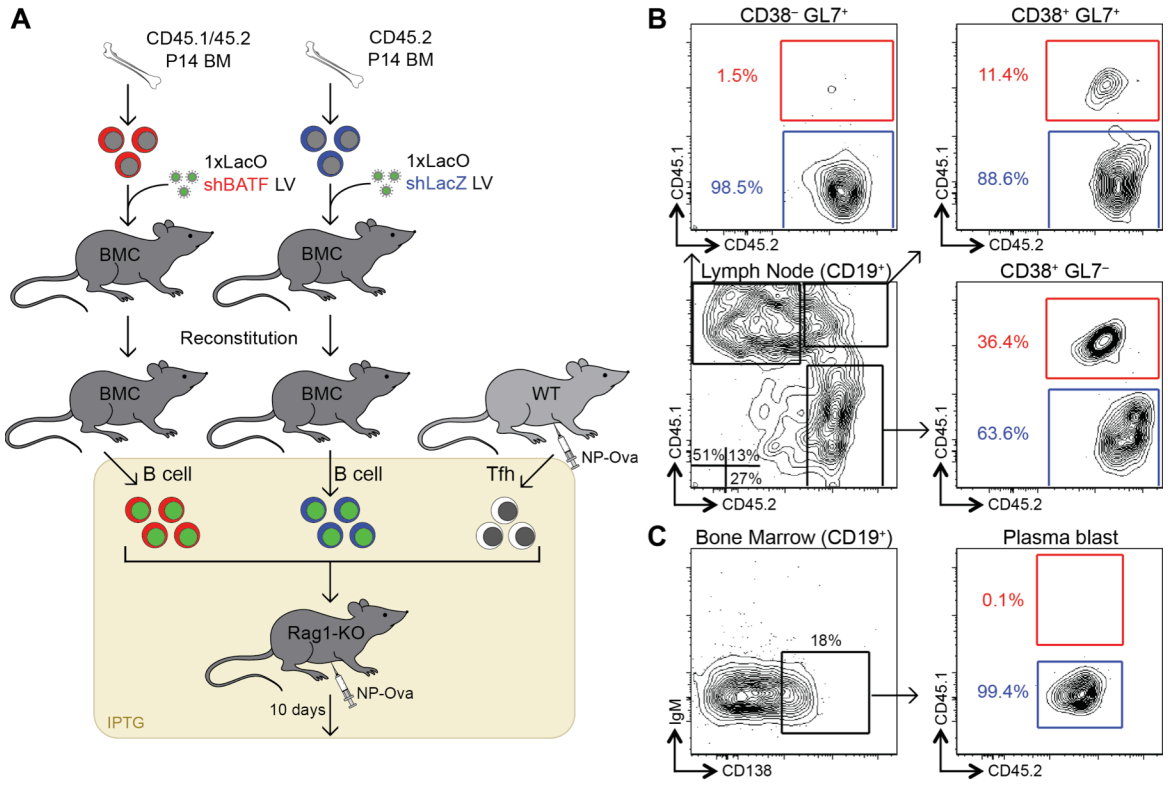

Figure 4.10. Bone marrow chimeras generated with inducible shRNA-transduced LSK enables functional gene perturbation in B cells. (**A**) Experimental schema describing the three-cell transfer experiment to assess the importance of BATF in B cells *in vivo*. (**B**) The subsets of $CD19^+$ B cells present in draining (inguinal) lymph node of NP-OVA-immunized Rag1-KO mice at d10 following immunization and B cell and Tfh cell transfer (bottom left). The relative abundance of $CD45.1^+/CD45.2^+$ shBATF B cells (red gate) and $CD45.1^+$ shLacZ B cells (blue gate) in each of the subsets shown on the right and top. (**C**) Gating of plasma blasts in bone marrow at d10 p.i. (left) and the relative abundance of shBATF and shLacZ plasma blasts shown on the right.

was a complete absence of shBATF cells in the plasma blast compartment in the bone marrow as all developing plasma blasts were derived from the shLacZ carrying control cells (Figure 4.10C). Thus, this platform also enables gene perturbation in mature naive B cells and confirmed the known role for BATF in B cell differentiation.

Discussion

We have developed a strategy to inducibly silence gene expression in unperturbed hematopoietic cells *in vivo* using RNAi. We used this system to show that BATF is required to initiate effector $CD8^+$ T cell differentiation, but becomes dispensable after 72h. Our findings suggest that this experimental approach can accelerate the understanding of how effector and memory T cell responses are regulated.

The ability to manipulate gene expression using RNAi is a powerful tool with which to investigate gene function in the immune system. Silencing gene expression in T cells (158, 177, 178) or dendritic cells (175, 183) has allowed the rapid identification of regulators of cell differentiation and activation. Viral vectors expressing shRNA molecules are a frequently used delivery method for RNAi. However, this approach is limited by the inability to deliver viral vectors to quiescent T cells. As a result, most approaches have used activation *in vitro* (176, 178) or *in vivo* (158, 177) of T cells to achieve efficient transduction. We show that both of these approaches profoundly alter the underlying transcriptional and functional state of naive T cells. T cells stimulated *in vitro*, even in the absence of TCR stimulation show up-regulation of TFs and cytotoxic molecules including granzymes, suggesting that many aspects of effector differentiation are initiated by these manipulations even before viral infection and gene knockdown can occur. In contrast, our experimental system permits the inducible knockdown of genes in quiescent, unperturbed naive T cells, allowing the events that occur during the initial minutes to hours after T cell stimulation to be interrogated. Thus our approach now allows functional genomic studies in transduction-refractory hematopoietic cells *in vivo* without *ex vivo* perturbation to examine numerous clinically relevant immune processes reliably modeled in the mouse.

We used this system to analyze the role of BATF in the differentiation of effector $CD8^+$ T cells immediately after antigen encounter in a mouse model of acute viral infection. BATF plays a profound role in the differentiation of many immune cell types (184). It is

required for the development of T_{FH} (183) and Th17 (185, 186) cells and for class switching in B cells (183), suggesting it plays an essential role in cell differentiation in diverse lineages. Recent studies in $CD8^+$ T cells show that loss of BATF deregulates a suite of TFs, cytokine sensors, and metabolic pathways, leading to a profound defect in effector $CD8^+$ T cell differentiation (174). However, while these studies show that BATF is required to initiate effector $CD8^+$ differentiation, it remains unclear whether BATF is also needed to maintain the differentiation program once underway. We now show that BATF is required at or immediately after antigen encounter but rapidly becomes dispensable for subsequent development of effector $CD8^+$ T cells until at least d8 p.i. These results suggest that BATF is required to initiate, but not maintain, the effector differentiation program.

Our findings are consistent with several prior observations that support a role for BATF in the initial commitment to an effector state. First, the defects in effector $CD8^+$ T cells that lack BATF are evident within 72 - 96 hours of antigen encounter (174), suggesting that BATF plays a critical function in the earliest hours of effector differentiation. Second, BATF associates with its binding sites within 24 hours of T cell activation, which suggests that it could play a regulatory role as early as hours after antigen encounter (174). Finally, during Th17 differentiation, loss of BATF results in decreased chromatin accessibility at some regions normally bound by BATF, which suggests that BATF may play a role as a pioneer TF (186). Pioneer TFs can regulate the chromatin structure at critical regulatory regions to enable the subsequent binding of other TFs (187). Our findings of a transient role for BATF at the initiation of effector differentiation are consistent with its function as pioneer TF. For instance, BATF may function at the start of effector $CD8^+$ differentiation to increase chromatin accessibility to other effector TFs that cement the differentiation state. Because BATF is no longer required for effector differentiation by 72h, other TFs, such as T-bet and STAT proteins, may be responsible for maintaining effector differentiation once underway (188).

Analysis of the temporal role of candidate effector TFs, either with our inducible RNAi system, or with conditional knockout strains will help test this hypothesis.

Our previous studies of BATF in $CD8^+$ differentiation showed that *Batf*$^{-/-}$ but not *Batf*$^{+/-}$ T cells displayed a defect in effector differentiation suggesting that the remaining expression of BATF in *Batf*$^{+/-}$ is sufficient for proper T cell activation (174). In this context, it is perhaps surprising that 68% knockdown of BATF (Figure 4.7E) was associated with such marked impairment of effector differentiation while the BATF heterozygous mouse did not show such a defect. Possibly, compensatory mechanisms or adaptation may be in effect when the cells are lacking half of genomic BATF throughout their development. Alternatively, there may be a critical "dose" of BATF that is required for normal effector $CD8^+$ T cell differentiation: knockdown of BATF reduces protein abundance below this threshold while heterozygous deletion does not.

Our findings help provide a mechanism to explain why effector $CD8^+$ T cell differentiation can be initiated by brief, transient TCR activation. Studies of temporally-limited antigen-exposure have shown that as little as 4h of TCR stimulation can initiate cell division and induce cytolytic function in naive $CD8^+$ T cells (33), and only 20h of stimulation can initiate a self-sustaining program of effector and memory $CD8^+$ T cell differentiation that is cell autonomous (35, 165). Although the ultimate fate of effector cells is strongly modulated by antigen persistence, inflammation and the cytokine milieu, these studies indicate that $CD8^+$ T cells encounter an irreversible decision point within hours of antigen encounter (35). Our findings suggest that transcriptional regulation by BATF may be one component of that decision point. BATF may launch differentiation by irreversibly engaging the effector transcriptional program within the first 24h of stimulation.

Although we have demonstrated this experimental approach to investigate the role of BATF in the early commitment events in effector $CD8^+$ T cell differentiation, we anticipate that this approach can also be extended to discover genes that regulate differentiation and

longevity of memory $CD8^+$ T cells, and the mechanisms that lead to $CD8^+$ T cell exhaustion. Moreover, the use of a bone marrow chimeric system results in transduction of all hematopoietic-derived lineages with the inducible shRNA vector. This makes analysis of gene function in other cell types that are refractory to viral transduction such as naive $CD4^+$ T cells or B cells is now equally feasible. As such, this approach can be used to provide insights into important clinical questions such as how protective immune responses to vaccination are generated, or how T cell dysfunction arises in chronic viral infection or cancer. Concerns regarding off-target effects of RNAi need to be considered with this system as for any RNAi-based experiment. However, the use of this strategy should provide a broadly useful tool for interrogating gene function in unperturbed cells of hematopoietic origin.

Materials and Methods

Mice

P14 TCR Tg mice were used as described (110). Wildtype C57BL/6J, Ly5.1 (CD45.1), and Thy1.1 mice were purchased from Jackson Laboratories (Bar Harbor, ME). All mice were used according to the Harvard Medical School Standing Committee on Animals and National Institutes of Animal Healthcare Guidelines. Animal protocols were approved by the Harvard Medical School Standing Committee on Animals.

Generation of bone marrow chimeras

Bone marrow was isolated and red blood cells lysed using ACK lysis buffer (Gibco). LSK cells were enriched using anti-CD117 microbeads (Miltenyi Biotech) and then sorted using a BD FACSAria cytometer. Sorted cells were plated overnight in StemSpan SFEM medium (StemCell Technologies) with 100 µg/ml recombinant stem cell factor, thrombopoietin, IL-7, and Flt3-ligand (PeproTech). Cells were then spin-infected with lentiviral supernatants at 650 x g for 90 min at 37°C on non-treated plates that had been coated overnight at 4°C with 100 µg/ml RetroNectin (Takara Bio). Fresh medium was added after 1h and cells were rested overnight. Viral stocks were titrated to ensure the majority of cells were infected with a single virus. The cells were then washed in PBS (Gibco) and 50,000 cells were injected intravenously in recipient mice that had been irradiated with 2 doses of 600 cGy, 3 hours apart.

Lentivirus production

293T cells were seeded in DMEM with 10% FBS. The following day, the cells were transfected with shRNA construct pLKO.1-TRC005 or 1xLacO (which is now available from Sigma-Aldrich under the name “MISSION 1X LacO Inducible”) with the packaging plasmids

Pax2 (gag, pol) and VSV-G using TransIT-LT1 (Mirus Bio) or ExGen 500 (Thermo Scientific Fermentas). Viral supernatants were collected 48h – 72h later.

In vitro knockdown of GFP

A stable GFP-expressing Jurkat cell line was constructed by cloning GFP^+ cells, following transduction with PGK-eGFP lentivirus. GFP-Jurkat cells were transduced with shRNA targeting GFP under a constitutive (pLKO.1-TRC0056, with puromycin resistance) or inducible (1xLacO, with Thy1.1 reporter) promoters. Varying doses of dioxane-free IPTG (Promega) was added at the indicated concentrations and durations. The fraction of remaining GFP was assessed with Accuri C6 flow cytometer (BD Biosciences) as the GFP MFI of the transduced relative to the GFP MFI of the untransduced GFP^+ Jurkat cells, both normalized to the fluorescence of the unmanipulated (non-GFP-expressing) Jurkat cells.

T cell transfers and infections

$CD8^+$ T cells from bone marrow chimeric animals were isolated using the $CD8^+$ T cell isolation kit II for magnetic separation (Miltenyi Biotech) and then GFP^+ $CD8^+$ congenic cells were sorted using BD FACSAria cytometer. P14 $CD8^+$ T cells (10^4 – 10^6 cell/animal) were injected in recipient mice i.v., which were subsequently infected *intraperitonealy* with 2 x 10^5 p.f.u. LCMV Armstrong. For influenza infections, the mice were anesthetized with 2.5% Avertin and infected with 0.5 LD_{50} H1N1 *Influenza* virus (PR8), engineered to express GP_{33-41} peptide of LCMV (PR8-GP33)(93), intranasally. Both viruses were a generous gift of Dr. E. John Wherry (University of Pennsylvania School of Medicine, Philadelphia, PA).

Flow cytometry and cell sorting

Single cell suspensions from spleen or bone marrow were stained with combinations of anti-CD8α (53-6.7), anti-CD4 (RM4-5), anti-B220 (RA3-6B2), anti-CD11b (M1/70), anti-

CD11c (N418), anti-CD44 (IM7), anti-CD62L (MEL-14), anti-CD45.1 (A20), anti-CD45.2 (104), anti-Thy1.1 (OX-7), anti-Thy1.2 (30-H12), anti-CD25 (PC61), anti-CD27 (LG.3A10), anti-CD122 (TM-β1), anti-CD127 (A7R34), anti-CXCR3 (CXCR3-173), anti-TNF-α (MP6-XT22), anti-IFNγ (XMG1.2), anti-Granzyme B (GB11), anti-ICOS (15F9), anti-CD19 (6D5), CD69 (H1.2F3), anti-Fas (Jo2), and anti-GL7 (GL7) (all from BioLegend), anti-T-bet (O4-46) and anti-Ki-67 (B56) (from BD Biosciences), anti-Eomes (Dan11mag) (from eBioscience), and anti-KLRG1 (2F1) (from Abcam). For CXCR5, biotinylated anti-CXCR5 (2G8; BD Biosciences) was used, followed by streptavidin-Brilliant Violet 421 (Biolegend). Poly-Caspase activity was detected using FLICA Vybrant-FAM Assay Kit (Life Technologies). For intracellular cytokine staining, splenocytes were first stimulated with 0.5 μg/ml GP_{33-41} (Genscript) or no peptide for 5 hours at 37°C in the presence of GolgiPlug (BD Biosciences). For intracellular staining of cytokines and Ki-67, the cells were surface stained, fixed/permeabilized, and intracellularly stained using the Foxp3/Transcription Factor Staining Buffer Set (eBioscience) as directed by the manufacturer. To assess T cell proliferation, mice were injected with 2 mg BrdU i.p. 16h prior to analysis and BrdU incorporation was detected using the FITC BrdU Flow Kit (BD Pharmingen).

LSK cells were sorted from CD117-enriched bone marrow cells stained with CD117 (ACK2), Sca-1 (D7) and a lineage antibody cocktail which included biotin-labeled anti-CD5 (53-7.3), anti-Gr1 (RB6-8C5), anti-B220 (RA3-6B2), anti-CD3e (145-2C11), anti-CD11b (M1/70), anti-Ter-119 (Ter-119) and detected with fluorophore conjugated Streptavidin (all from BioLegend). Data were acquired using LSR II or Accuri C6 (BD Biosciences) cytometers and analyzed with FlowJo software (v9.7.2, TreeStar).

shRNA construct generation

Target sequences of the shRNA used are: shBatf 1 (CCGCAAAGAGATCAAACAGCT), shBatf 2 (CTGGACAAGTATTGAACACAA), shBatf 3

(GAGCTCAAGTACTTCACATCA), shLacZ (CCGTCATAGCGATAACGAGTT), shRFP (GCTTCAAGTGGGAGCGCGTGA), shGFP (ACAACAGCCACAACGTCTATA). Cloning methods can be found at http://www.broadinstitute.org/rnai/public/. Briefly, complementary oligos (IDT) were annealed and ligated into AgeI and EcoRI digested pLKO.1-TRC005 and 1xLacO (both obtained from The RNAi Consortium, Broad Institute). All ligated constructs were sequence verified for the presence of the correct shRNA by Sanger sequencing.

Gene expression analysis

Congenically marked $CD8^+$ T cells were sorted from chimeric mice or following *in vitro* stimulation. Cells were lysed and RNA was extracted using RNeasy Plus Mini Kit (Qiagen). cDNA was generated using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and analyzed on an ABI 7500 Fast real-time quantitative PCR instrument using gene Taqman probesets from Applied Biosystems.

Immunoblotting

$CD8^+$ T cells stimulated *in vitro* with aCD3/aCD28 were lysed in modified Laemmli buffer (60 mM Tris-HCl [pH 7.2], 10% glycerol, and 2% sodium dodecyl sulfate [SDS]) containing 1 U/ml of DNase (Benzonase nuclease; Novagen) and Complete protease inhibitor cocktail (Roche) for 30 min at 4°C. Protein concentration in the lysates was estimated with a bicinchoninic acid assay (Thermo Scientific), and 75 mg of each lysate was subjected to SDS-polyacrylamide gel electrophoresis and western blotting on nitrocellulose membranes as previously described (189). The primary antibodies used were b-actin (1:5000 dilution, Abcam, ab8227) and BATF (1:500 dilution, Brookwood Biomedical, PAB4003). Densitometry was performed using ImageJ software (NIH).

Generation of follicular T helper cells and B cell transfer

To generate follicular T helper (Tfh) cells, C57BL/6J mice were immunized subcutaneously in flanks with 100 µg NP18-OVA (Biosearch Technologies) emulsified in an emulsion of *Mycobacterium tuberculosis* strain H37RA complete Freund's adjuvant (Difco) at 1:1 ratio. $CD4^+$ $CXCR5^+$ $ICOS^+$ Tfh cells were isolated from the draining (inguinal) lymph nodes 7-8 days after immunization. B cells from bone marrow chimeric animals generated as described above were first enriched using anti-CD19 microbeads (Miltenyi Biotech) and then sorted for GFP^+ $CD19^+$ cells. shBATF and shLacZ-integrated B cells were mixed at a 1:1 ratio. A mixture of 2×10^6 polyclonal B cells and 2.5×10^5 Tfh cells was injected i.v. in Rag1-deficient mice (Jackson Labs). These mice were subsequently immunized with NP18-OVA as described above. Single cell suspensions from inguinal lymph node, spleen, and bone marrow were analyzed using flow cytometry 10 days later immunization.

Acknowledgments

I would like to acknowledge Glenn Cowley, John Doench, Ozan Alkan, and David Root (The RNAi consortium, Broad Institute of MIT and Harvard) for sharing pLKO.1-TRC005 and 1xLacO vectors and protocols for vector cloning and lentivirus production. Immunoblotting data was generated with the help of R. Anthony Barnitz (Haining Lab, Dana-Farber Cancer Institute) and the *in vivo* experiments interrogating B cell biology were performed with the help of Peter T. Sage (Sharpe Lab, Harvard Medical School). E. John Wherry (University of Pennsylvania Perelman School of Medicine) generously provided the LCMV-Armstrong for infections.

Chapter 5. *In vivo* RNAi screen identifies TGIF1 as a novel negative regulator of $CD8^+$ T cell effector function

Introduction

Several transcription factors have important roles in regulating $CD8^+$ T cell activation, proliferation, and differentiation (6, 19, 20, 39, 74, 174, 190, 191). Numerous additional genes have been suggested to contribute to this process based on the kinetics of their expression (23, 75, 163, 190); yet technical limitations make functional characterization of many of these gene products challenging. While generating germline knockouts for putative regulatory genes of interest is possible for small numbers of genes, it is not feasible to assess the functional role of hundreds of putative regulators in primary T cells *in vivo*. Additionally, potential lethality associated with some gene knockout animals forces researchers to pursue the generation of conditional knockouts or fetal liver chimeric animals. Furthermore, many genes such as Notch may control both in early T cell development in the thymus as well as in effector and memory generation from naive T cells (192-195).

In an effort to establish functional roles for novel genes in particular biological processes, many researchers have used the power of RNAi screens (196). This technology relies on the RNA-induced silencing complex (RISC) degrading the targeted gene's mRNA to which a complementary small-interfering RNA (siRNA) is bound (197, 198). The siRNA is either delivered directly through electroporation of cells or through a retroviral vector that stably integrates in the genome and expresses shRNA, which is then cleaved by Dicer to generate siRNA. Due to technical limitations, most screens have been performed *in vitro*, as it is often challenging to introduce siRNA-transfected or retrovirally transduced cells into an animal for examining the role of siRNA-targeted genes *in vivo*. In cancer biology, RNAi screens have been successful to identify genes important in malignant cells using *in vivo*

RNAi screens by injecting shRNA-bearing tumor cell lines in mice (199-201). However, screens in primary immune cells have been more challenging.

An RNAi-based screening approach is very appealing for deciphering the functional roles of genes in $CD8^+$ T cell effector differentiation and memory development. However, there are two main difficulties in approaching T cell memory formation with established screening tools: (i) T cells are very difficult to transduce without activation which can itself perturb $CD8^+$ T cell differentiation; and (ii) physiological effector differentiation and memory development cannot be fully recapitulated physiologically *in vitro*, requiring screens to be performed *in vivo*.

Although shRNA have been successfully introduced into mouse T cells, all of these methods rely on activation of T cells prior to transduction (19, 98, 174, 178, 179, 202-204). There are several potential difficulties with this experimental approach including: (a) the delay in knocking down target genes that are up-regulated immediately following activation makes assessing their phenotype difficult; (b) excess stimulation with cytokines or through TCR potentially leads to use of non-physiological states of T cells in screening experiment; (c) viral vectors can activate innate sensors in various splenic cell populations; and (d) injection of dead or dying cells, a frequent side effect of viral transduction.

For these reasons, we have used the novel *in vivo* RNAi platform described in the previous chapter of this thesis to screen pools of shRNAs targeting curated sets of genes of interest. This system enables generation of rigorously naive $CD8^+$ T cells with genomically integrated inducible shRNA that enable us to systematically interrogate the functional role of genes in effector differentiation *in vivo*.

We used the inducible shRNA approach in a functional genomic screening approach to identify genes that control the effector differentiation of primary naive $CD8^+$ T cells into effector $CD8^+$ T cells responding to a viral infection. Due to limitations, such a screening approach cannot be done on a genome-wide level, primarily because of the bottlenecks in

the progression of shRNA diversity during the development from an LSK cell into a mature naive $CD8^+$ T cell (205). These include both technical limitations, such as potential loss of certain shRNAs in the process of LSK lentiviral transduction, as well as biological limitations, as not all LSK cells will engraft in irradiated recipients, and only a subset of those eventually contribute to the seeding of the thymus with thymic progenitors that eventually give rise to mature T cells (206, 207). For these reasons, we performed a targeted screen that focused on a defined functional class of proteins confirmed to be expressed by $CD8^+$ T cells based on transcriptomic data and whose expression is up-regulated following cell activation. Specifically, we focused on transcription factors and DNA-binding proteins for several reasons: 1) have a profound influence on cellular state and function; 2) they often have a more profound effect on cell biology as they regulate numerous target genes; 3) partial loss of function induced by RNAi can cause a phenotype which is not always true for other classes of proteins, such as enzymes, where a minor amount of remaining protein can continue to exhibit significant enzymatic activity; 4) their biology is difficult to perturb with conventional methods such as small molecules or monoclonal antibodies; and 5) knockout mice exist for many of these proteins, enabling follow-up functional studies.

The *in vivo* RNAi screen measured the effects of ~80 genes on effector P14 $CD8^+$ T cell proliferation following activation by LCMV Armstrong. We scored the functional relevance of genes by the relative enrichment or preferential loss of shRNAs targeting respective genes at d8 p.i. compared to their representation at the time of P14 cell transfer and LCMV infection (d0). We identified several known and potentially novel regulators of $CD8^+$ T cells activation having both agonistic and antagonistic effects on this process.

Single gene validation experiments led us to focus on a potentially novel regulator of T cell function, TGF-β-induced factor homeobox 1 also known as. 5'-TG-3'-interacting factor (TGIF1) for several reasons. First, its expression is up-regulated in both $CD8^+$ and $CD4^+$ T cells responding to acute and chronic infection in early as well as late stages of infection (75,

76). Second, the only characterization of this transcription factor has been performed in embryonic neural development, hematopoiesis, and tumorigenesis (208-210), but its function in the immune system has not yet been interrogated. Limited work on TGIF1 suggests that it may interact with several genes well-appreciated in T cell biology such as, Smad2 downstream of TGF-β signaling (211), retinoic X receptor alpha (RXRα) in retinoic acid signaling pathway (212-214), c-Jun (215), and HDAC1 (211, 216-218). Our data suggest an unappreciated role for TGIF1 in $CD8^+$ T cells where in negatively regulating cell activation and proliferation, and in skewing the differentiation of effector $CD8^+$ T cells. TGIF1 knockdown in effector $CD8^+$ T cells decreased the relative abundance of $KLRG1^+$ $CD127^-$ SLECs and increased $KLRG1^-$ $CD127^+$ MPECs. Similar trends were also observed in conditional TGIF1-knockout mice in which TGIF1 is selectively absent in peripheral $CD8^+$ T cells. Preliminary data additionally suggest that TGIF1 may be directly regulating one of the main drivers of SLEC differentiation, Blimp-1 downstream of IL-2 signaling.

We have thus established the first *in vivo* RNAi screen in unmanipulated primary $CD8^+$ T cells and showed that it can be applied to screen various additional groups of genes in $CD8^+$ T cells as well as any hematopoietic cell that are refractory to conventional transduction protocols, such as B cells, NK cells, and NKT cells. Additionally, we showed the usefulness of the platform and found a previously unappreciated role of transcription factor TGIF1 as a negative regulator of $CD8^+$ T cell activation and a regulator of effector $CD8^+$ T cell differentiation.

Results

Existing public data enable prioritization of potential novel regulators of CD8$^+$ T cells function

To select genes for our focused RNAi screen, we analyzed TF transcript abundance in previously published publicly accessible genome-wide expression microarray data from naive and virus-specific CD8$^+$ and CD4$^+$ T cells responding to either LCMV-Armstrong or Clone 13 at early activation and expansion phases (days 6 and 8 p.i.), at memory or exhaustion stage, respectively (day 30 p.i.) (75, 76) (GSE41870), as well as studies comparing short-lived effector cells (SLEC) and memory progenitor effector cells (MPEC) at early stages following LCMV-Armstrong (39, 40) (GSE8678, GSE10230). To select genes, we first transformed the mouse microarray datasets in the space of human ortholog genes by Affymetrix probe mapping from Mouse 430.2 Array probes into Human U133A Array probes and subsequently the corresponding official gene symbols. We then extracted confirmed annotated transcription factors (219) and added additional manually-annotated DNA binding molecules (Nir Yosef, UC Berkeley, unpublished data) to compile a list of 2671 genes that could serve as potential transcriptional regulators. We further filtered for genes that were readily detected at the transcript level in T cells at any stage (maximum Log2-transformed probe value above 6.0), leaving 1467 genes. From this list, we identified groups of differentially expressed genes (fold change ≥ 1.10) in the pairwise comparisons of CD8$^+$ and CD4$^+$ T cells at day 6 vs. day 0, day 8 vs. day 0, day 30 vs. day 0 in LCMV-Armstrong and LCMV-Clone 13 infection, and in MPEC versus SLEC, resulted in a total 271 genes (Supplemental Table 5.1). These genes were then prioritized using lexicographical ordering (220) where the weight was placed in the following order: 1.) CD8$^+$ T cells at day 6 and day 8 in LCMV-Armstrong; 2.) CD8$^+$ at day 6 and day 8 in LCMV-Clone 13; 3.) CD8$^+$ T cells at day 30 in LCMV-Armstrong; 4.) CD8$^+$ at day 30 LCMV-Clone 13; 5.) genes differentially expressed in MPEC and SLEC CD8$^+$ T cells, and 6.) CD4$^+$ T cells in the course of LCMV

infections as in $CD8^+$ T cells (Supplemental Table 5.1). We used this prioritization to build a pool of shRNA that may be most applicable to interrogate $CD8^+$ T cell differentiation in the context of acute viral infection but that may also be subsequently used to screen exhaustion state of $CD8^+$ T cells in chronic infections or to interrogate the biology of $CD4^+$ T cells.

In this way, we compiled a list of ~80 genes we hypothesized may be most important in effector differentiation of $CD8^+$ T cells based on their transcriptional behavior in mouse T cells (Supplemental Table 5.2). This list included known T cell fate regulators (*Tbx21, Prdm1,* and *Ikzf, Runx, Stat,* and *Irf* family members), genes important for overall effector T cell differentiation (*Fos, Nfkbib, Batf*), genes important for proliferation (*E2f* family), as well as numerous genes with no known function in T cells (*Zfp* family of genes, *Tgif1, Rfx7,* etc.). We included 3-5 different shRNAs targeting each gene, selecting shRNAs with validated efficient knockdown (The RNAi Consortium, Broad Institute of MIT and Harvard, unpublished data) and separated the shRNAs into four non-overlapping pools. Each pool contained shRNAs targeting 13-21 genes and included 25 control shRNAs targeting irrelevant genes such as red fluorescent protein (RFP), luciferase (Luc) and LacZ, to contain in a total of 77-91 shRNAs per pool (Supplemental Table 5.2). This distribution of experimental and control gene enables robust analysis for functional effect beyond stochastic events in T cell activation and proliferation

In vivo RNAi system in primary $CD8^+$ T cells identifies novel regulators of effector activation

In vivo RNAi screen was set up to identify regulators of effector $CD8^+$ T cell activation. As a readout of this process, we used cell accumulation that is reflective of both proliferation and survival of T cells following activation *in vivo*. To circumvent the effects of the antigen specificity of different T cell clones responding to LCMV Armstrong infection, we used P14 TCR transgenic T cells. These T cells were generated in bone marrow chimeras from P14

TCR transgenic LSK cells transduced with a pool of different shRNA-1xLacO-GFP integrating lentiviruses described in Chapter 4. This approach enabled reliable acquisition of inducible shRNA-integrated P14 $CD8^+$ T cells following reconstitution. These shRNA-integrated P14 T cells were transferred into mice infected with LCMV-Armstrong. The transcription of shRNA in $CD8^+$ T cells was induced by adding IPTG in drinking water of bone marrow chimeric mice three days prior to cell transfer and was maintained in LCMV infected mice throughout the experiment.

At the beginning of the experiment, a portion of these shRNA-integrated P14 cells were set aside to enumerate representation of all shRNAs in the starting population of cells. The rest of the cells were either stimulated *in vitro* with anti-CD3 and anti-CD28 in the presence of IL-2 or were transferred in a separate cohort of CD45.2 wild-type mice that were subsequently infected with LCMV-Armstrong. The relative distribution of hairpins was assessed at days 3 and 6 for *in vitro* stimulated cultures and in transferred $CD8^+$ T cells extracted from the spleens 8 days following cell transfer and LCMV infection (Figure 5.1).

The fate of a single cell is variable even when the TCR is fixed due to external factors such as the timing or the microenvironment of T cell activation (221-223). However, previous reports suggest that tracking ~50 cells reproduces the behavior of the whole population (221, 222). To avoid stochastic expansion or loss of particular clones regardless of the function of the shRNA knockdown, we transferred 10,000 cells with a maximum of 100 different shRNAs resulting in each hairpin being represented on average in 100 cells at the beginning of the experiment. Indeed, the distribution of control shRNAs in a pool did not change significantly throughout the *in* vivo or *in vitro* expansion of cells relative to their distribution in the starting population of cells (Figure 5.2A). In contrast, the overall distribution of experimental shRNAs, did change from day 0 to days following *in vitro* and *in vivo* stimulation (Figure 5.2B). Interestingly, the changes induced by shRNAs *in vitro* were minimal compared to the reproducible changes across the different biological replicates *in vivo*, highlighting the

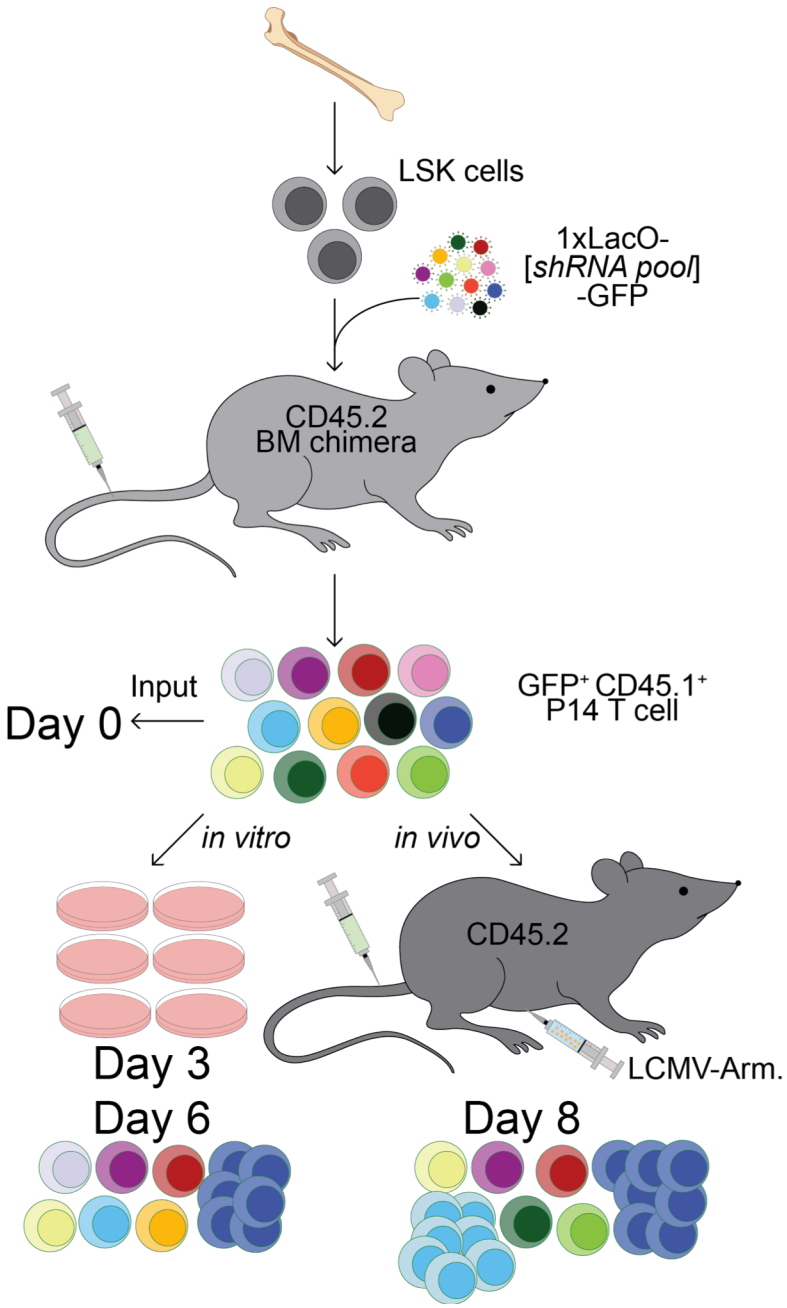

Figure 5.1. A schema describing the experimental setup for the *in vivo* RNAi screen in CD8$^+$ T cells. Bone marrow chimeric mice were generated using congenic P14 LSK cells that were transduced with a pool of ~80 different GFP-expressing shRNA lentiviruses. Following complete reconstitution, the chimeric mice were treated with IPTG for 3 days and sacrificed to isolate GFP$^+$ CD8$^+$ P14 cells. These cells were either 1) sequenced to capture the shRNA representation at d0, 2) stimulated *in vitro* for 3 or 6 days, or 3) transferred into congenic mice that were infected with LCMV-Armstrong and collected 8 days p.i. for quantification of relative abundance of the different shRNAs.

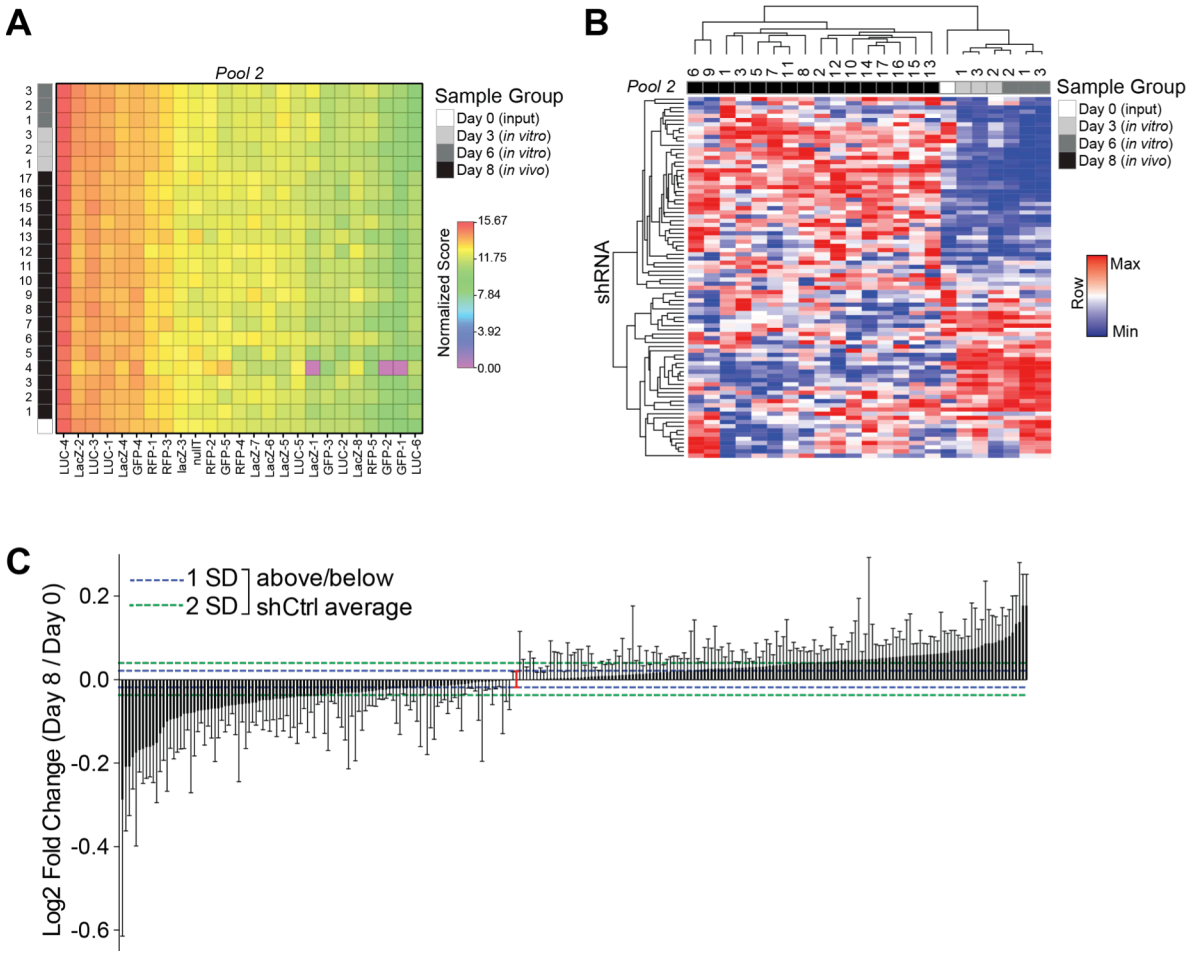

Figure 5.2. *In vivo* activation and proliferation of CD8$^+$ T cells changes the distribution of experimental, but not control, shRNAs. (**A**) Relative representation of control shRNAs at d3, d6 (*in vitro*) or d8 (*in vivo*), ranked by their abundance at d0. (**B**) Unsupervised bi-clustering of all experimental shRNAs in a representative screening pool. (**C**) The Log2-transformed fold change of the relative abundance of experimental shRNAs (black) at d8 compared to d0 across. Each column represents 1 shRNA targeting 1 gene and the error bars represent SD across biological replicates. The average and SD of control shRNA are represented in red. The blue and green dotted lines represent 1 SD and 2 SD above and below the average of control shRNAs.

importance of performing such experiments in a physiologically relevant *in vivo* setting (Figure 5.2B).

We identified several shRNAs that scored highly as significantly enriched or depleted relative to the distribution of control hairpins, and were preferentially lost (corresponding to positive regulators of $CD8^+$ T cell activation, Figure 5.2C, bottom left), or were enriched (corresponding to negative regulators, Figure 5.2C, top right). The ranked list included shRNAs targeting positive regulators that were preferentially lost from the $CD8^+$ T cell population at d8 p.i. (Figure 5.2C, bottom left), as well as shRNAs targeting negative regulators of $CD8^+$ T cell activation that were enriched at d8 p.i. (Figure 5.2C, top right) (Supplemental Table 5.3). Positive regulators included both known genes, such as *Fos (224-226)*, *Irf7 (227)*, and *HnrpII* (228, 229), as well as potentially novel agonists of T cell activation and/or proliferation following stimulation, such as *Hhex*, *Zik1*, and *Chaf1b*. Additionally, the experiment identified known negative regulators of $CD8^+$ T cell responses, such as *Nfkbib* (230, 231) and *Spi1* (232) as well as potential novel regulators including *Tgif1*, *Wdhd1*, and *Zfp414*.

Regulators identified in the screen can be validated and represent robust biology

To validate the functional role of potentially novel regulators of $CD8^+$ T cell activation and confirm that the *in vivo* RNAi screen using inducible shRNA yields biologically meaningful data, we performed P14 transfer experiments testing one shRNA at a time. We transferred mixtures of P14 $CD8^+$ T cells with integrated shLacZ or an experimental shRNA as illustrated in Figure 4.8A. We tested two type 1 interferon inducible genes, *Irf1* and *Irf7 (227, 233, 234)* with a single shRNA each and three potentially novel negative regulators of $CD8^+$ T cell activation – *Tgif1*, *Wdhd1*, and *Zfp414* – using 3-5 unique shRNAs (Figure 5.3A). The data analyzing a single shRNA for a single gene per mouse mostly corroborated the results obtained from the *in vivo* RNAi screen, where *Irf1* and *Irf7* knock-downs in $CD8^+$ T

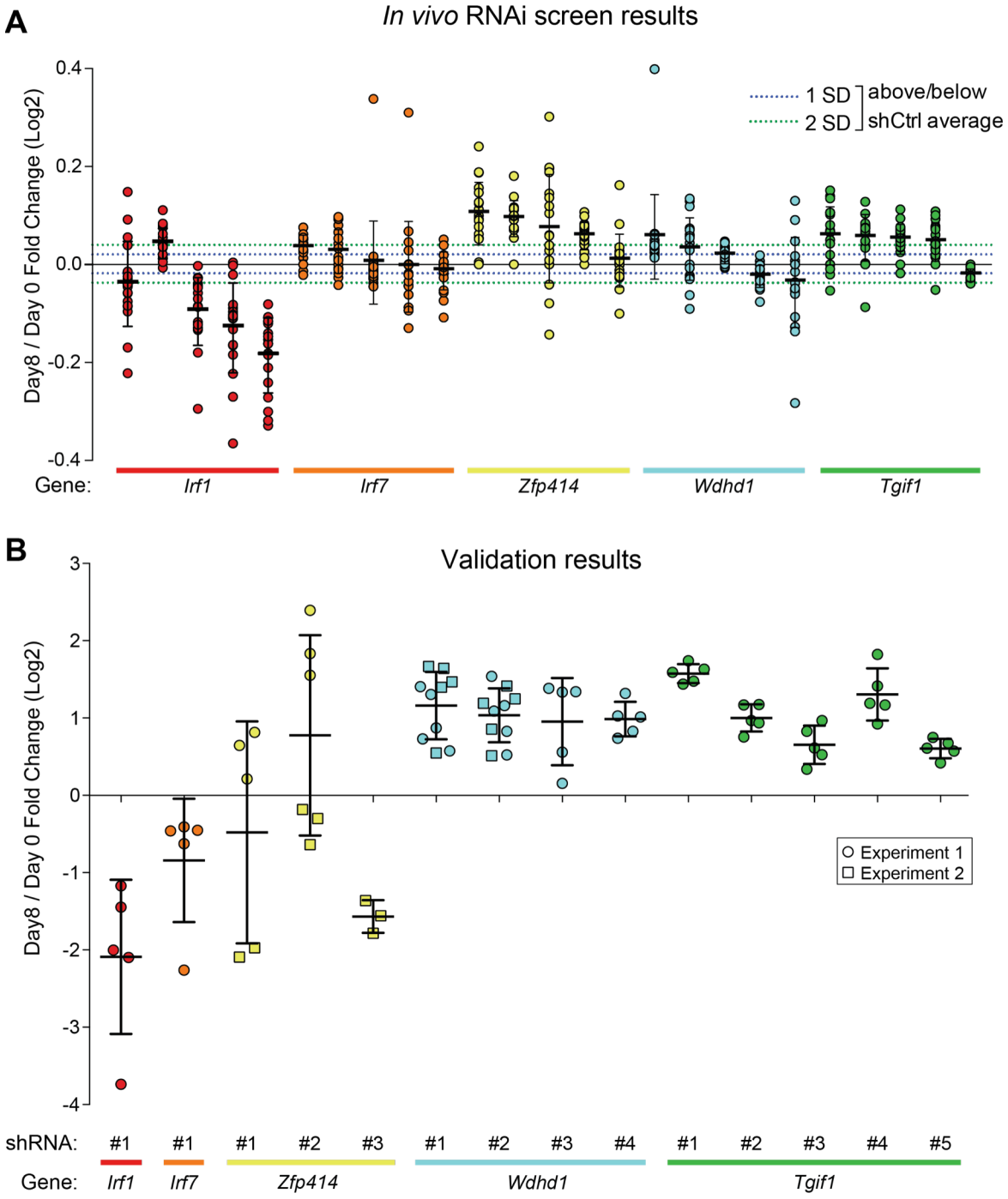

Figure 5.3. The functional significance of several genes identified in the screen is confirmed in validation experiments. (**A**) Results from *in vivo* RNAi pooled screen that are validated in B. (**B**) P14 competition experiments using 1:1 injection mixtures of shGene:shLacZ carrying P14 cells. The ratio of cells was analyzed at d8 post transfer into mice and LCMV-Armstrong infection and normalized to the ratio at d0.

cells resulted in a competitive disadvantage relative to control cells. In contrast, *Tgif1* and *Wdhd1* knock-downs both resulted in expansion of $CD8^+$ T cells carrying the shRNA relative to the control cells (Figure 5.3). We noted, however, that not all genes completely reproduced the results from the screen with *Zfp414*-targetting shRNAs having more distribution between offering advantage or disadvantage to activated $CD8^+$ T cells (Figure 5.3B). We also observed a drastically increased effect size when studying each hairpin individually compared to when it was examined as a part of a larger pool, suggesting a potential dilution effect of signal that was occurring in a pooled setting (Figure 5.3, A and B). However, we were able to validate the functional effects of many genes identified in the screen.

Knock-down of TGIF1 leads to a profound increase in the accumulation of effector $CD8^+$ T cells following LCMV Armstrong infection

Based on the validation data, we focused on TGIF1 due to the strong consistency of the effect size observed with multiple shRNAs targeting *TGIF1* and the known interaction of proteins, such as Smad2 (211), c-Jun (215), RXRα (212), and HDAC1 (211, 217, 218). All of those proteins have well appreciated roles in T cell biology, suggesting further that TGIF1 might regulate $CD8^+$ T cell function. We first confirmed that TGIF1 is up-regulated following activation in both transcript (Figure 5.4A) as well as protein abundance (Figure 5.4B). We also validated the effective knock-down efficiency of four separate shRNAs targeting *TGIF1* upon addition of IPTG in activated primary $CD8^+$ T cells using the bone-marrow chimeric system described above (Figure 5.4C).

We tested the ability of these *TGIF1*-targeting shRNAs to affect $CD8^+$ T cell activation and effector differentiation. We mixed congenically marked shTGIF1 and shLacZ-carrying naive $CD8^+$ T cells at a 1:1 ratio and transferred 10,000 cells in total into wild-type mice that were subsequently infected with LCMV Armstrong. We noticed a marked out-competition of

shLacZ-carrying cells by the shTGIF1-carrying $CD8^+$ T cells at d8 p.i. in spleen (Figure 5.4, D and F). This effect was also observed in the blood and was noticeable as early as d5 p.i. (Figure 5.4, E and F). To investigate if the accumulation in the spleen was due to alteration of trafficking, we also analyzed the composition of effector $CD8^+$ T cells in the lung and observed a similar increase as in the periphery (Figure 5.4F, middle). These results suggest a role for TGIF1 in the negative regulation of $CD8^+$ T cells during LCMV Armstrong infection.

TGIF1 may affect $CD8^+$ effector T cell differentiation to MPEC and SLEC subsets through regulation of Blimp-1

Effector $CD8^+$ T cell responses to viral infections are heterogeneous and there are at least two subsets that have been well-characterized in the literature and are known to be transcriptionally regulated (39-42, 45-47). These are $KLRG1^+$ $CD127^-$ short-lived effector cells (SLEC) that expand robustly following activation but rapidly contract following resolution of the infection, and $KLRG1^-$ $CD127^+$ memory progenitor effector cells (MPEC) that expand less initially but are the subset that can persist longer and eventually gives rise to the memory compartment of the $CD8^+$ T cells. TGIF1 knock-down in P14 CD8+ T cells in LCMV Armstrong infected mice decreased the relative abundance of SLECs and increased the MPECs (Figure 5.5A) as compared to control cells, suggesting that this gene may be affecting the transcriptional program orchestrating $CD8^+$ T cell effector differentiation.

To enable deeper analysis of TGIF1 function in $CD8^+$ T cells, we generated a conditional knock-out mouse, in which TGIF1 is deleted selectively in peripheral $CD8^+$ T cells. This was accomplished by crossing the previously described $TGIF1^{flox/flox}$ strain (208) to a mouse strain where Cre recombinase expression is regulated by the *CD8a* E8I-promoter that selectively expressed in peripheral $CD8^+$ T Cells (235). $CD8^+$ T cells in these mice did not exhibit any differences such as alteration of the proportion of cells that express CD44 and CD62L or other surface receptors at the baseline state. Additionally, the TGIF1-KO virus-

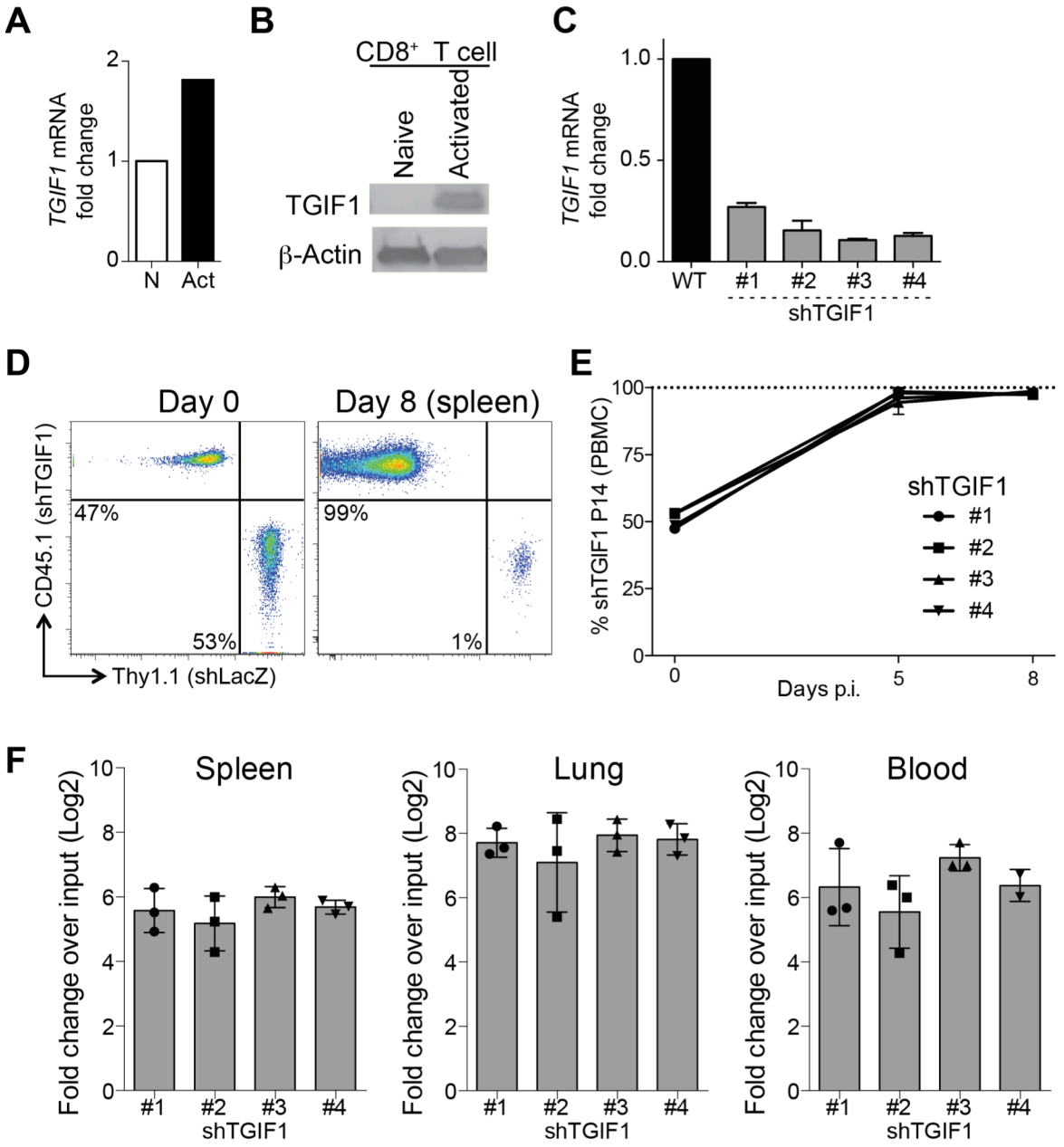

Figure 5.4. TGIF1 is up-regulated following activation and negatively affects CD8$^+$ T cell abundance in viral infection. (**A, B**) TGIF1 mRNA (A) and protein levels (B) are up-regulated in CD8$^+$ T cells following 3 day in vitro stimulation. (**C**) Efficiency of TGIF1 knock-down in day 3 stimulated 1xLacO shRNA-integrated CD8$^+$ T cells incubated with IPTG. (**D**) Representative FACS plot of shTGIF1 and shLacZ P14 cells at the day of injection (do, left) and at d8 following cell transfer in LCMV-Armstrong infected mice. (**E**) Kinetics of the relative expansion of shTGIF1 P14 cells in blood. (**F**) Relative expansion of shTGIF1 P14 CD8$^+$ T cells in the spleen, lung, and blood at d8 p.i.

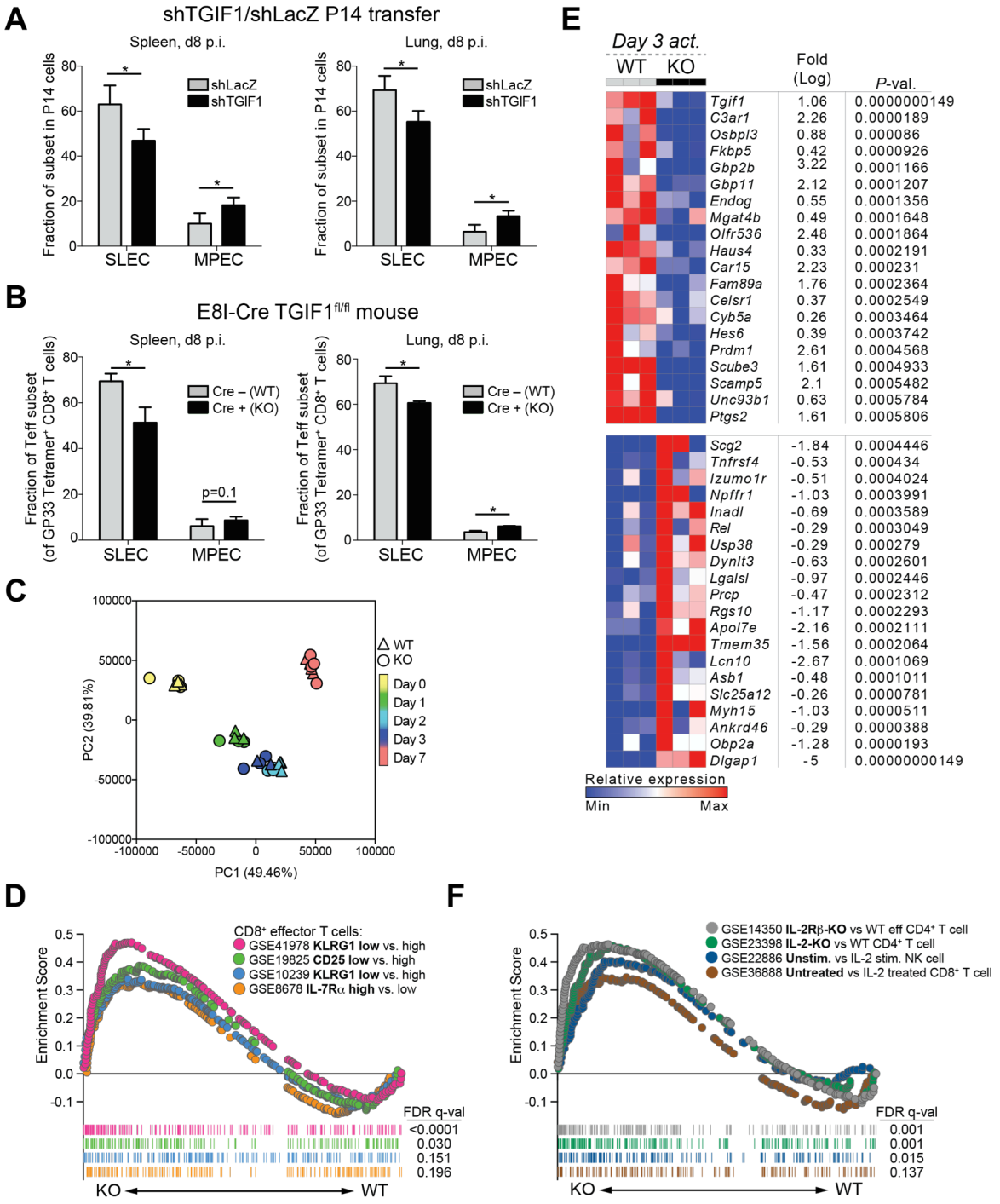

Figure 5.5. TGIF1 influences effector CD8$^+$ T cell differentiation. (**A, B**) Relative abundance of SLEC and MPEC subsets in shTGIF1 or shLacZ P14 cells (A) and E8I-Cre+ (KO) or E8I-Cre– GP_{33} tetramer$^+$ CD8$^+$ T cells (B) at d8 post LCMV-Armstrong infection. (**C**) Principal component analysis using top 5000 most variable genes based on mean absolute deviation (MAD). (**D**) GSEA enrichment plots of d3 *in* vitro activated KO and WT cells using public MPEC/SLEC gene sets. (**E**) Heatmap showing top 20 up- and down-regulated genes in WT vs. KO cells at d3 *in vitro* activated cells. (**F**) GSEA enrichment plots of d3 *in* vitro activated KO and WT cells using public gene sets related to IL-2 treated immune cells.

specific $CD8^+$ T cells did not expand significantly more in response to LCMV-Armstrong infection compared to wild-type cells (data not shown). However, following LCMV-Armstrong infection, we observed an effect similar to TGIF1 knock-down, where the TGIF1-deficient $CD8^+$ T cells produced relatively fewer SLECs and more MPECs in both the spleen and the lung (Figure 5.5B).

To understand the differences between the wild-type and TGIF1-deleted $CD8^+$ T cells, we performed RNA-sequencing in naive $CD8^+$ T cells, *in vitro* activated $CD8^+$ T cells, and virus-specific (GP_{33-41} $tetramer^+$) $CD8^+$ T cells at d8 following LCMV Armstrong infection. The differences between the two genotypes were not robust (Figure 5.5C). However, gene-set enrichment analysis (GSEA) using the ImmuneSigDB collection of immunologic gene expression signature that we generated (described in detail in the Chapter 6) (78, 236) enabled us to confirm genome-wide level of transcriptional differences between the TGIF1-KO and wild-type activated $CD8^+$ T cells. The TGIF1-KO cells were enriched for the MPEC gene signatures previously reported in the literature (Figure 5.5D) (39, 40, 46, 237). When we surveyed the specific differentially expressed genes, we noticed that a key driver of $CD8^+$ T cell SLEC differentiation, Blimp-1, failed to be up-regulated in the absence of TGIF1 (Figure 5.5E).

Blimp-1, encoded by *Prdm1* gene, is one of the key regulators of $CD8^+$ effector T cell differentiation toward MPEC or SLEC subsets. Higher Blimp-1 expression leads to differentiation of $CD8^+$ T cells toward the more terminally differentiated SLECs (20, 73, 74). To examine whether TGIF1 may be directly regulating Blimp-1, we performed TGIF1 chromatin immunoprecipitation (ChIP-Seq). Our preliminary data suggests the presence of TGIF1 binding at the promoter region of the *Prdm1* gene (data not shown), providing a mechanism whereby TGIF1 directly regulates its function. This is consistent with previous reports identifying direct interaction of TGIF1 with the *Prdm1* locus in mixed lineage leukemia (MLL)-rearranged acute myeloid leukemia (AML) (238). In $CD8^+$ T cells Blimp-1 activity is

enhanced by IL-2 signaling to promote SLEC differentiation (20, 73). TGIF1-KO $CD8^+$ T cells enriched for sets of genes up-regulated in IL-2Rβ– or IL-2–knockout cells compared to wild-type cells and sets of genes down-regulated in cells treated with IL-2 compared to untreated cells (Figure 5.5F) (239-242), suggesting that TGIF1 may be more generally interfering with transcriptional programing induced by IL-2 in T cells.

While further experimentation is needed to confirm this finding and understand the mechanism in detail, these data point to at a potential novel role for TGIF1 in influencing $CD8^+$ T cell effector differentiation by affecting IL-2 induced changes and potentially directly regulating the transcription factor Blimp-1.

Discussion

High-throughput perturbation of genes to study the immune system is appealing as it has the potential to systematically survey the function of genes in immune cells at scale. Substantial strides have been made toward that goal in the recent years, yet most studies still predominantly relies on *in vitro* studies of easily-transducable immune cells, such as dendritic cells (183) or by manipulating the cells *ex vivo* to enable viral integration followed by transfer *in vivo* (19, 98, 178, 179, 243). While these approaches have yielded valuable biological insight, they are not able to interrogate the functional of cells that occur early following activation. Furthermore, some approaches rely on the non-physiologically conditions, such as high concentrations of anti-CD3/anti-CD28 antibodies, or cytokines that induced homeostatic proliferation. The approach described herein thus describes a novel platform that enables functional screening of genes *in vivo* using inducible shRNA in unperturbed cells of the hematopoietic system.

The inducible nature of our screening platform additionally offers the opportunity to investigate the roles of specific genes at different stages of the immune response. In the context of acute viral infections, for example, genes can be transiently suppressed during the activation with short exposure to IPTG, but then restored in their expression capacity at the contraction or memory stages. Similarly, shRNA expression may be selectively induced by IPTG at the memory stage to interrogate the genes important in the maintenance of memory T cells. Analogous approach can be used to identify novel genes driving T cell exhaustion in the context of persistent infection or in a tumor model. With the recent therapeutic successes of approaches to revive T cell exhaustion with checkpoint blockade, our approach may accelerate the discovery of novel targets that can be pursued in clinical development.

Using the *in vivo* RNAi screen, we identified TGIF1 as a potential novel negative regulator of $CD8^+$ T cell activation. This transcriptional regulator has been previously recognized to cause holoprocencephaly type 4 in humans when mutated (244),

However its up-regulation in $CD8^+$ T cells following activation and its reported binding partners and suggest a potentially interesting additional role for this gene in regulating $CD8^+$ T cells. We observed that TGIF1 knockdown results in enhanced proliferation of P14 $CD8^+$ T cells in LCMV Armstrong infected mice. Additionally, the effector differentiation of these cells is altered with increased skewing toward the MPEC subset and away from the SLEC. Similar differentiation effects were observed in mice with selective deletion of TGIF1 in peripheral $CD8^+$ T cells. Gene expression analysis of wild-type and TGIF1-deleted activated $CD8^+$ T cells also revealed that these cells are enriched for genes up-regulated in MPEC and genes that are down-regulated in conditions of increased IL-2 signaling. While further detailed examination of this is needed to confirm the mechanism of regulation, our preliminary data suggest that TGIF1 may be binding to the promoter region of Blimp-1 gene and directly inducing or enhancing its expression. It is encouraging that another study examining MLL-ALL observed a similar TGIF1 peak in the Blimp-1 promoter region using ChIP-Seq and overexpression of TGIF1 in these cells lead to the increase of Blimp-1 expression (238). It is not know at this point whether TGIF1 may be binding to this region directly by itself or whether it interacts with it through cofactors. Because Blimp-1 has a role in other immune cells beyond $CD8^+$ T cells, these findings could have implications for the regulation of $CD4^+$ T cell subset differentiation such as Tfh (245) and Th17 cells (246) as well as B cells (247, 248). Additional research on this transcriptional regulators can thus have a larger impact on the immunology field beyond uncovering its role in the biology of $CD8^+$ T cells.

Our data highlights a much more profound role in $CD8^+$ T cells when TGIF1 is knocked-down than when it is knocked-out. Several potential reasons could be responsible for this inconsistency. First, some of the effect size may be masked when TGIF1 is lacking throughout the life of a mature $CD8^+$ T cell and potentially changing cells and perhaps leading them to employ genetic compensatory mechanisms occasionally described in the knockout animals (249). For example, TGIF2, which has been shown in the past to work

similarly to TGIF1 and arose from gene duplication, may have compensatory function (218, 250, 251). Additionally, the partial loss of the protein abundance in an inducibly immediately prior to activation of cells may reveal a phenotype otherwise masked by other mechanisms. Lastly, the nature of the experiment involving P14 competition in LCMV infection could be amplifying the effect size because it is measuring the relative abundance of shTGIF1-integrated P14 $CD8^+$ T cells compared to the control P14 T cells. As such, effects observed in the cells with the gene knockdown may be masked in the knockout setting. Such effects may therefore be better examined in models with a known role for TGF-β such as chronic LCMV Clone 13 infection or using tumor models (133, 252).

In this study, we leveraged the biology of RNA interference to perturb the expression of genes through degradation of their mRNA. Due to the pooled nature of the screen that relies on the tracking the shRNA representation after multiple rounds of division from day 0 to day 8, only a permanent integration in the genome is a viable option, and thus transient transfection would not enable such recovery of data. Additional approaches, however, have been recently described to alter the genome of the cells, including the CRISPR/Cas9 system, TALENs, and Zinc finger nucleases (ZFNs) (253). Our platform can be readily combined with CRISPR/Cas9 technologies to allow rapid, inducible generation of knock-out or knock-in unmanipulated immune cells of any type. While this study identified a novel regulator of T cell biology, it thus simultaneously opens doors for a systematic interrogation of genes in immune cells and identification of important, potentially drugable targets on immune cells to impact human health.

Materials and Methods

Mice

P14 TCR Tg mice were used as described (110). Wildtype C57BL/6J, Ly5.1 (CD45.1), and Thy1.1 mice were purchased from Jackson Laboratories (Bar Harbor, ME). TGIF1$^{fl/fl}$ have been generated previously (208) and were bred to the E8I-Cre mice (254). All mice were used according to the Harvard Medical School Standing Committee on Animals and National Institutes of Animal Healthcare Guidelines. Animal protocols were approved by the Harvard Medical School Standing Committee on Animals.

Generation of bone marrow chimeras

Bone marrow was isolated and red blood cells lysed using ACK lysis buffer (Gibco). LSK cells were enriched using anti-CD117 microbeads (Miltenyi Biotech) and then sorted using a BD FACSAria cytometer. Sorted cells were plated overnight in StemSpan SFEM medium (StemCell Technologies) with 100 μg/ml recombinant murine stem cell factor (SCF), thrombopoietin (TPO), IL-7, and Flt3-ligand (PeproTech). Cells were then spin-infected with lentiviral supernatants at 650 x g for 90 min at 37°C on non-treated plates that had been coated overnight at 4°C with 100 mg/ml RetroNectin (Takara Bio). Fresh medium was added after 1 h and cells were rested overnight. Viral stocks were titrated to ensure the majority of cells were infected with a single virus. The cells were then washed in PBS (Gibco) and 50,000 cells were injected intravenously in recipient mice that had been irradiated with 2 doses of 600 cGy, 3 hours apart.

Pooled library generation

The shRNA libraries in 1xLaO vector were constructed in a pooled fashion from existing TRC005 vectors containing desired shRNA. Pooled TRC005 plasmids were digested

with EcoRI and AgeI, cut shRNA oligonucleotides isolated, and ligated into an EcoRI/AgeI-cut 1xLacO vector.

Lentivirus production

293T cells were seeded in DMEM with 10% FBS. The following day, the cells were transfected with shRNA construct pLKO.1-TRC005 or 1xLacO (available from Sigma-Aldrich under the name "MISSION 1X LacO Inducible") with the GFP reporter, and with the packaging plasmids Pax2 (gag, pol) and VSV-G using TransIT-LT1 (Mirus Bio) or ExGen 500 (Thermo Scientific Fermentas). Viral supernatants were collected 48h – 72h later.

T cell transfers and infections

$CD8^+$ T cells from bone marrow chimeric animals were isolated using the $CD8^+$ T cell isolation kit II for magnetic separation (Miltenyi Biotech) and then GFP^+ $CD8^+$ congenic cells were sorted using BD FACSAria cytometer. P14 $CD8^+$ T cells (10^4 – 10^6 cells/animal) were injected in recipient mice i.v., which were subsequently infected *intraperitonealy* with 2 x 10^5 p.f.u. LCMV Armstrong. For influenza infections, the mice were anesthetized with 2.5% Avertin and infected with 0.5 LD_{50} H1N1 *Influenza* virus (PR8), engineered to express GP_{33-41} peptide of LCMV (PR8-GP33)(93), intranasally. Both viruses were a generous gift of Dr. E. John Wherry (University of Pennsylvania School of Medicine, Philadelphia, PA).

Flow cytometry and cell sorting

Single cell suspensions from spleen or bone marrow were stained with combinations of anti-CD8α (53-6.7), anti-CD44 (IM7), anti-CD62L (MEL-14), anti-CD45.1 (A20), anti-CD45.2 (104), anti-Thy1.1 (OX-7), anti-Thy1.2 (30-H12), anti-CD25 (PC61), anti-CD127 (A7R34) (all from BioLegend), and anti-KLRG1 (2F1) (from Abcam).

LSK cells were sorted from CD117-enriched bone marrow cells stained with CD117

(ACK2), Sca-1 (D7) and a lineage antibody cocktail which included biotin-labeled anti-CD5 (53-7.3), anti-Gr1 (RB6-8C5), anti-B220 (RA3-6B2), anti-CD3e (145-2C11), anti-CD11b (M1/70), anti-Ter-119 (Ter-119) and detected with fluorophore conjugated Streptavidin (all from BioLegend). Data were acquired using LSR II or Accuri C6 (BD Biosciences) cytometers and analyzed with FlowJo software (v9.7.2, TreeStar).

shRNA construct generation

For validation and TGIF1 analysis, single gene-targeting constructs were generated in the 1xLacO vector. Target sequences of the shRNA used in validation experiments are:

shTGIF1 - #1 (CGAATGTTTCTTGGTAGTTTC),

shTGIF1 - #2 (GATGGCAAGAGATGCATTATT),

shTGIF1 - #3 (AGTACAGATGTACCGCAAATA),

shTGIF1 - #4 (ATTTCAGAAGCTAGCTCTATT),

shTGIF1 - #5 (TAGTGGATGTTGCACTCAAAC),

shWDHD1 - #1 (TGATTATGAGGAGAGCATTAA),

shWDHD1 - #2 (TCCTTCGACTGTTCACTATTG),

shWDHD1 - #3 (GTCTCCCTGTGGGCAGTATTT),

shWDHD1 - #4 (AGAGCAGCAGGAACTCTTAAT),

shZFP414 - #1 (GTTCGTGATCTAGCACAGCAT),

shZFP414 - #2 (CTTCAAGCATCTGCATGTTTG),

shZFP414 - #3 (CTACTTCAAGTGTGAGAATTG),

shIRF1 (GGCTAGAGATGCAGATTAATT),

shIRF7 (CTTCGACTTCAGCACTTTCTT),

shLacZ (CCGTCATAGCGATAACGAGTT),

shRFP (GCTTCAAGTGGGAGCGCGTGA).

Complete list of shRNAs and target sequences is listed in Supplemental Table 5.4.

Cloning methods can be found at http://www.broadinstitute.org/rnai/public/. Briefly, complementary oligos (IDT) were annealed and ligated into AgeI and EcoRI digested 1xLacO-GFP vector. All ligated constructs were sequence verified for the presence of the correct shRNA by Sanger sequencing.

Immunoblotting

$CD8^+$ T cells stimulated *in vitro* with 4 μg/ml of plate-bound anti-CD3 and anti-CD28 with 100 U/ml of rhIL-2 (Peprotech) were lysed in modified Laemmli buffer (60 mM Tris-HCl [pH 7.2], 10% glycerol, and 2% SDS) containing 1 U/ml of DNase (Benzonase nuclease; Novagen) and Complete protease inhibitor cocktail (Roche) for 30 min at 4°C. Protein concentration in the lysates was estimated with a bicinchoninic acid assay (Thermo Scientific), and 75 μg of each lysate was subjected to SDS-polyacrylamide gel electrophoresis and western blotting on nitrocellulose membranes as previously described (189). The primary antibodies used were anti-β-actin (1:5000 dilution, Abcam, clone ab8227) and anti-TGIF1 (1:500 dilution, Santa Cruz Biotechnology, clone H-172).

Gene expression analysis

Wild-type or congenically marked $CD8^+$ T cells were sorted from chimeric mice or following *in vitro* stimulation with 4 μg/ml plate-bound anti-CD3 (clone 145-2C11) and anti-CD28 (clone 37.51) (BioXcell) and 100 U/ml recombinant human IL-2 (Peprotech). Cells were lysed 12 h – 5 days following activation and RNA was extracted using RNeasy Plus Mini Kit (Qiagen). cDNA was generated using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and analyzed on an ABI 7500 Fast real-time quantitative PCR instrument using gene Taqman probesets from Applied Biosystems.

For RNA sequencing, wild-type and E8I-Cre TGIF1$^{fl/fl}$ CD8$^+$ T cells were collected at day 0 (MACS-enriched for naive CD8$^+$ T cells), or collected at days 1, 2, or 3 following *in vitro* stimulation with 4 μg/ml of plate-bound anti-CD3 and anti-CD28, or were sorted on splenic GP_{33-41}-tetramer$^+$ CD44$^+$ CD8$^+$ T cells at d8 following LCMV-Armstrong infection. The cells were lysed in RLT Plus lysis buffer (Qiagen) supplemented with 1% β-mercaptoethanol (Sigma). RNA was extracted according to manufacturer's instructions (Qiagen RNeasy Plus Micro kit). Next generation sequencing libraries were generated according to manufacturer's instructions using the Next Ultra RNA kit for Illumina (New England Biolabs). Libraries were sequenced on Illumina HiSeq 2500 by paired-end 30bp sequencing, to an average depth of 17 million reads per sample.

Reads were aligned to the mouse genome (Mm10) transcriptome assembly using TopHat2 (255), and aligned bam files were aggregated and reduced to feature counts using Rsubread (256). Counts were quantile normalized and genes were filtered using a cutoff of >10 counts in >2 samples. We then used limma voom to calculate differential gene expression (257). Genes were ranked based on log-transformed p-values with a sign correction based on the log fold-change, and these ranked lists were used to perform preranked gene set enrichment analysis (78).

Chromatin immunoprecipitation (ChIP) sequencing

CD8$^+$ T cells were activated *in* vitro with plate-bound anti-CD3 and anti-CD28 in the presence of 100 U/ml recombinant human IL-2 (Peprotech). Three days following activation, 60 × 10^6 cells were fixed with 1% formaldehyde for 10 min at 37°C, washed twice in ice-cold PBS, pelleted, and 'flash frozen' and stored at −80°C for later processing as described previously (174). Specifically, the fixed cells were resuspended in 1440 ml of lysis buffer (0.5% SDS, 50 mM Tris, pH 8, 10 mM EDTA, 1× Complete protease inhibitor (Roche)).

Chromatin was sheared using the E210 Ultrasonicator (Covaris) and microTUBEs (Covaris). Each microTUBE (120 ml of lysate) was sonicated with six treatments of 60 seconds each with the following settings: intensity, 5; 'duty cycle', 10%; 200 cycles per burst. This sonication resulted in most DNA fragments being of 200–250 base pairs in length. The sonicated lysates were collected and centrifuged for 10 min at 4°C. Supernatants were collected and four parts of dilution buffer (1.25% Triton X-100, 12.5 mM Tris, pH 8, 187.5 mM NaCl, 1×Complete protease inhibitor) were added. Dynabeads Protein G (Life Technologies) pre-bound to 1, 5, and 10 μg of anti-TGIF1 antibody (Santa Cruz Biotechnology, H-172) were used for overnight immunoprecipitation. Precipitated immunocomplexes were then washed, for 5 min, as follows: first with low-salt buffer (0.1% SDS, 1% Triton X-100, 20 mM Tris, pH 8, 2 mM EDTA, 150 mM NaCl and 1× Complete protease inhibitor), then with high-salt buffer (0.1% SDS, 1% Triton X-100, 20 mM Tris, pH 8, 2 mM EDTA, 500 mM NaCl and 1× Complete protease inhibitor), twice with LiCl buffer (0.7% sodium deoxycholate, 1% NP-40, 20 mM Tris, pH 8, 1 mM EDTA, 500 mM LiCl and 1× Complete protease inhibitor) and lastly with Tris-EDTA buffer (with protease inhibitor). The ChIP immunocomplexes were eluted from the beads twice in elution buffer (1% SDS and 0.1 M $NaHCO_3$) with continuous agitation for 30 min. The eluates were pooled and then were incubated overnight at 65°C for reversal of the formaldehyde crosslinks. The eluates were then treated for 2 h at 37°C with 200 mg RNase A (Qiagen) and 40 mg proteinase K (Life Technologies). ChIP DNA fragments were purified with a MinElute Reaction Cleanup kit (Qiagen).

The DNA fragments were then prepared for high-throughput Illumina sequencing with a NEBNext ChIP-Seq Library Master Mix Set for Illumina kit (New England BioLabs) and NEBNext Multiplex Oligos for Illumina (Index Primers 1-12) kit (New England BioLabs) according to a modified manufacturer's protocol. For the ChIP or input samples, 10–50 ng of DNA was used for the preparation of the sequencing library. The ends of DNA fragments

were repaired, given dA tails, and ligated to Illumina adaptors according to the kit instructions. DNA fragments of 150–600 base pair length were selected with Pippin Prep 2% Agarose Gel Cassettes and the Pippin Prep DNA Size Selection System (Sage Science). DNA was purified with a QIAquick PCR Purification Kit (Qiagen). The DNA was then amplified by 12 cycles of PCR with NEBNext Multiplex Oligos. Amplified DNA was purified with DNA to Agencourt AMPure XP beads (Beckman Coulter) a 1:1 ratio. The 'multiplexed' DNA libraries were sequenced on the Illumina HiSeq2000. Sequencing reads were aligned to the mouse NCBI37/mm9 reference genome (National Center for Biotechnology Information) with Bowtie software for the alignment of short DNA sequences (258), and sorted with the SAMTools format for storing large nucleotide sequence alignments and utilities for manipulating alignments (Sequence Alignment/Map) (259). Identical duplicated reads were removed so the analysis examined only unique reads) using Picard utilities for the manipulation of SAM files.

SPP and the MACS2 (v2) method (260, 261) were both used for peak finding. Regions from both peak identifiers were combined with mergeBed of the BEDTools utilities for comparing genomic features (262). Quality rating for the merged regions was assigned on the basis of the maximum significance from either source. We used the library generated with 5 µg of anti-TGIF1 antibody, as it was the only one that revealed peaks over the input.

Acknowledgments

I would like to acknowledge Glenn Cowley and David Root (Broad Institute of MIT and Harvard) for performing the cloning of shRNA libraries in 1xLacO vector and for their assistance with shRNA library sequencing and analysis for the functional *in vivo* screening. RNA-sequencing was performed and analyzed with the assistance of Kathleen Yates and Kevin Bi (Haining Lab, Dana-Farber Cancer Institute), TGIF1 chromatin immunoprecipitation sequencing (ChIP-Seq) was performed with Michael DiIorio and Hsiao-Wei Tsao (Haining Lab, Dana-Farber Cancer Institute). Anne van der Leun (Haining Lab, Dana-Farber Cancer Institute) assisted with cloning and validating various shRNA vectors. Madeleine Lemieux (Bioinfo) assisted with ChIP-Seq analysis.

Chapter 6. Compendium of immune signatures identifies conserved and species-specific biology in response to inflammation

Parts of this chapter have previously been published in:

Jernej Godec, Yan Tan, Arthur Liberzon, Pablo Tamayo, Atul J. Butte, Jill P. Mesirov, W. Nicholas Haining. A novel compendium of immune signatures identifies both conserved and species-specific biology in the mouse and human response to inflammation. *Immunity*. 2016 Jan 19;44(1):194-206. (PMID: 26795250)

Introduction

Experiments in both human cells and mouse models have been used to discover many of the mechanisms by which the immune system functions. Identifying aspects of immunobiology that are evolutionarily conserved between humans and mouse models is useful because it can reveal mechanisms of fundamental importance to both species. Moreover, it can provide reassurance that information gleaned from mouse models will be applicable to the human condition. This is crucial, as much of immunobiology cannot be examined physiologically in humans due to inaccessibility of certain tissues or cell types or the difficulty in recapitulating complex biological milieu *in vitro*. However, considerable controversy exists as to the degree to which mouse models can recapitulate events occurring in immunologic disease states in humans (263-268). These concerns have extended to the analysis of genome-wide analysis of mRNA levels where analyses of the same datasets from mouse and human sepsis reached opposite conclusions regarding the degree of cross-species similarity (79, 80). Contradictory findings have also been reported in the comparison of gene expression across a range of human and mouse tissues (269, 270).

One of the challenges in identifying similarities between gene expression datasets is that major changes in the cell state can be associated with relatively small alterations in the expression level of a relatively large numbers of genes. Analysis of co-regulated changes in *sets* of functionally related genes, rather than *individual* genes, has therefore become an important strategy to identify subtle, but biologically meaningful, differences in gene expression (77, 78, 271). This is a particularly useful approach when analyzing samples in which experimental variability (such as those collected from heterogeneous human subjects) or evolutionary divergence (such as comparisons between species) add experimental "noise" to gene expression profiles. Several approaches for testing for the enrichment of gene sets have been developed, including Gene Set Enrichment analysis (GSEA) (78). GSEA has been made more powerful by the availability of curated collections of gene expression signatures extracted from a variety of sources including published experimental datasets. The largest of these collections, the Molecular Signatures Database (MSigDB), contains more than 8,000 signatures (272). However, only a small fraction of these gene sets pertain to immune processes and cell types.

We now report the creation of ImmuneSigDB, a compendium of ~5,000 well-annotated signatures generated by analysis of 389 published studies of cell states and perturbations in the mouse and human immune systems. Using this collection of signatures, we demonstrated that signatures of cell differentiation in lymphoid cells and endotoxin stimulation in myeloid cells are highly conserved between humans and mouse models. Moreover, analysis of the transcriptional response to sepsis in human samples and mouse models showed that there was highly significant conservation of gene expression between the species when measured at the gene set level. However, in addition to the conserved transcriptional programs, we also identify species-specific differences in the biological processes associated with sepsis. These findings help interpret contradictory observations regarding the extent of evolutionary conservation in the transcriptional response to sepsis.

ImmuneSigDB will enable the detailed analysis of cross-species gene expression that is critical to establishing which biological processes are conserved and which are not, thus allowing mouse models to better inform our understanding of human disease.

Results

Generating a compendium of gene signatures curated from immune expression profiles.

We generated a comprehensive compendium of gene sets pertaining to immune biology. The term “gene-set” in this study refers to groups of genes identified by selecting either up- or down-regulated genes in comparisons of gene expression profiles of interest. We identified and uniformly analyzed 389 published studies in the immunology literature that included genome-wide expression profiling data (outlined in Figure 6A). We selected studies to analyze based on immunological key words in the title or abstract followed by additional manual review. We prioritized studies published in immunology journals of broad interest (Supplemental Table 6.1). We identified the corresponding publicly available datasets in the NCBI Gene Expression Omnibus (GEO) and, for uniformity, focused on studies performed on Affymetrix platforms (Supplemental Table 6.2) that included three or more biological replicates. Each study was reviewed to identify and annotate the biology represented and to define meaningful pairwise comparisons that would create biologically useful gene sets. For example, an individual study may include a single comparison, such as stimulated versus unstimulated cells, or can have multiple comparisons, as is the case where several cell types were subjected to different culture conditions and analyzed at several time points. In such cases, only meaningful pairwise comparisons, rather than all possible comparisons, were made (Figure 6.2).

The raw expression data obtained from each GEO study was pre-processed uniformly (see Methods). We identified and extracted differentially expressed genes (see Methods and Figure 6.1A), which comprised the gene sets for the ImmuneSigDB collection. These sets represent genes coordinately up- or down-regulated in many major immune cell types (Figure 6.1B) either in their baseline state or following a range of genetic or chemical perturbations. ImmuneSigDB included data from healthy human subjects, patients with

immune or non-immune diseases, and mouse models. Mapping orthologous genes to a common identifier allowed us to include both human (n=135) and mouse (n=254) studies (Figure 6.1C). From these studies, we identified 2,436 meaningful comparisons and extracted 4,872 gene sets of up- or down-regulated genes comprising the ImmuneSigDB (see Methods). The number of gene sets identified per published study ranged from one comparison, (e.g., representing an activated vs. unperturbed state or knock-out vs. wild-type cell) to over 50 (e.g., often representing several cell types cultured in different conditions for varying amounts of time) (Figure 6.1D). Particular biological conditions over-represented in the literature, such as those related to T cell biology, are correspondingly over-represented in ImmuneSigDB, with slightly fewer gene sets from myeloid cells and B cells (Figure 6.1B). ImmuneSigDB is publicly available at www.msigdb.org.

ImmuneSigDB expands the biological coverage of the MSigDB

We compared the gene sets generated from immune cells (ImmuneSigDB) with those in gene sets in the MSigDB collection. MSigDB is a curated collection of gene sets generated from published gene expression studies that are generally not from the immunology literature (272). We measured overlap in constituent genes between each gene set in the ImmuneSigDB and all the other MSigDB collections and found that only a small minority of gene sets significantly overlapped (Figure 6.1, E and F), suggesting that ImmuneSigDB added a large amount of new transcriptional information. A small subset of gene sets in ImmuneSigDB and MSigDB were highly similar (0.64% of gene sets with $P<10^{-8}$) and these could be clustered into three groups related to proliferation, inflammation or Type 1 interferon

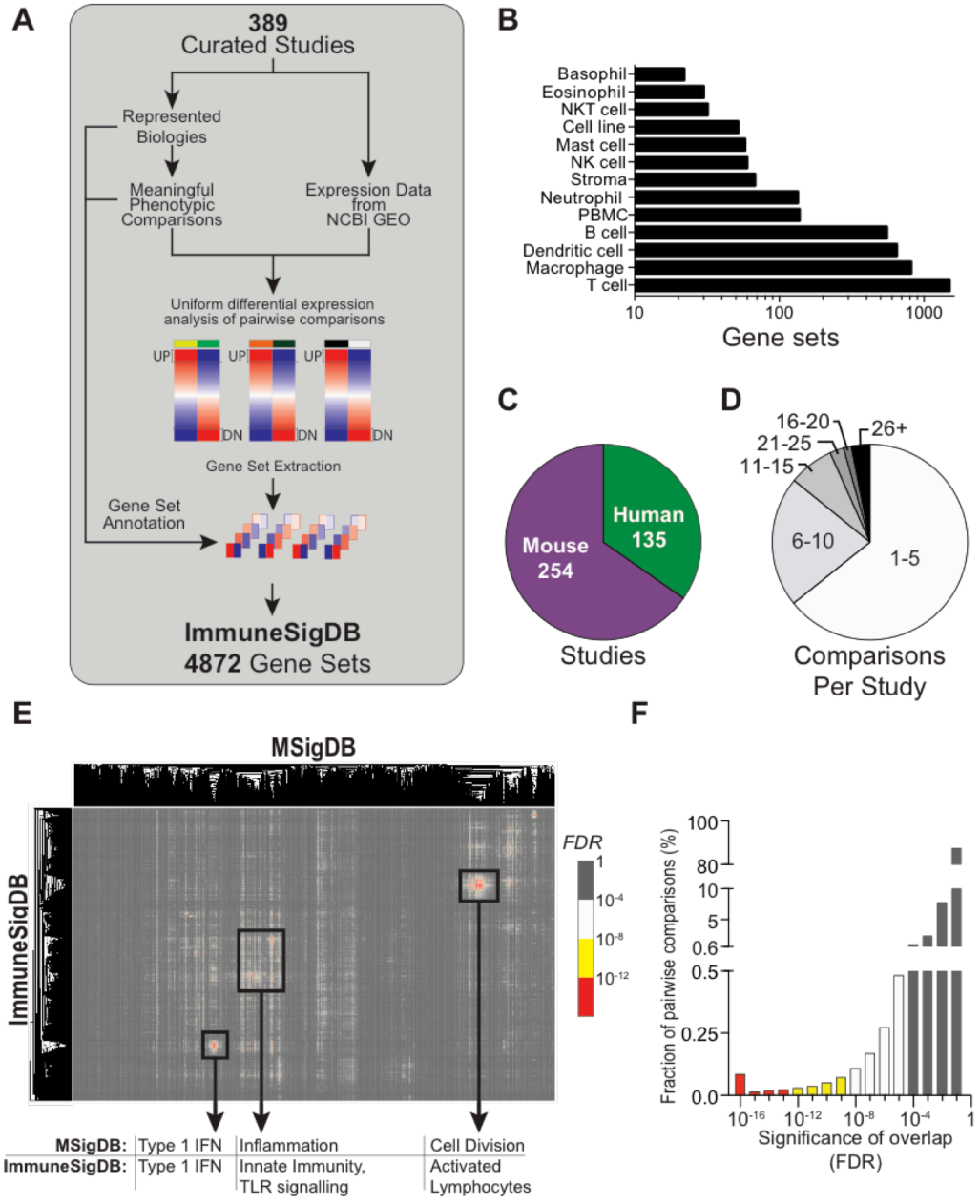

Figure 6.1. ImmuneSigDB collection is derived from re-analysis of published data. (A) A schematic of the ImmuneSigDB pipeline. (B) Number of gene sets corresponding to major immune lineages or cell lines and (C) species of origin contained in ImmuneSigDB. (D) Number of pairwise comparisons made per each study used in ImmuneSigDB. (E) Overlap in gene set membership in ImmuneSigDB with MSigDB gene sets. Heatmap indicates False Discovery Rate (FDR) values of each pairwise comparison between gene sets. Highlighted are representative biological processes in each of the significantly overlapping clusters of gene sets. (F) Distribution of the FDR ranges of significance across all pair-wise comparisons of gene set membership.

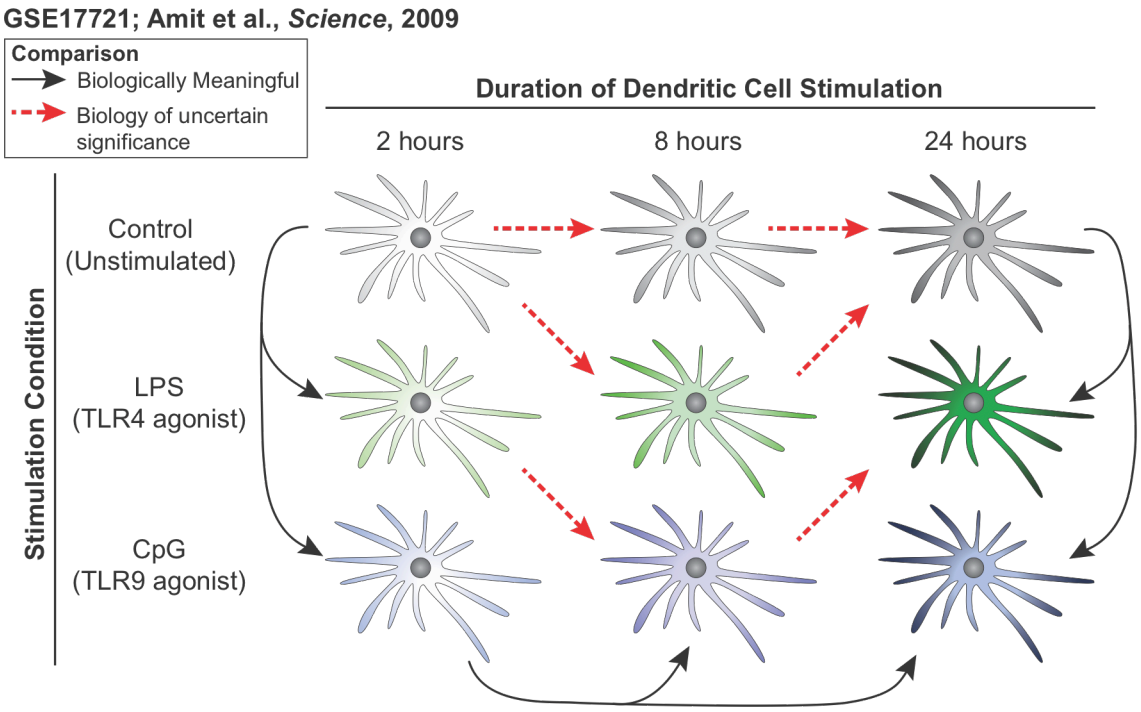

Figure 6.2. Selection of biologically meaningful comparisons. Schematic of an example of comparisons that are biologically meaningful (black arrows) or difficult to interpret (red arrows) in one of the studies included in the ImmuneSigDB (GSE17721).

response (Figure 6.1E). This suggested that with the exception of these core biological processes, gene sets derived from immune cell expression profiles contain genes distinct from non-immune-related gene expression profiles that previously predominated the MSigDB.

We performed an analogous analysis of pairwise overlaps in gene membership between gene sets *within* ImmuneSigDB. While most were unique, we found a larger number of gene sets with significant overlap (1.46% with $P<10^{-8}$) within ImmuneSigDB than between ImmuneSigDB and MSigDB (Figure 6.3). These gene sets largely related to lineage-specific signatures shared between datasets generated from similar types of cells. Therefore, ImmuneSigDB has minimal overlap with MSigDB and provides new gene sets describing immune biology.

ImmuneSigDB provides a complementary resource to existing immune module collections.

Several groups have previously created collections of gene modules in the immune system. In studies by Chaussabel et al. (273) and Li et al. (274), existing studies of gene expression profiles in human peripheral blood mononuclear cell (PBMC) or whole blood were analyzed to identify modules of co-regulated genes to aid in the analysis of gene expression profiles from immune cells. Several features distinguish ImmuneSigDB from either of these collections (summarized in Supplemental Table 6.3). First, ImmuneSigDB was generated by direct comparison of the genes that were up- or down-regulated in two known sample classes from each study. This allowed the published study to serve as a source of comprehensive annotation of each gene set, in contrast to either of the module collections that were generated by analysis of aggregated pools of samples, limiting the direct experimental annotation of each module. Second, ImmuneSigDB was considerably larger than either module collection (Supplemental Table 6.3). Third, ImmuneSigDB included data

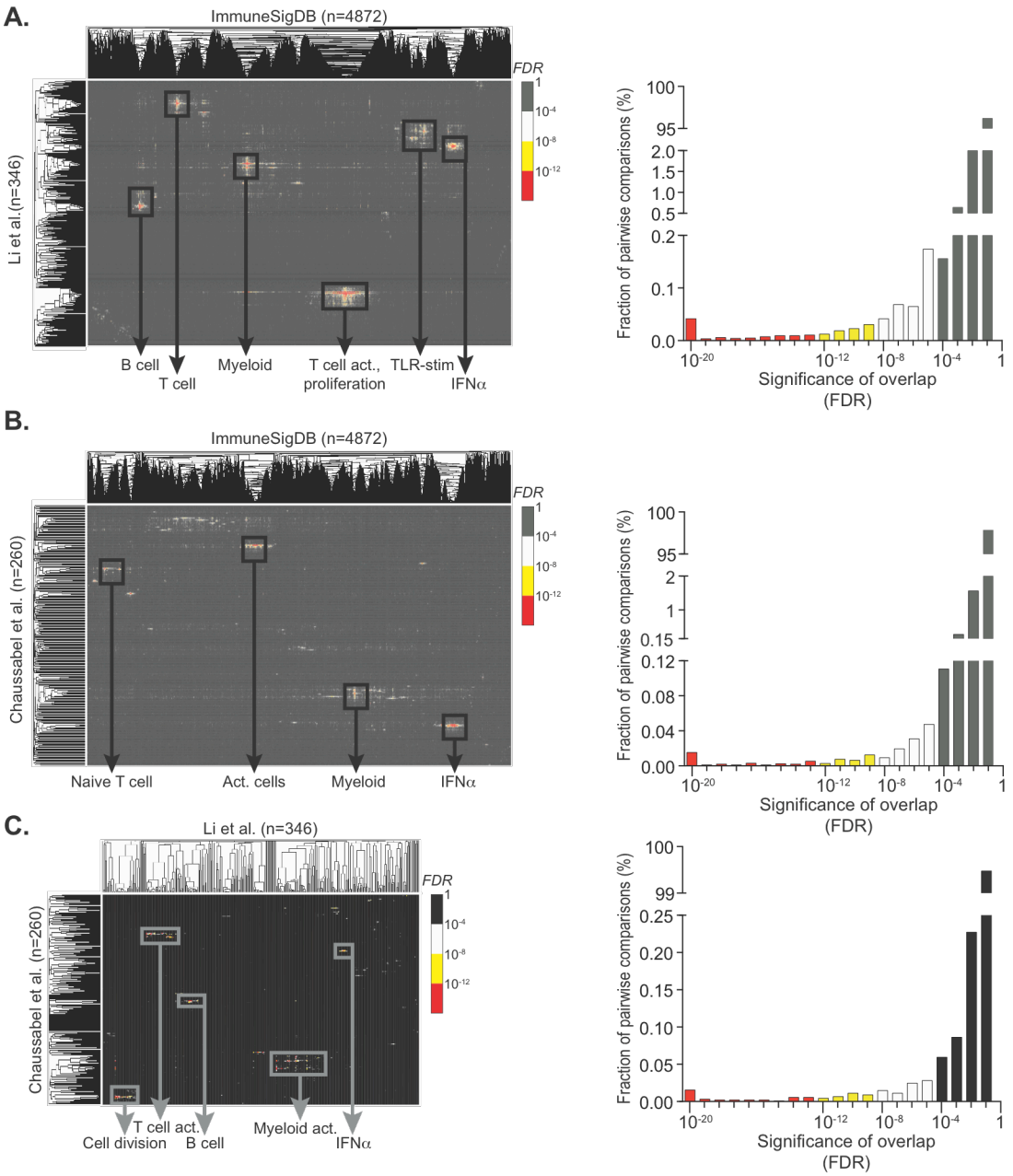

Figure 6.3. ImmuneSigDB provides unique gene sets not represented in previous immune signature collections. (A) Overlap in gene set membership between ImmuneSigDB and gene signature collection developed by Li et al. Heatmap indicates significance of overlap indicated by FDR values (right). Highlighted are representative biological processes or cell lineages in each of the significantly overlapping clusters of gene sets/modules. Summary of FDR ranges of all pairwise overlaps are shown to the right. (B and C) Analysis as in (A) showing overlapping gene memberships between immune gene modules defined by Chaussabel et al. and ImmuneSigDB (B) and between the Li et al. and Chaussabel et al. collections (C).

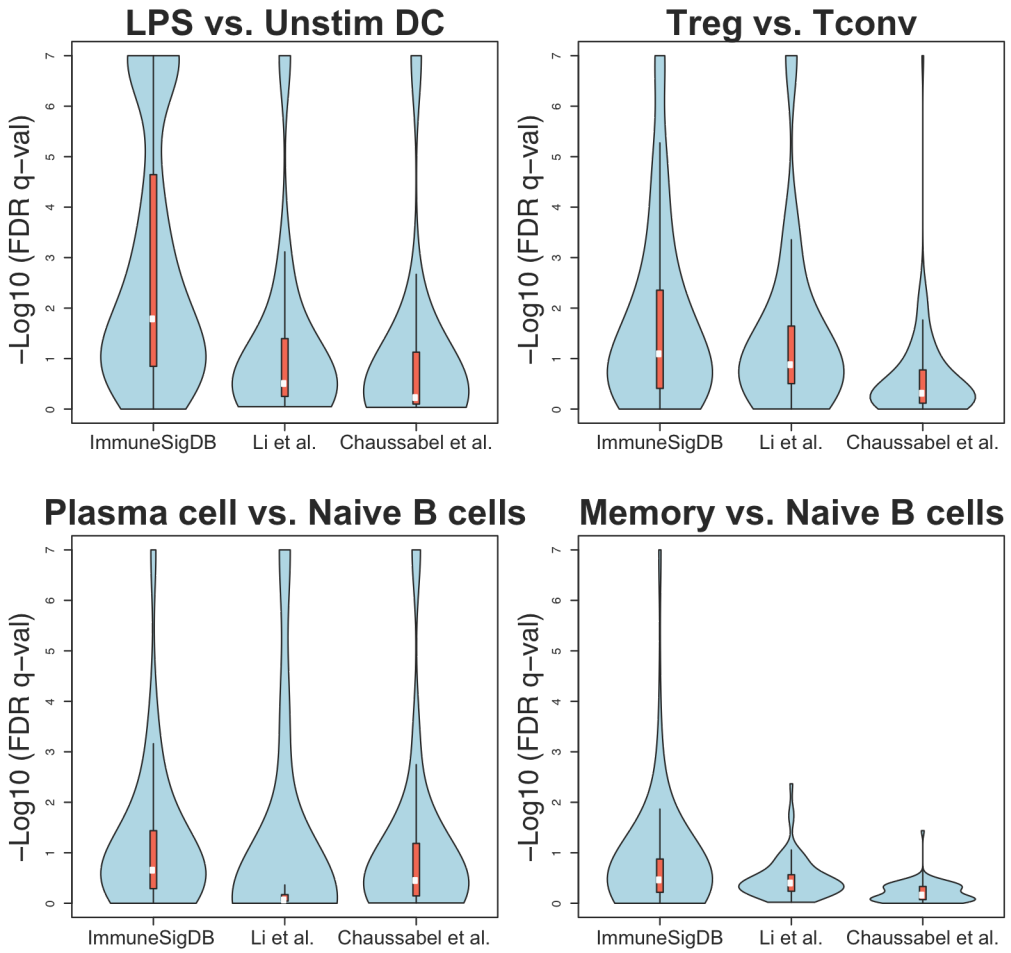

Figure 6.4. GSEA using ImmuneSigDB or module collections. Significance of enrichment of gene sets contained in ImmuneSigDB or modules in collections by Li et al or Chaussabel et al. Each of the three collections was used to analyze four publically available datasets of gene expression datasets from: LPS vs. unstimulated DC (top left); Tregs vs conventional $CD4^+$ T cells (top right); plasma cells vs. naïve B cells (bottom left); and memory B cells vs. naïve B cells (bottom right).

from both mouse models and humans, and from 13 cell or tissue types, rather than solely from human PBMC and whole blood profiles.

To compare directly the gene-sets in ImmuneSigDB with the module collections of Chaussabel and Li, we measured overlap in constituent genes between each gene set in the ImmuneSigDB and all modules in either the Chaussabel or Li collections (Figure 6.3, B and C). We found that only a small fraction of ImmuneSigDB gene sets significantly overlapped with either collection (0.06% and 0.18% with FDR of $<10^{-8}$ for Chaussabel and Li, respectively), suggesting that the gene-sets within ImmuneSigDB and the modules in the Chaussabel and Li collections were largely distinct. The small number of significantly overlapping gene-sets and/or modules contained genes predominantly related to immune cell lineages (e.g., T cell or myeloid) or to the response to interferon-α (IFN-α) or Toll-like receptor (TLR) ligands. Interestingly, the overlap between modules contained in the Chaussabel and Li collections was similarly limited (Figure 6.3C), suggesting analysis of immune expression profiles using each of the three collections could provide complementary information.

Finally, we performed GSEA using four published datasets in human immune cells (LPS stimulated DC, Tregs, plasma B cells, and memory B cells) to compare the results using ImmuneSigDB with the module collections by Chaussabel and Li (Figure 6.4). A larger number of ImmuneSigDB gene sets were significantly enriched (even after correcting for multiple hypothesis testing) in each of the four data sets than with either the Chaussabel or Li collections. Moreover, inspection of the top 20 most enriched gene sets from ImmuneSigDB and modules from the Chaussabel or Li collections illustrates the extensive biological annotations (including links to the original studies) available for each ImmuneSigDB gene set (Supplemental Tables 6.4 – 6.7). Thus analysis with ImmuneSigDB provides a resource for the analysis of gene expression data in the immune system that is complementary to existing collections.

Enrichment of ImmuneSigDB gene sets recapitulates known lineage-specific differences in mouse and human hematopoietic cell lineages.

We next tested whether enrichment analysis of gene expression using ImmuneSigDB could recapitulate known differences in lineage-specific gene expression within the immune system. We analyzed a large, publicly available dataset of gene expression profiles from the Immunological Genome Project (ImmGen) consisting of immune cell types and cell states in mice (275) using a single sample version of GSEA (ssGSEA) (276). In this approach, gene sets are tested for enrichment in the list of genes in a single sample ranked by absolute expression rather than by comparison with another sample. The resulting ssGSEA scores provide an estimate of the degree of enrichment of each ImmuneSigDB gene set in each individual sample in the dataset. In this way we generate a dataset containing as rows the profiles of enrichment of each ImmuneSigDB gene set and as columns the samples. Unsupervised hierarchical clustering of samples from four distinct immune cell types – dendritic cells, B cells, ab T cells, and stem cells – in the space of gene set enrichment scores revealed near-perfect clusters of the respective cell types (Figure 6.5). Within each lineage, subgroups, such as naive T cells or memory T cells also were clustered accurately together. Similarly, accurate clustering of different lineages was observed when we analyzed human-derived cells in the Differentiation Map (DMAP) (Figure 6.6) (277). Strikingly, hematopoietic stem cells were accurately distinguished from other lineages despite the fact that very few (1.62%) gene sets derived from stem cells were included in the ImmuneSigDB. This suggests that HSCs are characterized by differential expression of gene sets related to biological processes shared with immune cells (30).

We noted that distinct clusters of gene sets showed differential enrichment in specific cell lineages (Figure 6.5A, Clusters 1–5). We characterized these gene set clusters by determining the relative frequencies of genes shared by the gene sets in these clusters

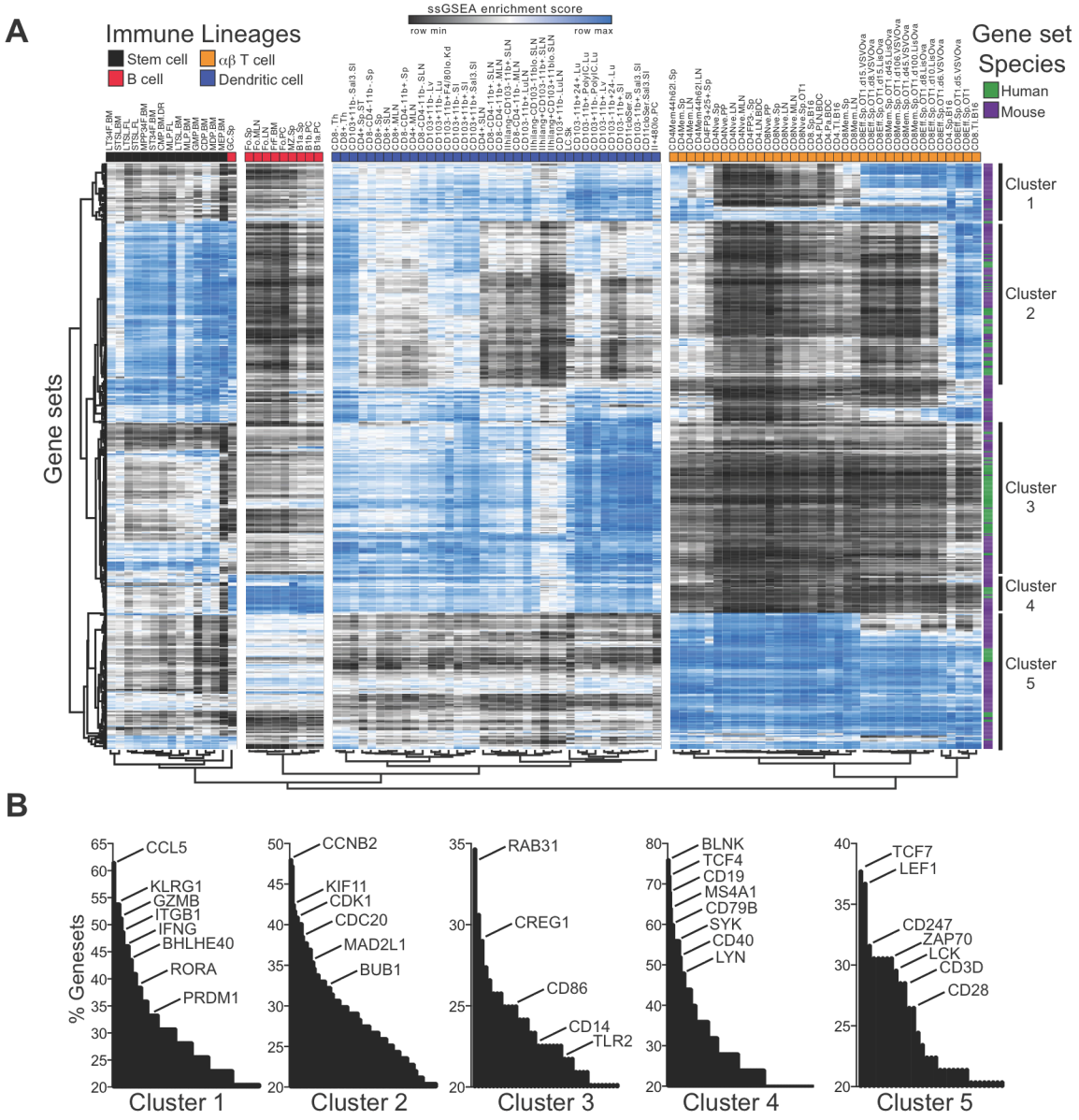

Figure 6.5. Mouse immune lineages are accurately clustered using ImmuneSigDB enrichments. (A) Unsupervised bi-clustering of ssGSEA values using ImmuneSigDB in samples of four representative mouse immune lineages. Hierarchical clustering of the 10% of gene sets with highest mean absolute deviation is shown. Species of origin of gene sets indicated by green (human) and purple (mouse) bars on the right. Major branches of the gene set dendrogram clusters are indicated by the numbered black bars on the right. (B) Distribution of genes contained in gene sets in the same gene set dendrogram clusters as indicated in (A).

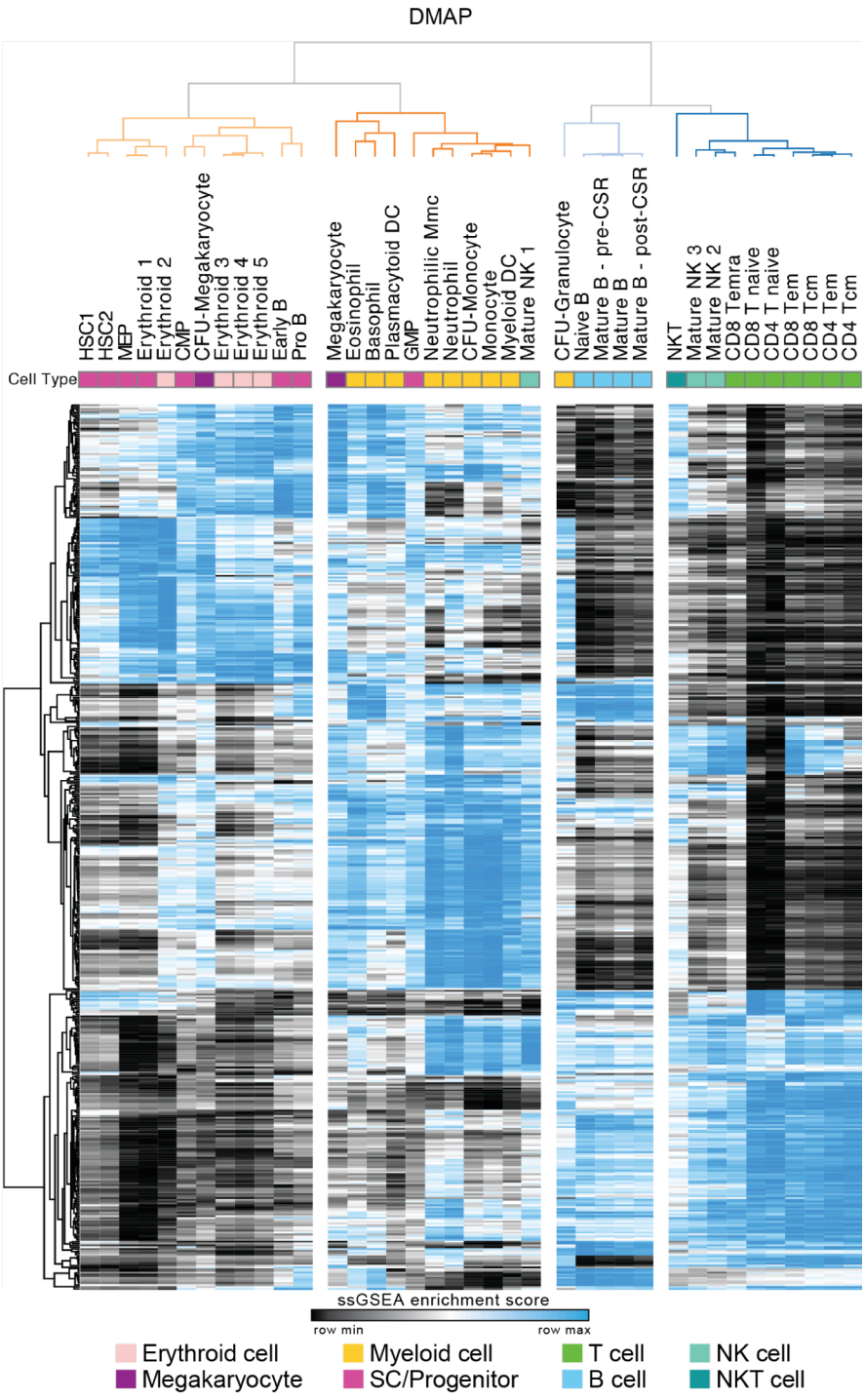

Figure 6.6. Human hematopoietic lineages are accurately clustered using ImmuneSigDB enrichments. (A) Unsupervised bi-clustering of ssGSEA values using ImmuneSigDB in cell subsets represented in human differentiation map (DMAP) dataset. Clustering was performed in gene sets with the top 10% highest mean absolute deviation.

(Figure 6.5B). For example, gene sets in Cluster 1, which predominantly distinguished effector and memory T cells from naive T cells, most commonly included genes encoding effector molecules such as granzyme B (*GZMB*), IFN-γ (*IFNG*) as well as Blimp1 (*PRDM1*) and integrin beta 1 (*ITGB1*), and were predominantly derived from expression profiles of effector and memory $CD8^+$ T cells in the context of viral infection and anti-tumoral responses (data not shown). Cluster 5, which predominantly distinguished T cells from other cell lineages, included T cell genes such as transcription factors TCF7 and LEF1 as well as components of T cell receptor signaling, CD3z (*CD247*), CD3d (*CD3D*), ZAP70, and Lck and included most gene sets derived from comparing T cells to other immune cell types (Figure 6.5B). Stem cells showed strong enrichment of gene sets in Cluster 2 whose predominant genes play a dominant role in regulating cell cycle (Figure 6.5B). B cells and dendritic cells were distinguished by a separate cluster of gene sets that included known genes representing those lineages.

We noted that each gene set cluster that distinguished different lineages included both human and mouse gene sets (Figure 6.5, purple and green bars), suggesting that similar patterns of enrichment were observed using gene sets derived from expression profiles from tissues of both species. Thus, ImmuneSigDB robustly clusters human and mouse immune lineages based on whole transcriptome enrichments of both mouse and human-derived gene sets.

Analogous cell types and contexts in mice and humans show common patterns of gene expression.

We, and others, have previously used GSEA to show that the transcriptional profiles from memory and exhausted $CD8^+$ T cells are highly concordant between mouse and human datasets (129, 130, 152). To test whether this similarity in gene expression between species is also observed for other cell states we extended this analysis to other cell types and

perturbations included in ImmuneSigDB. We focused on four separate biological comparisons (Figure 6.7) where analysis of gene expression had been made in analogous cell types or perturbations in both human and mouse immune cells. This allowed us to test whether sets of genes differentially expressed in mouse immune cells showed enrichment in profiles from the analogous comparisons in humans and vice versa (278).

We first identified 15 studies (6 mouse; 9 human) in which the transcriptional response to lipopolysaccharide (LPS) stimulation had been profiled in myeloid cells; each study had been used to generate a gene set in ImmuneSigDB. We selected one human dataset and generated a ranked list of genes differentially expressed following LPS stimulation. We then performed GSEA using gene sets from the other 14 mouse and human data sets. We found that both human- and mouse-derived gene sets showed highly significant enrichment (FDR<0.001), suggesting a strong conservation in the transcriptional response to LPS between the two species (Figure 6.7A). Gene sets derived from studies on human cells tended to show slightly higher enrichment scores than those generated from mouse cells.

We then selected additional cross-species comparisons that were represented by multiple datasets within the ImmuneSigDB. We found cross-species similarity in the gene expression profiles of comparisons of plasma versus naive B cells, memory B cells versus naive B cells, and regulatory T (Treg) versus conventional T cells (Tconv) (FDR<0.001, Figure 6.7, B – D and Supplemental Tables 6.8 – 6.11). Furthermore, we observed that the biology was not just conserved to the same extent but in some cases mouse-derived gene sets were more strongly enriched in human datasets than other human gene sets, as depicted by the peak height of the respective graphs of their GSEA enrichment scores. These findings indicate that components of the transcriptional signatures of LPS stimulation and some T and B cell differentiation programs are similar in humans and mouse models.

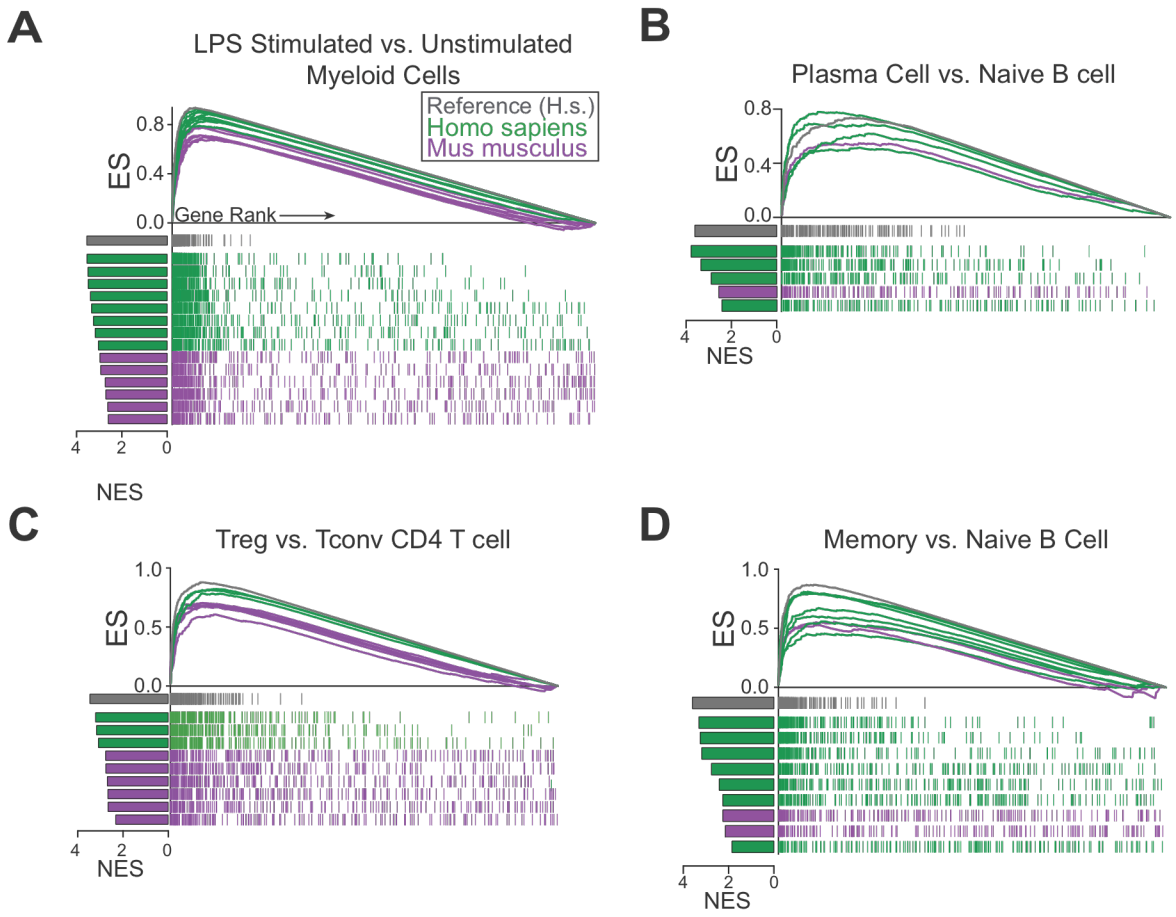

Figure 6.7. Transcriptional programs are conserved across mouse and human immune lineages. (A) GSEA of a randomly selected human study comparing LPS-stimulated and unstimulated dendritic cells using ImmuneSigDB gene sets derived from the study itself (grey) or gene sets from other mouse (purple) or human (green) datasets of LPS-stimulated myeloid cells. Mountain plots show all genes ranked by differential expression in sepsis versus control conditions on the X-axis, and the curves indicate cumulative enrichment (measured by enrichment score on the Y-axis). The ticks below the line correspond to the position of genes in each gene set. (B-D) Analysis as in (A) for three additional cell differentiation states: plasma cells (B), Tregs (C), and memory B cells (D). All gene sets shown are significantly enriched (FDR < 0.001).

Blood cells from human and mouse sepsis share conserved biology reflected in their transcriptomes

As we observed common patterns of gene expression in these cross-species comparisons, we next studied more a complex transcriptional dataset from human sepsis and the corresponding mouse models to test whether ImmuneSigDB could resolve similarities or differences between human and mouse transcriptional profiles. Recent studies have analyzed the transcriptional response to sepsis in multiple datasets of gene expression profiles obtained from human PBMC samples or from mouse models (79, 80). However, these studies have differed in their conclusions regarding the degree of similarity between species. We reasoned that analysis with ImmuneSigDB might allow a more detailed analysis of immune signatures elicited by the sepsis response in both species.

We began by using ImmuneSigDB to compare the similarity in gene expression in human and mouse datasets included in the previous studies of the transcriptional responses to sepsis. We selected, at random, a pair of human and mouse studies in which peripheral blood cell gene expression was measured in sepsis versus control conditions (PBMC following sepsis in human (GSE9960) and mouse (GSE19668)) (279, 280). We first identified genes up-regulated following sepsis in each study, and tested whether that signature was enriched in the corresponding profile of sepsis versus control in the other species.

We observed strong enrichment of the set of sepsis-induced genes derived from the mouse study in the human dataset (FDR<0.001, Figure 6.8A, left). Similarly, we found that a gene set comprised of genes up-regulated in human PBMC samples in sepsis versus control was strongly enriched in the mouse sepsis gene expression profile (FDR=0.002, Figure 6.8A, right). This internal comparison suggests that there is marked similarity between the genes up-regulated by sepsis in humans and in a mouse model.

Next, we identified similarity in gene expression in the sepsis response by testing for enrichment of all gene sets in ImmuneSigDB gene sets in the same pair of human and

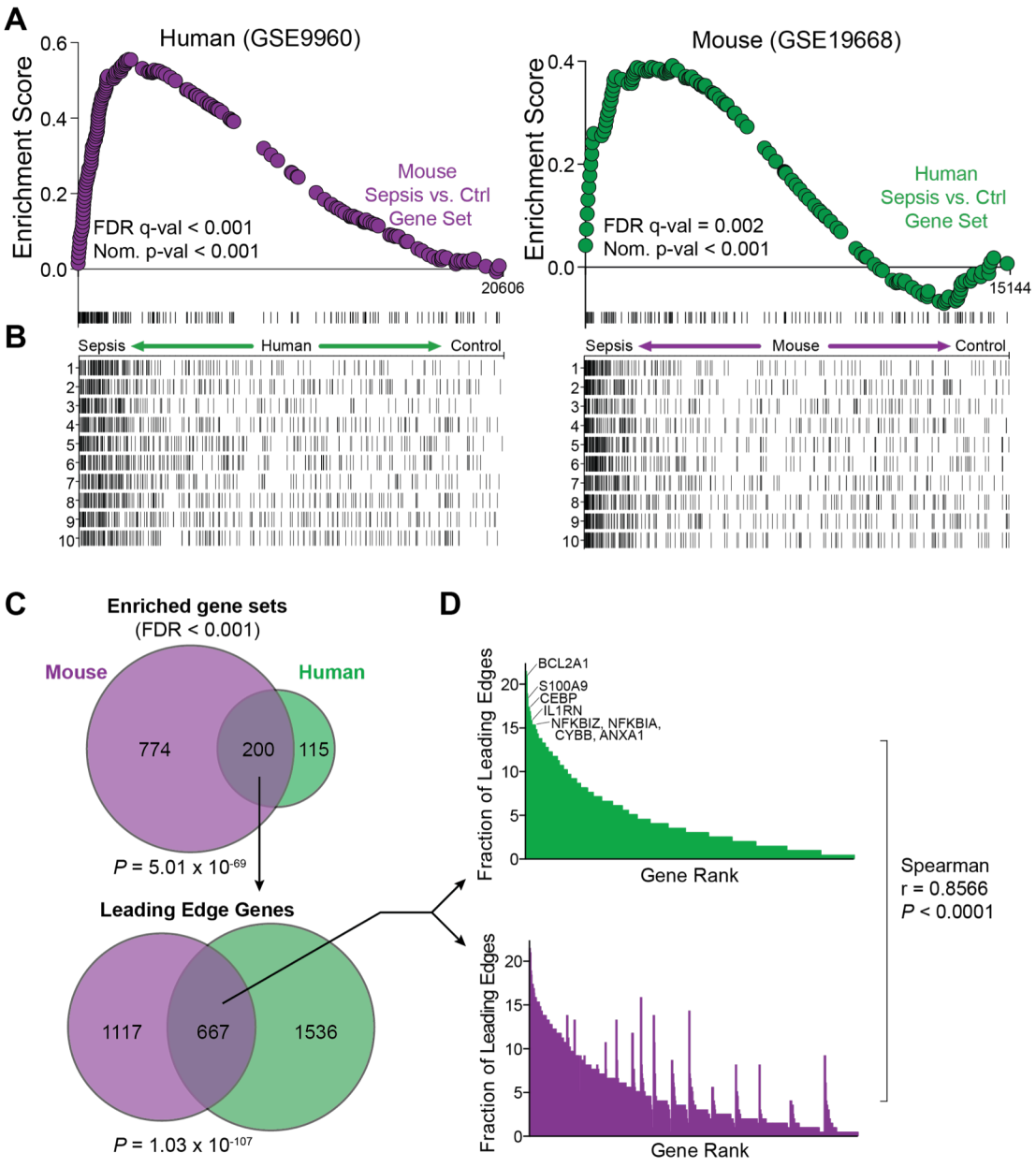

Figure 6.8. The transcriptional response to sepsis is conserved in humans and mouse models. (A and B) GSEA of the set of genes up-regulated in mouse sepsis (GSE19668, C57BL/6) in the ranked list of genes up-regulated in human sepsis (GSE9960, Gram negative infection) (A, left); and of the corresponding human sepsis gene set enriched in rank ordered list of genes up-regulated in mouse sepsis (A, right). Mountain plots indicate cumulative enrichment, and (B) ticks below the line correspond to the position of genes in the 10 most enriched gene sets from ImmuneSigDB in the rank order of genes up-regulated in sepsis versus control conditions (X-axis). (C) Venn diagram showing overlap in the identity of significantly enriched ImmuneSigDB gene sets in mouse (purple) or human (green) sepsis dataset (top) and the number of shared leading edge genes in the gene sets enriched in both species (bottom). Statistical significance calculated by the hypergeometric test. (D) Frequency of the genes occurring in the leading edge of a gene sets enriched in human (green) and mouse (purple) sepsis datasets. Statistical significance of the similarity in gene rank is calculated by the Spearman test.

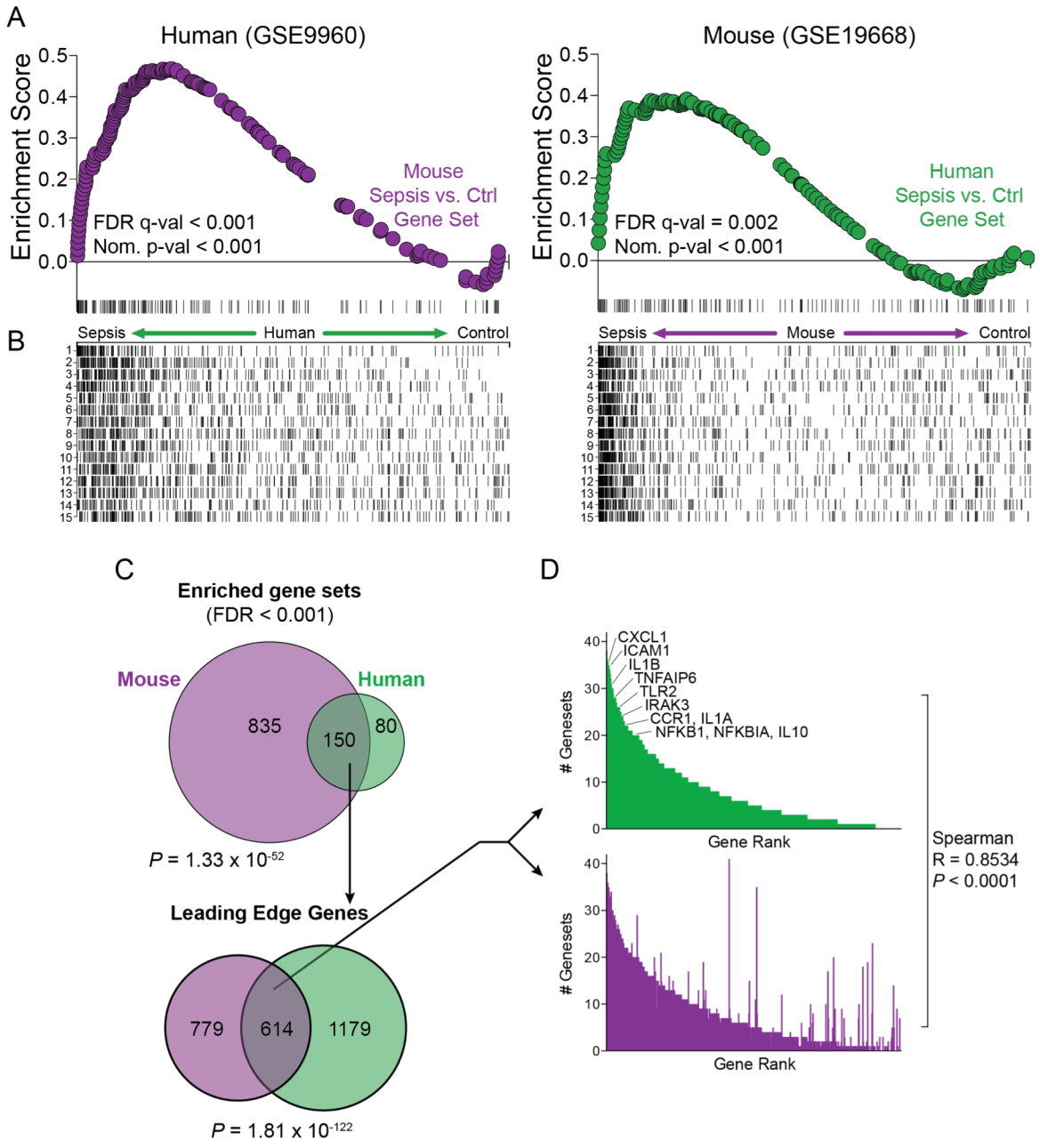

Figure 6.9. Human and mouse blood cells in gram positive sepsis undergo similar transcriptional response. (A and B) GSEA of gene sets of genes up-regulated in Gram positive human (GSE9960) or mouse (GSE19668, C57BL/6) in the opposite species – mouse gene set enriched in the human dataset (left) and the human gene set enriched in mouse dataset (right). Mountain plots indicate cumulative enrichment, and (B) ticks below the line correspond to the position of genes in the 10 most enriched gene sets from ImmuneSigDB in the rank order of genes up-regulated in sepsis versus control conditions (X-axis). (C) Number of significantly enriched gene sets mouse (purple) or human (green) dataset. Statistical significance was calculated by the hypergeometric test. (D) Distribution of the number of gene sets including the shared genes in the leading edges of common enriched gene of mouse and human sepsis dataset enrichments. Statistical significance is calculated by the Spearman test.

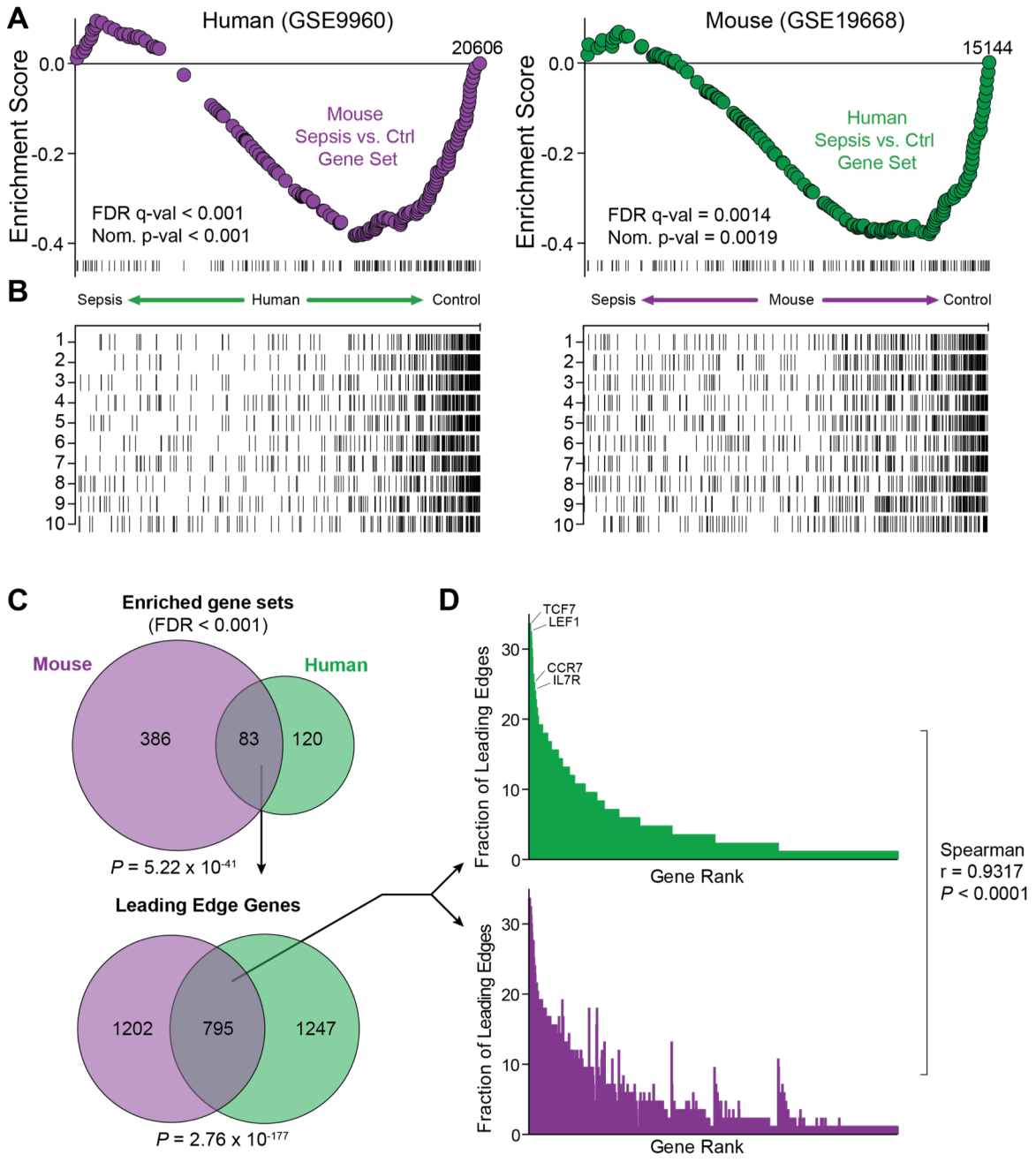

Figure 6.10. Human and mouse blood cells in gram positive sepsis undergo similar transcriptional response. (A and B) GSEA of gene sets of down-regulated in Gram negative bacteria-induced sepsis in human (GSE9960) or gram positive sepsis mouse (GSE19668, C57BL/6) in the opposite species – mouse gene set enriched in the human dataset (left) and the human gene set enriched in mouse dataset (right). Mountain plots indicate cumulative enrichment, and (B) ticks below the line correspond to the position of genes in the 10 most enriched gene sets from ImmuneSigDB in the rank order of genes down-regulated in sepsis versus control conditions (X-axis). (C) Number of significantly enriched gene sets mouse (purple) or human (green) dataset. Statistical significance was calculated by the hypergeometric test. (D) Distribution of the number of gene sets including the shared genes in the leading edges of common enriched gene of mouse and human sepsis dataset enrichments. Statistical significance is calculated by the Spearman test.

mouse studies. We compared the ImmuneSigDB gene sets that were significantly enriched in the gene expression profiles of the human and mouse gram negative and/or positive sepsis response (Figure 6.8B and Figure 6.9). We observed marked similarity in ImmuneSigDB gene sets that were enriched the sepsis-induced signatures in each species (P=5.01x10^{-69}, Figure 6.8, B and C, and Supplemental Tables 6.12 and 6.13).

To identify which genes in the gene sets that were enriched in both species were "driving" the enrichment of the shared gene sets, we focused on the "leading edge" of enrichment. Leading edge genes in a gene set enrichment analysis are those that contribute most to the enrichment of a particular gene set and include the most significantly up-regulated genes in a given gene set. We found that the leading edges of gene sets that were enriched in both species were similar (Spearman r = 0.857, P<0.0001; Figure 6.8D, 6.10D) indicating that the strong enrichment of shared gene sets is due to the up-regulation of similar genes. We found the same results when we performed the same set of analyses using a pair of human and mouse datasets where both were from gram-positive sepsis or when we analyzed gene sets enriched in down-regulated genes in sepsis compared to control (Figure 6.9 and Figure 6.10, Supplemental Tables 6.14). These data demonstrate a high degree of concordance in gene sets that are enriched following sepsis in humans and mouse models.

Identifying species-specific components of transcriptional responses induced by sepsis in human and mouse.

We noted that while many gene sets in ImmuneSigDB were enriched in both species, there were also many gene sets enriched in one species but not the other (Figure 6.8C and 6.10). This suggested that in addition to similarities in the sepsis response, there may be species-specific differences in the transcriptional signatures of sepsis. In order to identify the biological basis for these species-specific differences, we devised an analytic approach,

termed Leading Edge Metagene (LEM) analysis, to identify main biological "themes" in groups of ImmuneSigDB gene sets enriched in the sepsis datasets. We introduce LEM here and describe it in more detail elsewhere (see Methods). LEM analysis is a novel method to identify the groups of co-regulated genes – that we term metagenes – that are highly enriched in multiple gene sets in a comparison of interest (such as sepsis vs. control).

For LEM analysis, we first considered all gene sets that were significantly enriched in each dataset of sepsis versus control comparison group (FDR<0.001). We then filtered the genes in these enriched gene sets to include only leading edge genes (Figure 6.11A, top and middle). These leading edge genes represent the subset of genes in the group of enriched gene-sets that drive the enrichment score with respect to up-regulation in the sepsis phenotype. We then used non-negative matrix factorization (NMF) (281-284) to identify groups of genes that are members of multiple gene sets (Figure 6.12). NMF analysis therefore identifies groups of genes – which we term metagenes – that are members of the leading edge of multiple gene sets that are enriched in the transcriptional response to sepsis (Figure 6.11A, bottom).

LEM analysis of the gene sets enriched in human sepsis (316 gene sets) and mouse sepsis (974 gene sets) studied in Figure 6.8 identified three metagenes that were correlated with the sepsis response in each study. Individual metagenes were strongly overrepresented for genes related to distinctive biological processes as annotated by GO terms and Reactome (99, 285) (Figure 6.11B and Figure 6.13). For instance, in the human sepsis response, we identified a metagene with an overrepresentation of genes involved in mitosis (P=4.9 x 10^{-22}) such as *CCNA2*, *BUB1*, and *KIF11.* A second metagene was enriched for genes related to phagocytosis (P=2.02 x 10^{-13}; *LAMP2*, *NCF4*, and *ATPV0B*) and a third metagene was enriched for genes related to inflammation (P=3.7 x 10^{-4}; *IL1A*, *NFKB1*, and *CCL20*) (Figure 6.11B and Figure 6.12). Overlap between the metagene gene memberships and specific GO terms revealed one predominant biological process in each (Figure 6.11B).

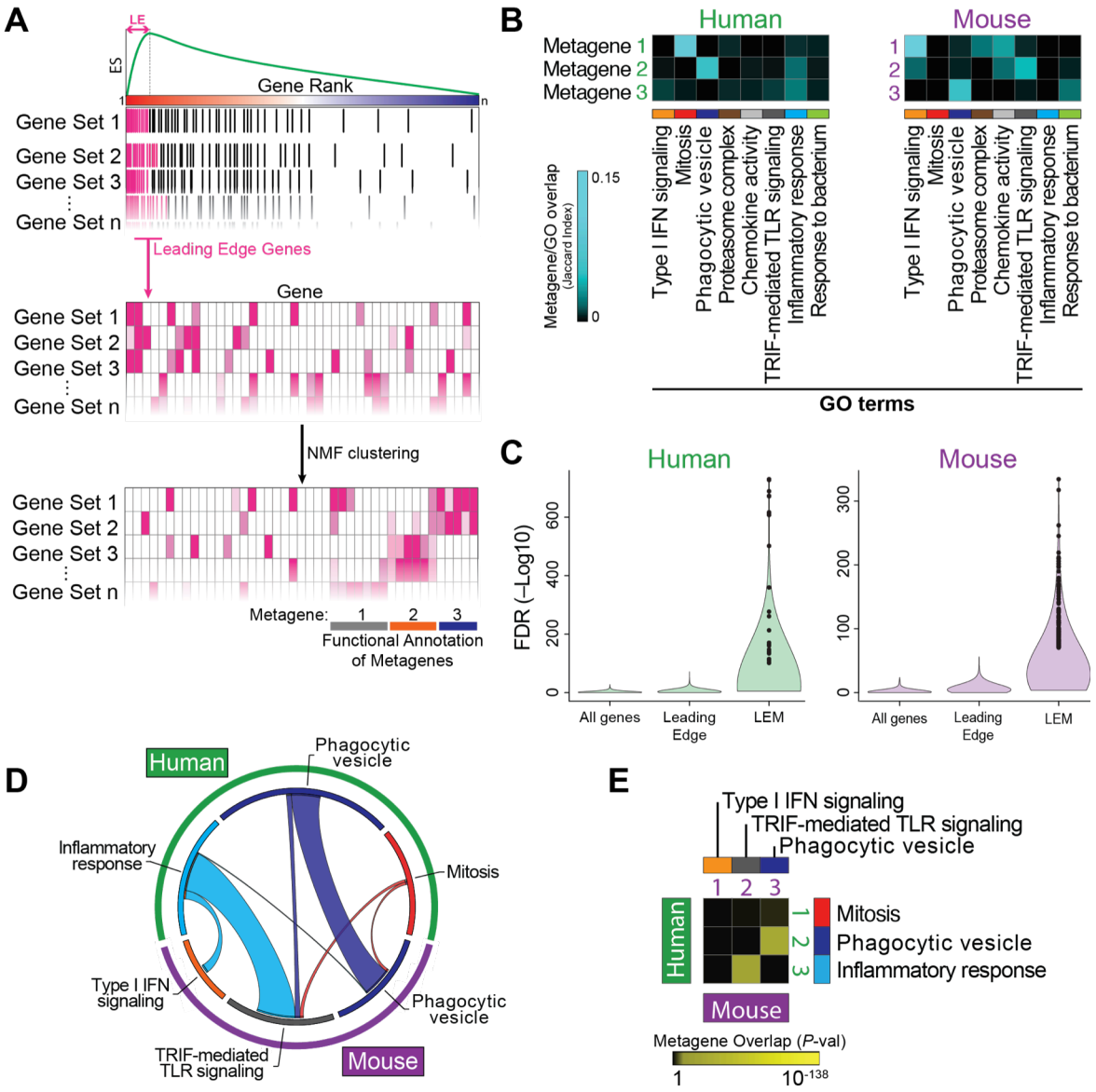

Figure 6.11. Leading edge clustering using non-negative matrix factorization (NMF) identifies metagenes representing distinct biological processes. (A) A schematic of the process by which leading edge metagenes are identified. (B) Biological annotation of metagenes identified in the studies analyzed in Figure 6.8 generated using GO terms. (C) Violin plots showing *P* values of significance of GO Term overlaps with human (left) and mouse (right) sepsis metagenes (LEM), or equivalent-size samples of leading edge genes, or randomly selected genes. (D) Circos plot of the relative size and overlap of metagenes in mouse (purple, outer segment) and human (green, outer segment) sepsis datasets. Relative number of genes in metagenes is indicated by segment length of the inner circle. Thickness of the ribbon corresponds to the relative number of genes shared between metagenes in the two species. (E) Heatmap of *P*-values corresponding to significance of overlap in pairwise comparison of metagene gene membership (yellow, highly significant; black, not significant). Statistical significance of the overlap calculated by hypergeometric test.

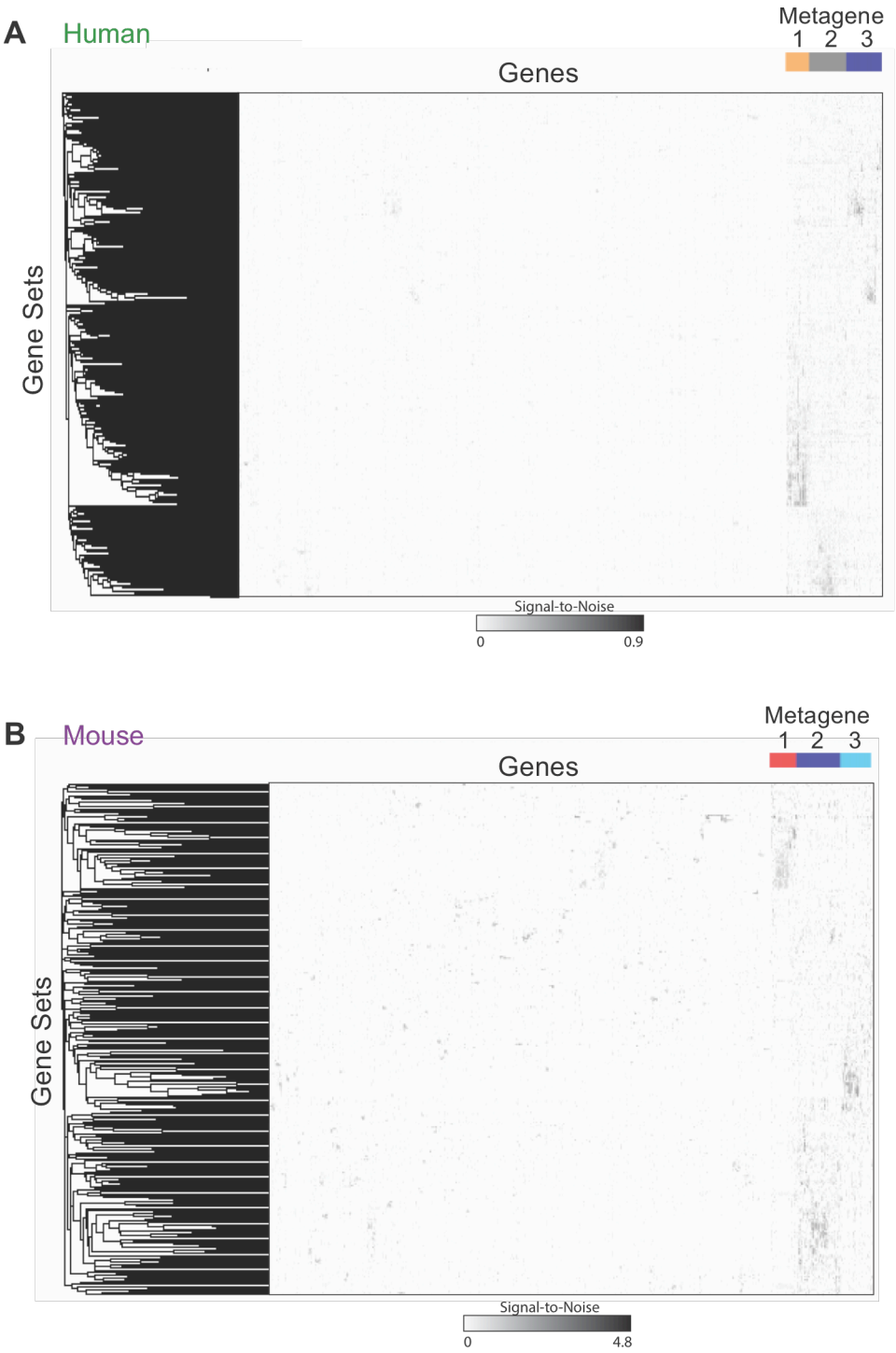

Figure 6.12. Leading Edge Metagene (LEM) analysis identifies commonly represented and co-regulated in immune biological processes. Heatmap representing the sparse matrix of all leading edge genes (FDR<0.001) in GSEA analysis of human (A) and mouse sepsis (B) as in Figure 6.11. Gene sets were clustered using Pearson correlation and genes ordered based on their LEM membership, annotated above.

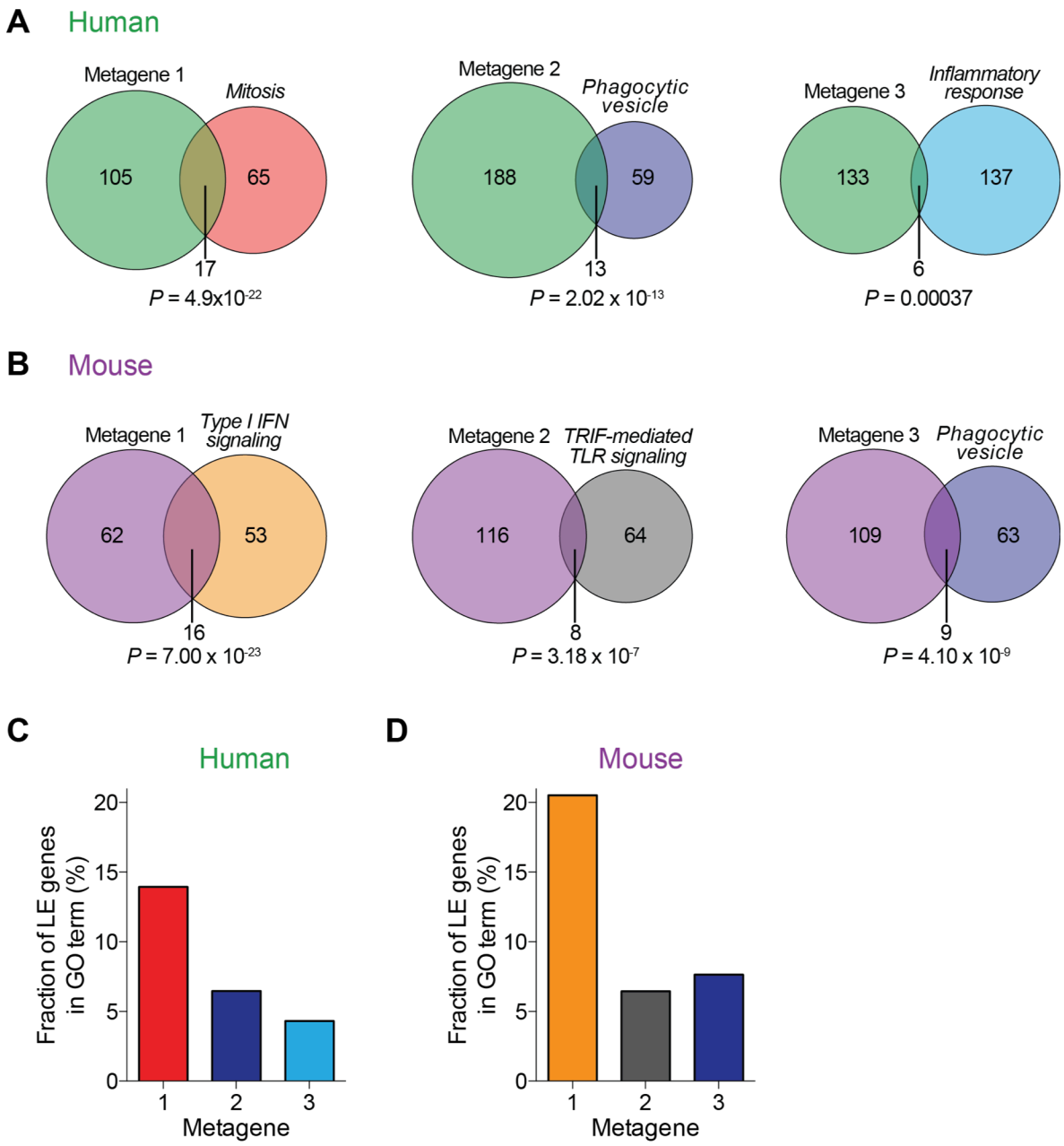

Figure 6.13. LEM analysis identifies metagenes that are unique, yet overlapping with Gene Ontology Terms. (A and B) Overlap in the genes contained in LEMs defined for human (A) and mouse (B) sepsis described in Figure 6.8 and the predominant GO terms. Numbers of genes are indicated in the Venn diagrams, and the statistical significance of each overlap was assessed using hypergeometric test in the space of 20606 (A) and 15183 (B) total genes, based on annotated unique genes or human orthologs. (C and D) Fraction of LEM genes that are contained in the predominant GO term for human (C) and mouse (D) studies.

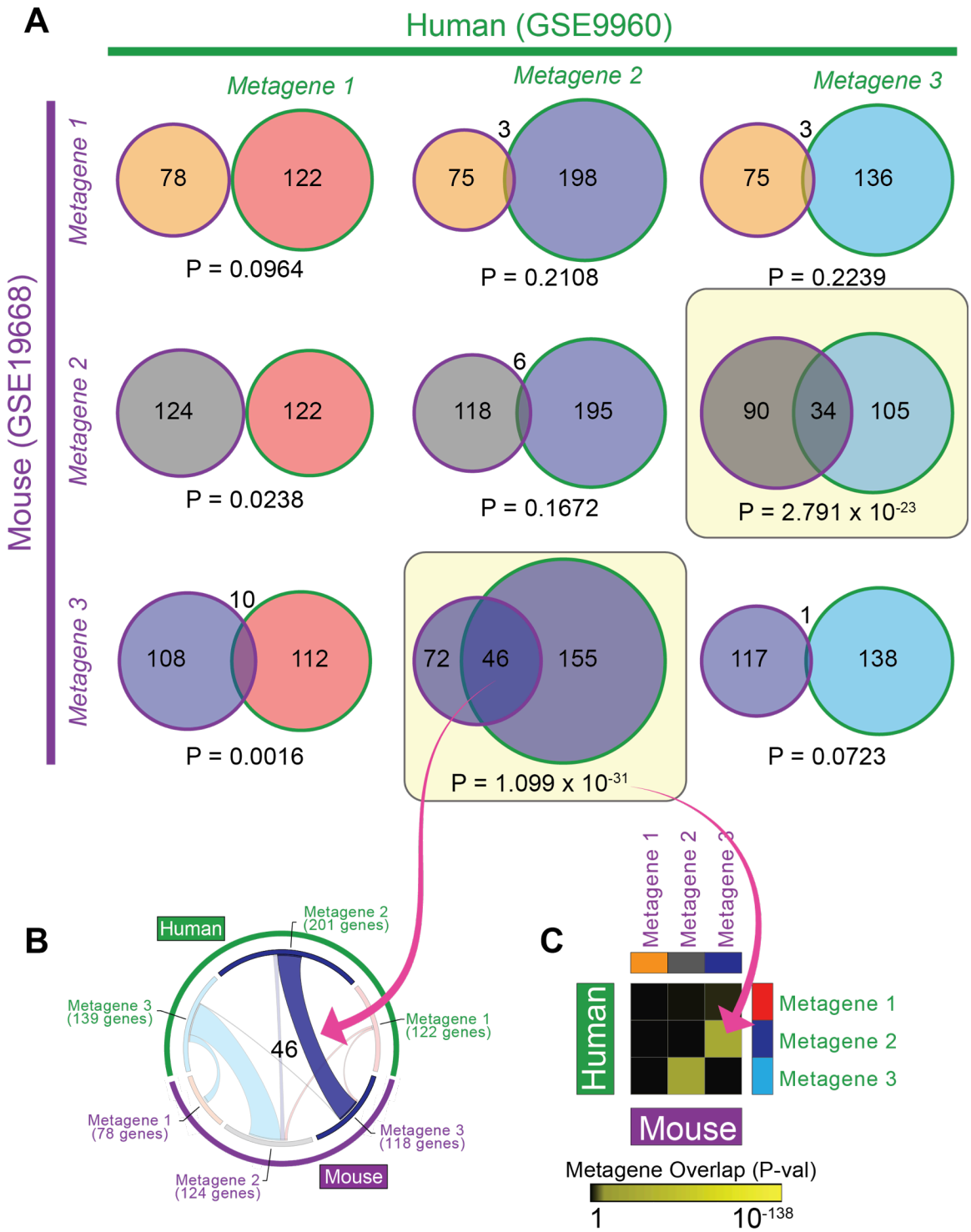

Figure 6.14. Illustration describing metagene overlap representation. (A) Venn diagrams representing absolute gene overlap of each human metagene with each mouse metagene. Numbers represent the number of unique and overlapping genes in each comparison. Statistical significance was assessed using Hypergeometric text in the space of all shared genes from the two datasets (n=12,634). The relative overlap of each human metagene with each mouse metagene is represented in Circos plot (B) and the significance of each overlap is represented in a heat map form (C).

However, the while each metagene was significantly enriched for one predominant GO term, only 5 – 15% of genes contained in each metagene overlapped with genes in the predominant GO term (Figure 6.11B). This suggests that the genes contained in each LEM are related to recognizable biological processes, but that the metagenes represent discrete modules of genes that overlap with but are distinct from GO term categories.

We reasoned that metagenes would provide a more "refined" list of functionally related genes than their parental gene-sets. We therefore tested whether leading edge metagenes were more highly enriched for genes related to biological processes (again as annotated by overlap with GO terms) than their parental gene sets (Figure 6.11C). We tested the set of 3 leading edge metagenes for overlap with the collection of GO annotated gene lists, and determined the significance of each GO term's overlap. We compared the *P* values generated by GO term overlap with the set of genes comprising each metagene with an equivalent number of genes randomly sampled from the original pool of leading edge genes, or from all genes in the genome. We found that the significance of GO term overlap was much higher in the leading edge metagenes than in the original leading edge genes or in a random set of genes. LEM analysis therefore is an effective strategy to both identify major biological processes active in a phenotype of interest, and simplify the list of 315 and 974 enriched gene sets in human and mouse, respectively, to a core set of 3 metagenes in each organism that correspond to major biological themes.

We next compared the similarity between metagenes identified in the sepsis response in humans with those in mouse sepsis models. We visualized the pairwise overlap in genes in each metagene using a Circos plot (Figure 6.11D), and determined the significance of the overlap for each pairwise comparison of mouse and human metagenes (Figure 6.11E and 6.15) (286). We found striking cross-species similarities for some but not all metagenes. For example, a metagene annotated as "Phagocytic Vesicle" correlated with both the human and mouse sepsis response and contained a very similar set of genes

(hypergeometric test P=1.09x10^{-31}, dark blue ribbon, Figure 6.11D). Similarly there was a highly significant overlap in the metagene annotated as “inflammatory response” in the human dataset and “TRIF-mediated TLR signaling” in the mouse model (P=2.79x10^{-23}).

However, we also identified metagenes that were not conserved between humans with sepsis and the mouse model. For example, a metagene enriched for genes pertaining to cell cycle (“mitosis” GO term) in humans did not share a corresponding metagene in the mouse model. In the mouse, a type 1 interferon signaling metagene overlapped with very few human metagenes. This analysis approach using ImmuneSigDB suggests that while some biological processes are strongly conserved between these two human and mouse datasets (e.g., phagocytosis, TLR mediated inflammatory response), other biological components are not (e.g. mitosis).

Global shared and species-specific biological processes can be identified using ImmuneSigDB and NMF clustering

We next extended this approach to six datasets of sepsis versus control conditions from three independent studies in humans and from four comparisons in two mouse studies. We identified between three or four metagenes in each study providing a total of 35 metagenes present in the collected group of sepsis studies (Figure 6.15A). We annotated each metagene based on enrichment of GO Terms (Figure 6.15A, right) and evaluated the significance of pair-wise overlap in the genes present in each metagene.

We found that in almost every study, there was at least one metagene induced by sepsis that showed a highly significant overlap (indicated in yellow in the heatmap, Figure 6.15A) with metagenes from every other sepsis study, regardless of species of origin. One study that proved an exception was the human study GSE9960, which studied a response to mixed infection, and showed relatively little overlap with any mouse study. However the metagenes identified in that study also showed limited overlap with metagenes from other

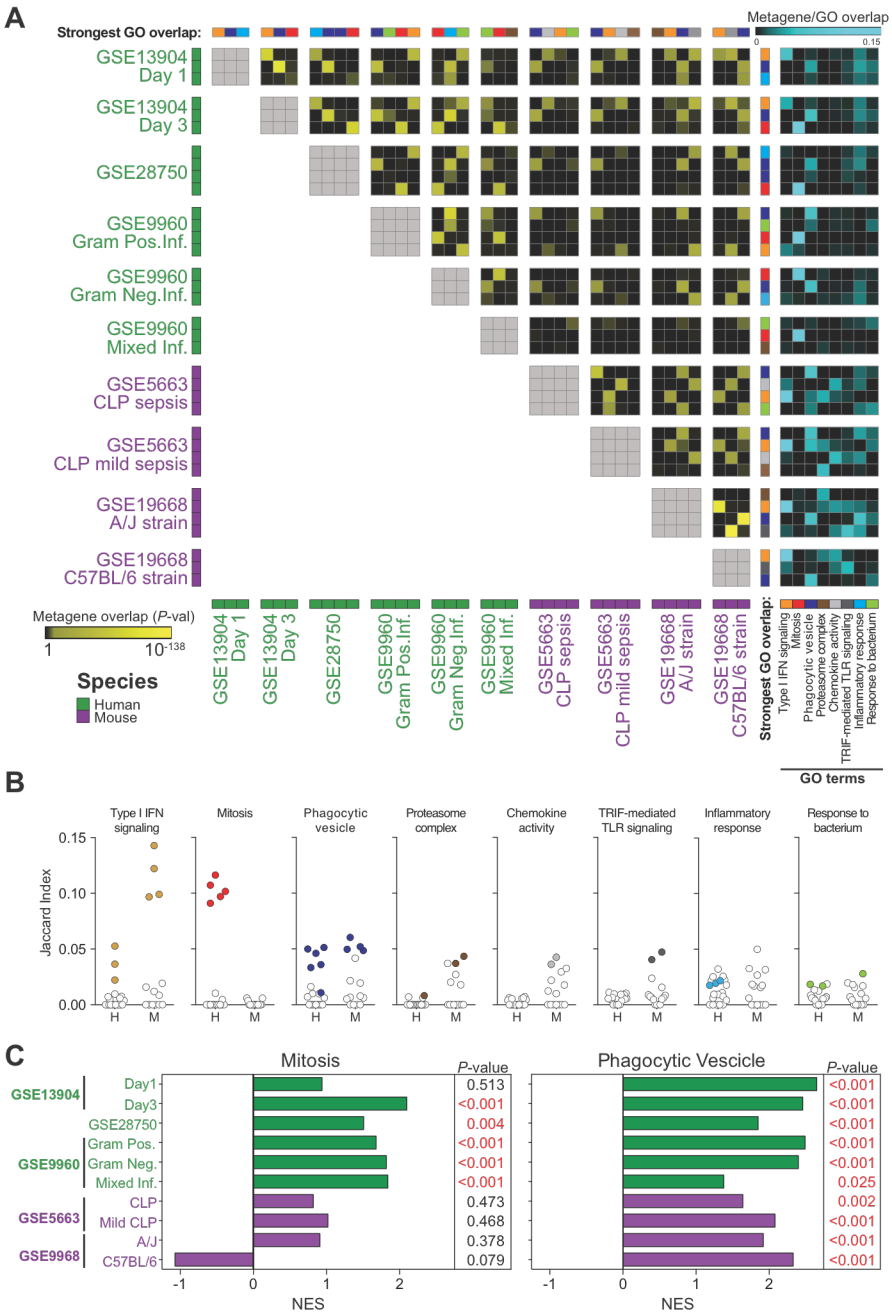

Figure 6.15. ImmuneSigDB identifies shared and unique biology in mouse and human sepsis studies. (A) Pairwise overlaps of all metagenes from mouse (purple bars) and human (green bars) sepsis studies. Heatmap indicates *P*-values corresponding to significance of overlap between each metagene (small squares) in each study (larger squares; yellow, highly significant; black, not significant). The biological annotation of each metagene is based on the significance of enrichment of the GO term indicated (blue, large overlap; black, no overlap) (right). The most significantly enriched GO term annotating each metagene is indicated by the key in lower right. (B) Jaccard index representing the extent of overlap of metagenes from human (H) and mouse (M) studies. Colored are metagenes that are annotated with the respective biological process as in (A). (C) Enrichment scores of biological processes that are species-specific (e.g. mitosis, left) or shared (e.g. phagocytic vesicle, right) in the human (green bars) and mouse (purple bars) sepsis datasets. Significance of the enrichment of the named biological process in each data set is indicated by the *P*-values on the right.

human studies, suggesting that that transcriptional response contained in that study may represent a different type of biological response to the other human and mouse studies.

In addition to these strongly conserved metagenes, we also found that there were metagenes induced by sepsis that had a striking species-specific distribution. For example, the phagocytic vesicle metagene was either present or strongly overlapped with a metagene present in every dataset, both mouse and human. In contrast, the Mitosis metagene was much more specific to human datasets with no significant overlap in mouse studies (Figure 6.15B). To confirm these results, we tested the significance of enrichment of two GO terms – mitosis and phagocytic vesicle. When looking at the whole genome transcriptional changes, we indeed observed that mitosis was represented exclusively in human cells in sepsis while phagocytic vesicle process was represented in both species, as we predicted based on the LEM analysis (Figure 6.15C). These data reveal context-specific transcriptional modules induced by sepsis in humans and mice, and also highlight the distinct transcriptional components of the response present in one species but not the other.

Discussion

We analyzed expression profiles from 389 published studies of mouse and human immune cells to generate a collection of curated gene signatures corresponding to cell states and perturbations in the immune system. This collection of almost 5,000 genes sets contains substantial biological information that was not currently contained in existing collections. We used this new compendium to show that transcriptional signatures induced by LPS stimulation in dendritic cells, and transcriptional programs of T cell and B cell differentiation were highly conserved between humans and mouse models. Moreover, we used ImmuneSigDB to analyze expression profiles from patients and mouse models of sepsis and showed highly significant overlap, suggesting that components of the transcriptional response to sepsis were highly conserved between species. However, we also find that there are substantial species-specific differences, both in enriched gene sets and their component metagenes, in sepsis response signatures, suggesting that not all biological processes induced by sepsis evident at the transcriptional level in humans are present in mouse models and vice versa. These findings suggest that ImmuneSigDB provides a useful tool for detecting subtle patterns of similarity and difference in large-scale datasets of gene expression from cells and tissues in the immune system.

Several studies have directly compared the transcriptional programs in the human and mouse immune systems. We, and others, previously identified conserved patterns of gene expression that change during the differentiation of memory T and B cells, and in exhausted to $CD8^+$ T cells (129, 130, 152). A recent comparison of gene expression in seven immune cell groups from humans and mice also found a highly significant degree of similarity in global patterns of expression and in the putative transcriptional regulators of these genes (287). However, in that study, while the majority of genes showed a pattern of expression that was highly correlated between species, 30 – 50% of genes did not show significant correlation between species. Two recent studies of the sepsis datasets analyzed in the

present study reached opposite conclusions regarding the degree of similarity between the mouse and human response to sepsis (79, 80). Thus, the degree of conservation of transcriptional signatures in the mouse and human immune systems remains controversial (263-270).

Our analysis using ImmuneSigDB suggests that there are both conserved *and* species-specific transcriptional programs induced by sepsis in the immune system. Overall, the transcriptional program shows highly significant similarity between sepsis in the human and mouse. Specifically, analysis of the leading edge metagenes across human (6 comparisons) and mouse sepsis datasets (4 comparisons) found that in many of the datasets from both species there was coordinate up-regulation of metagenes involved in interferon-response and phagocytic processes. This suggests that features of the sepsis response such as interferon release and neutrophilia are shared between species.

However, we also show that many gene sets are enriched in only one species, and metagenes related to mitosis were highly enriched in sepsis-induced profiles in humans but were not significantly enriched in the mouse model. Thus, it is likely that while some components of the sepsis response are highly conserved between species, there is also substantial divergence in the biological processes detected by transcriptional profiling each. Detailed analysis of the transcriptional features of the mouse and human immune systems is therefore required to substantiate conclusions regarding the conservation of a particular biology of interest in two datasets. Whether the differences we observe are due to inherent biological differences between the two species remains unclear. For example, it is possible that the mitotic signature is present in human, but not mouse, because the exact timing of the initiation of activation of immune cells in humans with sepsis is not precisely known and may be more variable compared to tightly controlled, narrow window of induction of sepsis in mouse models.

Our compendium adds to a growing list of collections of transcriptionally co-regulated genes in the immune system. In the human immune system, several studies have identified groups of co-regulated gene modules from expression profiles derived from blood samples representing a range of states of health and disease (273, 274). This modular approach to the analysis of gene expression can aid interpretation of gene expression profiles, increase robustness, and facilitate analyses that span multiple datasets. However, ImmuneSigDB is distinct from those previously described in several respects. Studies by Chaussabel and by Li have focused on identifying collections of genes – termed modules – that tend to vary in expression in a coordinate fashion across a reference set of expression profiles (273, 274). Defining modules based on network reconstruction across hundreds or thousands of experimental conditions makes it difficult to associate a particular module with a defined cell state or perturbation that usually results in its up- or down-regulation. In contrast, the annotations describing each gene set in the ImmuneSigDB include all the experimental details from a published manuscript, allowing a more transparent connection between gene set and biology. Moreover, ImmuneSigDB was designed for use with GSEA, because each gene set contains either up- or down-regulated genes only, rather than a combination of both as can appear in Chaussabel or Li modules, which may limit the use of the latter collections in analyses such as GSEA. Finally, each collection of previously-published modules was defined in a single species (humans), making the generalizability of these compendia to other species hard to predict.

The ImmuneSigDB collection differs in another important respect from previous module collections. The studies by Chaussabel et al. and Li et al. were designed to identify *non-overlapping* modules of gene expression. However, ImmuneSigDB contains gene sets derived from experimental perturbations that are likely to induce multiple biological processes, each of which might be represented by sub-signatures in a given gene set. Moreover, several gene sets may contain the transcriptional correlate of the same biological

processes. For some analytic purposes, it might be useful to have a single gene set that includes the multiple biological processes that are initiated by the complex stimulus such as receptor-ligand engagement, or cell differentiation. However, for other applications, such as the analysis that we conducted of the sepsis datasets, a more "atomic" approach maybe preferred. We have therefore developed an analytic approach to extract non-redundant leading edge metagenes, from the experimentally-derived gene expression profiles.

Analysis with ImmuneSigDB using GSEA or GSEA combined with a leading-edge metagene analysis may therefore provide the systems immunologist with a useful resource for the analysis of gene expression in the immune system.

Materials and Methods

ImmuneSigDB Generation

We surveyed the immunology literature and identified published studies that included human or mouse microarray Affymetrix gene expression data in NCBI Gene Expression Omnibus (GEO, www.ncbi.nlm.nih.gov/geo/). We downloaded the corresponding datasets from GEO (288). When available, raw microarray data in the form of the CEL files were normalized by the Robust Multichip Average (RMA) (289) using justRMA function from the R Bioconductor package affy (version 1.40.0) (290). When CEL files were absent, we downloaded processed expression data from GEO by means of R GEO query package (version 2.28.0) (291). We mapped Affymetrix probe set identifiers to human gene symbols using the Collapse Dataset tool (max probe algorithm) of the GSEA program (292). We used ortholog gene assignments from Mouse Genome Informatics. The specific mappings were retrieved from the MGI web site on 14 April 2012 and contained 17,827 human - mouse ortholog gene pairings. Phenotype classes were assigned manually according to the original sample annotations and based on review of meaningful biological comparisons (Figure 6.2). We implemented a pipeline in R, which combined processed microarray data with the phenotype annotations and produced standard formatted files (.gct and .cls) for each comparison as needed.

For each two-class comparison, the genes were ranked according to an information-based similarity metric (RNMI) (293) from top up-regulated to bottom down-regulated genes in the two groups. Gene sets comprised genes differentially expressed with an FDR < 0.02, and a maximum number of genes was set at 200 (i.e. all gene sets had at most 200 differentially expressed genes). This way we generated two gene sets from each assigned biological comparison of two groups – “Group_A_vs_Group_B_UP” and “Group_A_vs_Group_B_DN”, for the top up-regulated and bottom down-regulated genes,

respectively, identified for the genes most different in the samples in Group A compared to the samples in Group B. The resource is accessible as the C7 collection at www.msigdb.org.

Gene set enrichment analysis (GSEA)

GSEA was performed as described previously (78, 271). To analyze transcriptional data from Immunological Genome Project (ImmGen) (275) and Differentiation Map (DMAP) (277), we used single sample GSEA (ssGSEA) as described previously (276, 294), to create a matrix in which columns represented individual samples and rows corresponded to gene sets, and the values represented the single sample ssGSEA score of each gene set in each sample. We averaged the biological replicates and filtered this matrix to include only the top 10% of gene sets based on Mean Absolute Deviation (MAD) across sample types and bi-clustered using 1-Pearson Correlation.

Leading Edge Metagene (LEM) Analysis

We developed an approach to identify groups of genes – termed *leading edge metagenes* (LEM) – that are both associated with a phenotype of interest, and shared between multiple gene sets enriched in that phenotypic comparison (See Experimental Methods). We reasoned that groups of genes that are co-regulated in the phenotype of interest and also present in multiple gene sets are likely to represent the core sub-signatures of genes related to distinct biological processes or pathways. Our approach leverages the notion of the *leading edge genes* in a GSEA analysis, which are the genes whose expression profile is most highly correlated with the phenotype distinction in a comparison of biological states and thus drives the GSEA enrichment statistic. LEM analysis identifies groups of genes (metagenes) that are common to multiple gene sets returned in a GSEA result, and strongly correlated with the phenotype of interest.

First we perform GSEA using the ImmuneSigDB in a two-class comparison of interest (e.g., sepsis vs. control). GSEA yields an enrichment score to quantify the overrepresentation of a gene set (e.g. genes coordinately up- or down-regulated in previous experiments) at the top or bottom of a ranked list of genes (e.g. generated by differential expression of in a comparison of interest). The leading edge of each enriched gene set is defined as the subset of genes with positive contribution to the enrichment score before it reaches its peak; i.e. those that are most correlated with the phenotype of interest.

We then consolidate the leading-edges of the *m* top-scoring gene sets into a sparse *n by m* matrix *M,* where the number of rows is the cardinality of the union of genes from all the leading-edges in the *m* top gene sets, and the columns correspond to the genes in the *m* enriched gene sets. The value of each entry in the matrix is the signal to noise ratio of the corresponding gene between two conditions in comparison (eq. 1) and 0 if the gene is not in the leading edge of that gene set. A large signal to noise ratio indicates a significant difference in expressions of the corresponding genes between the two conditions.

$$s2n = \frac{\mu_A - \mu_B}{\sigma_A - \sigma_B} \qquad (1)$$

To identify clusters in this matrix, we use non-matrix factorization (281-284) to yield two matrices, *W* and *H*. *W* matrix is a low-dimensional representation of the *M* matrix and each dimension of *W* is a linear combination of *n* genes, called a metagene. The entries in the *H* matrix represent the quantity of each metagene required to reconstruct each of the M gene sets. The coefficient in *W* matrix can be viewed as the contribution of each gene to the corresponding metagene. Inspection of the *W* matrix shows that in each metagene, the coefficients of most genes are usually very small, and only a small number of genes have a coefficient significantly larger than 0. As each metagene is a positive linear combination of all the genes, a small coefficient indicates negligible contribution to the metagene. Thus the next step of our algorithm is to filter out genes with small coefficients in each metagene. To do

that, we first assume that the background distribution of coefficients fulfills an exponential distribution. We set a filtering threshold at the 95% quantile of the fitted exponential distribution and set all coefficients below this to zero.

As each gene can contribute to more than one metagene we next need to assign each gene to a single metagene. The assignment of genes to metagenes uses the following rules: 1) if one gene has no contribution to any of the metagenes, it will be defined as not in any metagene; 2) each gene with a coefficient above the threshold (defined above) will be assigned to the metagene where it has the largest coefficient. Each metagene is annotated with a biological “theme” based on the Jaccard overlap of its constituent genes with GO categories (99).

Acknowledgments

I would like to acknowledge Arthur Liberzon (Mesirov Lab, Broad Institute of MIT and Harvard) for assistance with automation of the re-analysis and processing of studies accessed from GEO to obtain gene sets. Yan Tan (Haining and Mesirov Labs, Dana-Farber Cancer Institute and Broad Institute of MIT and Harvard) developed the Leading Edge Metagene Analysis software used in this study.

Chapter 7. Conclusions

Deep understanding of $CD8^+$ T cell biology is key for rational design of novel medical interventions such as cellular immunity-inducing vaccines as well as immunotherapies for cancer, infectious diseases, and autoimmunity. This thesis investigates the biology involved in $CD8^+$ T cell activation, contraction, memory differentiation, and exhaustion. The work presented herein advances the understanding of the mechanisms of known molecules – PD-1 and BATF – in $CD8^+$ T cell memory development and initial steps of T cell activation, showcases the development and utility of a new experimental platform that enables systematic interrogation of gene function in any hematopoietic cell, and highlights the novel functions for molecules without previous appreciation in $CD8^+$ T cells – CD39 and TGIF1.

T cell activation is a tightly regulated process of signaling events from the surface collaboratively inducing transcriptional changes driving T cell differentiation. While the stimulatory signals through TCR/CD28 have been long-appreciated in T cell activation, the role of co-inhibitory surface receptors in this process is becoming increasingly appreciated. PD-1 and other co-inhibitory receptors such as CTLA-4, CD244 (2B4), Tim-3, and Lag-3 are often associated with T cell exhaustion, but are also transiently up-regulated following the activation of T cells (295). Their expression and engagement of respective ligands can thus act as a negative regulation to balance the activation signals that T cells receive. We show here that PD-1 engagement on $CD8^+$ T cells is important for the optimal generation of $CD8^+$ memory T following viral infection in mice. However, PD-1 is not absolutely essential for memory generation considering that virus-specific memory $CD8^+$ T cells are present in influenza-immune mice. These mice are also protected from a lethal secondary challenge with influenza, albeit with increased morbidity and delayed resolution of infection. These findings highlight the complexity of the different factors that affect $CD8^+$ memory T cell generation, such as strength and duration of signals through the TCR and co-stimulation, the signals from the environment including cytokines and other soluble mediators (6, 27, 36). Our

data underscore the importance of negative regulation in addition to positive signals during T cell activation for optimal generation of T cell memory. It is possible that additional co-inhibitory receptors up-regulated following T cell activation may also contribute to the early events that shape the formation of memory T cells. Additionally, the resulting T cell clones in the memory pool may differ in absence of co-inhibitory signals. For example, without PD-1, high affinity TCRs may over-activate T cells and lead to activation-induced cell death (AICD). Similarly, lower affinity clones that may normally be below a set threshold of total activation in the presence of co-inhibition may be allowed to expand further to contribute to the memory T cell pool. Understanding of the long-term effects of engaging or blocking co-inhibitory receptors on T cells during different stages of T cell activation is important, as modulation of many of these pathways enters the clinic to be evaluated for their immunotherapeutic effects. Importantly, PD-1 has already transformed the therapy of many cancers with unprecedented effects (53, 54, 296). While some studies suggest lasting effect even when patients are off therapy (297), other reports indicate that cancer can reoccur following anti-PD-1 therapy (298). While emerging reports highlight the importance of PD-1 axis on memory T cells in cancer (299), it is thus important to further elucidate the effects that PD-1 and other co-inhibitory receptors may have at different stages of T cell differentiation on the generation, quality, and maintenance of T cell memory to understand how blockade of these receptors may affect the durability of anti-tumor immunity.

While PD-1 is now a mainstay in treating various cancers, the therapeutic potential of PD-1 modulation was observed in the context of chronic viral infection (48, 51, 134, 136, 300, 301). This is due to the similar biological process of $CD8^+$ T cell exhaustion that occurs in the context of persistent antigen exposure during a chronic infection or cancer. Similar parallels may hold true for the ectonucleotidase CD39. CD39 has a similar expression pattern as PD-1 as it is up-regulated following cell activation but becomes differentially expressed at high levels on the exhausted $CD8^+$ T cells. Additionally, the purinergic pathway has an important

role in suppressing immune responses in the tumor microenvironment (302-306). Preclinical studies inhibiting the second enzyme in the break-down process of extracellular ATP to free adenosine – CD73 – have enhanced anti-tumoral immunity (302-305). Targeting CD39 may provide similar reduction of extracellular adenosine to reduce immune suppression. However, targeting CD39 instead of CD73 may potentially provide additional benefits such as local accumulation of extracellular ATP. Through binding the P2X purinoreceptors, extracellular ATP can additionally activate the immune system and thus serve as an adjuvant in the tumor microenvironment. Targeting multiple components of the purinergic pathway or adding CD39 antagonism to established checkpoint blockade may further provide additive effects. This is particularly important since uncontrolled death of cells is abundant in tumors resulting in excessive release of extracellular ATP. This ATP accumulation in absence of CD39, together with extracellular adenosine reduction, may synergize to boost anti-tumor T cell, NK cell, macrophage, and DC responses, while also antagonizing the proliferation and survival of the tumor cells (307-309).

It is also intriguing that in our preliminary experiments, TGF-β1 acted as the strongest inducer of CD39 expression in $CD8^+$ T cells *in vitro*. TGF-β is well appreciated to play an important suppressive role in the setting of chronic viral infection (133) and tumors (310-312) where $CD8^+$ T cells become exhausted and may thus up-regulate CD39. Additionally, CD39 is up-regulated by the hypoxia inducible factor 1 alpha (HIF-1α), which is also appreciated to be induced and function in both chronic viral infection (313) as well as in tumors (314-317), providing another potential mechanism that may contribute to the high CD39 expression on $CD8^+$ T cells associated with exhaustion. Our data further suggest that exhausted $CD8^+$ T cells may be able to act on other to suppress other arms of the immune system.

As discussed previously, CD39 can be expressed on a variety of cell types such as endothelial, epithelial, as well as myeloid and lymphoid lineages. Yet its expression on $CD8^+$ T cells seems to be restricted to exhausted cells. While germline knockout mice suffer from

severe immunopathology following LCMV Clone 13 infection, it is not yet clear what the functional effect of lacking CD39 on $CD8^+$ T cells would be. One hypothesis is that CD39 on Treg cells may play an important role in protecting mice from immunopathology. However, $CD4^+$ T cells are routinely depleted in established models of LCMV Clone 13 infection without inducing mortality, suggesting that additional cells are responsible for the fatal immunopathology we observed in the CD39-KO mice. Additional experiments using conditional knockout mice, in which CD39 is only absent in $CD8^+$ T cells (E8I-Cre x $CD39^{fl/fl}$), or $CD8^+$ T cell transfers, using WT and CD39-KO P14 TCR-transgenic cells, will have to be performed to understand the functional contribution of CD39 on $CD8^+$ T cells and its cell-intrinsic effects on $CD8^+$ T cells. Thus, a further examination of the exact mechanism that governs CD39 activity in both chronic viral infection and tumor challenge may lead to novel therapies that could combat those devastating diseases.

In addition to studying the activity of previously-identified molecules, such as PD-1, BATF and CD39, we also sought to discover new regulators of $CD8^+$ T cell differentiation. To do this, we used RNAi screening to examine many potentially novel genes that function in T cells. In developing our screening platform, we considered features that would allow us to study differentiation in the most physiological setting. The complex signals that T cells receive in the various tissues in the course of activation *in vivo* are difficult to accurately mimic *in vitro*, leading us to use *in vivo* models. In addition, established approaches for perturbing genes in T cells using retroviruses all relied on a type of prior stimulation and thus initiating differentiation of T cells prior to experimentation. To circumvent these limitations we established a novel *in vivo* RNAi screening platform using IPTG-inducible shRNAs delivered to HSCs that we used to but reconstitute bone marrow chimeras, which enables us to generate un-manipulated, but genetically perturbable immune cells, including naive T cells, for *in vivo* studies.

To validate the experimental system we studied the function of BATF in $CD8^+$ T cells. While the role for BATF in $CD8^+$ T cells was first established in the context of T cell exhaustion, we have subsequently shown that it is also essential for proper effector $CD8^+$ T cell differentiation (174). This was shown using BATF-deficient P14 cells, which, when transferred together with WT P14 cells into an LCMV infected mouse, are drastically outcompeted by the WT cells. We observed a similar defect using the P14 cells with inducible shRNA targeting BATF. However, by leveraging the inducible nature of the *in vivo* shRNA experimental platform, we were able to determine that BATF is important in the initial programing of T cells but is no longer essential once the differentiation is underway. This is consistent with other reports suggesting that BATF may act as a pioneer transcription factor required to bind to genomic regions in order for other transcription factors to guide differentiation of cells (186, 318). To ensure the maximum efficiency of gene knock-down in most experiments, we exposed inducible shRNA-bearing $CD8^+$ T cells to IPTG at least 3 days prior to experimentation. This enabled accumulation of the shRNA transcripts by the time of cell activation to suppress any specific transcription induced by the activation. Using this approach, we were able to determine efficient silencing of BATF transcript *in vivo*. However, it is more challenging to assess the exact kinetics of protein down-regulation when BATF was knocked down at later time points, due to the insufficient number of cells obtainable at different time points for biochemical analyses. Thus, we used data generated *in vitro* to confirm that BATF protein is also down-regulated in addition to the mRNA. Importantly, each protein may have a different half-life, which must be taken into account when assessing the kinetics of potentially transient contributions of genes in a given biological process.

In an effort to better understand the transcriptional circuitry underlying $CD8^+$ T cell differentiation program, we used this experimental platform to systematically interrogate the potential functional role of ~80 transcriptional regulators. We identified several new candidate

genes and validated a role for TGIF1 in $CD8^+$ effector T cell differentiation. We show that TGIF1 may be interfering with the IL-2-mediated induction of Blimp-1 to decrease the formation of SLECs and enhance formation of MPECs.

This experimental platform enables generation of all immune cells with integrated shRNAs. We used it to also confirm the role of BATF in B cells where it important for proper germinal center B cell and plasma cell differentiation (181, 182). Thus, this experimental platform may be suitable for modulating a variety of types of immune cells that are refractory to viral transduction and/or are difficult to culture and transfer to study *in vivo*. This platform may thus enable systematic functional interrogation of genes in any hematopoietic–derived cell *in vivo*, such as the natural killer cells and other types of innate lymphoid cells, various subsets of dendritic cells, and others. We initially used an inducible shRNA promoter to avoid developmental effects of gene knock-down during the course of hematopoiesis. However, this approach provides the opportunity to perturb genes selectively at specific stages of a cell life cycle. For example, we can selectively perturb genes during the contraction phase of $CD8^+$ T cells to examine what genes may be enhancing the transition into memory cells. Additionally, we can allow the T cells to become memory or exhausted cells and interrogate the function of genes responsible to maintain the $CD8^+$ T cells in these states. While we predominantly used this system to examine T cell biology in the context of viral infection, the immune cells derived in this way can be used in established models to examine other immune processes such as tumor immunity and autoimmunity.

To improve our ability to extract biologically meaningful information from large immune gene expression datasets, we generated ImmuneSigDB – a collection of ~5000 gene sets corresponding to most differentially expressed genes of immune cells in various states, tissues, genotypes or after various perturbations. We used it together with a novel approach we developed, leading edge metagene (LEM) analysis, to examine mouse and human transcriptional responses to sepsis in blood. We uncovered conserved and species-

specific aspects of biology that is reflected in altered transcriptional programs triggered by sepsis. While this analysis examined transcriptomes of mouse and human cells obtained from blood of septic animals, there are additional experimental differences that may be responsible any time the biology in human patients is compared to mouse models. These include the genetic homogeneity and lack of other immune perturbations throughout the life of laboratory mice as well as the much better controlled, yet sometimes artificial, experimental systems to model human disease. LEM analysis provides a novel tool to highlight core bricks of biology underlying specific biological phenotypes and is being successfully applied with ImmuneSigDB to other studies, like examining the immune system of laboratory mice compared to pet store mice and humans (319). However, gene set collections such as ImmuneSigDB are more widely used in conventional GSEA. While we only generated this resource recently, it has already been widely adopted by the scientific community and has been successfully used to analyze various datasets and publish results (320-329).

Together, the body of work presented in this thesis advances our knowledge of major states of $CD8^+$ T cell biology, uncovering novel mechanisms underlying $CD8^+$ T cell function, as well as identifying potential novel therapeutic targets that may be transformative in creating better vaccines, treating infections, or fighting cancer.

Bibliography

1. Harty, J.T., Tvinnereim, A.R., and White, D.W. 2000. CD8+ T cell effector mechanisms in resistance to infection. *Annu Rev Immunol* 18:275-308.

2. Wong, P., and Pamer, E.G. 2003. CD8 T cell responses to infectious pathogens. *Annu Rev Immunol* 21:29-70.

3. Zhang, N., and Bevan, M.J. 2011. CD8(+) T cells: foot soldiers of the immune system. *Immunity* 35:161-168.

4. MacIver, N.J., Michalek, R.D., and Rathmell, J.C. 2013. Metabolic regulation of T lymphocytes. *Annu Rev Immunol* 31:259-283.

5. Kanno, Y., Vahedi, G., Hirahara, K., Singleton, K., and O'Shea, J.J. 2012. Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. *Annu Rev Immunol* 30:707-731.

6. Kaech, S.M., and Cui, W. 2012. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol* 12:749-761.

7. Gray, S.M., Kaech, S.M., and Staron, M.M. 2014. The interface between transcriptional and epigenetic control of effector and memory CD8(+) T-cell differentiation. *Immunol Rev* 261:157-168.

8. Acuto, O., and Michel, F. 2003. CD28-mediated co-stimulation: a quantitative support for TCR signalling. *Nat Rev Immunol* 3:939-951.

9. Chen, L., and Flies, D.B. 2013. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol* 13:227-242.

10. Curtsinger, J.M., Schmidt, C.S., Mondino, A., Lins, D.C., Kedl, R.M., Jenkins, M.K., and Mescher, M.F. 1999. Inflammatory cytokines provide a third signal for activation of naive CD4+ and CD8+ T cells. *J Immunol* 162:3256-3262.

11. Valenzuela, J., Schmidt, C., and Mescher, M. 2002. The roles of IL-12 in providing a third signal for clonal expansion of naive CD8 T cells. *J Immunol* 169:6842-6849.

12. Badovinac, V.P., Tvinnereim, A.R., and Harty, J.T. 2000. Regulation of antigen-specific CD8+ T cell homeostasis by perforin and interferon-gamma. *Science* 290:1354-1358.

13. Wherry, E.J., Teichgraber, V., Becker, T.C., Masopust, D., Kaech, S.M., Antia, R., von Andrian, U.H., and Ahmed, R. 2003. Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat Immunol* 4:225-234.

14. Khanolkar, A., Fuller, M.J., and Zajac, A.J. 2004. CD4 T cell-dependent CD8 T cell maturation. *J Immunol* 172:2834-2844.

15. Bachmann, M.F., Schwarz, K., Wolint, P., Meijerink, E., Martin, S., Manolova, V., and Oxenius, A. 2004. Cutting edge: distinct roles for T help and CD40/CD40 ligand in regulating differentiation of proliferation-competent memory CD8+ T cells. *J Immunol* 173:2217-2221.

16. Kolumam, G.A., Thomas, S., Thompson, L.J., Sprent, J., and Murali-Krishna, K. 2005. Type I interferons act directly on CD8 T cells to allow clonal expansion and memory formation in response to viral infection. *J Exp Med* 202:637-650.

17. Lang, K.S., Recher, M., Navarini, A.A., Harris, N.L., Lohning, M., Junt, T., Probst, H.C., Hengartner, H., and Zinkernagel, R.M. 2005. Inverse correlation between IL-7 receptor expression and CD8 T cell exhaustion during persistent antigen stimulation. *Eur J Immunol* 35:738-745.

18. Cui, W., Joshi, N.S., Liu, Y., Meng, H., Kleinstein, S.H., and Kaech, S.M. 2014. TLR4 Ligands Lipopolysaccharide and Monophosphoryl Lipid A Differentially Regulate Effector and Memory CD8+ T Cell Differentiation. *J Immunol* 192:4221-4232.

19. Cui, W., Liu, Y., Weinstein, J.S., Craft, J., and Kaech, S.M. 2011. An interleukin-21-interleukin-10-STAT3 pathway is critical for functional maturation of memory CD8+ T cells. *Immunity* 35:792-805.

20. Xin, A., Masson, F., Liao, Y., Preston, S., Guan, T., Gloury, R., Olshansky, M., Lin, J.X., Li, P., Speed, T.P., et al. 2016. A molecular threshold for effector CD8 T cell differentiation controlled by transcription factors Blimp-1 and T-bet. *Nat Immunol*.

21. Wherry, E.J. 2011. T cell exhaustion. *Nat Immunol* 12:492-499.

22. Wherry, E.J., and Ahmed, R. 2004. Memory CD8 T-cell differentiation during viral infection. *J Virol* 78:5535-5545.

23. 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* 27:670-684.

24. Wherry, E.J., and Kurachi, M. 2015. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* 15:486-499.

25. Cui, W., and Kaech, S.M. 2010. Generation of effector CD8+ T cells and their conversion to memory T cells. *Immunol Rev* 236:151-166.

26. Barber, D.L., Wherry, E.J., and Ahmed, R. 2003. Cutting edge: rapid in vivo killing by memory CD8 T cells. *J Immunol* 171:27-31.

27. Laidlaw, B.J., Craft, J.E., and Kaech, S.M. 2016. The multifaceted role of CD4(+) T cells in CD8(+) T cell memory. *Nat Rev Immunol* 16:102-111.

28. Weninger, W., Crowley, M.A., Manjunath, N., and von Andrian, U.H. 2001. Migratory properties of naive, effector, and memory CD8(+) T cells. *J Exp Med* 194:953-966.

29. Weninger, W., Manjunath, N., and von Andrian, U.H. 2002. Migration and differentiation of CD8+ T cells. *Immunol Rev* 186:221-233.

30. Luckey, C.J., Bhattacharya, D., Goldrath, A.W., Weissman, I.L., Benoist, C., and Mathis, D. 2006. Memory T and memory B cells share a transcriptional program of self-renewal with long-term hematopoietic stem cells. *Proc Natl Acad Sci U S A* 103:3304-3309.

31. Williams, M.A., Holmes, B.J., Sun, J.C., and Bevan, M.J. 2006. Developing and maintaining protective CD8+ memory T cells. *Immunol Rev* 211:146-153.

32. Williams, M.A., and Bevan, M.J. 2007. Effector and memory CTL differentiation. *Annu Rev Immunol* 25:171-192.

33. van Stipdonk, M.J., Lemmens, E.E., and Schoenberger, S.P. 2001. Naïve CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nature Immunology* 2:423-429.

34. Bevan, M.J., and Fink, P.J. 2001. The CD8 response on autopilot. *Nat Immunol* 2:381-382.

35. Kaech, S.M., and Ahmed, R. 2001. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naïve cells. *Nature Immunology* 2:415-422.

36. Hand, T.W., and Kaech, S.M. 2009. Intrinsic and extrinsic control of effector T cell survival and memory T cell development. *Immunol Res* 45:46-61.

37. Joshi, N.S., and Kaech, S.M. 2008. Effector CD8 T cell development: a balancing act between memory cell potential and terminal differentiation. *J Immunol* 180:1309-1315.

38. Parish, I.A., and Kaech, S.M. 2009. Diversity in CD8(+) T cell differentiation. *Curr Opin Immunol* 21:291-297.

39. Joshi, N.S., Cui, W., Chandele, A., Lee, H.K., Urso, D.R., Hagman, J., Gapin, L., and Kaech, S.M. 2007. Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. *Immunity* 27:281-295.

40. Sarkar, S., Kalia, V., Haining, W.N., Konieczny, B.T., Subramaniam, S., and Ahmed, R. 2008. Functional and genomic profiling of effector CD8 T cell subsets with distinct memory fates. *J Exp Med* 205:625-640.

41. Huster, K.M., Busch, V., Schiemann, M., Linkemann, K., Kerksiek, K.M., Wagner, H., and Busch, D.H. 2004. Selective expression of IL-7 receptor on memory T cells identifies early CD40L-dependent generation of distinct CD8+ memory T cell subsets. *Proc Natl Acad Sci U S A* 101:5610-5615.

42. Kaech, S.M., Tan, J.T., Wherry, E.J., Konieczny, B.T., Surh, C.D., and Ahmed, R. 2003. Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat Immunol* 4:1191-1198.

43. Kalia, V., Sarkar, S., Gourley, T.S., Rouse, B.T., and Ahmed, R. 2006. Differentiation of memory B and T cells. *Curr Opin Immunol* 18:255-264.

44. Yuzefpolskiy, Y., Baumann, F.M., Kalia, V., and Sarkar, S. 2015. Early CD8 T-cell memory precursors and terminal effectors exhibit equipotent in vivo degranulation. *Cell Mol Immunol* 12:400-408.

45. Pipkin, M.E., Sacks, J.A., Cruz-Guilloty, F., Lichtenheld, M.G., Bevan, M.J., and Rao, A. 2010. Interleukin-2 and Inflammation Induce Distinct Transcriptional Programs that Promote the Differentiation of Effector Cytolytic T Cells. *Immunity* 32:79-90.

46. Kalia, V., Sarkar, S., Subramaniam, S., Haining, W.N., Smith, K.A., and Ahmed, R. 2010. Prolonged interleukin-2Ralpha expression on virus-specific CD8+ T cells favors terminal-effector differentiation in vivo. *Immunity* 32:91-103.

47. Belz, G.T., and Masson, F. 2010. Interleukin-2 tickles T cell memory. *Immunity* 32:7-9.

48. 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* 439:682-687.

49. Paley, M.A., Kroy, D.C., Odorizzi, P.M., Johnnidis, J.B., Dolfi, D.V., Barnett, B.E., Bikoff, E.K., Robertson, E.J., Lauer, G.M., Reiner, S.L., et al. 2012. Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science* 338:1220-1225.

50. Pauken, K.E., and Wherry, E.J. 2015. SnapShot: T Cell Exhaustion. *Cell* 163:1038-1038 e1031.

51. Day, C.L., Kaufmann, D.E., Kiepiela, P., Brown, J.A., Moodley, E.S., Reddy, S., Mackey, E.W., Miller, J.D., Leslie, A.J., DePierres, C., et al. 2006. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* 443:350-354.

52. Fuller, M.J., Callendret, B., Zhu, B., Freeman, G.J., Hasselschwert, D.L., Satterfield, W., Sharpe, A.H., Dustin, L.B., Rice, C.M., Grakoui, A., et al. 2013. Immunotherapy of chronic hepatitis C virus infection with antibodies against programmed cell death-1 (PD-1). *Proc Natl Acad Sci U S A* 110:15001-15006.

53. Brahmer, J.R., Tykodi, S.S., Chow, L.Q., Hwu, W.J., Topalian, S.L., Hwu, P., Drake, C.G., Camacho, L.H., Kauh, J., Odunsi, K., et al. 2012. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* 366:2455-2465.

54. Topalian, S.L., Hodi, F.S., Brahmer, J.R., Gettinger, S.N., Smith, D.C., McDermott, D.F., Powderly, J.D., Carvajal, R.D., Sosman, J.A., Atkins, M.B., et al. 2012. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 366:2443-2454.

55. Brown, K.E., Freeman, G.J., Wherry, E.J., and Sharpe, A.H. 2010. Role of PD-1 in regulating acute infections. *Curr Opin Immunol* 22:397-401.

56. Majumdar, S., and Aggarwal, B.B. 2003. Adenosine suppresses activation of nuclear factor-kappaB selectively induced by tumor necrosis factor in different cell types. *Oncogene* 22:1206-1218.

57. Ernst, P.B., Garrison, J.C., and Thompson, L.F. 2010. Much ado about adenosine: adenosine synthesis and function in regulatory T cell biology. *J Immunol* 185:1993-1998.

58. Whiteside, T.L., Mandapathil, M., and Schuler, P. 2011. The role of the adenosinergic pathway in immunosuppression mediated by human regulatory T cells (Treg). *Curr Med Chem* 18:5217-5223.

59. Whiteside, T.L. 2014. Regulatory T cell subsets in human cancer: are they regulating for or against tumor progression? *Cancer Immunol Immunother* 63:67-72.

60. Junger, W.G. 2011. Immune cell regulation by autocrine purinergic signalling. *Nat Rev Immunol* 11:201-212.

61. Deaglio, S., Dwyer, K.M., Gao, W., Friedman, D., Usheva, A., Erat, A., Chen, J.F., Enjyoji, K., Linden, J., Oukka, M., et al. 2007. Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J Exp Med* 204:1257-1265.

62. Dwyer, K.M., Deaglio, S., Gao, W., Friedman, D., Strom, T.B., and Robson, S.C. 2007. CD39 and control of cellular immune responses. *Purinergic Signal* 3:171-180.

63. Kansas, G.S., Wood, G.S., and Tedder, T.F. 1991. Expression, distribution, and biochemistry of human CD39. Role in activation-associated homotypic adhesion of lymphocytes. *J Immunol* 146:2235-2244.

64. Moncrieffe, H., Nistala, K., Kamhieh, Y., Evans, J., Eddaoudi, A., Eaton, S., and Wedderburn, L.R. 2010. High Expression of the Ectonucleotidase CD39 on T Cells from the Inflamed Site Identifies Two Distinct Populations, One Regulatory and One Memory T Cell Population. In *The Journal of Immunology*. 134-143.

65. Pulte, D., Furman, R.R., Broekman, M.J., Drosopoulos, J.H., Ballard, H.S., Olson, K.E., Kizer, J.R., and Marcus, A.J. 2011. CD39 expression on T lymphocytes correlates with severity of disease in patients with chronic lymphocytic leukemia. *Clin Lymphoma Myeloma Leuk* 11:367-372.

66. Boer, M.C., van Meijgaarden, K.E., Bastid, J., Ottenhoff, T.H., and Joosten, S.A. 2013. CD39 is involved in mediating suppression by Mycobacterium bovis BCG-activated human CD8(+) CD39(+) regulatory T cells. *Eur J Immunol* 43:1925-1932.

67. Matloubian, M., Somasundaram, T., Kolhekar, S.R., Selvakumar, R., and Ahmed, R. 1990. Genetic basis of viral persistence: single amino acid change in the viral glycoprotein affects ability of lymphocytic choriomeningitis virus to persist in adult mice. *J Exp Med* 172:1043-1048.

68. 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* 67:7340-7349.

69. Wherry, E.J., Blattman, J.N., Murali-Krishna, K., van der 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* 77:4911-4927.

70. Heery, D.M., Pierrat, B., Gronemeyer, H., Chambon, P., and Losson, R. 1994. Homo- and heterodimers of the retinoid X receptor (RXR) activated transcription in yeast. *Nucleic Acids Res* 22:726-731.

71. Kurokawa, R., DiRenzo, J., Boehm, M., Sugarman, J., Gloss, B., Rosenfeld, M.G., Heyman, R.A., and Glass, C.K. 1994. Regulation of retinoid signalling by receptor polarity and allosteric control of ligand binding. *Nature* 371:528-531.

72. Bertolino, E., Reimund, B., Wildt-Perinic, D., and Clerc, R.G. 1995. A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem* 270:31178-31188.

73. Boulet, S., Daudelin, J.F., and Labrecque, N. 2014. IL-2 induction of Blimp-1 is a key in vivo signal for CD8+ short-lived effector T cell differentiation. *J Immunol* 193:1847-1854.

74. Rutishauser, R.L., Martins, G.A., Kalachikov, S., Chandele, A., Parish, I.A., Meffre, E., Jacob, J., Calame, K., and Kaech, S.M. 2009. Transcriptional repressor Blimp-1 promotes CD8(+) T cell terminal differentiation and represses the acquisition of central memory T cell properties. *Immunity* 31:296-308.

75. Doering, T.A., Crawford, A., Angelosanto, J.M., Paley, M.A., Ziegler, C.G., and Wherry, E.J. 2012. Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory. *Immunity* 37:1130-1144.

76. Crawford, A., Angelosanto, J.M., Kao, C., Doering, T.A., Odorizzi, P.M., Barnett, B.E., and Wherry, E.J. 2014. Molecular and transcriptional basis of CD4(+) T cell dysfunction during chronic infection. *Immunity* 40:289-302.

77. Haining, W.N., and Wherry, E.J. 2010. Integrating genomic signatures for immunologic discovery. *Immunity* 32:152-161.

78. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550.

79. Seok, J., Warren, H.S., Cuenca, A.G., Mindrinos, M.N., Baker, H.V., Xu, W., Richards, D.R., McDonald-Smith, G.P., Gao, H., Hennessy, L., et al. 2013. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* 110:3507-3512.

80. Takao, K., and Miyakawa, T. 2014. Genomic responses in mouse models greatly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A*.

81. Francisco, L.M., Sage, P.T., and Sharpe, A.H. 2010. The PD-1 pathway in tolerance and autoimmunity. *Immunol Rev* 236:219-242.

82. Pardoll, D.M. 2012. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 12:252-264.

83. Page, D.B., Postow, M.A., Callahan, M.K., Allison, J.P., and Wolchok, J.D. 2014. Immune modulation in cancer with antibodies. *Annu Rev Med* 65:185-202.

84. Frebel, H., Nindl, V., Schuepbach, R.A., Braunschweiler, T., Richter, K., Vogel, J., Wagner, C.A., Loffing-Cueni, D., Kurrer, M., Ludewig, B., et al. 2012. Programmed

death 1 protects from fatal circulatory failure during systemic virus infection of mice. *J Exp Med* 209:2485-2499.

85. Allie, S.R., Zhang, W., Fuse, S., and Usherwood, E.J. 2011. Programmed death 1 regulates development of central memory CD8 T cells after acute viral infection. *J Immunol* 186:6280-6286.

86. Fuse, S., Tsai, C.Y., Molloy, M.J., Allie, S.R., Zhang, W., Yagita, H., and Usherwood, E.J. 2009. Recall responses by helpless memory CD8+ T cells are restricted by the up-regulation of PD-1. *J Immunol* 182:4244-4254.

87. Xu, D., Fu, H.H., Obar, J.J., Park, J.J., Tamada, K., Yagita, H., and Lefrancois, L. 2013. A potential new pathway for PD-L1 costimulation of the CD8-T cell response to Listeria monocytogenes infection. *PLoS One* 8:e56539.

88. Rowe, J.H., Johanns, T.M., Ertelt, J.M., and Way, S.S. 2008. PDL-1 blockade impedes T cell expansion and protective immunity primed by attenuated Listeria monocytogenes. *J Immunol* 180:7553-7557.

89. Talay, O., Shen, C.H., Chen, L., and Chen, J. 2009. B7-H1 (PD-L1) on T cells is required for T-cell-mediated conditioning of dendritic cell maturation. *Proc Natl Acad Sci U S A* 106:2741-2746.

90. Yao, S., Wang, S., Zhu, Y., Luo, L., Zhu, G., Flies, S., Xu, H., Ruff, W., Broadwater, M., Choi, I.H., et al. 2009. PD-1 on dendritic cells impedes innate immunity against bacterial infection. *Blood* 113:5811-5818.

91. Erickson, J.J., Gilchuk, P., Hastings, A.K., Tollefson, S.J., Johnson, M., Downing, M.B., Boyd, K.L., Johnson, J.E., Kim, A.S., Joyce, S., et al. 2012. Viral acute lower respiratory infections impair CD8+ T cells through PD-1. *J Clin Invest* 122:2967-2982.

92. Decman, V., Laidlaw, B.J., Dimenna, L.J., Abdulla, S., Mozdzanowska, K., Erikson, J., Ertl, H.C., and Wherry, E.J. 2010. Cell-intrinsic defects in the proliferative response of antiviral memory CD8 T cells in aged mice upon secondary infection. *J Immunol* 184:5151-5159.

93. Mueller, S.N., Langley, W.A., Li, G., Garcia-Sastre, A., Webby, R.J., and Ahmed, R. 2010. Qualitatively different memory CD8+ T cells are generated after lymphocytic choriomeningitis virus and influenza virus infections. *J Immunol* 185:2182-2190.

94. Flynn, K.J., Belz, G.T., Altman, J.D., Ahmed, R., Woodland, D.L., and Doherty, P.C. 1998. Virus-specific CD8+ T cells in primary and secondary influenza pneumonia. *Immunity* 8:683-691.

95. Liang, S., Mozdzanowska, K., Palladino, G., and Gerhard, W. 1994. Heterosubtypic immunity to influenza type A virus in mice. Effector mechanisms and their longevity. *J Immunol* 152:1653-1661.

96. Francisco, L.M., Salinas, V.H., Brown, K.E., Vanguri, V.K., Freeman, G.J., Kuchroo, V.K., and Sharpe, A.H. 2009. PD-L1 regulates the development, maintenance, and function of induced regulatory T cells. *J Exp Med* 206:3015-3029.

97. Chow, L.M., and Baker, S.J. 2006. PTEN function in normal and neoplastic growth. *Cancer Lett* 241:184-196.

98. Araki, K., Turner, A.P., Shaffer, V.O., Gangappa, S., Keller, S.A., Bachmann, M.F., Larsen, C.P., and Ahmed, R. 2009. mTOR regulates memory CD8 T-cell differentiation. *Nature* 460:108-112.

99. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25-29.

100. Xiao, Y., Yu, S., Zhu, B., Bedoret, D., Bu, X., Francisco, L.M., Hua, P., Duke-Cohan, J.S., Umetsu, D.T., Sharpe, A.H., et al. 2014. RGMb is a novel binding partner for PD-L2 and its engagement with PD-L2 promotes respiratory tolerance. *J Exp Med*.

101. Butte, M.J., Keir, M.E., Phamduy, T.B., Sharpe, A.H., and Freeman, G.J. 2007. Programmed death-1 ligand 1 interacts specifically with the B7-1 costimulatory molecule to inhibit T cell responses. *Immunity* 27:111-122.

102. Yokosuka, T., Takamatsu, M., Kobayashi-Imanishi, W., Hashimoto-Tane, A., Azuma, M., and Saito, T. 2012. Programmed cell death 1 forms negative costimulatory microclusters that directly inhibit T cell receptor signaling by recruiting phosphatase SHP2. *J Exp Med* 209:1201-1217.

103. Zinselmeyer, B.H., Heydari, S., Sacristan, C., Nayak, D., Cammer, M., Herz, J., Cheng, X., Davis, S.J., Dustin, M.L., and McGavern, D.B. 2013. PD-1 promotes immune exhaustion by inducing antiviral T cell motility paralysis. *J Exp Med* 210:757-774.

104. Freeman, G.J., Long, A.J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., Fitz, L.J., Malenkovich, N., Okazaki, T., Byrne, M.C., et al. 2000. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J Exp Med* 192:1027-1034.

105. 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. *Mol Cell Biol* 25:9543-9553.

106. Latchman, Y., Wood, C.R., Chernova, T., Chaudhary, D., Borde, M., Chernova, I., Iwai, Y., Long, A.J., Brown, J.A., Nunes, R., et al. 2001. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. *Nat Immunol* 2:261-268.

107. Patsoukis, N., Brown, J., Petkova, V., Liu, F., Li, L., and Boussiotis, V.A. 2012. Selective effects of PD-1 on Akt and Ras pathways regulate molecular components of the cell cycle and inhibit T cell proliferation. *Sci Signal* 5:ra46.

108. Keir, M.E., Freeman, G.J., and Sharpe, A.H. 2007. PD-1 regulates self-reactive CD8+ T cell responses to antigen in lymph nodes and tissues. *J Immunol* 179:5064-5070.

109. Keir, M.E., Liang, S.C., Guleria, I., Latchman, Y.E., Qipo, A., Albacker, L.A., Koulmanda, M., Freeman, G.J., Sayegh, M.H., and Sharpe, A.H. 2006. Tissue expression of PD-L1 mediates peripheral T cell tolerance. *J Exp Med* 203:883-895.

110. Pircher, H., Burki, K., Lang, R., Hengartner, H., and Zinkernagel, R.M. 1989. Tolerance induction in double specific T-cell receptor transgenic mice varies with antigen. *Nature* 342:559-561.

111. Laidlaw, B.J., Decman, V., Ali, M.A., Abt, M.C., Wolf, A.I., Monticelli, L.A., Mozdzanowska, K., Angelosanto, J.M., Artis, D., Erikson, J., et al. 2013. Cooperativity between CD8+ T cells, non-neutralizing antibodies, and alveolar macrophages is important for heterosubtypic influenza virus immunity. *PLoS Pathog* 9:e1003207.

112. Blazar, B.R., Taylor, P.A., Panoskaltsis-Mortari, A., and Vallera, D.A. 1998. Rapamycin inhibits the generation of graft-versus-host disease- and graft-versus-leukemia-causing T cells by interfering with the production of Th1 or Th1 cytotoxic cytokines. *J Immunol* 160:5355-5365.

113. Lechner, F., Wong, D.K., Dunbar, P.R., Chapman, R., Chung, R.T., Dohrenwend, P., Robbins, G., Phillips, R., Klenerman, P., and Walker, B.D. 2000. Analysis of successful immune responses in persons infected with hepatitis C virus. *J Exp Med* 191:1499-1512.

114. Kroy, D.C., Ciuffreda, D., Cooperrider, J.H., Tomlinson, M., Hauck, G.D., Aneja, J., Berger, C., Wolski, D., Carrington, M., Wherry, E.J., et al. 2014. Liver environment and HCV replication affect human T-cell phenotype and expression of inhibitory receptors. *Gastroenterology* 146:550-561.

115. Duraiswamy, J., Ibegbu, C.C., Masopust, D., Miller, J.D., Araki, K., Doho, G.H., Tata, P., Gupta, S., Zilliox, M.J., Nakaya, H.I., et al. 2011. Phenotype, function, and gene expression profiles of programmed death-1(hi) CD8 T cells in healthy human adults. *J Immunol* 186:4200-4212.

116. Buggert, M., Tauriainen, J., Yamamoto, T., Frederiksen, J., Ivarsson, M.A., Michaelsson, J., Lund, O., Hejdeman, B., Jansson, M., Sonnerborg, A., et al. 2014. T-bet and Eomes are differentially linked to the exhausted phenotype of CD8+ T cells in HIV infection. *PLoS Pathog* 10:e1004251.

117. Kurktschiev, P.D., Raziorrouh, B., Schraut, W., Backmund, M., Wachtler, M., Wendtner, C.M., Bengsch, B., Thimme, R., Denk, G., Zachoval, R., et al. 2014. Dysfunctional CD8+ T cells in hepatitis B and C are characterized by a lack of antigen-specific T-bet induction. *J Exp Med* 211:2047-2059.

118. Blackburn, S.D., Shin, H., Freeman, G.J., and Wherry, E.J. 2008. Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. *Proc Natl Acad Sci U S A* 105:15016-15021.

119. Kaczmarek, E., Koziak, K., Sevigny, J., Siegel, J.B., Anrather, J., Beaudoin, A.R., Bach, F.H., and Robson, S.C. 1996. Identification and characterization of CD39/vascular ATP diphosphohydrolase. *J Biol Chem* 271:33116-33122.

120. Borsellino, G., Kleinewietfeld, M., Di Mitri, D., Sternjak, A., Diamantini, A., Giometto, R., Hopner, S., Centonze, D., Bernardi, G., Dell'Acqua, M.L., et al. 2007. Expression of ectonucleotidase CD39 by Foxp3+ Treg cells: hydrolysis of extracellular ATP and immune suppression. *Blood* 110:1225-1232.

121. Junger, W.G. 2011. Immune cell regulation by autocrine purinergic signalling. In *Nature Publishing Group*: Nature Publishing Group. 201-212.

122. Zarek, P.E., Huang, C.T., Lutz, E.R., Kowalski, J., Horton, M.R., Linden, J., Drake, C.G., and Powell, J.D. 2008. A2A receptor signaling promotes peripheral tolerance by inducing T-cell anergy and the generation of adaptive regulatory T cells. *Blood* 111:251-259.

123. Huang, S., Apasov, S., Koshiba, M., and Sitkovsky, M. 1997. Role of A2a extracellular adenosine receptor-mediated signaling in adenosine-mediated inhibition of T-cell activation and expansion. *Blood* 90:1600-1610.

124. Lokshin, A., Raskovalova, T., Huang, X., Zacharia, L.C., Jackson, E.K., and Gorelik, E. 2006. Adenosine-mediated inhibition of the cytotoxic activity and cytokine production by activated natural killer cells. *Cancer Res* 66:7758-7765.

125. Robson, S.C., Kaczmarek, E., Siegel, J.B., Candinas, D., Koziak, K., Millan, M., Hancock, W.W., and Bach, F.H. 1997. Loss of ATP diphosphohydrolase activity with endothelial cell activation. *J Exp Med* 185:153-163.

126. Papanikolaou, A., Papafotika, A., Murphy, C., Papamarcaki, T., Tsolas, O., Drab, M., Kurzchalia, T.V., Kasper, M., and Christoforidis, S. 2005. Cholesterol-dependent lipid assemblies regulate the activity of the ecto-nucleotidase CD39. *J Biol Chem* 280:26406-26414.

127. Wu, Y., Sun, X., Kaczmarek, E., Dwyer, K.M., Bianchi, E., Usheva, A., and Robson, S.C. 2006. RanBPM associates with CD39 and modulates ecto-nucleotidase activity. *Biochem J* 396:23-30.

128. Kasprowicz, V., Schulze Zur Wiesch, J., Kuntzen, T., Nolan, B.E., Longworth, S., Berical, A., Blum, J., McMahon, C., Reyor, L.L., Elias, N., et al. 2008. High level of PD-1 expression on hepatitis C virus (HCV)-specific CD8+ and CD4+ T cells during acute HCV infection, irrespective of clinical outcome. *J Virol* 82:3154-3160.

129. Quigley, M., Pereyra, F., Nilsson, B., Porichis, F., Fonseca, C., Eichbaum, Q., Julg, B., Jesneck, J.L., Brosnahan, K., Imam, S., et al. 2010. Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF. *Nat Med* 16:1147-1151.

130. Baitsch, L., Baumgaertner, P., Devevre, E., Raghav, S.K., Legat, A., Barba, L., Wieckowski, S., Bouzourene, H., Deplancke, B., Romero, P., et al. 2011. Exhaustion of tumor-specific CD8(+) T cells in metastases from melanoma patients. *J Clin Invest* 121:2350-2360.

131. Blackburn, S.D., Shin, H., Haining, W.N., Zou, T., Workman, C.J., Polley, A., Betts, M.R., Freeman, G.J., Vignali, D.A., and Wherry, E.J. 2009. Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat Immunol* 10:29-37.

132. Enjyoji, K., Sevigny, J., Lin, Y., Frenette, P.S., Christie, P.D., Esch, J.S., 2nd, Imai, M., Edelberg, J.M., Rayburn, H., Lech, M., et al. 1999. Targeted disruption of cd39/ATP diphosphohydrolase results in disordered hemostasis and thromboregulation. *Nat Med* 5:1010-1017.

133. Tinoco, R., Alcalde, V., Yang, Y., Sauer, K., and Zuniga, E.I. 2009. Cell-intrinsic transforming growth factor-beta signaling mediates virus-specific CD8+ T cell deletion and viral persistence in vivo. *Immunity* 31:145-157.

134. Trautmann, L., Janbazian, L., Chomont, N., Said, E.A., Gimmig, S., Bessette, B., Boulassel, M.R., Delwart, E., Sepulveda, H., Balderas, R.S., et al. 2006. Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. *Nat Med* 12:1198-1202.

135. Jenabian, M.A., Seddiki, N., Yatim, A., Carriere, M., Hulin, A., Younas, M., Ghadimi, E., Kok, A., Routy, J.P., Tremblay, A., et al. 2013. Regulatory T cells negatively affect IL-2 production of effector T cells through CD39/adenosine pathway in HIV infection. *PLoS Pathog* 9:e1003319.

136. Petrovas, C., Casazza, J.P., Brenchley, J.M., Price, D.A., Gostick, E., Adams, W.C., Precopio, M.L., Schacker, T., Roederer, M., Douek, D.C., et al. 2006. PD-1 is a regulator of virus-specific CD8+ T cell survival in HIV infection. *J Exp Med* 203:2281-2292.

137. Pita-Lopez, M.L., Gayoso, I., DelaRosa, O., Casado, J.G., Alonso, C., Munoz-Gomariz, E., Tarazona, R., and Solana, R. 2009. Effect of ageing on CMV-specific CD8 T cells from CMV seropositive healthy donors. *Immun Ageing* 6:11.

138. Rey, J., Giustiniani, J., Mallet, F., Schiavon, V., Boumsell, L., Bensussan, A., Olive, D., and Costello, R.T. 2006. The co-expression of 2B4 (CD244) and CD160 delineates a subpopulation of human CD8+ T cells with a potent CD160-mediated cytolytic effector function. *Eur J Immunol* 36:2359-2366.

139. Shin, H., Blackburn, S.D., Intlekofer, A.M., Kao, C., Angelosanto, J.M., Reiner, S.L., and Wherry, E.J. 2009. A role for the transcriptional repressor Blimp-1 in CD8(+) T cell exhaustion during chronic viral infection. *Immunity* 31:309-320.

140. Migueles, S.A., Laborico, A.C., Shupert, W.L., Sabbaghian, M.S., Rabin, R., Hallahan, C.W., Van Baarle, D., Kostense, S., Miedema, F., McLaughlin, M., et al. 2002. HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. *Nat Immunol* 3:1061-1068.

141. Shin, H., Blackburn, S.D., Blattman, J.N., and Wherry, E.J. 2007. Viral antigen and extensive division maintain virus-specific CD8 T cells during chronic infection. *J Exp Med* 204:941-949.

142. Cox, A.L., Mosbruger, T., Lauer, G.M., Pardoll, D., Thomas, D.L., and Ray, S.C. 2005. Comprehensive analyses of CD8+ T cell responses during longitudinal study of acute human hepatitis C. *Hepatology* 42:104-112.

143. Zajac, A.J., Blattman, J.N., Murali-Krishna, K., Sourdive, D.J., Suresh, M., Altman, J.D., and Ahmed, R. 1998. Viral immune evasion due to persistence of activated T cells without effector function. *J Exp Med* 188:2205-2213.

144. 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* 362:758-761.

145. Seddiki, N., Brezar, V., and Draenert, R. 2014. Cell exhaustion in HIV-1 infection: role of suppressor cells. *Curr Opin HIV AIDS* 9:452-458.

146. Toth, I., Le, A.Q., Hartjen, P., Thomssen, A., Matzat, V., Lehmann, C., Scheurich, C., Beisel, C., Busch, P., Degen, O., et al. 2013. Decreased frequency of CD73+CD8+ T cells of HIV-infected patients correlates with immune activation and T cell exhaustion. *J Leukoc Biol* 94:551-561.

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

148. 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. *J Exp Med* 160:521-540.

149. Leisner, C., Loeth, N., Lamberth, K., Justesen, S., Sylvester-Hvid, C., Schmidt, E.G., Claesson, M., Buus, S., and Stryhn, A. 2008. One-pot, mix-and-read peptide-MHC tetramers. *PLoS One* 3:e1678.

150. Lazarowski, E.R., Tarran, R., Grubb, B.R., van Heusden, C.A., Okada, S., and Boucher, R.C. 2004. Nucleotide release provides a mechanism for airway surface liquid homeostasis. *J Biol Chem* 279:36855-36864.

151. Chen, Y., Corriden, R., Inoue, Y., Yip, L., Hashiguchi, N., Zinkernagel, A., Nizet, V., Insel, P.A., and Junger, W.G. 2006. ATP release guides neutrophil chemotaxis via P2Y2 and A3 receptors. *Science* 314:1792-1795.

152. Haining, W.N., Ebert, B.L., Subrmanian, A., Wherry, E.J., Eichbaum, Q., Evans, J.W., Mak, R., Rivoli, S., Pretz, J., Angelosanto, J., et al. 2008. Identification of an evolutionarily conserved transcriptional signature of CD8 memory differentiation that is shared by T and B cells. *J Immunol* 181:1859-1868.

153. Monti S, T.P., Mesirov J, Golub T. 2003. Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. *Machine Learning* 52:91-118.

154. Liberzon, A. 2014. A description of the Molecular Signatures Database (MSigDB) Web site. *Methods Mol Biol* 1150:153-160.

155. Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5:e13984.

156. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pages, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. 2006. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 313:1960-1964.

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

158. Joshi, N.S., Cui, W., Chandele, A., Lee, H.K., Urso, D.R., Hagman, J., Gapin, L., and Kaech, S.M. 2007. Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. *Immunity* 27:281-295.

159. Nakaya, H.I., Wrammert, J., Lee, E.K., Racioppi, L., Marie-Kunze, S., Haining, W.N., Means, A.R., Kasturi, S.P., Khan, N., Li, G.M., et al. 2011. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* 12:786-795.

160. Stary, G., Olive, A., Radovic-Moreno, A.F., Gondek, D., Alvarez, D., Basto, P.A., Perro, M., Vrbanac, V.D., Tager, A.M., Shi, J., et al. 2015. VACCINES. A mucosal vaccine against Chlamydia trachomatis generates two waves of protective memory T cells. *Science* 348:aaa8205.

161. Appay, V., Douek, D.C., and Price, D.A. 2008. CD8+ T cell efficacy in vaccination and disease. *Nat Med* 14:623-628.

162. Koup, R.A., and Douek, D.C. 2011. Vaccine design for CD8 T lymphocyte responses. *Cold Spring Harb Perspect Med* 1:a007252.

163. Kaech, S.M., Hemby, S., Kersh, E., and Ahmed, R. 2002. Molecular and functional profiling of memory CD8 T cell differentiation. *Cell* 111:837-851.

164. Mercado, R., Vijh, S., Allen, S.E., Kerksiek, K., Pilip, I.M., and Pamer, E.G. 2000. Early programming of T cell populations responding to bacterial infection. *Journal of immunology (Baltimore, Md : 1950)* 165:6833-6839.

165. van Stipdonk, M.J.B., Hardenberg, G., Bijker, M.S., Lemmens, E.E., Droin, N.M., Green, D.R., and Schoenberger, S.P. 2003. Dynamic programming of CD8+ T lymphocyte responses. *Nature Immunology* 4:361-365.

166. Wong, P., and Pamer, E.G. 2001. Cutting edge: antigen-independent CD8 T cell proliferation. *Journal of immunology (Baltimore, Md : 1950)* 166:5864-5868.

167. Chang, J.T., Palanivel, V.R., Kinjyo, I., Schambach, F., Intlekofer, A.M., Banerjee, A., Longworth, S.A., Vinup, K.E., Mrass, P., Oliaro, J., et al. 2007. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* 315:1687-1691.

168. Intlekofer, A.M., Takemoto, N., Wherry, E.J., Longworth, S.A., Northrup, J.T., Palanivel, V.R., Mullen, A.C., Gasink, C.R., Kaech, S.M., Miller, J.D., et al. 2005. Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin. *Nature Immunology* 6:1236-1244.

169. Pearce, E.L., Mullen, A.C., Martins, G.A., Krawczyk, C.M., Hutchins, A.S., Zediak, V.P., Banica, M., DiCioccio, C.B., Gross, D.A., Mao, C.-A., et al. 2003. Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science (New York, NY)* 302:1041-1043.

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

171. Shin, H., Blackburn, S.D., Intlekofer, A.M., Kao, C., Angelosanto, J.M., Reiner, S.L., and Wherry, E.J. 2009. A Role for the Transcriptional Repressor Blimp-1 in CD8+ T Cell Exhaustion during Chronic Viral Infection. *Immunity* 31:309-320.

172. Cannarile, M.A., Lind, N.A., Rivera, R., Sheridan, A.D., Camfield, K.A., Wu, B.B., Cheung, K.P., Ding, Z., and Goldrath, A.W. 2006. Transcriptional regulator Id2 mediates CD8+ T cell immunity. *Nature Immunology* 7:1317-1325.

173. Cruz-Guilloty, F., Pipkin, M.E., Djuretic, I.M., Levanon, D., Lotem, J., Lichtenheld, M.G., Groner, Y., and Rao, A. 2009. Runx3 and T-box proteins cooperate to establish the transcriptional program of effector CTLs. *Journal of Experimental Medicine* 206:51-59.

174. Kurachi, M., Barnitz, R.A., Yosef, N., Odorizzi, P.M., Diiorio, M.A., Lemieux, M.E., Yates, K., Godec, J., Klatt, M.G., Regev, A., et al. 2014. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8(+) T cells. *Nat Immunol* 15:373-383.

175. Amit, I., Garber, M., Chevrier, N., Leite, A.P., Donner, Y., Eisenhaure, T., Guttman, M., Grenier, J.K., Li, W., Zuk, O., et al. 2009. Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326:257-263.

176. Yang, L., Boldin, M.P., Yu, Y., Liu, C.S., Ea, C.K., Ramakrishnan, P., Taganov, K.D., Zhao, J.L., and Baltimore, D. 2012. miR-146a controls the resolution of T cell responses in mice. *J Exp Med* 209:1655-1670.

177. Araki, K., Turner, A.P., Shaffer, V.O., Gangappa, S., Keller, S.A., Bachmann, M.F., Larsen, C.P., and Ahmed, R. 2009. mTOR regulates memory CD8 T-cell differentiation. *Nature* 460:108-112.

178. Zhou, P., Shaffer, D.R., Alvarez Arias, D.A., Nakazaki, Y., Pos, W., Torres, A.J., Cremasco, V., Dougan, S.K., Cowley, G.S., Elpek, K., et al. 2014. In vivo discovery of immunotherapy targets in the tumour microenvironment. *Nature* 506:52-57.

179. Chen, R., Belanger, S., Frederick, M.A., Li, B., Johnston, R.J., Xiao, N., Liu, Y.C., Sharma, S., Peters, B., Rao, A., et al. 2014. In vivo RNA interference screens identify regulators of antiviral CD4(+) and CD8(+) T cell differentiation. *Immunity* 41:325-338.

180. Jackson, A.L., Burchard, J., Schelter, J., Chau, B.N., Cleary, M., Lim, L., and Linsley, P.S. 2006. Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 12:1179-1187.

181. Ise, W., Kohyama, M., Schraml, B.U., Zhang, T., Schwer, B., Basu, U., Alt, F.W., Tang, J., Oltz, E.M., Murphy, T.L., et al. 2011. The transcription factor BATF controls the global regulators of class-switch recombination in both B cells and T cells. *Nat Immunol* 12:536-543.

182. Ubel, C., Sopel, N., Graser, A., Hildner, K., Reinhardt, C., Zimmermann, T., Rieker, R.J., Maier, A., Neurath, M.F., Murphy, K.M., et al. 2014. The activating protein 1 transcription factor basic leucine zipper transcription factor, ATF-like (BATF), regulates lymphocyte- and mast cell-driven immune responses in the setting of allergic asthma. *J Allergy Clin Immunol* 133:198-206 e191-199.

183. Chevrier, N., Mertins, P., Artyomov, M.N., Shalek, A.K., Iannacone, M., Ciaccio, M.F., Gat-Viks, I., Tonti, E., DeGrace, M.M., Clauser, K.R., et al. 2011. Systematic Discovery of TLR Signaling Components Delineates Viral-Sensing Circuits. *Cell* 147:853-867.

184. Murphy, T.L., Tussiwand, R., and Murphy, K.M. 2013. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat Rev Immunol* 13:499-509.

185. Schraml, B.U., Hildner, K., Ise, W., Lee, W.L., Smith, W.A., Solomon, B., Sahota, G., Sim, J., Mukasa, R., Cemerski, S., et al. 2009. The AP-1 transcription factor Batf controls T(H)17 differentiation. *Nature* 460:405-409.

186. Ciofani, M., Madar, A., Galan, C., Sellars, M., Mace, K., Pauli, F., Agarwal, A., Huang, W., Parkurst, C.N., Muratet, M., et al. 2012. A validated regulatory network for Th17 cell specification. *Cell* 151:289-303.

187. Magnani, L., Eeckhoute, J., and Lupien, M. 2011. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet* 27:465-474.

188. Vahedi, G., Takahashi, H., Nakayamada, S., Sun, H.W., Sartorelli, V., Kanno, Y., and O'Shea, J.J. 2012. STATs shape the active enhancer landscape of T cell populations. *Cell* 151:981-993.

189. Barnitz, R.A., Wan, F., Tripuraneni, V., Bolton, D.L., and Lenardo, M.J. 2010. Protein kinase A phosphorylation activates Vpr-induced cell cycle arrest during human immunodeficiency virus type 1 infection. *J Virol* 84:6410-6424.

190. Best, J.A., Blair, D.A., Knell, J., Yang, E., Mayya, V., Doedens, A., Dustin, M.L., Goldrath, A.W., and Immunological Genome Project, C. 2013. Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nat Immunol* 14:404-412.

191. Dominguez, C.X., Amezquita, R.A., Guan, T., Marshall, H.D., Joshi, N.S., Kleinstein, S.H., and Kaech, S.M. 2015. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *J Exp Med* 212:2041-2056.

192. Laky, K., Fleischacker, C., and Fowlkes, B.J. 2006. TCR and Notch signaling in CD4 and CD8 T-cell development. *Immunol Rev* 209:274-283.

193. Eagar, T.N., Tang, Q., Wolfe, M., He, Y., Pear, W.S., and Bluestone, J.A. 2004. Notch 1 signaling regulates peripheral T cell activation. *Immunity* 20:407-415.

194. Backer, R.A., Helbig, C., Gentek, R., Kent, A., Laidlaw, B.J., Dominguez, C.X., de Souza, Y.S., van Trierum, S.E., van Beek, R., Rimmelzwaan, G.F., et al. 2014. A central role for Notch in effector CD8(+) T cell differentiation. *Nat Immunol* 15:1143-1151.

195. Mathieu, M., Duval, F., Daudelin, J.F., and Labrecque, N. 2015. The Notch signaling pathway controls short-lived effector CD8+ T cell differentiation but is dispensable for memory generation. *J Immunol* 194:5654-5662.

196. Hannon, G.J., and Rossi, J.J. 2004. Unlocking the potential of the human genome with RNA interference. *Nature* 431:371-378.

197. Hannon, G.J. 2002. RNA interference. *Nature* 418:244-251.

198. Wilson, R.C., and Doudna, J.A. 2013. Molecular mechanisms of RNA interference. *Annu Rev Biophys* 42:217-239.

199. Rudalska, R., Dauch, D., Longerich, T., McJunkin, K., Wuestefeld, T., Kang, T.W., Hohmeyer, A., Pesic, M., Leibold, J., von Thun, A., et al. 2014. In vivo RNAi screening identifies a mechanism of sorafenib resistance in liver cancer. *Nat Med* 20:1138-1146.

200. Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P., et al. 2008. An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* 135:852-864.

201. Miller, P.G., Al-Shahrour, F., Hartwell, K.A., Chu, L.P., Jaras, M., Puram, R.V., Puissant, A., Callahan, K.P., Ashton, J., McConkey, M.E., et al. 2013. In Vivo RNAi screening identifies a leukemia-specific dependence on integrin beta 3 signaling. *Cancer Cell* 24:45-58.

202. Zhou, P., and Wucherpfennig, K.W. 2014. Discovering cancer immunotherapy targets in vivo. *Oncoimmunology* 3:e28500.

203. Wan, Q., Kozhaya, L., Imberg, K., Mercer, F., Zhong, S., Krogsgaard, M., and Unutmaz, D. 2013. Probing the effector and suppressive functions of human T cell subsets using antigen-specific engineered T cell receptors. *PLoS One* 8:e56302.

204. Zhong, S., Malecek, K., Perez-Garcia, A., and Krogsgaard, M. 2010. Retroviral transduction of T-cell receptors in mouse T-cells. *J Vis Exp*.

205. Godec, J., Cowley, G.S., Barnitz, R.A., Alkan, O., Root, D.E., Sharpe, A.H., and Haining, W.N. 2015. Inducible RNAi in vivo reveals that the transcription factor BATF is required to initiate but not maintain CD8+ T-cell effector differentiation. *Proc Natl Acad Sci U S A* 112:512-517.

206. Naik, S.H., Perie, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R.J., and Schumacher, T.N. 2013. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* 496:229-232.

207. Foss, D.L., Donskoy, E., and Goldschneider, I. 2001. The importation of hematogenous precursors by the thymus is a gated phenomenon in normal adult mice. *J Exp Med* 193:365-374.

208. Shen, J., and Walsh, C.A. 2005. Targeted disruption of Tgif, the mouse ortholog of a human holoprosencephaly gene, does not result in holoprosencephaly in mice. *Mol Cell Biol* 25:3639-3647.

209. Zhang, M.Z., Ferrigno, O., Wang, Z., Ohnishi, M., Prunier, C., Levy, L., Razzaque, M., Horne, W.C., Romero, D., Tzivion, G., et al. 2015. TGIF governs a feed-forward network that empowers Wnt signaling to drive mammary tumorigenesis. *Cancer Cell* 27:547-560.

210. Yan, L., Womack, B., Wotton, D., Guo, Y., Shyr, Y., Dave, U., Li, C., Hiebert, S., Brandt, S., and Hamid, R. 2013. Tgif1 regulates quiescence and self-renewal of hematopoietic stem cells. *Mol Cell Biol* 33:4824-4833.

211. Wotton, D., Lo, R.S., Lee, S., and Massague, J. 1999. A Smad transcriptional corepressor. *Cell* 97:29-39.

212. Bartholin, L., Powers, S.E., Melhuish, T.A., Lasse, S., Weinstein, M., and Wotton, D. 2006. TGIF inhibits retinoid signaling. *Mol Cell Biol* 26:990-1001.

213. Allie, S.R., Zhang, W., Tsai, C.Y., Noelle, R.J., and Usherwood, E.J. 2013. Critical role for all-trans retinoic acid for optimal effector and effector memory CD8 T cell differentiation. *J Immunol* 190:2178-2187.

214. Guo, Y., Lee, Y.C., Brown, C., Zhang, W., Usherwood, E., and Noelle, R.J. 2014. Dissecting the role of retinoic acid receptor isoforms in the CD8 response to infection. *J Immunol* 192:3336-3344.

215. Pessah, M., Prunier, C., Marais, J., Ferrand, N., Mazars, A., Lallemand, F., Gauthier, J.M., and Atfi, A. 2001. c-Jun interacts with the corepressor TG-interacting factor (TGIF) to suppress Smad2 transcriptional activity. *Proc Natl Acad Sci U S A* 98:6198-6203.

216. Tschismarov, R., Firner, S., Gil-Cruz, C., Goschl, L., Boucheron, N., Steiner, G., Matthias, P., Seiser, C., Ludewig, B., and Ellmeier, W. 2014. HDAC1 controls CD8+ T cell homeostasis and antiviral response. *PLoS One* 9:e110576.

217. Boucheron, N., Tschismarov, R., Goschl, L., Moser, M.A., Lagger, S., Sakaguchi, S., Winter, M., Lenz, F., Vitko, D., Breitwieser, F.P., et al. 2014. CD4(+) T cell lineage integrity is controlled by the histone deacetylases HDAC1 and HDAC2. *Nat Immunol* 15:439-448.

218. Lee, B.K., Shen, W., Lee, J., Rhee, C., Chung, H., Kim, K.Y., Park, I.H., and Kim, J. 2015. Tgif1 Counterbalances the Activity of Core Pluripotency Factors in Mouse Embryonic Stem Cells. *Cell Rep* 13:52-60.

219. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* 10:252-263.

220. Harzheim, E. 2005. *Ordered Sets*. New York: Springer.

221. Gerlach, C., Rohr, J.C., Perie, L., van Rooij, N., van Heijst, J.W., Velds, A., Urbanus, J., Naik, S.H., Jacobs, H., Beltman, J.B., et al. 2013. Heterogeneous differentiation patterns of individual CD8+ T cells. *Science* 340:635-639.

222. Buchholz, V.R., Flossdorf, M., Hensel, I., Kretschmer, L., Weissbrich, B., Graf, P., Verschoor, A., Schiemann, M., Hofer, T., and Busch, D.H. 2013. Disparate individual fates compose robust CD8+ T cell immunity. *Science* 340:630-635.

223. Tubo, N.J., Pagan, A.J., Taylor, J.J., Nelson, R.W., Linehan, J.L., Ertelt, J.M., Huseby, E.S., Way, S.S., and Jenkins, M.K. 2013. Single naive CD4+ T cells from a

diverse repertoire produce different effector cell types during infection. *Cell* 153:785-796.

224. Jain, J., Valge-Archer, V.E., and Rao, A. 1992. Analysis of the AP-1 sites in the IL-2 promoter. *J Immunol* 148:1240-1250.

225. Moore, J.P., Menzel, G.E., Hesketh, T.R., and Metcalfe, J.C. 1988. C-fos gene activation in murine thymocytes by a mechanism independent of protein kinase C or a Ca2+ signal. *FEBS Lett* 233:64-68.

226. Chiu, R., Boyle, W.J., Meek, J., Smeal, T., Hunter, T., and Karin, M. 1988. The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell* 54:541-552.

227. Zhou, S., Cerny, A.M., Fitzgerald, K.A., Kurt-Jones, E.A., and Finberg, R.W. 2012. Role of interferon regulatory factor 7 in T cell responses during acute lymphocytic choriomeningitis virus infection. *J Virol* 86:11254-11265.

228. Wu, Z., Jia, X., de la Cruz, L., Su, X.C., Marzolf, B., Troisch, P., Zak, D., Hamilton, A., Whittle, B., Yu, D., et al. 2008. Memory T cell RNA rearrangement programmed by heterogeneous nuclear ribonucleoprotein hnRNPLL. *Immunity* 29:863-875.

229. Cho, V., Mei, Y., Sanny, A., Chan, S., Enders, A., Bertram, E.M., Tan, A., Goodnow, C.C., and Andrews, T.D. 2014. The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA. *Genome Biol* 15:R26.

230. Matsumoto, M., Yamada, T., Yoshinaga, S.K., Boone, T., Horan, T., Fujita, S., Li, Y., and Mitani, T. 2002. Essential role of NF-kappa B-inducing kinase in T cell activation through the TCR/CD3 pathway. *J Immunol* 169:1151-1158.

231. Lin, X., O'Mahony, A., Mu, Y., Geleziunas, R., and Greene, W.C. 2000. Protein kinase C-theta participates in NF-kappaB activation induced by CD3-CD28 costimulation through selective activation of IkappaB kinase beta. *Mol Cell Biol* 20:2933-2940.

232. Chang, H.C., Han, L., Jabeen, R., Carotta, S., Nutt, S.L., and Kaplan, M.H. 2009. PU.1 regulates TCR expression by modulating GATA-3 activity. *J Immunol* 183:4887-4894.

233. Prinz, M., and Kalinke, U. 2010. New lessons about old molecules: how type I interferons shape Th1/Th17-mediated autoimmunity in the CNS. *Trends Mol Med* 16:379-386.

234. Yarilina, A., Park-Min, K.H., Antoniv, T., Hu, X., and Ivashkiv, L.B. 2008. TNF activates an IRF1-dependent autocrine loop leading to sustained expression of chemokines and STAT1-dependent type I interferon-response genes. *Nat Immunol* 9:378-387.

235. Maekawa, Y., Minato, Y., Ishifune, C., Kurihara, T., Kitamura, A., Kojima, H., Yagita, H., Sakata-Yanagimoto, M., Saito, T., Taniuchi, I., et al. 2008. Notch2 integrates signaling by the transcription factors RBP-J and CREB1 to promote T cell cytotoxicity. *Nat Immunol* 9:1140-1147.

236. Godec, J., Tan, Y., Liberzon, A., Tamayo, P., Bhattacharya, S., Butte, A.J., Mesirov, J.P., and Haining, W.N. 2016. Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity* 44:194-206.

237. Knell, J., Best, J.A., Lind, N.A., Yang, E., D'Cruz, L.M., and Goldrath, A.W. 2013. Id2 influences differentiation of killer cell lectin-like receptor G1(hi) short-lived CD8+ effector T cells. *J Immunol* 190:1501-1509.

238. Willer, A., Jakobsen, J.S., Ohlsson, E., Rapin, N., Waage, J., Billing, M., Bullinger, L., Karlsson, S., and Porse, B.T. 2015. TGIF1 is a negative regulator of MLL-rearranged acute myeloid leukemia. *Leukemia* 29:1018-1031.

239. Sharma, R., Sharma, P.R., Kim, Y.C., Leitinger, N., Lee, J.K., Fu, S.M., and Ju, S.T. 2011. IL-2-controlled expression of multiple T cell trafficking genes and Th2 cytokines in the regulatory T cell-deficient scurfy mice: implication to multiorgan inflammation and control of skin and lung inflammation. *J Immunol* 186:1268-1278.

240. Yu, A., Zhu, L., Altman, N.H., and Malek, T.R. 2009. A low interleukin-2 receptor signaling threshold supports the development and homeostasis of T regulatory cells. *Immunity* 30:204-217.

241. Lin, J.X., Li, P., Liu, D., Jin, H.T., He, J., Ata Ur Rasheed, M., Rochman, Y., Wang, L., Cui, K., Liu, C., et al. 2012. Critical Role of STAT5 transcription factor tetramerization for cytokine responses and normal immune function. *Immunity* 36:586-599.

242. Abbas, A.R., Baldwin, D., Ma, Y., Ouyang, W., Gurney, A., Martin, F., Fong, S., van Lookeren Campagne, M., Godowski, P., Williams, P.M., et al. 2005. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun* 6:319-331.

243. Kao, C., Oestreich, K.J., Paley, M.A., Crawford, A., Angelosanto, J.M., Ali, M.-A.A., Intlekofer, A.M., Boss, J.M., Reiner, S.L., Weinmann, A.S., et al. 2011. Transcription factor T-bet represses expression of the inhibitory receptor PD-1 and sustains virus-specific CD8+ T cell responses during chronic infection. *Nature Immunology* 12:663-671.

244. Gripp, K.W., Wotton, D., Edwards, M.C., Roessler, E., Ades, L., Meinecke, P., Richieri-Costa, A., Zackai, E.H., Massague, J., Muenke, M., et al. 2000. Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. *Nat Genet* 25:205-208.

245. Johnston, R.J., Poholek, A.C., DiToro, D., Yusuf, I., Eto, D., Barnett, B., Dent, A.L., Craft, J., and Crotty, S. 2009. Bcl6 and Blimp-1 are reciprocal and antagonistic regulators of T follicular helper cell differentiation. *Science* 325:1006-1010.

246. Jain, R., Chen, Y., Kanno, Y., Joyce-Shaikh, B., Vahedi, G., Hirahara, K., Blumenschein, W.M., Sukumar, S., Haines, C.J., Sadekova, S., et al. 2016. Interleukin-23-Induced Transcription Factor Blimp-1 Promotes Pathogenicity of T Helper 17 Cells. *Immunity* 44:131-142.

247. Shapiro-Shelef, M., Lin, K.I., McHeyzer-Williams, L.J., Liao, J., McHeyzer-Williams, M.G., and Calame, K. 2003. Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* 19:607-620.

248. Shaffer, A.L., Lin, K.I., Kuo, T.C., Yu, X., Hurt, E.M., Rosenwald, A., Giltnane, J.M., Yang, L., Zhao, H., Calame, K., et al. 2002. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* 17:51-62.

249. Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Holper, S., Kruger, M., and Stainier, D.Y. 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* 524:230-233.

250. Hu, Y., Yu, H., Shaw, G., Renfree, M.B., and Pask, A.J. 2011. Differential roles of TGIF family genes in mammalian reproduction. *BMC Dev Biol* 11:58.

251. Taniguchi, K., Anderson, A.E., Sutherland, A.E., and Wotton, D. 2012. Loss of Tgif function causes holoprosencephaly by disrupting the SHH signaling pathway. *PLoS Genet* 8:e1002524.

252. Gorelik, L., and Flavell, R.A. 2001. Immune-mediated eradication of tumors through the blockade of transforming growth factor-beta signaling in T cells. *Nat Med* 7:1118-1122.

253. Gaj, T., Gersbach, C.A., and Barbas, C.F., 3rd. 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* 31:397-405.

254. Maekawa, Y., Minato, Y., Ishifune, C., Kurihara, T., Kitamura, A., Kojima, H., Yagita, H., Sakata-Yanagimoto, M., Saito, T., Taniuchi, I., et al. 2008. Notch2 integrates signaling by the transcription factors RBP-J and CREB1 to promote T cell cytotoxicity. *Nature Immunology* 9:1140-1147.

255. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36.

256. Liao, Y., Smyth, G.K., and Shi, W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41:e108.

257. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47.

258. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

259. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.

260. Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* 26:1351-1359.

261. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137.

262. Quinlan, A.R., and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841-842.

263. Davis, M.M. 2008. A prescription for human immunology. *Immunity* 29:835-838.

264. van der Worp, H.B., Howells, D.W., Sena, E.S., Porritt, M.J., Rewell, S., O'Collins, V., and Macleod, M.R. 2010. Can animal models of disease reliably inform human studies? *PLoS Med* 7:e1000245.

265. Hackam, D.G., and Redelmeier, D.A. 2006. Translation of research evidence from animals to humans. *JAMA* 296:1731-1732.

266. Rice, J. 2012. Animal models: Not close enough. *Nature* 484:S9.

267. Warren, H.S., Tompkins, R.G., Moldawer, L.L., Seok, J., Xu, W., Mindrinos, M.N., Maier, R.V., Xiao, W., and Davis, R.W. 2014. Mice are not men. *Proc Natl Acad Sci U S A*.

268. Shay, T., Lederer, J.A., and Benoist, C. 2014. Genomic responses to inflammation in mouse models mimic humans: We concur, apples to oranges comparisons won't do. *Proc Natl Acad Sci U S A*.

269. Lin, S., Lin, Y., Nery, J.R., Urich, M.A., Breschi, A., Davis, C.A., Dobin, A., Zaleski, C., Beer, M.A., Chapman, W.C., et al. 2014. Comparison of the transcriptional landscapes between human and mouse tissues. *Proc Natl Acad Sci U S A* 111:17224-17229.

270. Gilad Y, M.-M.O. 2015. 10.12688/f1000research.6536.1. *F1000Research* 4.

271. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. 2003. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267-273.

272. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., and Mesirov, J.P. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27:1739-1740.

273. Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., et al. 2008. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 29:150-164.

274. Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D.S., Johnson, S.E., Milton, A., Rajam, G., et al. 2014. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 15:195-204.

275. Heng, T.S., Painter, M.W., and Immunological Genome Project, C. 2008. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* 9:1091-1094.

276. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462:108-112.

277. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144:296-309.

278. Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 37:48-55.

279. Tang, B.M., McLean, A.S., Dawes, I.W., Huang, S.J., and Lin, R.C. 2009. Gene-expression profiling of peripheral blood mononuclear cells in sepsis. *Crit Care Med* 37:882-888.

280. Ahn, S.H., Deshmukh, H., Johnson, N., Cowell, L.G., Rude, T.H., Scott, W.K., Nelson, C.L., Zaas, A.K., Marchuk, D.A., Keum, S., et al. 2010. Two genes on A/J chromosome 18 are associated with susceptibility to Staphylococcus aureus infection by combined microarray and QTL analyses. *PLoS Pathog* 6:e1001088.

281. Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A* 101:4164-4169.

282. Lee, D.D., and Seung, H.S. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401:788-791.

283. Lee, S. 2000. Algorithms for Non-negative Matrix Factorization. In *Proceedings of Neural Information Processing Systems*. p. 556-562.

284. Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W., and Mesirov, J.P. 2007. Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc Natl Acad Sci U S A* 104:5959-5964.

285. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. 2011. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 39:D691-697.

286. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639-1645.

287. Shay, T., Jojic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., Regev, A., et al. 2013. Conservation and divergence in the transcriptional programs of the human and mouse immune systems. *Proc Natl Acad Sci U S A* 110:2946-2951.

288. Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. 2007. NCBI GEO: mining tens of millions of expression profiles--database and tools update. *Nucleic Acids Res* 35:D760-765.

289. Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185-193.

290. Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307-315.

291. Davis, S., and Meltzer, P.S. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23:1846-1847.

292. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J.P. 2007. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* 23:3251-3253.

293. Abazeed, M.E., Adams, D.J., Hurov, K.E., Tamayo, P., Creighton, C.J., Sonkin, D., Giacomelli, A.O., Du, C., Fries, D.F., Wong, K.K., et al. 2013. Integrative radiogenomic profiling of squamous cell lung cancer. *Cancer Res* 73:6289-6298.

294. Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. 2006. GenePattern 2.0. *Nat Genet* 38:500-501.

295. Henrickson, S.E., Perro, M., Loughhead, S.M., Senman, B., Stutte, S., Quigley, M., Alexe, G., Iannacone, M., Flynn, M.P., Omid, S., et al. 2013. Antigen availability determines CD8(+) T cell-dendritic cell interaction kinetics and memory fate decisions. *Immunity* 39:496-507.

296. Postow, M.A., Callahan, M.K., and Wolchok, J.D. 2015. Immune Checkpoint Blockade in Cancer Therapy. *J Clin Oncol* 33:1974-1982.

297. Topalian, S.L., Sznol, M., McDermott, D.F., Kluger, H.M., Carvajal, R.D., Sharfman, W.H., Brahmer, J.R., Lawrence, D.P., Atkins, M.B., Powderly, J.D., et al. 2014. Survival, durable tumor remission, and long-term safety in patients with advanced melanoma receiving nivolumab. *J Clin Oncol* 32:1020-1030.

298. Lipson, E.J., Sharfman, W.H., Drake, C.G., Wollner, I., Taube, J.M., Anders, R.A., Xu, H., Yao, S., Pons, A., Chen, L., et al. 2013. Durable cancer regression off-treatment and effective reinduction therapy with an anti-PD-1 antibody. *Clin Cancer Res* 19:462-468.

299. Ribas, A., Shin, D.S., Zaretsky, J., Frederiksen, J., Cornish, A., Avramis, E., Seja, E., Kivork, C., Siebert, J., Kaplan-Lefko, P., et al. 2016. PD-1 Blockade Expands Intratumoral Memory T Cells. *Cancer Immunol Res* 4:194-203.

300. Freeman, G.J., Wherry, E.J., Ahmed, R., and Sharpe, A.H. 2006. Reinvigorating exhausted HIV-specific T cells via PD-1-PD-1 ligand blockade. *J Exp Med* 203:2223-2227.

301. Urbani, S., Amadei, B., Tola, D., Massari, M., Schivazappa, S., Missale, G., and Ferrari, C. 2006. PD-1 expression in acute hepatitis C virus (HCV) infection is associated with HCV-specific CD8 exhaustion. *J Virol* 80:11398-11403.

302. Stagg, J., Beavis, P.A., Divisekera, U., Liu, M.C., Moller, A., Darcy, P.K., and Smyth, M.J. 2012. CD73-deficient mice are resistant to carcinogenesis. *Cancer Res* 72:2190-2196.

303. Stagg, J., Divisekera, U., Duret, H., Sparwasser, T., Teng, M.W., Darcy, P.K., and Smyth, M.J. 2011. CD73-deficient mice have increased antitumor immunity and are resistant to experimental metastasis. *Cancer Res* 71:2892-2900.

304. Allard, B., Pommey, S., Smyth, M.J., and Stagg, J. 2013. Targeting CD73 enhances the antitumor activity of anti-PD-1 and anti-CTLA-4 mAbs. *Clin Cancer Res* 19:5626-5635.

305. Loi, S., Pommey, S., Haibe-Kains, B., Beavis, P.A., Darcy, P.K., Smyth, M.J., and Stagg, J. 2013. CD73 promotes anthracycline resistance and poor prognosis in triple negative breast cancer. *Proc Natl Acad Sci U S A* 110:11091-11096.

306. Ohta, A., Gorelik, E., Prasad, S.J., Ronchese, F., Lukashev, D., Wong, M.K., Huang, X., Caldwell, S., Liu, K., Smith, P., et al. 2006. A2A adenosine receptor protects tumors from antitumor T cells. *Proc Natl Acad Sci U S A* 103:13132-13137.

307. Bastid, J., Cottalorda-Regairaz, A., Alberici, G., Bonnefoy, N., Eliaou, J.F., and Bensussan, A. 2013. ENTPD1/CD39 is a promising therapeutic target in oncology. *Oncogene* 32:1743-1751.

308. Antonioli, L., Pacher, P., Vizi, E.S., and Hasko, G. 2013. CD39 and CD73 in immunity and inflammation. *Trends Mol Med* 19:355-367.

309. Gessi, S., Merighi, S., Sacchetto, V., Simioni, C., and Borea, P.A. 2011. Adenosine receptors and cancer. *Biochim Biophys Acta* 1808:1400-1412.

310. Shalapour, S., and Karin, M. 2015. Immunity, inflammation, and cancer: an eternal fight between good and evil. *J Clin Invest* 125:3347-3355.

311. Massague, J. 2008. TGFbeta in Cancer. *Cell* 134:215-230.

312. Thomas, D.A., and Massague, J. 2005. TGF-beta directly targets cytotoxic T cell functions during tumor evasion of immune surveillance. *Cancer Cell* 8:369-380.

313. Doedens, A.L., Phan, A.T., Stradner, M.H., Fujimoto, J.K., Nguyen, J.V., Yang, E., Johnson, R.S., and Goldrath, A.W. 2013. Hypoxia-inducible factors enhance the effector responses of CD8(+) T cells to persistent antigen. *Nat Immunol* 14:1173-1182.

314. Doedens, A.L., Stockmann, C., Rubinstein, M.P., Liao, D., Zhang, N., DeNardo, D.G., Coussens, L.M., Karin, M., Goldrath, A.W., and Johnson, R.S. 2010. Macrophage expression of hypoxia-inducible factor-1 alpha suppresses T-cell function and promotes tumor progression. *Cancer Res* 70:7465-7475.

315. Palazon, A., Goldrath, A.W., Nizet, V., and Johnson, R.S. 2014. HIF transcription factors, inflammation, and immunity. *Immunity* 41:518-528.

316. Nizet, V., and Johnson, R.S. 2009. Interdependence of hypoxic and innate immune responses. *Nat Rev Immunol* 9:609-617.

317. Labiano, S., Palazon, A., and Melero, I. 2015. Immune response regulation in the tumor microenvironment by hypoxia. *Semin Oncol* 42:378-386.

318. Glasmacher, E., Agrawal, S., Chang, A.B., Murphy, T.L., Zeng, W., Vander Lugt, B., Khan, A.A., Ciofani, M., Spooner, C.J., Rutz, S., et al. 2012. A genomic regulatory element that directs assembly and function of immune-specific AP-1-IRF complexes. *Science* 338:975-980.

319. Lalit K. Beura, S.E.H., Kevin Bi, Jason M. Schenkel, Oludare A. Odumade, Kerry A. Casey, Emily A. Thompson, --Kathryn A. Fraser, Ali Filali-Mouhim, Rafick P. Sekaly, Marc K. Jenkins, Vaiva Vezys, W. Nicholas Haining, Stephen C. Jameson, David Masopust. 2016. Recapitulating adult human immune traits in laboratory mice by normalizing environment. *Nature* In Press.

320. Wilkinson, J.M., Gunvaldsen, R.E., Detmer, S.E., Dyck, M.K., Dixon, W.T., Foxcroft, G.R., Plastow, G.S., and Harding, J.C. 2015. Transcriptomic and Epigenetic Profiling of the Lung of Influenza-Infected Pigs: A Comparison of Different Birth Weight and Susceptibility Groups. *PLoS One* 10:e0138653.

321. Shannon, C.P., Balshaw, R., Ng, R.T., Wilson-McManus, J.E., Keown, P., McMaster, R., McManus, B.M., Landsberg, D., Isbel, N.M., Knoll, G., et al. 2014. Two-stage, in silico deconvolution of the lymphocyte compartment of the peripheral whole blood transcriptome in the context of acute kidney allograft rejection. *PLoS One* 9:e95224.

322. Brusic, V., Gottardo, R., Kleinstein, S.H., Davis, M.M., and committee, H.s. 2014. Computational resources for high-dimensional immune analysis from the Human Immunology Project Consortium. *Nat Biotechnol* 32:146-148.

323. Vishwanathan, S.A., Burgener, A., Bosinger, S.E., Tharp, G.K., Guenthner, P.C., Patel, N.B., Birse, K., Hanson, D.L., Westmacott, G.R., Henning, T.R., et al. 2015. Cataloguing of Potential HIV Susceptibility Factors during the Menstrual Cycle of Pig-Tailed Macaques by Using a Systems Biology Approach. *J Virol* 89:9167-9177.

324. Visconte, V., Tabarroki, A., Zhang, L., Parker, Y., Hasrouni, E., Mahfouz, R., Isono, K., Koseki, H., Sekeres, M.A., Sauntharajah, Y., et al. 2014. Splicing factor 3b subunit 1 (Sf3b1) haploinsufficient mice display features of low risk Myelodysplastic syndromes with ring sideroblasts. *J Hematol Oncol* 7:89.

325. Forn-Cuni, G., Varela, M., Fernandez-Rodriguez, C.M., Figueras, A., and Novoa, B. 2015. Liver immune responses to inflammatory stimuli in a diet-induced obesity model of zebrafish. *J Endocrinol* 224:159-170.

326. Micci, L., Ryan, E.S., Fromentin, R., Bosinger, S.E., Harper, J.L., He, T., Paganini, S., Easley, K.A., Chahroudi, A., Benne, C., et al. 2015. Interleukin-21 combined with ART reduces inflammation and viral reservoir in SIV-infected macaques. *J Clin Invest* 125:4497-4513.

327. Lin, S.J., Gagnon-Bartsch, J.A., Tan, I.B., Earle, S., Ruff, L., Pettinger, K., Ylstra, B., van Grieken, N., Rha, S.Y., Chung, H.C., et al. 2015. Signatures of tumour immunity distinguish Asian and non-Asian gastric adenocarcinomas. *Gut* 64:1721-1731.

328. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 1:417-425.

329. Katayama, S., Skoog, T., Jouhilahti, E.M., Siitonen, H.A., Nuutila, K., Tervaniemi, M.H., Vuola, J., Johnsson, A., Lonnerberg, P., Linnarsson, S., et al. 2015. Gene expression analysis of skin grafts and cultured keratinocytes using synthetic RNA normalization reveals insights into differentiation and growth control. *BMC Genomics* 16:476.

Publications

1. **Jernej Godec**, Yan Tan, Arthur Liberzon, Pablo Tamayo, Atul J. Butte, Jill P. Mesirov, W. Nicholas Haining. A novel compendium of immune signatures identifies both conserved and species-specific biology in the mouse and human response to inflammation. *Immunity*. 2016 Jan 19;44(1):194-206. (PMID: 26795250)

2. Prakash K. Gupta*, **Jernej Godec***, David Wolski*, Emily Adland, Kathleen Yates, Cormac Cosgrove, Carola Ledderose, Wolfgang G. Junger, Simon C. Robson, E. John Wherry, Galit Alter, Philip J. R. Goulder, Paul Klenerman, Arlene H. Sharpe, Georg M. Lauer, W. Nicholas Haining. CD39 expression identifies terminally exhausted $CD8^+$ T cells. *PLOS Pathogens*. 2015 Oct 20;11(10):e1005177. (PMID: 26485519). *Co-first author.

3. Neil Ruparelia, **Jernej Godec**, Regent Lee, Joshua T. Chai, Erica Dall'Armellina, Debra McAndrew, Janet E. Digby, J. Colin Forfar, Bernard D. Prendergast, Rajesh K Kharbanda, Adrian P. Banning, Stefan Neubauer, Craig A. Lygate, Keith M. Channon, W. Nicholas Haining, Robin P. Choudhury. Acute myocardial infarction activates distinct inflammatory and proliferation pathways in circulating monocytes, prior to recruitment, identified through conserved transcriptional responses in mice and humans. *European Heart Journal.* 2015 Aug 1;36(29):1923-34. (PMID: 25982896).

4. **Jernej Godec**, Glenn S. Cowley, R. Anthony Barnitz, Ozan Alkan, David E. Root, Arlene H. Sharpe, W. Nicholas Haining. Inducible RNAi *in vivo* reveals that BATF is required to initiate but not to maintain CD8+ T cell effector differentiation. *Proceedings of the National Academy of Sciences USA*. 2015 Jan 13;112(2):512-7. (PMID: 25548173)

5. Peter T. Sage, David Alvarez, **Jernej Godec**, Ulrich von Andrian, Arlene H. Sharpe. Memory-like Properties of Circulating T Follicular Regulatory and Helper Cells. *Journal of Clinical Investigation*. 2014, 124(12):5191-204. (PMID: 25347469)

6. Yan Wang, **Jernej Godec**, Khadija Ben-Aissa, Kairong Cui, Keji Zhao, Alexandra B. Pucsek, Yun Kyung Lee, Casey T. Weaver, Ryoji Yagi, Vanja Lazarevic. T-bet and Runx transcription factors are required for the ontogeny of pathogenic IFNγ-producing Th17 cells. *Immunity*. 2014, 40(3):355-66. (PMID: 24530058)

7. Makoto Kurachi, R. Anthony Barnitz, Nir Yosef, Pamela M. Odorizzi, Michael A. DiIorio, Madeleine E. Lemieux, Kathleen Yates, **Jernej Godec,** Martin G. Klatt, Aviv Regev, E. John Wherry, W. Nicholas Haining. BATF operates as an essential differentiation checkpoint in early effector $CD8^+$ T cells. *Nature Immunology*. 2014, 15(4):373-83. (PMID: 24584090)

8. Adlai R. Grayson, Erica M. Walsh, Michael J. Cameron, **Jernej Godec**, Todd Ashworth, Jessica M. Ambrose, Alexandra B. Aserlind, Hongfang Wang, Gerard Evan, Michael J. Kluk, James E. Bradner, Jon C. Aster, Christopher A. French. MYC, a downstream target of BRD-NUT, is necessary and sufficient for the blockade of differentiation in NUT midline carcinoma. *Oncogene*. 2013, 33(13):1736-42 (PMID:23604113)

Supplemental Information

Supplemental Figure 3.1. Clinical characteristics of the subjects with HCV infection.

PatientID	Gender	Chronic/resolver	Viral load	Genotype	ALT
00-010	M	Resolver	50	undetectable	39
00-023	F	Chronic	300	Type 1	18
01-00E	F	Resolver		undetectable	
01-021	F	Chronic	2500000	Type 1	
02-003	F	Chronic	18700	Type 1	123
02-00Z	M	Resolver		No test/unknown	
05-00Y	F	Resolver	50	undetectable	14
06-00K	F	Resolver	600	Type 1	124
06-00L	M	Resolver	50	Type 3	82
06-042	M	Chronic	615	Type 1	
07-00S	F	Resolver	615	No test/unknown	212
07-00Z	F	Resolver		No test/unknown	375
07-032	M	Chronic	16300	Type 2	565
07-052	F	Chronic	219	Type 1	
07-080	F	Resolver	21500	No test/unknown	316
07-082	F	Chronic	261000	Type 1	
08-024	F	Chronic	700000	Type 1	411
08-027	M	Chronic	3838	No test/unknown	58
09-00B	M	Chronic	217000	Type 1	354
09-033	F	Resolver	700000	Type 1	2299
09-037	M	Chronic	600	Type 1	32
10-00H	F	Chronic	704	Type 1	107
10-00M	F	Resolver		No test/unknown	
10-034	M	Chronic		Type 3	58
10-048	F	Chronic	321000	Type 1	179
10-054	F	Chronic	1130	Type 1	209
10-062	F	Chronic	2587650	Type 3	109
10-078	M	Resolver	89200	Type 3	875
10-094	M	Chronic	822000	Type 3	146
10-106	F	Resolver	19347	No test/unknown	217
11-00M	M	Chronic	8150000	Type 1	371
11-014	M	Resolver	3150	Type 2	129
11-017	F	Chronic	25431	Type 1	481
12-043	M	Resolver	61602	No test/unknown	692
12-055	M	Resolver	2311	Type 1	47
12-088	F	Chronic		Type 1	
12-103	F	Chronic	432	Type 3	44
12-108	M	Chronic	42000	Type 1	81
12-181	F	Chronic	346000	Type 3	657
13-022	M	Chronic	2260000	Type 2	587
13-024	M	Chronic	241	Type 1	205
13-066	F	Chronic	43	No test/unknown	118
99-021	M	Resolver		No test/unknown	
BR-1264	F	Resolver	50	Type 1	17
BR-277	F	Resolver	50	Type 1	13
BR-3000	M	Resolver	47272	Type 1	36
BR-3012	M	Resolver	50	Type 1	
BR-320	F	Resolver	50	Type 1	15
BR-3497	M	Chronic		Type 1	67
BR-3821	F	Resolver	50	Type 1	
BR-554	F	Chronic (resolves 3 years into chronic	64497	Type 1	112
BR-599	F	Resolver	1000	Type 1	16
BR-84	M	Resolver	50	undetectable	12
BR-903	M	Chronic	3341	Type 3	45
BR-994	M	Chronic		Type 1	102
BR-1036	F	Resolver	1000	Type 1	4
BR-1144	F	Resolver	1000	Type 1	2
BR-949	F	Chronic	70047	Type 1	36
BR-54	F	Chronic		Type 1	237
14-134	M	Chronic	258000	No test/unknown	198
14-140	M	Chronic		Type 1	

Supplemental Table 3.2. Clinical characteristics of the subjects with HIV infection.

Patient ID	Gender	Progressor/Controller	ON/OFF Rx	Viral Load	CD4 Count
254567	M	Chronic	OFF	1823	606
350103	F	Chronic	OFF	431	625
350534	M	Chronic	OFF	24500	154
359260	M	Chronic	OFF	10322	541
384682	M	Chronic	ON	147	510
387879	M	Chronic	OFF	14600	677
403998	F	Chronic	OFF	2100	877
128019	M	Viraemic Controllers	OFF	unknown	unknown
186089	M	Viraemic Controllers	OFF	82	740
237983	F	Viraemic Controllers	OFF	189	1232
270245	M	Viraemic Controllers	OFF	15	unknown
302225	M	Viraemic Controllers	OFF	65	484
711950	M	Viraemic Controllers	OFF	300	700
732751	M	Viraemic Controllers	OFF	1860	1550
255675	M	Elite Controllers	OFF	103	963
269198	M	Elite Controllers	OFF	unknown	unknown
285297	F	Elite Controllers	OFF	118	1246
321797	M	Elite Controllers	OFF	unknown	unknown
831969	F	Elite Controllers	OFF	unknown	unknown
R060	M	Chronic	OFF	117934	480
R086	M	Chronic	OFF	172886	410
R089	M	Chronic	OFF	44000	680
R046	M	Chronic	OFF	28445	910
R050	M	Chronic	OFF	20210	440
R041	M	Chronic	OFF	8435	320
R017	M	Chronic	OFF	172886	410
N034	M	Chronic	OFF	44000	680
R134	M	Chronic	OFF	500000	430
N012	M	Chronic	OFF	36695	?
N090	F	Chronic	OFF	3362	490
N104	M	Chronic	OFF	4533	390
OX019	F	Chronic	OFF	42912	740
R051	M	Chronic	OFF	500000	560
R069	M	Chronic	OFF	63257	450
N004	M	Chronic	OFF	500000	430
N093	F	Chronic	OFF	2216	700
OX034	M	Chronic	OFF	124153	430
H005	M	Chronic	OFF	747	640
H033	M	Chronic	OFF	8036	430
R103	F	Chronic	OFF	8435	320

Supplemental Table 3.3. The complete list of MHC-peptide multimers used in Chapter 3.

Virus	Multimer type	MHC	Peptide used	Antigen derived from	Supplier
HCV	Pentamer	A*02:01	GIDPNIRTGV	HCV NS3 1273-1082	Proimmune
HCV	Pentamer	A*02:01	KLVALGINAV	HCV NS3 1406-1415	Proimmune
HCV	Pentamer	A*01:01	ATDALMTGY	HCV NS3 1435-1443	Proimmune
HCV	Pentamer	B*40:01	REISVPAEIL	HCV NS5a 2266-2275	Proimmune
HCV	Pentamer	B*07:02	GPRLGVRAT	HCV core 41-49	Proimmune
HCV	Pentamer	A*02:01	VLSDFKTWL	HCV NS5a 1987-1995	Proimmune
HCV	Pentamer	A*02:01	YPYRLWHYPC	HCV E2 610-619	Proimmune
HCV	Pentamer	B*27:01	ARMILMTHF	HCV core 470-478	Proimmune
HCV	Pentamer	A*02:01	CINGVCWTV	HCV NS3 1073-1081	Proimmune
CMV	Dextramer	A*02:01	NLVPMVATC	HCMV pp65	Immudex
EBV	Dextramer	A*02:01	GLCTLVAML	EBV BMLF-1	Immudex
HIV	Tetramer	A*24:02	RYPLTFGW	Nef RW8	Custom made
HIV	Tetramer	B*57:01	KAFSPEVIPMF	Gag KF11	Custom made
HIV	Tetramer	B*14:02	DRFYKTLRA	Gag DA9	Custom made
HIV	Tetramer	B*35:01	HPVHAGPIA	Gag HA9	Custom made
HIV	Tetramer	B*14:02	DRFYKTLRA	Gag DA9	Custom made
HIV	Dextramer	A*02:01	SLYNTVATL	Gag SL9	Immudex
HIV	Pentamer	B*07:02	TPQDLNTML	Gag TL9	Proimmune
HIV	Dextramer	A*02:01	SLYNTVATL	Gag SL9	Immudex
HIV	Dextramer	B*57:01	KAFSPEVIPMF	Gag KF11	Immudex
HIV	Tetramer	B*08:01	EIYKRWII	Gag EI8	Custom made
HIV	Tetramer	B*35:01	VPLRPMTY	Nef VY8	Beckman
HIV	Dextramer	B*07:02	GPGHKARVL	Gag GL9	Immudex
LCMV	Tetramer	H-2Db	SGVENPGGYCL	GP276-286	Dr. E. John Wherry

Supplemental Table 5.1. The list of prioritized genes with potential role in T cells responding to a viral infection.

	Relative Score			GSE41870												GSE8678		GSE10239	
				Acute CD8			Chronic CD8			Acute CD4			Chronic CD4			d6/7 Acute		d4.5 Acute	
	Total	CD8	CD4	d6	d8	d30	d6	d8	d30	d6	d8	d30	d6	d8	d30	CD127hi	CD127lo	KLRG1int	KLRG1hi
BATF	45811	45500	311	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
CSDA	45811	45500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
NFKBIA	45811	45500	311	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
PRDM1	45811	45500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
BHLHE40	44811	44500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
TBX21	44811	44500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
EZH2	44800	44500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	1	0	1
H2AFX	44300	44000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0	1
EPAS1	44110	44000	110	1	1	1	1	1	0	1	0	1	0	0	0	0	1	0	0
E2F2	44100	44000	100	1	1	0	1	1	0	1	0	0	0	0	0	0	1	1	0
E2F8	43810	43500	310	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0
POLE	43801	43500	301	1	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0
HMGB3	43800	43500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	1	0	0
TCF19	43750	43500	250	1	1	0	1	1	1	1	1	0	0	1	0	0	1	0	0
TFDP1	43300	43000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0	0
HMGB2	43100	43000	100	1	1	0	1	1	0	1	0	0	0	0	0	0	1	0	0
ATF6	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
HOPX	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
IRF4	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
NFIL3	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
NR4A1	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
NR4A2	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
NR4A3	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
TRPS1	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
RBM47	36810	36500	310	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
TRIM37	36800	36500	300	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0
DEPDC1B	36600	36500	100	1	1	0	1	0	0	1	0	0	0	0	0	0	1	0	1
ZEB2	36500	36500	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
SPIC	36310	36000	310	1	1	1	1	1	0	1	1	1	1	1	0	0	0	0	0
WHSC1	35811	35500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0
FOXM1	35800	35500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0
HHEX	35700	35500	200	1	1	0	1	1	1	0	1	0	1	1	0	0	0	0	0
SPI1	35600	35500	100	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0
ZBTB32	35301	35000	301	1	1	0	1	1	0	1	1	0	1	1	1	0	0	0	0
APITD1	35300	35000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0
IGF2BP3	35300	35000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0
IRF8	35300	35000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0
CHAF1B	35200	35000	200	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0
MYBL2	35200	35000	200	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0
HIF1A	34311	34000	311	1	0	0	1	1	0	1	1	1	1	1	1	1	0	0	1
HNRPLL	33751	33500	251	1	0	0	1	1	1	1	1	0	1	0	1	0	1	0	0
WDHD1	33300	33000	300	1	0	0	1	1	0	1	1	0	1	1	0	0	1	0	0
ID2	27311	27000	311	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0
FOSB	26811	26500	311	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
TGIF1	26811	26500	311	1	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
RNPS1	26600	26500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0
CREM	26561	26500	61	1	0	1	1	1	1	0	0	1	1	0	1	0	0	0	0
KLF10	25811	25500	311	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0
MBNL3	25750	25500	250	1	0	0	1	0	0	1	1	0	1	0	0	0	1	0	0
MXD1	25550	25500	50	1	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0
XBP1	25510	25500	10	1	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0
ETV6	25311	25000	311	1	0	0	1	1	0	1	1	1	1	1	1	0	0	0	0
NFE2	25150	25000	150	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0	0
IRF7	24801	24500	301	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0	0
EOMES	19511	19500	11	1	0	0	0	0	1	0	0	1	0	0	1	0	1	0	1
ATF3	19311	19000	311	1	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0
NUP43	18600	18500	100	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0
JDP2	18250	18000	250	0	1	0	0	1	1	1	1	0	0	1	0	0	0	0	0
TAF9	18100	18000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
NFYB	18000	18000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
MXD3	17600	17500	100	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
RDBP	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
ZIK1	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
ZNF367	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
CARHSP1	17500	17500	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
MAFB	17500	17500	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
GABPB1	16811	16500	311	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
ZNF642	16551	16500	51	0	0	1	1	1	1	0	0	0	1	0	1	0	0	0	0

PBX3	15811	15500	311	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0
ZNF414	15600	15500	100	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0
STAT3	15551	15500	51	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0
MXI1	15500	15500	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
REL	15311	15000	311	0	0	0	1	1	0	1	1	1	1	1	1	0	0	0	0
MEF2C	15000	15000	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
MAFF	11811	11500	311	1	0	1	0	0	1	1	1	1	1	1	1	0	0	0	0
SMAD3	11000	11000	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
WDR5	11000	11000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
BHLHE41	10550	10500	50	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
KLF12	10500	10500	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
SMARCC1	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TCF7L2	10100	10000	100	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
E2F1	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E2F6	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NFKBIB	9311	9000	311	0	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0
FOSL2	9111	9000	111	0	0	1	0	1	1	0	0	1	1	1	1	0	0	0	0
EGR1	9010	9000	10	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1
FOS	9001	9000	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
CBFB	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
CREBBP	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
EEF1E1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
EIF3M	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
ELF1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
IFI35	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
JUND	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
KLF2	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
MLL4	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
NMI	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
RBMS1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
RFX7	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
SF3B1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
SP3	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
STAT1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
TCF7	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
TRIM21	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
TRIM28	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
WDR74	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
ZSCAN21	9000	9000	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
KLF11	8500	8500	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
ZNF280B	8500	8500	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
PLEK	45811	45500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1
ITGAL	45311	45000	311	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1
CASP4	44811	44500	311	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
CCNA2	44811	44500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1
CDKN2C	44811	44500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1
RPA2	44811	44500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0
CLIC4	44801	44500	301	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0
LIG1	44801	44500	301	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	0
NCAPG2	44801	44500	301	1	1	0	1	1	1	1	1	0	1	1	1	0	1	0	1
CDC6	44800	44500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	1	1	0
LGALS3	44800	44500	300	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	0
FBXO5	44650	44500	150	1	1	0	1	1	1	1	0	0	0	1	0	0	1	0	1
DBF4	44600	44500	100	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0
LGALS1	44311	44000	311	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	0
MCM6	44301	44000	301	1	1	0	1	1	0	1	1	0	1	1	1	0	1	1	0
MCM5	44300	44000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	1	0
MCM7	44300	44000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	1	0
APAF1	43811	43500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0
CCNB2	43811	43500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	0
XDH	43801	43500	301	1	1	0	1	1	1	1	1	0	1	1	1	0	1	0	0
ECT2	43300	43000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0	0
MCM3	43300	43000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0	0
MCM4	43300	43000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0	0
CDC20	43250	43000	250	1	1	0	1	1	0	1	1	0	0	1	0	0	1	0	0
LITAF	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
MYBL1	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
POLA1	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
RORA	36811	36500	311	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
CD74	36810	36500	310	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0
UHRF1	36300	36000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	0	1	0

DTL	35811	35500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0
HELLS	35811	35500	311	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0
LPXN	35811	35500	311	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0
IFIH1	35801	35500	301	1	0	1	1	1	1	1	1	0	1	1	1	1	0	0	1
ANLN	35800	35500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0
BARD1	35800	35500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0
BRIP1	35800	35500	300	1	1	0	1	1	1	1	1	0	1	1	0	0	0	0	0
CCNE2	35650	35500	150	1	1	0	1	1	1	1	0	0	0	1	0	0	0	0	0
LANCL2	35311	35000	311	1	1	0	1	1	0	1	1	1	1	1	1	0	0	0	0
LMAN1	35301	35000	301	1	1	0	1	1	0	1	1	0	1	1	1	0	0	0	0
ATAD2	35300	35000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0
BRCA1	35300	35000	300	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0
MCM8	35250	35000	250	1	1	0	1	1	0	1	1	0	0	1	0	0	0	0	0
PSMD13	35250	35000	250	1	1	0	1	1	0	1	1	0	0	1	0	0	0	0	0
RFC2	35250	35000	250	1	0	1	1	1	0	1	1	0	1	0	0	0	1	1	0
CCNF	35100	35000	100	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0
E2F7	35100	35000	100	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0
CRIP1	34310	34000	310	0	1	0	1	1	0	1	1	1	1	1	0	0	1	0	1
LMAN2	34250	34000	250	1	0	1	1	1	0	1	1	0	1	0	0	0	1	0	0
MCM2	34250	34000	250	1	0	0	1	1	0	1	1	0	0	1	0	0	1	1	0
CLIC1	33310	33000	310	1	0	0	1	1	0	1	1	1	1	1	0	0	1	0	0
RFC4	33300	33000	300	1	0	0	1	1	0	1	1	0	1	1	0	0	1	0	0
BLM	33250	33000	250	1	0	0	1	1	0	1	1	0	1	0	0	0	1	0	0
RPA3	33100	33000	100	1	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0
HDGF	28600	28500	100	1	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0
HIST1H1C	27010	27000	10	0	1	1	0	1	1	0	0	1	0	0	0	0	1	0	0
CCNE1	26800	26500	300	1	0	0	1	0	0	1	1	0	1	1	0	0	1	1	0
SIVA1	26300	26000	300	1	0	0	1	1	0	1	1	0	1	1	0	0	0	1	0
GSTO1	26000	26000	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0
POGK	26000	26000	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0
EIF4H	25810	25500	310	1	0	0	1	0	0	1	1	1	1	1	0	0	1	0	0
BRD9	25600	25500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0
TOPBP1	25600	25500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0
INTS12	25500	25500	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
PLEKHA2	25311	25000	311	1	0	0	1	1	0	1	1	1	1	1	1	0	0	0	0
PLEKHB2	25301	25000	301	1	0	0	1	1	0	1	1	0	1	1	1	0	0	0	0
BRCA2	25300	25000	300	1	0	0	1	1	0	1	1	0	1	1	0	0	0	0	0
PLD4	25200	25000	200	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0
KPNA3	25100	25000	100	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0
SHC1	25100	25000	100	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0
DDX1	19100	19000	100	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
RFC5	19100	19000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0
TCEB3	18661	18500	161	0	1	1	0	1	0	0	1	1	1	0	1	0	0	0	0
SSRP1	18200	18000	200	1	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0
AAAS	18100	18000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
RANBP1	18100	18000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
CFH	18000	18000	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
CSTF3	18000	18000	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
ZYX	18000	18000	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
PTBP1	17811	17500	311	1	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0
RAPGEF5	17811	17500	311	1	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0
CORO1C	17700	17500	200	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0
TTC9C	17700	17500	200	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
RPAP3	17650	17500	150	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0
COPS7B	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
CSRP1	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
FANCM	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
LMO2	17600	17500	100	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
MYD88	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
NCBP2	17600	17500	100	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
POLE3	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
THOC6	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
YARS	17600	17500	100	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
CPSF4	17500	17500	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
MEFV	17500	17500	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
RAD18	17500	17500	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
RNF121	17500	17500	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
ISG20	16301	16000	301	0	0	0	0	1	1	1	1	0	1	1	1	1	0	0	0
UHRF2	16211	16000	211	0	0	0	0	1	1	1	0	1	1	1	1	1	0	0	0
BIN3	15810	15500	310	0	0	0	1	0	0	1	1	1	1	1	0	0	1	0	0
GTF2F1	15600	15500	100	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0

NSMCE2	15500	15500	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
MTX1	11100	11000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
BOLA3	10500	10500	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
CORO1B	10200	10000	200	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
RPA1	10200	10000	200	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
BUB3	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GTF2F2	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
HDLBP	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
NUP133	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
POLR3K	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
PSMD7	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
RBBP4	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
SKP2	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
STIP1	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TRAIP	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TTF2	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
USP39	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
UTP6	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
XRCC6	10100	10000	100	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
GTF2E2	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
KBTBD8	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MTX2	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NEDD1	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PLRG1	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PPARG	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPAG1	10000	10000	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SRBD1	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TAF12	10000	10000	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WIPI2	10000	10000	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DHX40	9500	9500	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0
DDX3X	9111	9000	111	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	1
APEX1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
CASP8	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
CCNG2	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
CLTC	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
DEK	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
GPS1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
H2AFY	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
KLHDC3	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
MRPL2	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
PARP14	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
PARP9	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
PFDN2	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
PHB2	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
POLD1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
PSMD14	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
RNF138	9000	9000	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
RPS3	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
RUVBL1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
SAMD9L	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
SAMHD1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
SAT1	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
TRAF7	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
VARS	9000	9000	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0
RBL2	8500	8500	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0
RNF125	8311	8000	311	0	0	0	0	0	0	1	1	1	1	1	1	1	0	0	0
CLGN	8101	8000	101	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0
ANAPC5	8100	8000	100	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
ST13	8100	8000	100	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
RNF19A	8051	8000	51	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0
PWP1	9100	9000	100	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0

Supplemental Table 5.2. The list of genes selected for the *in vivo* RNAi screen in CD8+ T cells and their representation in respective pools.

	Gene	# Available shRNA	# shRNA with KD validation >70%	No KD data available	# shRNA selected for the screen
POOL 1					
1	Whsc1	9	8	0	3
2	Zfp414	5	5	0	3
3	E2f1	6	5	0	3
4	Pole	5	4	0	3
5	Nr4a2	4	4	0	3
6	Wdhd1	7	4	0	3
7	Tgif1	5	4	0	3
8	Mbnl3	4	4	0	3
9	Mxd1	5	4	0	3
10	Trim28	5	1	4	5
11	Rnps1	3	1	2	3
12	Wdr5	3	1	2	3
13	Xbp1	4	2	0	2
14	Pbx3	2	2	0	2
15	Sfpi1	5	0	5	5
16	Hmgb3	8	0	8	5
17	Irf7	5	0	5	5
18	Epas1	3	0	3	3
19	Igf2bp3	3	0	3	3
20	Hif1a	4	3	0	3
		Control:	25	Experimental:	66
POOL 2					
1	Jdp2	7	4	1	3
2	Stat3	2	2	0	2
3	E2f6	2	2	0	2
4	Mxd3	4	4	0	3
5	Gabpb1	5	4	0	3
6	Maff	4	4	0	3
7	Klf12	5	4	0	3
8	Nr4a1	5	3	0	3
9	Hhex	5	3	0	3
10	Chaf1b	3	3	0	3
11	Mybl2	4	3	1	3
12	Hnrpll	4	3	0	3
13	Hopx	5	0	5	5
14	Etv6	5	3	0	3
15	Cbfb	2	2	0	2
16	Sp3	5	2	0	2
17	Irf1	5	2	0	2
18	Zscan21	6	0	6	5
19	Zik1	5	0	5	5
20	Zfp367	5	0	5	5
21	Pparg	4	4	0	3
		Control:	25	Experimental:	66
POOL 3					
1	Nup43	3	3	0	3
2	Carhsp1	4	3	0	3
3	Mafb	3	3	0	3
4	Crebbp	5	3	0	3
5	Klf2	5	3	0	3
6	Nmi	3	3	0	3
7	Rfx7	5	3	0	3
8	Wdr74	4	3	0	3
9	Klf10	13	1	8	5
10	Tcf7l2	8	1	7	5
11	Smad3	5	1	4	5
12	Nfkbib	3	2	0	2
13	Egr1	10	2	0	2
14	Nfyb	5	2	0	2
15	Fos	7	2	0	2
16	Mxi1	5	0	5	5
17	Rel	5	0	5	5
18	Nfe2	4	0	4	4
19	Rdbp	4	0	4	4
20	Atf3	3	3	0	3
		Control:	25	Experimental:	68
POOL 4					
1	Ifi35	2	1	0	4
2	Stat1	5	1	0	4
3	Tcf7	3	1	0	4
4	Bmi1	1	1	0	4
5	Mef2c	0	0	0	4
6	Runx3	0	0	0	4
7	Ikzf1	4	0	0	4
8	Ikzf3	5	0	0	4
9	Runx2	2	0	2	4
10	Ikzf2	6	0	0	4
11	Eomes	2	0	0	4
12	Irf4	5	0	5	4
13	Tbx21	2	0	2	4
		Control:	25	Experimental:	52

Supplemental Table 5.3. The relative distribution of shRNAs in $CD8^+$ T cells in the *in vivo* RNAi screen at day 8 relative to day 0.

Construct Barcode	Gene ID	Pool #	Input	Day 8 in vivo	Fold change	Fold change (Log2)
CGTCTCTAGTGCCAACTTTAT	Fos	Pool 4	8.113333245	5.906050264	0.727943755	-0.458101112
CAGAGCTGAAGACGCAGATAG	Jdp2	Pool 3	11.04336138	9.591594781	0.868539429	-0.203336751
CTGCGTTTGCAGCCGATTTAC	Etv6	Pool 3	9.798179596	8.524506901	0.870009252	-0.200897352
GATGTTAGCCCGGACACTTTC	Irf1	Pool 3	12.3942304	10.91228868	0.880432937	-0.183714977
GAGGATCATTAACGACTATAA	Crebbp	Pool 4	14.19123317	12.63493957	0.890334154	-0.167581195
GCCTCCCTGATGAATAAAGAT	Pparg	Pool 3	12.03383539	10.74490919	0.892891488	-0.163443238
TATGCACTTTGTCGCTATTAA	Hif1a	Pool 2	10.00183683	8.932817868	0.893117737	-0.163077722
TGGACGCTGCTACTGGCAAAT	Rdbp	Pool 4	12.47463108	11.16987282	0.895407066	-0.159384392
GCCCAGTGAATGTCTAGCAAA	E2f6	Pool 3	12.67006986	11.34655069	0.89553971	-0.15917069
GCTCCACACTATGAAGACATT	Pparg	Pool 3	10.25863233	9.194951841	0.896313617	-0.157924481
TCTAGCAGGATTCGCTGTAAC	Zfp367	Pool 3	11.13404555	10.0023003	0.898352738	-0.154646065
TGCCGGTGACGTGCAACAATT	Nr4a1	Pool 3	10.07383653	9.112822914	0.904603017	-0.144643287
GTTGCATAGAGAGGGTCTTAT	Irf1	Pool 3	12.85981605	11.78193849	0.916182505	-0.12629308
TGGAAGAAGAGAACCACAAAC	Xbp1	Pool 2	12.22980596	11.21689871	0.917177161	-0.124727665
CGGTTTGATGTGTCACTGGTA	E2f6	Pool 3	11.93210417	11.03913116	0.925162151	-0.11222185
CCAGGAGCAAGACCGTGTTTA	Irf7	Pool 2	10.85264033	10.11421757	0.93195916	-0.101661359
CGTACCTCTGTAACTTATGTT	Rnps1	Pool 2	13.72597261	12.845693	0.935867597	-0.095623658
GGCTAGAGATGCAGATTAATT	Irf1	Pool 3	12.99586852	12.17382372	0.936745682	-0.094270672
CCTTCTACATCGACGACATCT	Hhex	Pool 3	10.80574818	10.14300342	0.938667387	-0.091314059
GCCTTGGACGAGTTGATTAAA	E2f6	Pool 3	11.55997195	10.8538769	0.938918965	-0.090927446
CGTCATGTGTGCTGCATTATT	Hnrpll	Pool 3	11.75755368	11.07801069	0.94220371	-0.085889082
GCTGTCAATGTTGTACGCTTT	Chaf1b	Pool 3	14.53238433	13.74247767	0.945645075	-0.080629291
GTTTCTACACGAGCGTCAAAG	Zik1	Pool 3	13.62124132	12.89011992	0.946324906	-0.079592499
CTTGTACCAACACGACAATAA	Mbnl3	Pool 2	8.595557907	8.135495273	0.946476699	-0.079361106
CTGTGCCTAGAGAAGCTAAAG	Mxd1	Pool 2	13.14142877	12.45249324	0.947575295	-0.07768751
CACTCAATGGAGCTGATATAT	Hnrpll	Pool 3	13.44678021	12.77587129	0.950106352	-0.073839081
GCGTTTCACTGTGGTCTATTG	Nup43	Pool 4	14.90368148	14.18372623	0.951692791	-0.071432151
GAATCCTCTTTACCCAATTAC	Hnrpll	Pool 3	12.71330079	12.10838808	0.95241891	-0.07033183
GCCAACTTCAATCGCATCATT	Pole	Pool 2	13.64145561	12.99416533	0.952549765	-0.070133628
GCCGGGAATATGTCGACTTAG	Cbfb	Pool 3	11.05131124	10.53759638	0.953515484	-0.068671729
GACTACACTAAACCTTATTTG	Hnrpll	Pool 3	13.14949313	12.54454807	0.9539948	-0.067946692
AGTCCAAGGATCGATCTAAAG	Rnps1	Pool 2	11.2645334	10.76057461	0.955261459	-0.066032435
CCTGGTGCTCATGATACTAAA	Whsc1	Pool 2	10.77789127	10.30662394	0.956274626	-0.0645031
TGGAGAAGAAGTTCGAGACTC	Hhex	Pool 3	10.7945653	10.33050613	0.957009925	-0.063394208
ATCATCGTCTCCGGATGTAAA	Rfx7	Pool 4	11.3550628	10.880171	0.958177968	-0.061634454
TCCGTATGAGCTTCGTCAAAG	Smad3	Pool 4	12.17965945	11.67138196	0.958268334	-0.0614984
CATAGTTATTCACAGGTTATT	Rfx7	Pool 4	17.51046145	16.78952426	0.958828202	-0.060655752
ATGACCAGTGTGCAATGATTA	Zik1	Pool 3	12.91507606	12.39762892	0.959934642	-0.058991913
GCCCTTTACCACAGTTGATTT	Pparg	Pool 3	13.69855053	13.16401506	0.960978683	-0.057423666
GAGTACCTCCAGTCTGATGAG	Zfp367	Pool 3	12.02204764	11.5747391	0.96279265	-0.054702967
CGACAGCATGAAGTCGTTCTT	Chaf1b	Pool 3	13.70266927	13.19625321	0.963042525	-0.054328591
ACAGAACCGTGAGGTACTTTA	Wdr74	Pool 4	12.15643842	11.71299295	0.96352176	-0.053610846
CCAGAAAGAGCTTGGATATTT	Whsc1	Pool 2	15.07533689	14.53173879	0.963941231	-0.052982904
TGGAGAGGATGGCCGAATAAA	Nup43	Pool 4	10.37653105	10.00301722	0.964003979	-0.052888993
AGCTCCAACTCCTCCCGATAA	Rnps1	Pool 2	10.93234862	10.54576039	0.964638135	-0.051940249
CTCACTCCTGGAGCATGTTAA	E2f1	Pool 2	13.17640699	12.73906451	0.966808669	-0.048697686
GGCCCTGAGCTGGACTTATAT	Maff	Pool 3	13.37598152	12.94141633	0.967511529	-0.047649242
TCCTAGGAATCCCAGATTATT	Crebbp	Pool 4	10.02987755	9.705318984	0.967640825	-0.047456457
ACCTGCATCGTGCGCACAGAT	Jdp2	Pool 3	12.63582045	12.23176021	0.968022635	-0.046887312
TTGACTAATTGAGGTGTTAAT	Nfyb	Pool 4	11.9339361	11.55574937	0.968309975	-0.046459139
TATGCCTCCATGCTGCCATAC	Mxd1	Pool 2	11.59389388	11.23683698	0.969203021	-0.045129193
CATGTGCCTTTAAGCCTATAG	Nr4a1	Pool 3	12.39295653	12.01539748	0.969534384	-0.044636031
CCATCACCTTCCAGTAGTAAT	Hopx	Pool 3	12.44126233	12.07368231	0.970454765	-0.043267128
ACAATGACTGCAGGCATTAAT	Sp3	Pool 3	11.26692104	10.93621347	0.97064792	-0.042980009
CCCATTCCTCATCCGTCAAAT	Hif1a	Pool 2	13.07169949	12.69760068	0.971381012	-0.041890809
TAAGCCTCATCCTCATGATTT	Rel	Pool 4	10.57930444	10.28598108	0.972273853	-0.040565372
CCACATAACTTAAACGCTAAT	Pbx3	Pool 2	15.34672191	14.93045367	0.972875755	-0.039672523
CCAGAATCTCTCAAGCGTGAA	Mybl2	Pool 3	11.41734751	11.10884784	0.972979742	-0.039518327
GCTAGAAGTTAATGGGAGAAA	Mbnl3	Pool 2	15.78292153	15.36019177	0.973216001	-0.039168055
ACTTCAACAAGGTCAACAAGC	Hopx	Pool 3	13.64915649	13.29029218	0.973707949	-0.038438975
GCATCCTTTGTCTCACCAATT	Atf3	Pool 4	13.41745011	13.06511173	0.973740287	-0.038391062
AGGGCTGATCTGGATCAATAA	Irf1	Pool 3	10.74277747	10.47173221	0.974769535	-0.036866932
AGAGCAGCAGGAACTCTTAAT	Wdhd1	Pool 2	10.38773289	10.13295	0.975472714	-0.035826576
CAGACGCAGAAATGGTTTAAG	Hopx	Pool 3	11.17747047	10.90351307	0.975490215	-0.035800693
GCCATCTCTTATGTTGTAATT	Pbx3	Pool 2	14.6778306	14.32407481	0.975898632	-0.035196793
CTTGCACATGAAGCGACACAT	Klf2	Pool 4	7.89875559	7.709423237	0.976030104	-0.03500245

GTTTCTGCAGAGGGAGTCAGA	Jdp2	Pool 3	11.6903222	11.43337556	0.978020569	-0.032063288
GGTTGAGAACCAGGAGTTAAG	Xbp1	Pool 2	10.57468747	10.36079976	0.979773614	-0.029479655
GCAGGTATCTTGAGAAGCCAA	Stat3	Pool 3	11.12737581	10.90241806	0.979783397	-0.029465251
TCTAATATCCCACGGTCTAAT	Rfx7	Pool 4	12.86796814	12.60819651	0.979812537	-0.029422343
GGCTGAAAGTCCGGCTCAAAT	Maff	Pool 3	12.19348149	11.94749009	0.979825992	-0.029402533
TCGGGAGTGACGAGGGTTATT	Mxi1	Pool 4	8.832671601	8.659087027	0.980347444	-0.028634951
CGTCATTCTGTTGTGGAAGAT	Chaf1b	Pool 3	13.64093607	13.38589744	0.981303436	-0.027228783
ACTATCAGATGACGCAAATAA	Rel	Pool 4	13.65206244	13.40130989	0.981632625	-0.026744896
TTACCTTTGTACCAATCAATA	Sp3	Pool 3	12.84515586	12.61057368	0.981737693	-0.026590487
TCAGCACCAAGACCATAATAG	Zik1	Pool 3	11.81003835	11.59509089	0.981799597	-0.026499521
TGATTATGAGGAGAGCATTAA	Wdhd1	Pool 2	14.09166698	13.84486426	0.98248591	-0.025491377
TCAGAACCATGACGAAGATAA	Etv6	Pool 3	12.22753303	12.01456406	0.982582834	-0.025349059
CCCTTATCACATGTGTGGATT	Wdr74	Pool 4	14.55456348	14.31565379	0.983585238	-0.023878012
ATTTCAGAAGCTAGCTCTATT	Tgif1	Pool 2	14.78615828	14.55104331	0.984098982	-0.023124664
GCTGGATTGATCTCCACAGAT	Cbfb	Pool 3	14.29750129	14.07173115	0.98420912	-0.022963211
CACTTTGGTATCCCACACCTA	Pole	Pool 2	12.95455756	12.76129549	0.985081538	-0.021684949
CTGTCCAATGTCAACCGGAAT	Smad3	Pool 4	9.307669626	9.170534194	0.985266405	-0.021414228
GTAAGGTAGTGTTAACTATAT	Hmgb3	Pool 2	14.78861009	14.57277378	0.985405234	-0.021210961
AGTACAGATGTACCGCAAATA	Tgif1	Pool 2	10.97529127	10.81997344	0.985848409	-0.02056227
ATCAGTTACTGTGCGATATTA	Nup43	Pool 4	9.332745584	9.205238892	0.986337708	-0.019846405
AGAAGGAACATTGCAGAGCTA	Atf3	Pool 4	12.60170466	12.43330245	0.986636554	-0.019409357
GCAGCGTTAGAGAACGACAAA	Wdr5	Pool 2	13.92401266	13.74573724	0.987196548	-0.018590745
CCCAACCAAGTACCCTCAGAA	Wdr74	Pool 4	11.9330254	11.78219988	0.987360664	-0.018350925
GTGACCTTAGATAGCGTTAAT	Klf12	Pool 3	12.64246284	12.48308953	0.987393809	-0.018302496
GTGGAGATGATACCCAACATA	Nr4a2	Pool 2	13.8648281	13.69922977	0.988056229	-0.017334948
CCTGAGTTGAATTATCAGCTT	Stat3	Pool 3	14.92392446	14.75423704	0.98862984	-0.016497643
GCTGTTTGGAGACTGGCTATT	Irf7	Pool 2	13.7221665	13.57739025	0.989449461	-0.015302075
GCAATTGGTGGCTCATCATTT	Sp3	Pool 3	11.78535929	11.66246544	0.989572329	-0.015122936
CAATTAGAAACGACCAATATT	Fos	Pool 4	14.08680826	13.9475392	0.990113511	-0.014334163
ACTAGAGAATCAGCTTATAAT	Klf12	Pool 3	13.96458689	13.82792429	0.990213631	-0.014188285
GTATCCAGGGTCCGAGGAAAT	Klf12	Pool 3	13.40763518	13.27662615	0.990228774	-0.014166223
TTCGCCCTGTGCACCATTATT	Pole	Pool 2	12.09503116	11.9834701	0.99077629	-0.01336875
CGCCCAGTCGTGCAGGTATAA	Mafb	Pool 4	10.92951839	10.83074021	0.990962257	-0.013097984
AGCGTCTGGCCAAGATGTTAC	Hhex	Pool 3	9.986854752	9.898832773	0.991186216	-0.01277197
AGCTCCCAAGAACTATGTATG	Mybl2	Pool 3	12.9207928	12.82480176	0.992570809	-0.010758069
GCTCGAAGAAGAACTCGAGAA	Cbfb	Pool 3	11.15714978	11.07565415	0.992695658	-0.010576612
GCTAGTTCAGAAGGCTGATCA	Zfp367	Pool 3	10.61033121	10.53425097	0.992829608	-0.010381955
TGTGGCTTTCAACATTTCAAA	Chaf1b	Pool 3	14.42690724	14.33043117	0.993312768	-0.009680039
GCAGAATATGAACATCGACAT	Nr4a2	Pool 2	14.39685967	14.30399956	0.993549974	-0.00933556
AGTCGACACAGCCTCGATATG	Hif1a	Pool 2	9.35744376	9.300130101	0.993875073	-0.008863574
CTATTGTGGACAAGATCTTTA	Nr4a1	Pool 3	11.07345626	11.01455002	0.994680411	-0.00769503
TGGATAGCGATATGGTCAATG	Hif1a	Pool 2	13.724144	13.65228542	0.994764076	-0.007573687
TTCGGAATGCAGATCTGTTAC	Hopx	Pool 3	9.899519147	9.849401069	0.994937322	-0.007322452
GCCAAGAAGTCCAAGAATCAT	E2f1	Pool 2	11.38163692	11.33502132	0.995904315	-0.005920959
TGCAATTCAGCAAGTAGTTAG	Gabpb1	Pool 3	15.95454502	15.90818533	0.997094264	-0.004198193
CACTCGGTTTGCTCAACAAAG	Mxi1	Pool 4	13.49130572	13.45547039	0.997343821	-0.003837155
CAATAGATAGTACAGGTATAT	Sp3	Pool 3	15.64810182	15.61885951	0.998131255	-0.002698551
CCCGATGGACAGCAAGTATTG	Gabpb1	Pool 3	11.56038466	11.53878781	0.998131822	-0.002697731
GTTGCTGACTATAAGTCTAAA	Hmgb3	Pool 2	14.79028892	14.76672698	0.998406931	-0.002300145
TTATCTTGCGCCAAGACAATT	Irf7	Pool 2	11.15543187	11.14170466	0.998769459	-0.001776388
TATTGCCGGAAAGCCTATAAG	Pole	Pool 2	12.89949276	12.88922694	0.999204169	-0.001148599
CTCTGCAGCAAGTTGCCTATC	Zfp414	Pool 2	8.040316714	8.03696461	0.999583088	-0.000601602
CTGCAGAACAGATGGTCATAG	E2f1	Pool 2	14.00678824	14.00870448	1.000136807	0.000197358
GAGACAACAGATGTAAGGTTA	Mybl2	Pool 3	14.1291085	14.1356342	1.000461862	0.000666172
GCGGAGACAGATCAACTTGAA	Fos	Pool 4	11.8384367	11.84747491	1.000763463	0.001101024
ATCAAGCCTGTAAGGACTATT	Carhsp1	Pool 4	14.85718092	14.87439707	1.001158777	0.001670793
TGTGTTTAGAACGCTTGAAAG	Mxi1	Pool 4	11.07614088	11.08904753	1.001165266	0.001680145
CCAACCTCAGACGACAATTTC	Crebbp	Pool 4	11.59792018	11.61390016	1.001377831	0.001986421
ACAGTGCTTCTGACCGGATTA	Nmi	Pool 4	11.94865502	11.96722144	1.00155385	0.002239992
TTGCCCTATATGCCGTATTTC	Pole	Pool 2	12.07267203	12.09210801	1.001609915	0.002320749
AGACATCGTGGAGAATTATTT	Trim28	Pool 2	14.6634751	14.68789023	1.00166503	0.002400133
CGCCTGGCATACCGATCTAAA	Nr4a1	Pool 3	10.73255931	10.75636133	1.00221774	0.00319598
GCTATGCCATTGGGAAATAAA	Rel	Pool 4	13.1863032	13.21626914	1.002272505	0.003274813
TATTGCCCTTGTTAGGTTTAA	Hmgb3	Pool 2	13.60646607	13.64800418	1.003052822	0.004397582
ACTTCGAAGGGTGCAACAAAG	Klf12	Pool 3	10.20156683	10.23284261	1.003065782	0.004416223
GACTCTTCCACACTCTTATTT	Maff	Pool 3	12.31903105	12.35712725	1.003092468	0.004454604
GCGTATGTAACCTATCAGGAT	E2f6	Pool 3	13.05043407	13.09146887	1.003144325	0.004529185

GATGTGCTTCCCTTATCAAAC	Sfpi1	Pool 2	11.98815873	12.03221358	1.003674863	0.00529199
TTGCACGAATCTAAGTTATTC	Maff	Pool 3	11.53204788	11.57587337	1.003800322	0.005472314
AGTAGGTCAAACGTGCTATTT	Carhsp1	Pool 4	11.16995597	11.21482223	1.004016691	0.005783253
TAGCGATCACTACTGGGATTT	Sfpi1	Pool 2	11.66149655	11.71318166	1.004432116	0.006380064
CCTAACTTTGTGGTTCCAGAT	Stat3	Pool 3	13.06526831	13.12585184	1.004636991	0.006674302
CCCACTCCTTCACAATCATAC	Whsc1	Pool 2	13.20007052	13.26177609	1.00467464	0.006728365
CCGTCTACAAAGGAGTCTGTA	Carhsp1	Pool 4	11.82254277	11.87844518	1.00472846	0.006805648
CATTGGGCAACAGAACATAAT	Gabpb1	Pool 3	10.11936348	10.16837917	1.004843752	0.006971188
CCTAAACAACGTGTTGGACTT	Klf2	Pool 4	14.83863913	14.92415197	1.005762849	0.008290169
CATCTACTGGCCACTTCAAAT	Klf10	Pool 4	11.38633836	11.45574139	1.00609529	0.008766953
CGATTTCTTAACTCCAGAGTT	Nr4a2	Pool 2	14.51100314	14.60650114	1.006581075	0.009463378
CACCAAGAGCTCGCACCTAAA	Klf2	Pool 4	13.54385152	13.63357766	1.006624861	0.009526135
CTACTTCAAGTGTGAGAATTG	Zfp414	Pool 2	10.16169567	10.22955181	1.006677639	0.009601774
CCACATGTTGAAGAGACTAAA	Mxd3	Pool 3	15.22109987	15.32336383	1.006718566	0.009660426
GCTAGAGCATAAGGTTTATCT	Wdr74	Pool 4	14.34566871	14.44394417	1.006850532	0.00984953
GGACTACAATCTGATTGTTAT	Trim28	Pool 2	13.92979974	14.03046557	1.007226653	0.010388365
TGAACTTACAGTCTAACAAAC	Klf12	Pool 3	12.66892208	12.76558362	1.007629815	0.010965716
CCCACGCATCAGACTTCTGTA	Carhsp1	Pool 4	12.93625352	13.04215712	1.008186574	0.011762647
GTTCAAAGTGGACAGCTCAAA	Zfp367	Pool 3	13.34517545	13.45512849	1.00823916	0.011837895
GATGATCTCAAACTTCGTTAA	Cbfb	Pool 3	11.23237341	11.34173449	1.009736239	0.013978485
GATTGCTATGCAGAACCAAAT	Gabpb1	Pool 3	14.45656029	14.59764629	1.009759306	0.014011442
ACGGAGGAACAAGCCCAAGAA	Mxd1	Pool 2	10.51163292	10.6169017	1.010014503	0.014376009
CAAGTTCATCTGCTGATAAAC	Wdr5	Pool 2	12.68808669	12.82038552	1.010427012	0.014965112
AGAGAGGCAGAGCACGGTTAT	Mxd3	Pool 3	12.16635457	12.29596753	1.010653393	0.015288304
TTGGGTCAGATGATCCCTATT	Trim28	Pool 2	12.83710815	12.98780134	1.011738874	0.016836983
CGTGAATTTCAGCGTGGAAAT	Mbnl3	Pool 2	9.1020766	9.209872763	1.01184303	0.016985498
CTGTTACAGTTACTCATATTA	Hnrpll	Pool 3	13.4727193	13.642211	1.012580363	0.018036412
ACGCTGACATCCGCAACATTG	Jdp2	Pool 3	11.85967482	12.0150837	1.013103975	0.018782246
CGTGTGAAGCAGTTATGATTT	Pbx3	Pool 2	9.452396338	9.582109415	1.013722772	0.019663164
TCCTTCGACTGTTCACTATTG	Wdhd1	Pool 2	14.7317192	14.93621981	1.013881653	0.019889261
GCCCTGGCAAAGCATTTGTAT	Pparg	Pool 3	12.75653512	12.93431562	1.013936426	0.019967197
GCGGAAACATGCGGCAAATTT	Wdr74	Pool 4	12.47588225	12.65539885	1.01438909	0.020611135
AGTTACCAGCTGGCTTAATAA	Nfyb	Pool 4	12.58600287	12.76798929	1.01445943	0.02071117
GACAGTCATCGGCCATGTTTA	Hhex	Pool 3	12.75698529	12.94868824	1.015027293	0.02151852
TACATCGGAGGAAGTCATTAT	Nfyb	Pool 4	13.53779891	13.74387739	1.015222451	0.021795879
CAGAGATGAGGGCGATGAATA	Nfkbib	Pool 4	12.83361124	13.03108939	1.015387575	0.022030511
ACGTGAACACCAAGTGCATTA	Smad3	Pool 4	10.51928549	10.69911579	1.017095296	0.024454858
AGCACTGGTGTGGTAGCTAAA	Trim28	Pool 2	14.29721398	14.54840021	1.017568893	0.025126473
CCATACCCTCTGTCCCAATAG	E2f1	Pool 2	13.52007218	13.76180954	1.017879886	0.025567328
TCGACATTAAAGCGACATTAA	Etv6	Pool 3	12.39434615	12.62472932	1.018587763	0.02657029
CAGTGTCTCCTGTCGAGAATA	Etv6	Pool 3	11.132381	11.34279534	1.018901109	0.027014035
TCAGACTCAGACCAAGAGGAT	Mxd3	Pool 3	13.28301792	13.53524108	1.018988393	0.027137618
CCAAAGAATGTGTTCAGGAAT	Nfyb	Pool 4	13.75832217	14.01984065	1.019008021	0.027165408
AGGATGGAATGTTGAGTATTT	Nup43	Pool 4	12.55925066	12.80341563	1.019441046	0.027778348
ACACCAGTCTCCGGACAACAA	Zfp367	Pool 3	12.49320019	12.73772046	1.019572269	0.027964039
CTGAATGATCACCGTAGAATA	Zik1	Pool 3	13.64458997	13.91822092	1.020054172	0.028645771
GCATCCAAACCTCTTCCTGAT	Nfkbib	Pool 4	14.38507612	14.67469409	1.020133225	0.028757574
GTGGCCACCTAGTGGAGTTAA	Irf7	Pool 2	12.00993584	12.25303942	1.020241872	0.028911218
CGGATGCCAGAGAATTCCAAA	Nmi	Pool 4	11.24415714	11.47344324	1.020391578	0.029122897
CATCGCTCTGAATAATGAGAA	Egr1	Pool 4	10.58916561	10.81158586	1.021004511	0.029989241
CTGTCAGAACTATAGATAATG	Nup43	Pool 4	13.20453101	13.48221107	1.02102915	0.030024055
GCACTCGGATACCTCTAACTA	Pbx3	Pool 2	12.62786323	12.89697901	1.021311268	0.030422627
GCAGTTGAGATATCTCCTCTA	Rnps1	Pool 2	14.06885186	14.37577383	1.021815709	0.03113502
CTGTGAGAGAAGAGGATATTC	Whsc1	Pool 2	12.79466463	13.0830052	1.022536001	0.032151637
CGGTTTACTACAAGCCCTCTT	Nr4a2	Pool 2	13.60860857	13.91646354	1.022622075	0.032273075
TTGATAGCCACTGTGATATTT	Wdr5	Pool 2	12.80480497	13.09799723	1.022897049	0.032660951
GAATGTGACCAAGGATCATAT	Rnps1	Pool 2	11.3790361	11.64709287	1.023557072	0.033591547
GGGCTTTCCCGCCTTTGTTTA	Wdhd1	Pool 2	11.30541629	11.58289939	1.024544263	0.034982313
GCTGAGTCTTTGTTTGGGTTT	Mafb	Pool 4	12.15780329	12.45860465	1.024741423	0.035259915
CGACTTTGATTTCAACTACAA	Stat3	Pool 3	12.80322687	13.12313401	1.024986446	0.035604832
GATCCAGTTGATCCGCATAAG	Irf7	Pool 2	11.47853566	11.7745359	1.025787282	0.03673159
GTCTCCCTGTGGGCAGTATTT	Wdhd1	Pool 2	13.15647691	13.49599261	1.025805974	0.036757878
GCATTAAGCAGGAACGAATTT	Etv6	Pool 3	12.90584942	13.23889875	1.025806076	0.036758021
CTTTCAAGTGAGCTCACAAAT	Nmi	Pool 4	11.88744665	12.19557994	1.025920898	0.036919499
CAGAATAAACACCTCTGCCAT	Atf3	Pool 4	13.64526695	14.01621156	1.027184856	0.038695837
CTACAGTAGAGAAGTTGTATG	Nfkbib	Pool 4	12.88337205	13.24177416	1.027818967	0.039586181
CCGTGGATGTAGCTAAGTATA	Klf10	Pool 4	13.22103991	13.59209642	1.028065607	0.039932334

TGGATTCCTTGTGCCTCATAT	Rdbp	Pool 4	10.71051759	11.01376782	1.028313313	0.0402799
ATGGAAGTTAGCAAGTTAAAT	Klf10	Pool 4	13.86040646	14.25334346	1.028349602	0.040330812
TTAGAGACACTAGGAGTAAAG	Smad3	Pool 4	13.20500263	13.5829706	1.028623089	0.040714443
CACTCCACTATCCACTATTAA	Egr1	Pool 4	13.09286651	13.46888737	1.028719522	0.040849688
ACCTCAATAAACCGGAGCCTA	Nfkbib	Pool 4	13.32571086	13.70926854	1.028783282	0.040939103
GACCCTGGAAGGGAAGTTAAA	Rdbp	Pool 4	12.13382863	12.49171955	1.0294953	0.041937244
GCCGTTCATTTCAACCGTGAT	Wdr5	Pool 2	12.81720945	13.20050314	1.029904612	0.042510724
GACTTGATTTGCATGGTATTG	Egr1	Pool 4	14.33359897	14.76311646	1.029965781	0.042596407
TACTCCGGGCTGCACTACTTA	Fos	Pool 4	12.24091905	12.61019864	1.030167635	0.042879121
TGGATATGGAAAGGGACATAA	Irf1	Pool 3	14.31659166	14.74992419	1.030267856	0.043019468
CTACTCCCAACACTGACATTT	Egr1	Pool 4	12.29720804	12.67034283	1.030343049	0.043124757
ATGAGAAACTGTTGGTATTTA	Sp3	Pool 3	11.99436612	12.37588319	1.031808023	0.04517457
CCCAGCTTCAGTCCGACTGAA	Mafb	Pool 4	10.87175853	11.22212681	1.032227378	0.045760801
GACCGATTGTATTTCTATAAG	Klf2	Pool 4	13.18007977	13.61333315	1.032871833	0.046661244
TCTGCTAAAGAACTCGGACAT	Mybl2	Pool 3	12.59963424	13.02317306	1.033615168	0.047699146
CACCGATTCCACTCAACTAAG	Smad3	Pool 4	14.30379018	14.79147564	1.034094842	0.048368508
GTAGCTATTCCCTCGACTTTA	Hhex	Pool 3	12.54197535	12.9704916	1.034166567	0.048468571
GCCAAACTATGCCCTGAAGTT	Wdr5	Pool 2	13.12222854	13.57589282	1.034572198	0.049034328
CGAATGTTTCTTGGTAGTTTC	Tgif1	Pool 2	11.78535865	12.19684724	1.034915236	0.04951261
AGCGACTTGAGTGCATCTAAT	Rel	Pool 4	9.38695703	9.719179247	1.035391897	0.050176933
TCTGGTTCCCTGGACGTTATC	Nr4a1	Pool 3	12.20421744	12.63695061	1.035457675	0.050268584
AGTCTTTGGAACGGATCAATG	Maff	Pool 3	12.22258708	12.65747574	1.035580738	0.050440036
CGGAGAAGAACCCACGGAAAT	Klf10	Pool 4	7.883751608	8.173332979	1.036731417	0.052042188
TTCCGCAGGTCAGACTCATTC	Rdbp	Pool 4	8.366353102	8.689640317	1.038641354	0.054697574
GATGGCAAGAGATGCATTATT	Tgif1	Pool 2	12.1866901	12.66213159	1.039013176	0.05521395
TCTACCCAGAAGGACCTAGTT	Xbp1	Pool 2	12.86870385	13.37366727	1.039239649	0.055528378
TAGTCAAAGTGCCATTCTTAA	Zik1	Pool 3	11.72313198	12.18366557	1.039284177	0.055590192
TCGGATGACTTGGTTACTTAC	Sfpi1	Pool 2	10.87179527	11.3037321	1.039730037	0.056208985
GAGATGGAGCGGATACGAATG	Mxi1	Pool 4	11.27071198	11.72097462	1.039949796	0.056513883
AGGCAGATAAAGTCCGATATG	Hmgb3	Pool 2	12.00150598	12.48449751	1.040244243	0.056922305
CTACAAGGTCAAGTGCGAGAA	Mafb	Pool 4	12.33561877	12.83366211	1.040374411	0.05710282
TAGTGGATGTTGCACTCAAAC	Tgif1	Pool 2	10.0352571	10.45279525	1.041607121	0.058811216
GGAGATCCTGGAGTACAACTT	Hopx	Pool 3	11.85126681	12.3598876	1.042916998	0.060624343
CTTCAAGCATCTGCATGTTTG	Zfp414	Pool 2	12.84035307	13.40097825	1.043661197	0.061653446
CTTCACTGCACGGGCCATATT	Hif1a	Pool 2	12.15294235	12.68762151	1.043995861	0.062115992
GACCTTACCTGTTCGTGAAAC	Fos	Pool 4	12.21586558	12.75467013	1.044106948	0.062269495
AGAAGATGGAGACGCAGAATA	Klf10	Pool 4	12.79691525	13.36267317	1.044210492	0.06241256
GAGCTATACCAACGTCCAATG	Sfpi1	Pool 2	11.49544396	12.01380885	1.045093072	0.06363143
GCACCAGTAGCAGATCAACTC	Mxd1	Pool 2	11.61394067	12.1436803	1.045612393	0.064348146
GAAGTGAAAGAAGCTATTATT	Rel	Pool 4	12.47337883	13.0520101	1.046389297	0.065419689
ACAGATGCACGTCCTCGATAC	Sfpi1	Pool 2	10.62760811	11.12064504	1.046392088	0.065423537
TCTTGGCTACCATCCATTAAG	Rfx7	Pool 4	7.749494668	8.119914149	1.047799179	0.067362236
AGAGCTGACTTTCAACTATAA	Whsc1	Pool 2	13.89075163	14.56138305	1.048278987	0.068022724
TCCGCAGCACTCAGACTATGT	Xbp1	Pool 2	10.48768957	11.00306614	1.049141096	0.069208715
GACCCTTTCAGTGCCACTTGT	Klf2	Pool 4	13.02086893	13.67190043	1.049999083	0.070388067
GCAGGATAATCAGGCATTAAT	Mxi1	Pool 4	12.27168123	12.88537636	1.050009051	0.070401764
GCGATGGTGGAGACGAGTTAT	Egr1	Pool 4	10.90974959	11.47594541	1.05189815	0.072995023
CAACGTCTTACCAACAGATTT	Nfyb	Pool 4	13.09856339	13.78002306	1.052025528	0.073169712
AGGCCTGTCTCTTTCGTTAAA	Xbp1	Pool 2	6.992335016	7.361602196	1.052810281	0.074245483
GCAAACCTCATGTACCACCTT	Mxd1	Pool 2	12.03540581	12.67240844	1.052927391	0.074405953
AGCAGCTCTGCCACTTCATAT	Nfkbib	Pool 4	9.422431048	9.946078957	1.055574608	0.078028552
GCAGTGAGGAACCAACGTAAA	Trim28	Pool 2	11.40957665	12.08194374	1.058930065	0.082607312
CCTTGTAAAGAGTCACTATAT	Rfx7	Pool 4	10.68098746	11.35982529	1.063555718	0.088895615
CTCAGCAGAGATGGACGATAT	Nmi	Pool 4	10.63132236	11.31023424	1.063859589	0.089307753
TGCTGCCAAGTGTCGAAACAA	Atf3	Pool 4	10.06438699	10.71732931	1.064876512	0.090686139
CTGGACAAGAAGTGACTGAAG	Jdp2	Pool 3	8.391186163	8.954903056	1.067179643	0.093803052
TAACTCTGGCCATAGCTTAAT	Crebbp	Pool 4	10.02817286	10.72016078	1.069004387	0.096267773
ACTTGAACCCTGCACCCTTTG	Zfp414	Pool 2	11.43054087	12.24885897	1.071590496	0.099753692
ACCCAGAGGTTCCCGTCAATT	Hmgb3	Pool 2	11.25159237	12.12376393	1.077515389	0.107708474
CTGAGCTGCACTGAGAGTTTC	Zfp414	Pool 2	10.11601629	10.92735838	1.080203716	0.111303416
AGGCCCTTGACTATCACTTTG	E2f1	Pool 2	10.06329949	10.94878203	1.087991274	0.121666985
CGACGAAGTCACCTATAAGAT	Carhsp1	Pool 4	10.40308096	11.42410024	1.098145855	0.135069684
GCAAGTCATCACGATAGTTAC	Gabpb1	Pool 3	11.01473504	12.47421247	1.132502273	0.179513946

Supplemental Table 5.4. The list of shRNAs and target sequences used in the RNAi screen.

Gene ID	Gene Symbol	Broad Clone ID	Target Sequence	Type
-10	*GFP*	TRCN0000231753	ACAACAGCCACAACGTCTATA	Control
-10	*GFP*	TRCN0000231762	CCACATGAAGCAGCACGACTT	Control
-10	*GFP*	TRCN0000231765	CCTACGGCGTGCAGTGCTTCA	Control
-10	*GFP*	TRCN0000231746	TGCCCGACAACCACTACCTGA	Control
-10	*GFP*	TRCN0000231745	TGACCCTGAAGTTCATCTGCA	Control
-15	*lacZ*	TRCN0000231738	CCGTCATAGCGATAACGAGTT	Control
-15	*lacZ*	TRCN0000231702	TCGTATTACAACGTCGTGACT	Control
-15	*lacZ*	TRCN0000072236	CCAACGTGACCTATCCCATTA	Control
-15	*lacZ*	TRCN0000231708	GCGCTAATCACGACGCGCTGT	Control
-15	*lacZ*	TRCN0000231710	CGCTAAATACTGGCAGGCGTT	Control
-15	*lacZ*	TRCN0000231706	GTCGGCTTACGGCGGTGATTT	Control
-15	*lacZ*	TRCN0000231722	CGCGATCGTAATCACCCGAGT	Control
-15	*lacZ*	TRCN0000231704	GCGTTGGCAATTTAACCGCCA	Control
-14	*Luciferase*	TRCN0000231730	AGAATCGTCGTATGCAGTGAA	Control
-14	*Luciferase*	TRCN0000231737	ATGTTTACTACACTCGGATAT	Control
-14	*Luciferase*	TRCN0000231707	CACTCGGATATTTGATATGTG	Control
-14	*Luciferase*	TRCN0000231739	AGTACTTCGAAATGTCCGTTC	Control
-14	*Luciferase*	TRCN0000231733	ACTTACGCTGAGTACTTCGAA	Control
-14	*Luciferase*	TRCN0000231742	ACGCTGAGTACTTCGAAATGT	Control
	nullT			Control
-12	*RFP*	TRCN0000231682	GCTTCAAGTGGGAGCGCGTGA	Control
-12	*RFP*	TRCN0000231725	ACTACACCATCGTGGAACAGT	Control
-12	*RFP*	TRCN0000231724	CCGTAATGCAGAAGAAGACCA	Control
-12	*RFP*	TRCN0000231687	TGCAGAAGAAGACCATGGGCT	Control
-12	*RFP*	TRCN0000231691	GAACGGCCACGAGTTCGAGAT	Control
223921	*Aaas*	TRCN0000241729	GAATTTGCCCAAGTAACTAAC	Experimental
223921	*Aaas*	TRCN0000241727	ATGCTCAGTTTCCACGCTTTA	Experimental
223921	*Aaas*	TRCN0000241728	CTGCTTCCCTGTGGTATTATT	Experimental
226641	*Atf6*	TRCN0000321327	GGCAAAGCAGCAGTCGATTAT	Experimental
226641	*Atf6*	TRCN0000374112	GTCCAATGACAAAGCTTTAAT	Experimental
226641	*Atf6*	TRCN0000350564	GCTGTCTGTGTGATGATAGTA	Experimental
53314	*Batf*	TRCN0000084768	CTGGACAAGTATTGAACACAA	Experimental
53314	*Batf*	TRCN0000084769	GCTCTCCGCAAAGAGATCAAA	Experimental
53314	*Batf*	TRCN0000084770	TCATCTGATGATGTGAGGAAA	Experimental
53314	*Batf*	TRCN0000084771	GAGGTGGTATACAGTGCCCAT	Experimental
53314	*Batf*	TRCN0000084772	CCGCAAAGAGATCAAACAGCT	Experimental
53314	*Batf*	TRCN0000235888	GTAGTGGTCAGAGGGACTTAA	Experimental
53314	*Batf*	TRCN0000235887	CTCTCCGCAAAGAGATCAAAC	Experimental
53314	*Batf*	TRCN0000235886	CTGGCAAACAGGACTCATCTG	Experimental
53314	*Batf*	TRCN0000235884	GAGCTCAAGTACTTCACATCA	Experimental
53314	*Batf*	TRCN0000235885	CCCTGACCTTCTGGACAAGAA	Experimental
12053	*Bcl6*	TRCN0000350566	CAACCTGAGGGAAGGCAATAT	Experimental
12053	*Bcl6*	TRCN0000321437	TGTCAAAGAGAAGGCTTTAAT	Experimental

12053	*Bcl6*	TRCN0000321365	GATGTTCTTCTCAACCTTAAT	Experimental
12053	*Bcl6*	TRCN0000084653	GCTGTCAAAGAGAAGGCTTTA	Experimental
12053	*Bcl6*	TRCN0000350566	CAACCTGAGGGAAGGCAATAT	Experimental
12053	*Bcl6*	TRCN0000321437	TGTCAAAGAGAAGGCTTTAAT	Experimental
12053	*Bcl6*	TRCN0000084657	TGAGCAGTTTAGAGCCCATAA	Experimental
12053	*Bcl6*	TRCN0000321365	GATGTTCTTCTCAACCTTAAT	Experimental
12053	*Bcl6*	TRCN0000084656	GAGAAGTGTAACCTGCACTTT	Experimental
12053	*Bcl6*	TRCN0000084654	CCGGCTCAATAATCTCGTGAA	Experimental
12053	*Bcl6*	TRCN0000321439	GGCAAGTCCCTAATGAGTATA	Experimental
12053	*Bcl6*	TRCN0000084655	CGGCCTGTTCTACAGTATCTT	Experimental
12053	*Bcl6*	TRCN0000321438	CAAGCCAGCCGGCTCAATAAT	Experimental
20893	*Bhlhe40*	TRCN0000226078	GTAGTGGTTTGGGCAAATTTC	Experimental
20893	*Bhlhe40*	TRCN0000218239	AGAACGTGTCAGCACAATTAA	Experimental
20893	*Bhlhe40*	TRCN0000226075	CATCTCAAACTTACTACTTTG	Experimental
20893	*Bhlhe40*	TRCN0000226076	TTACGTTGAAGCACGTGAAAG	Experimental
20893	*Bhlhe40*	TRCN0000081854	GCGGTTTACAAGCTGGTGATT	Experimental
20893	*Bhlhe40*	TRCN0000226078	GTAGTGGTTTGGGCAAATTTC	Experimental
20893	*Bhlhe40*	TRCN0000081853	GCGAGGTTACAGTGTTTATAT	Experimental
20893	*Bhlhe40*	TRCN0000218239	AGAACGTGTCAGCACAATTAA	Experimental
78653	*Bola3*	TRCN0000265294	GATGGTCAATCAGGCACTAAA	Experimental
78653	*Bola3*	TRCN0000250005	GACTGACTGCTCACGCTTAAA	Experimental
78653	*Bola3*	TRCN0000250006	ATCGAATCGGAAGAATTTAAA	Experimental
105246	*Brd9*	TRCN0000225739	CACCGAATGGTGTCCAATAAG	Experimental
105246	*Brd9*	TRCN0000225737	TGGACTTTGGCACGATGAAAG	Experimental
105246	*Brd9*	TRCN0000225738	ATAATGCGATGACGTACAATA	Experimental
105246	*Brd9*	TRCN0000225736	GCTCGTCCTACGAAGATTATA	Experimental
52502	*Carhsp1*	TRCN0000331494	CCCACGCATCAGACTTCTGTA	Experimental
52502	*Carhsp1*	TRCN0000301438	CCGTCTACAAAGGAGTCTGTA	Experimental
52502	*Carhsp1*	TRCN0000310872	ATCAAGCCTGTAAGGACTATT	Experimental
12448	*Ccne2*	TRCN0000313801	TCAGCCCTTGCATTATCATTG	Experimental
12448	*Ccne2*	TRCN0000313873	CCTCACCTCTGCCAGATTTAA	Experimental
12448	*Ccne2*	TRCN0000317391	CCACAGATGAGGTCAATACTT	Experimental
107995	*Cdc20*	TRCN0000280230	GCAGCAGAAACGACTTCGAAA	Experimental
107995	*Cdc20*	TRCN0000297357	CGGAATGACTACTACCTGAAT	Experimental
107995	*Cdc20*	TRCN0000280170	GCCGAACTCCTGGCAAATCTA	Experimental
12580	*Cdkn2c*	TRCN0000321369	GGAGTTCCAGGCTGATGTTAA	Experimental
12580	*Cdkn2c*	TRCN0000350567	GTCCGTTTCACTATCACTTAT	Experimental
12580	*Cdkn2c*	TRCN0000321440	TGGAAGAAATGAGGTCATTAG	Experimental
29876	*Clic4*	TRCN0000366442	CAGCGAAGTCAAGACGGATGT	Experimental
29876	*Clic4*	TRCN0000375065	ACTGGATGAGTACCTCAACTC	Experimental
29876	*Clic4*	TRCN0000375120	GAAGTGATGGTGAAAGCATTG	Experimental
26895	*Cops7b*	TRCN0000336386	TCCGGAAGAAAGATATCAATA	Experimental
26895	*Cops7b*	TRCN0000374705	ACCTCTTGGAGCAGTTCATTT	Experimental
26895	*Cops7b*	TRCN0000374706	AGGGAACTAGAAGACCTTATC	Experimental

54188	*Cpsf4*	TRCN0000317504	GCCATGTCTGTCCTTTCATTT	Experimental
54188	*Cpsf4*	TRCN0000317503	GTGTGAGTTCTTGCATGAATA	Experimental
54188	*Cpsf4*	TRCN0000313916	TGAGAGGACTGTGCAAGAAAG	Experimental
54188	*Cpsf4*	TRCN0000313917	GCTACAAGTGTGGTGAGAAAG	Experimental
228410	*Cstf3*	TRCN0000337863	GTCAGAGAAACCAAGGGTAAA	Experimental
228410	*Cstf3*	TRCN0000337867	ATACGATGAGTCGGCTTAAAC	Experimental
228410	*Cstf3*	TRCN0000337864	GCCAAGTTATTTAGTGACGAA	Experimental
27214	*Dbf4*	TRCN0000334054	CGCAGTCATGTGGACAAGTTT	Experimental
27214	*Dbf4*	TRCN0000334124	CTGCCGTGTATAAACTACTTT	Experimental
27214	*Dbf4*	TRCN0000334126	GACAGCTATATTCCTGAAGTT	Experimental
104721	*Ddx1*	TRCN0000309156	CAGAGGAGCATTGGAGTCATA	Experimental
104721	*Ddx1*	TRCN0000305216	GATGTGGTCTGAAGCTATTAA	Experimental
104721	*Ddx1*	TRCN0000305217	ACAACCAGATTCCGCAGATTA	Experimental
76843	*Dtl*	TRCN0000243406	GAACGGGTGGTCTTCACATTA	Experimental
76843	*Dtl*	TRCN0000243405	TGATGAAGCTGCCTACATTTG	Experimental
76843	*Dtl*	TRCN0000243404	GCACCAGCTGTCTAGTATATT	Experimental
13555	*E2f1*	TRCN0000374127	CTCACTCCTGGAGCATGTTAA	Experimental
13555	*E2f1*	TRCN0000312761	CCATACCCTCTGTCCCAATAG	Experimental
13555	*E2f1*	TRCN0000374126	CTGCAGAACAGATGGTCATAG	Experimental
13605	*Ect2*	TRCN0000336496	TGAGCAAGGAAATGATCATTA	Experimental
13605	*Ect2*	TRCN0000336437	GCAGTTGATGACTTTAGAAAT	Experimental
13605	*Ect2*	TRCN0000336440	TCGATTGTTTGATGTACTTAA	Experimental
14011	*Etv6*	TRCN0000231236	CAGTGTCTCCTGTCGAGAATA	Experimental
14011	*Etv6*	TRCN0000231235	TCAGAACCATGACGAAGATAA	Experimental
14011	*Etv6*	TRCN0000257202	CTGCGTTTGCAGCCGATTTAC	Experimental
14056	*Ezh2*	TRCN0000304506	GCACAAGTCATCCCGTTAAAG	Experimental
14056	*Ezh2*	TRCN0000301834	CGGCTCCTCTAACCATGTTTA	Experimental
14056	*Ezh2*	TRCN0000304505	ACTTGCCCACCTCGGAAATTT	Experimental
14235	*Foxm1*	TRCN0000349136	CGCTACTTGACATTGGACCAA	Experimental
14235	*Foxm1*	TRCN0000304362	ACTTAGAGAGGCCTATCAAAG	Experimental
14235	*Foxm1*	TRCN0000310953	GCGTATTTCCTTAGCTCATTA	Experimental
14390	*Gabpa*	TRCN0000304421	GCTACACCGACTACGATTAAA	Experimental
14390	*Gabpa*	TRCN0000304469	AGCTTAGTGTACAGGTAATTT	Experimental
14390	*Gabpa*	TRCN0000304508	ATGAACCAATAGGCAATTTAA	Experimental
14391	*Gabpb1*	TRCN0000233974	GCAAGTCATCACGATAGTTAC	Experimental
14391	*Gabpb1*	TRCN0000218841	GATTGCTATGCAGAACCAAAT	Experimental
14391	*Gabpb1*	TRCN0000218214	CATTGGGCAACAGAACATAAT	Experimental
14391	*Gabpb1*	TRCN0000233973	TGCAATTCAGCAAGTAGTTAG	Experimental
14873	*Gsto1*	TRCN0000366490	TACCCAGAGAAGAAGTTATTT	Experimental
14873	*Gsto1*	TRCN0000366550	AGACCTACCGCGAGTACTTAA	Experimental
14873	*Gsto1*	TRCN0000374877	AGAGGAGGGCATGGATAATTA	Experimental
68705	*Gtf2f2*	TRCN0000312862	GAACGAGGATCTTGCAAATAT	Experimental
68705	*Gtf2f2*	TRCN0000311880	GCTCCTAGAGAACACCCATTT	Experimental
68705	*Gtf2f2*	TRCN0000312860	GAGAAGCATCAGTACTATAAT	Experimental

68705	*Gtf2f2*	TRCN0000312908	TGCTAACCATCAGTACAATAT	Experimental
15270	*H2afx*	TRCN0000335511	CGAGTACCTCACTGCCGAGAT	Experimental
15270	*H2afx*	TRCN0000348628	TGCTGGCCTCATACCAGTTGA	Experimental
15270	*H2afx*	TRCN0000348629	GCGTCTTTGCTTCAGCTTGGT	Experimental
15242	*Hhex*	TRCN0000244318	AGCGTCTGGCCAAGATGTTAC	Experimental
15242	*Hhex*	TRCN0000235790	TGGAGAAGAAGTTCGAGACTC	Experimental
15242	*Hhex*	TRCN0000244318	AGCGTCTGGCCAAGATGTTAC	Experimental
15251	*Hif1a*	TRCN0000232222	TGGATAGCGATATGGTCAATG	Experimental
15251	*Hif1a*	TRCN0000232220	CCCATTCCTCATCCGTCAAAT	Experimental
15251	*Hif1a*	TRCN0000232223	TATGCACTTTGTCGCTATTAA	Experimental
50708	*Hist1h1c*	TRCN0000374722	TATTGGACTGCAAGGTGATTT	Experimental
50708	*Hist1h1c*	TRCN0000305598	ACGAACTGCCTTCCTTAATAT	Experimental
50708	*Hist1h1c*	TRCN0000305599	CAAAGAAGGCGAAGGTCACCA	Experimental
97165	*Hmgb2*	TRCN0000301407	GCTCAACATTAGCTTCAGTAT	Experimental
97165	*Hmgb2*	TRCN0000304269	AGAGCGACAAAGCTCGTTATG	Experimental
97165	*Hmgb2*	TRCN0000304268	TTGGAGATACTGCGAAGAAAC	Experimental
72692	*Hnrpll*	TRCN0000295035	GACTACACTAAACCTTATTTG	Experimental
72692	*Hnrpll*	TRCN0000295036	GAATCCTCTTTACCCAATTAC	Experimental
72692	*Hnrpll*	TRCN0000295034	CACTCAATGGAGCTGATATAT	Experimental
15902	*Id2*	TRCN0000229536	TGGACTGTGATACCGTTATTT	Experimental
15902	*Id2*	TRCN0000218289	TGAGCTTATGTCGAATGATAG	Experimental
15902	*Id2*	TRCN0000229534	GAAGGTGACCAAGATGGAAAT	Experimental
15902	*Id2*	TRCN0000229535	AGTACTCTGTGGCTAAATAAA	Experimental
15902	*Id2*	TRCN0000054388	CCTTCTGAGCTTATGTCGAAT	Experimental
15902	*Id2*	TRCN0000054390	GCTTATGTCGAATGATAGCAA	Experimental
15902	*Id2*	TRCN0000054391	CCCACTATCGTCAGCCTGCAT	Experimental
15902	*Id2*	TRCN0000229536	TGGACTGTGATACCGTTATTT	Experimental
15903	*Id3*	TRCN0000071438	GCAGCGTGTCATAGACTACAT	Experimental
15903	*Id3*	TRCN0000071439	TCTTAGCCTCTTGGACGACAT	Experimental
15903	*Id3*	TRCN0000071440	GCTGAGCTCACTCCGGAACTT	Experimental
71793	*Ints12*	TRCN0000241273	GCGTTGCTTGATGAGTCTTTG	Experimental
71793	*Ints12*	TRCN0000241274	TCGCTCAGTGAGTTGTGATAA	Experimental
71793	*Ints12*	TRCN0000241272	TTAAGCCATCCACGGTTATTT	Experimental
15900	*Irf8*	TRCN0000235797	GGACATTTCTGAGCCATATAA	Experimental
15900	*Irf8*	TRCN0000235799	ATCAACAGATCACCGTCTAAG	Experimental
15900	*Irf8*	TRCN0000235800	ATCCGAGAGCTGCAGCAATTC	Experimental
81703	*Jdp2*	TRCN0000374340	GTTTCTGCAGAGGGAGTCAGA	Experimental
81703	*Jdp2*	TRCN0000379069	CTGGACAAGAAGTGACTGAAG	Experimental
81703	*Jdp2*	TRCN0000366046	CAGAGCTGAAGACGCAGATAG	Experimental
81703	*Jdp2*	TRCN0000374274	ACCTGCATCGTGCGCACAGAT	Experimental
81703	*Jdp2*	TRCN0000376757	TGCACTTCCTGGAGGTGAAAC	Experimental
16597	*Klf12*	TRCN0000218857	TGAACTTACAGTCTAACAAAC	Experimental
16597	*Klf12*	TRCN0000218857	TGAACTTACAGTCTAACAAAC	Experimental
16597	*Klf12*	TRCN0000226309	ACTAGAGAATCAGCTTATAAT	Experimental

16646	*Kpna1*	TRCN0000324440	CCCAGCTCTTATTAGTATATT	Experimental
16646	*Kpna1*	TRCN0000324439	GCAACACAGAAGTTTAGGAAA	Experimental
16646	*Kpna1*	TRCN0000305813	CGAAACGGCTCAGGCATTAAT	Experimental
16648	*Kpna3*	TRCN0000366474	TGTAACACTAGAAGCTATATT	Experimental
16648	*Kpna3*	TRCN0000375209	GGAACGTCACATGGGTCATTG	Experimental
16648	*Kpna3*	TRCN0000366407	GAGCAAATACAGATGGTTATT	Experimental
71835	*Lancl2*	TRCN0000313109	TGTTGGTGAGACAGCTATTAA	Experimental
71835	*Lancl2*	TRCN0000312072	GCAGGATAAGAAGTATCTCTA	Experimental
71835	*Lancl2*	TRCN0000312071	GCTTCCTGATGAACTGCTGTA	Experimental
16842	*Lef1*	TRCN0000225787	TTGGTTAACGAGTCCGAAATC	Experimental
16842	*Lef1*	TRCN0000225788	GTAGCTGAGTGCACGCTAAAG	Experimental
56722	*Litaf*	TRCN0000234441	AGCAGCCTGTCTCCTTCTATG	Experimental
56722	*Litaf*	TRCN0000234443	CACTGTTTGGCTTGATCTATT	Experimental
56722	*Litaf*	TRCN0000234442	TGACCCAGCTGTCCTACAATG	Experimental
66890	*Lman2*	TRCN0000334067	CCTTCCTAGCTGTACGCTATT	Experimental
66890	*Lman2*	TRCN0000334068	CCTGTCTGACAACCACGACAT	Experimental
66890	*Lman2*	TRCN0000334135	GCTTCCTTAAGGACTGGGAAA	Experimental
17133	*Maff*	TRCN0000225792	AGTCTTTGGAACGGATCAATG	Experimental
17133	*Maff*	TRCN0000225793	TTGCACGAATCTAAGTTATTC	Experimental
17133	*Maff*	TRCN0000225791	GGCTGAAAGTCCGGCTCAAAT	Experimental
17191	*Mbd2*	TRCN0000039064	CGGCAAGATGATGCCTAGTAA	Experimental
17191	*Mbd2*	TRCN0000039066	CCTGAACACAACATTGCCAAT	Experimental
17191	*Mbd2*	TRCN0000039067	CGGATGAATGAACAACCACGT	Experimental
17191	*Mbd2*	TRCN0000039068	GAAGAGCGAGTCCAACAAGTA	Experimental
171170	*Mbnl3*	TRCN0000348500	CTTGTACCAACACGACAATAA	Experimental
171170	*Mbnl3*	TRCN0000348500	CTTGTACCAACACGACAATAA	Experimental
171170	*Mbnl3*	TRCN0000335179	GCTAGAAGTTAATGGGAGAAA	Experimental
17216	*Mcm2*	TRCN0000329158	TGTTGGCCATGTACCCTAAAT	Experimental
17216	*Mcm2*	TRCN0000329157	TCCAGTCTTTGCCACTATTAT	Experimental
17216	*Mcm2*	TRCN0000329229	TTTAGACAAAGAGCAACTTAG	Experimental
17216	*Mcm2*	TRCN0000375307	TTCAGCGTCATGCGGAGTATG	Experimental
17215	*Mcm3*	TRCN0000240613	GACTCACTGCTGTCTAGATTT	Experimental
17215	*Mcm3*	TRCN0000240616	GCATTCTCTGCGGACGATATA	Experimental
17215	*Mcm3*	TRCN0000240614	TGTCTCCCATCCTCCAGTAAC	Experimental
17217	*Mcm4*	TRCN0000305014	GGCAGTATTCCACTATATAAA	Experimental
17217	*Mcm4*	TRCN0000311212	GTGAACGTCACAGGCATATAT	Experimental
17217	*Mcm4*	TRCN0000315492	GCCGTGCTGAAAGACTACATT	Experimental
17218	*Mcm5*	TRCN0000312259	CCGCATTGAGAAGCAACTCAA	Experimental
17218	*Mcm5*	TRCN0000349880	TTGTATGTACAGAGGTAATAA	Experimental
17218	*Mcm5*	TRCN0000313276	ACTCACTCGCCGAGGTGATAT	Experimental
17220	*Mcm7*	TRCN0000339972	TAAATGTCTGGCAGGTTAATA	Experimental
17220	*Mcm7*	TRCN0000339971	GCGAAGATTTGAGTTGTATTT	Experimental
17220	*Mcm7*	TRCN0000339901	CCAGTATCCTTCTGAACTCAT	Experimental
17827	*Mtx1*	TRCN0000288205	CGTGCTGACCTATACCAGATT	Experimental

17827	*Mtx1*	TRCN0000295558	CCAGTGATGGGAAAGTCATTA	Experimental
17827	*Mtx1*	TRCN0000288270	CGCCAAGAACTATGTGGAAGT	Experimental
17119	*Mxd1*	TRCN0000348069	GCACCAGTAGCAGATCAACTC	Experimental
17119	*Mxd1*	TRCN0000347991	ACGGAGGAACAAGCCCAAGAA	Experimental
17119	*Mxd1*	TRCN0000333905	GCAAACCTCATGTACCACCTT	Experimental
17874	*Myd88*	TRCN0000301332	GCCAGCGAGCTAATTGAGAAA	Experimental
17874	*Myd88*	TRCN0000301404	CCTTGTTAGACCGTGAGGATA	Experimental
17874	*Myd88*	TRCN0000301406	CCTTTCACGTTCTCTACCATA	Experimental
59287	*Ncstn*	TRCN0000293129	CTCCTTCCACAATCGGTATTA	Experimental
59287	*Ncstn*	TRCN0000293130	CCTGAACCCTTAGGGAAGAAA	Experimental
59287	*Ncstn*	TRCN0000293074	CCCAAATGATGGGTTTGGTAT	Experimental
17988	*Ndrg1*	TRCN0000238071	ACCCGAGCAACCTACACTTAT	Experimental
17988	*Ndrg1*	TRCN0000257341	CTGGAGTCCTTCACCAGTTTG	Experimental
17988	*Ndrg1*	TRCN0000257343	AGGGTCTCGTGCTCATGAATG	Experimental
17997	*Nedd1*	TRCN0000248316	AGCAGACATGTGTCGATTTAA	Experimental
17997	*Nedd1*	TRCN0000248319	TTGGGCAGCGTTTCGGATAAT	Experimental
17997	*Nedd1*	TRCN0000248317	TCGGTCTCTTAAGGATCATAA	Experimental
18036	*Nfkbib*	TRCN0000304902	CAGAGATGAGGGCGATGAATA	Experimental
18036	*Nfkbib*	TRCN0000304842	CTACAGTAGAGAAGTTGTATG	Experimental
18036	*Nfkbib*	TRCN0000316391	GCATCCAAACCTCTTCCTGAT	Experimental
15370	*Nr4a1*	TRCN0000218931	CTATTGTGGACAAGATCTTTA	Experimental
15370	*Nr4a1*	TRCN0000234021	CATGTGCCTTTAAGCCTATAG	Experimental
15370	*Nr4a1*	TRCN0000234019	TCTGGTTCCCTGGACGTTATC	Experimental
15370	*Nr4a1*	TRCN0000218931	CTATTGTGGACAAGATCTTTA	Experimental
68501	*Nsmce2*	TRCN0000248754	CAGGTTCTACCCGTTACATAT	Experimental
68501	*Nsmce2*	TRCN0000248756	GTGAGTAGTGAGTACAGTATG	Experimental
68501	*Nsmce2*	TRCN0000248752	GAACTCTGATGCCGACTTTAA	Experimental
234865	*Nup133*	TRCN0000246817	GAGCAGCTTGTGGCACTAATT	Experimental
234865	*Nup133*	TRCN0000246815	GGAACGCCTACTCGGATATTC	Experimental
234865	*Nup133*	TRCN0000246814	GCAACACCTCACTGATCATTT	Experimental
69912	*Nup43*	TRCN0000247916	GCGTTTCACTGTGGTCTATTG	Experimental
69912	*Nup43*	TRCN0000247915	CTGTCAGAACTATAGATAATG	Experimental
69912	*Nup43*	TRCN0000247914	AGGATGGAATGTTGAGTATTT	Experimental
55982	*Paxip1*	TRCN0000248338	ATCCACGCCTGATCATCTATG	Experimental
55982	*Paxip1*	TRCN0000248335	TCCCGAGCAGATGTCGGATAA	Experimental
55982	*Paxip1*	TRCN0000248336	GACATCGACCCGCAGGTTATT	Experimental
83436	*Plekha2*	TRCN0000277296	CGCCTAACTCCATCTTGTCAA	Experimental
83436	*Plekha2*	TRCN0000285928	ACACCATTCTGCTTCGTTATC	Experimental
83436	*Plekha2*	TRCN0000277253	ACAACCTGTTTGAAATCATAA	Experimental
53317	*Plrg1*	TRCN0000332100	GCCCACAGCAATGAATTCTAT	Experimental
53317	*Plrg1*	TRCN0000332099	CCAGAGTTGATGCAAATCGTA	Experimental
53317	*Plrg1*	TRCN0000332032	CCTGGAAATCAGTGGTTCGTT	Experimental
18968	*Pola1*	TRCN0000287378	GCCAATCAGTTGGTGTAAATT	Experimental
18968	*Pola1*	TRCN0000287380	CCTGGATTTCAACAGTTTATA	Experimental

18968	*Pola1*	TRCN0000287379	CCAGTTTGTATCGTTGCAGTA	Experimental
18973	*Pole*	TRCN0000353501	TATTGCCGGAAAGCCTATAAG	Experimental
18973	*Pole*	TRCN0000328840	TTCGCCCTGTGCACCATTATT	Experimental
18973	*Pole*	TRCN0000328838	CACTTTGGTATCCCACACCTA	Experimental
12142	*Prdm1*	TRCN0000235772	GGACATGGAGGACGCTGATAT	Experimental
12142	*Prdm1*	TRCN0000235768	GGTTCGACATCAGCGACAATG	Experimental
12142	*Prdm1*	TRCN0000235772	GGACATGGAGGACGCTGATAT	Experimental
12142	*Prdm1*	TRCN0000235772	GGACATGGAGGACGCTGATAT	Experimental
12142	*Prdm1*	TRCN0000235768	GGTTCGACATCAGCGACAATG	Experimental
12142	*Prdm1*	TRCN0000235770	ACTATTGCCTAGCCATAATTA	Experimental
12142	*Prdm1*	TRCN0000235771	CATCTACTTCTACACTATTAA	Experimental
12142	*Prdm1*	TRCN0000084716	CGGAGCCATGAATCTCATTAA	Experimental
12142	*Prdm1*	TRCN0000084714	CCCTGCCAAGTTTACGCAATT	Experimental
12142	*Prdm1*	TRCN0000235769	TGTTGCCACCGTACGGCATTA	Experimental
12142	*Prdm1*	TRCN0000084715	CGTGGTAAGTAAGGAGTACAT	Experimental
12142	*Prdm1*	TRCN0000084717	CCGTCTACAGTAACCTCCTTA	Experimental
12142	*Prdm1*	TRCN0000084713	GCAACCTTTCTCTATGATAAT	Experimental
23997	*Psmd13*	TRCN0000314108	GACGGCAATCGGAGCTCTAAA	Experimental
23997	*Psmd13*	TRCN0000350044	TGGGAACTGTTCTTGCTATAA	Experimental
23997	*Psmd13*	TRCN0000317979	GCTGGATTTGCAGCAGATCAA	Experimental
19205	*Ptbp1*	TRCN0000295168	CACTATGGTTAACTACTATAC	Experimental
19205	*Ptbp1*	TRCN0000287704	CCAAAGCCTCTTTATTCTCTT	Experimental
19205	*Ptbp1*	TRCN0000287703	CTCAATGTCAAGTACAACAAT	Experimental
19385	*Ranbp1*	TRCN0000305639	GCCATCCGCTTCCTAAATGCT	Experimental
19385	*Ranbp1*	TRCN0000305693	CAAAGCTGTTCCGGTTTGCTT	Experimental
19385	*Ranbp1*	TRCN0000375879	GAGAATGCAGATGAGTCCAAC	Experimental
19646	*Rbbp4*	TRCN0000313825	ATTTGGGACACTCGTTCAAAC	Experimental
19646	*Rbbp4*	TRCN0000317347	GCGGAGAACATTTACAATGAT	Experimental
19646	*Rbbp4*	TRCN0000313850	GTTAGTCTTTGACCACTATAG	Experimental
72151	*Rfc5*	TRCN0000225749	ACGCCTTGAGACGAGTGATTG	Experimental
72151	*Rfc5*	TRCN0000218845	CTTCAGTTCGGATACACTTAT	Experimental
72151	*Rfc5*	TRCN0000225748	GACGACCGAGGGATCGATATT	Experimental
75212	*Rnf121*	TRCN0000366929	GTTTACCCTCTTTGGTCTTAA	Experimental
75212	*Rnf121*	TRCN0000375815	ATGCAATGGACTTTGGCATTT	Experimental
75212	*Rnf121*	TRCN0000375752	AGAAATGTGTGCAGACTATAT	Experimental
75212	*Rnf121*	TRCN0000376818	GGAGCTAGATGAGGTTGATAT	Experimental
19826	*Rnps1*	TRCN0000315607	CGTACCTCTGTAACTTATGTT	Experimental
19826	*Rnps1*	TRCN0000305274	AGTCCAAGGATCGATCTAAAG	Experimental
19826	*Rnps1*	TRCN0000309161	GCAGTTGAGATATCTCCTCTA	Experimental
68275	*Rpa1*	TRCN0000304604	AGCTATGAAGATTCGATTAAA	Experimental
68275	*Rpa1*	TRCN0000302173	GCCCTGAAGATCGCTAACAAA	Experimental
68275	*Rpa1*	TRCN0000302172	CGCATGATCTTATCGGCAAAT	Experimental
68240	*Rpa3*	TRCN0000317098	GCGACTCCTATAATTTCTAAT	Experimental
68240	*Rpa3*	TRCN0000313677	GAAGATACTAATCGCTTTGAT	Experimental

68240	*Rpa3*	TRCN0000317097	CCCTGTTTAAGGAAGATACTA	Experimental
71919	*Rpap3*	TRCN0000250034	ATTTACCTCCTATTCGAAATG	Experimental
71919	*Rpap3*	TRCN0000250035	GCCGCCTCTGTGTACATTATT	Experimental
71919	*Rpap3*	TRCN0000250037	TCCAGATGTATTCAATCAAAT	Experimental
218544	*Sgtb*	TRCN0000248790	CGACGACGACGATGACTATTT	Experimental
218544	*Sgtb*	TRCN0000248789	TCCTGAGAATGATTCGTATAA	Experimental
218544	*Sgtb*	TRCN0000248789	TCCTGAGAATGATTCGTATAA	Experimental
30954	*Siva1*	TRCN0000327611	CTTGTTCATCGTGCATGAGAT	Experimental
30954	*Siva1*	TRCN0000327529	GAGCGAAGATTGTTCCGTGAA	Experimental
30954	*Siva1*	TRCN0000340795	TCACATATCGAGCGAAGATTG	Experimental
26942	*Spag1*	TRCN0000336086	GCTAGATCCCGGAAACGTAAA	Experimental
26942	*Spag1*	TRCN0000353302	CAATCGAGCTCAGGCCGAAAT	Experimental
26942	*Spag1*	TRCN0000336087	GACGTCTCCAAGCCTACTAAT	Experimental
20833	*Ssrp1*	TRCN0000301240	CCGTCAGGGTATCATCTTTAA	Experimental
20833	*Ssrp1*	TRCN0000301162	CCTACCTTTCTACACCTGCAT	Experimental
20833	*Ssrp1*	TRCN0000301237	GCGTACATGCTGTGGCTTAAT	Experimental
20848	*Stat3*	TRCN0000071453	CCTAACTTTGTGGTTCCAGAT	Experimental
20848	*Stat3*	TRCN0000071454	CGACTTTGATTTCAACTACAA	Experimental
20848	*Stat3*	TRCN0000071456	CCTGAGTTGAATTATCAGCTT	Experimental
20848	*Stat3*	TRCN0000071457	GCAGGTATCTTGAGAAGCCAA	Experimental
66464	*Taf12*	TRCN0000287465	GCTACAGATCGCTGATGATTT	Experimental
66464	*Taf12*	TRCN0000294957	GCTTTACAGAGAAGCATATAT	Experimental
66464	*Taf12*	TRCN0000294918	CTCCAGTTTCTCATCTGTAAA	Experimental
66464	*Taf12*	TRCN0000294920	CACAAGGCTCCATGGCCAATA	Experimental
108143	*Taf9*	TRCN0000244259	ATCAGGCCTGAAGTACGTTAA	Experimental
108143	*Taf9*	TRCN0000244260	GGATTGAGCAGTGGGTCAAAG	Experimental
108143	*Taf9*	TRCN0000244387	ATAGGGTTGTCGATGAGTTAG	Experimental
57765	*Tbx21*	TRCN0000082043	CCCTGTCCAGTCAGTAACTTT	Experimental
57765	*Tbx21*	TRCN0000082044	GCTTCTAACACACACGTCTTT	Experimental
57765	*Tbx21*	TRCN0000082045	CCTCTTCTATCCAACCAGTAT	Experimental
57765	*Tbx21*	TRCN0000082046	CACAAGCCATTACAGGATGTT	Experimental
57765	*Tbx21*	TRCN0000082047	GCCCAAGGATATGATCTCACA	Experimental
57765	*Tbx21*	TRCN0000412904	ATGCCAGGGAACCGCTTATAT	Experimental
57765	*Tbx21*	TRCN0000415776	TTGGTCTGACACCTGTGTTAA	Experimental
21414	*Tcf7*	TRCN0000262741	AGAAATGCATTCGGTACTTAC	Experimental
21414	*Tcf7*	TRCN0000262742	AGAAGCCAGTCATCAAGAAAC	Experimental
21414	*Tcf7*	TRCN0000360414	TTCTCCACTCTACGAACATTT	Experimental
21414	*Tcf7*	TRCN0000360415	CCTCAATGCGTTCATGCTTTA	Experimental
21781	*Tfdp1*	TRCN0000304311	CCACATTCTACCAAACGAATC	Experimental
21781	*Tfdp1*	TRCN0000374166	AGGAGAAGAAGGAGATCAAAT	Experimental
21781	*Tfdp1*	TRCN0000301409	CCTGCAGCAAATTGCCTTCAA	Experimental
21815	*Tgif1*	TRCN0000233981	TAGTGGATGTTGCACTCAAAC	Experimental
21815	*Tgif1*	TRCN0000233980	AGTACAGATGTACCGCAAATA	Experimental
21815	*Tgif1*	TRCN0000218560	GATGGCAAGAGATGCATTATT	Experimental

21815	*Tgif1*	TRCN0000233979	ATTTCAGAAGCTAGCTCTATT	Experimental
386612	*Thoc6*	TRCN0000347107	AGCTAGAGTGATGCTTGTGTT	Experimental
386612	*Thoc6*	TRCN0000347104	TTGACATCATGACCCTGTATT	Experimental
386612	*Thoc6*	TRCN0000347105	TTCTGGCAGCTGGCAACAATT	Experimental
386612	*Thoc6*	TRCN0000347029	AGTCTGGAAGTACCCGAAATC	Experimental
22036	*Traip*	TRCN0000304883	ATGTGGTAAGAATAGGCTTTG	Experimental
22036	*Traip*	TRCN0000304882	TTATAGAACAGGGTCACATTG	Experimental
22036	*Traip*	TRCN0000316540	CGTGTCCCTCAAGAAAGAGTA	Experimental
68729	*Trim37*	TRCN0000349177	GCTTAGTTCAAGAAGTGGAAA	Experimental
68729	*Trim37*	TRCN0000304884	ACAGTGTAGTGCTCAATTAAT	Experimental
68729	*Trim37*	TRCN0000349178	GCACTGGTATATTACTCAGTT	Experimental
18140	*Uhrf1*	TRCN0000311096	CTGTAGCTCCAGTGCCGTTAA	Experimental
18140	*Uhrf1*	TRCN0000311100	AGGTGGTCATGGCCAACTATA	Experimental
18140	*Uhrf1*	TRCN0000302343	CACACACTCTTCGATTATGAT	Experimental
216987	*Utp6*	TRCN0000302654	CGCGGAGATTAAGGCTATCAT	Experimental
216987	*Utp6*	TRCN0000375691	GGCAGAAGAACGGATTGATTT	Experimental
216987	*Utp6*	TRCN0000302655	CCAAAGATCTACAGAAGGAAA	Experimental
218973	*Wdhd1*	TRCN0000366054	TCCTTCGACTGTTCACTATTG	Experimental
218973	*Wdhd1*	TRCN0000366051	GGGCTTTCCCGCCTTTGTTTA	Experimental
218973	*Wdhd1*	TRCN0000366052	AGAGCAGCAGGAACTCTTAAT	Experimental
218973	*Wdhd1*	TRCN0000365978	GTCTCCCTGTGGGCAGTATTT	Experimental
107823	*Whsc1*	TRCN0000218710	CCAGAAAGAGCTTGGATATTT	Experimental
107823	*Whsc1*	TRCN0000253040	CTGTGAGAGAAGAGGATATTC	Experimental
107823	*Whsc1*	TRCN0000226297	AGAGCTGACTTTCAACTATAA	Experimental
22433	*Xbp1*	TRCN0000232018	GGTTGAGAACCAGGAGTTAAG	Experimental
22433	*Xbp1*	TRCN0000232018	GGTTGAGAACCAGGAGTTAAG	Experimental
22433	*Xbp1*	TRCN0000232021	AGGCCTGTCTCTTTCGTTAAA	Experimental
22433	*Xbp1*	TRCN0000008421	CCAGGAGTTAAGAACACGCTT	Experimental
22433	*Xbp1*	TRCN0000232019	TCCGCAGCACTCAGACTATGT	Experimental
22433	*Xbp1*	TRCN0000232018	GGTTGAGAACCAGGAGTTAAG	Experimental
22433	*Xbp1*	TRCN0000232021	AGGCCTGTCTCTTTCGTTAAA	Experimental
22436	*Xdh*	TRCN0000319595	CACAATCCAGGATGCTATAAA	Experimental
22436	*Xdh*	TRCN0000319662	TTGAGTAATTCTGGGTAATTC	Experimental
22436	*Xdh*	TRCN0000317847	CGCCTTCAGAACAAGATCGTT	Experimental
14375	*Xrcc6*	TRCN0000321228	TGCTAGAGCTCGACCAGTTTA	Experimental
14375	*Xrcc6*	TRCN0000350563	AGCTCAGAAGCCCAGCCACTT	Experimental
14375	*Xrcc6*	TRCN0000321226	GCTCACTGTACCTACACTGAA	Experimental
328801	*Zfp414*	TRCN0000242169	CTACTTCAAGTGTGAGAATTG	Experimental
328801	*Zfp414*	TRCN0000242168	ACTTGAACCCTGCACCCTTTG	Experimental
328801	*Zfp414*	TRCN0000242170	CTTCAAGCATCTGCATGTTTG	Experimental

Supplemental Table 6.1. List of journals prioritized for the selection of studies used in ImmuneSigDB.

Journal
Nature
Science
Cell
Nature Immunology
Immunity
Journal of Experimental Medicine
Journal of Clinical Investigation
Cell Host and Microbe
Blood
Proc Natl Acad Sci USA
Current Opinion in Immunology
Trends in Immunology
Journal of Allergy and Clinical Immunology
Plos Pathogens
Plos One
Mucosal Immunology
Arthritis and Rheumatism
Seminars in Immunology
Autoimmunity
Journal of Immunology
European Journal of Immunology
Genes and Immunity
Infection and Immunity
Immunology and Cell Biology
Vaccine
Cytokine
Journal of Clinical Immunology
Immunology

Supplemental Table 6.2. Affymetrix gene expression microarray platforms used to generate the public datasets used in ImmuneSigDB.

Platform ID	Name	Organism
GPL1261	Mouse430_2	Mus musculus
GPL339	MOE430A	Mus musculus
GPL8321	Mouse430A_2	Mus musculus
GPL6246	MoGene-1_0-st	Mus musculus
GPL81	MG_U74Av2	Mus musculus
GPL570	HG-U133_Plus_2	Homo sapiens
GPL96	HG-U133A	Homo sapiens
GPL571	HG-U133A_2	Homo sapiens
GPL6244	HuGene-1_0-st	Homo sapiens
GPL8300	HG_U95Av2	Homo sapiens
GPL97	HG-U133B	Homo sapiens
GPL91	HG_U95A	Homo sapiens
GPL3921	HT_HG-U133A	Homo sapiens

Supplemental Table 6.3. Characteristics of ImmunSigDB and the collections generated by Li et al. and Chaussabel et al.

Parameter	ImmuneSigDB	Li	Chaussabel
Number of gene sets	4872	346	260
Fraction of gene sets annotated	100%	75%	15%
Number of Studies	389	540	9
Number of Samples	6283	32766	410
Species	2	1	1
Cells/tissue types	13	2	1

Supplemental Table 6.4. Top gene sets from ImmuneSigDB, Li et al. and Chaussabel et al. collections enriched in LPS-stimulated compared on unstimulated dendritic cells.

GSE14000 **GSEA of human dendritic cells from peripheral blood stimulated with LPS compared with unstimulated cells.**

ImmuneSigDB

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	GSE2706_UNSTIM_VS_8H_LPS_AND_R848_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with R848 and LPS in vitro for 8 hours compared to unstimulated cells.	200	< 1E-06	Peripheral Blood	Human	15995707
2	GSE2706_UNSTIM_VS_2H_LPS_AND_R848_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with R848 and LPS in vitro for 2 hours compared to unstimulated cells.	198	< 1E-06	Peripheral Blood	Human	15995707
3	GSE18791_UNSTIM_VS_NEWCATSLE_VIRUS_DC_10H_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with Newcastle virus in vitro for 10 hours compared to unstimulated cells.	199	< 1E-06	Peripheral Blood	Human	20164420
4	GSE2706_UNSTIM_VS_8H_R848_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with R848 for 8 hours compared to unstimulated cells.	200	< 1E-06	Peripheral Blood	Human	15995707
5	GSE2706_UNSTIM_VS_2H_LPS_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with LPS for 2 hours compared to unstimulated cells.	199	< 1E-06	Peripheral Blood	Human	15995707
6	GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_6H_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with Newcastle virus in vitro for 6 hours compared to control cells.	200	< 1E-06	Peripheral Blood	Human	15995707
7	GSE2706_UNSTIM_VS_8H_LPS_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with LPS for 8 hours compared to unstimulated cells.	199	< 1E-06	Peripheral Blood	Human	15995707
8	GSE2706_UNSTIM_VS_2H_R848_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with R848 for 2 hours compared to unstimulated cells.	199	< 1E-06	Peripheral Blood	Human	15995707
9	GSE4748_CTRL_VS_LPS_STIM_DC_3H_DN	Genes up-regulated in LPS stimulated monocyte-derived dendritic cells in vitro for 3 hours compared to unstimulated cells.	200	< 1E-06	Peripheral Blood	Human	16717116
10	GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_10H_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with Newcastle virus in vitro for 10 hours compared to control cells.	200	< 1E-06	Peripheral Blood	Human	20164420
11	GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_8H_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with Newcastle virus in vitro for 8 hours compared to control cells.	199	< 1E-06	Peripheral Blood	Human	20164420
12	GSE9988_LOW_LPS_VS_CTRL_TREATED_MONOCYTE_UP	Genes up-regulated in monocyte stimulated in vitro with low LPS compared to unstimulated cells.	198	< 1E-06	Peripheral Blood	Human	18292579
13	GSE16755_CTRL_VS_IFNA_TREATED_MAC_DN	Genes up-regulated in monocyte-derived macrophages stimulated in vitro with interferon alpha compared to unstimulated cells.	198	< 1E-06	Peripheral Blood	Human	19556424
14	GSE7509_UNSTIM_VS_IFNA_STIM_IMMATURE_DC_DN	Genes up-regulated in monocyte-derived immature dendritic cells stimulated in vitro with interferon alpha compared to unstimulated cells.	200	< 1E-06	Peripheral Blood	Human	17502666
15	GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_16H_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with Newcastle virus in vitro for 16 hours compared to unstimulated cells.	199	< 1E-06	Peripheral Blood	Human	20164420
16	GSE18791_CTRL_VS_NEWCASTLE_VIRUS_DC_12H_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with Newcastle virus in vitro for 12 hours compared to unstimulated cells.	200	< 1E-06	Peripheral Blood	Human	20164420
17	GSE9988_ANTI_TREM1_VS_LOW_LPS_MONOCYTE_DN	Genes up-regulated in monocyte stimulated in vitro with low LPS compared to anti-TREM1 antibodies incubation.	198	< 1E-06	Peripheral Blood	Human	15995707
18	GSE2706_R848_VS_R848_AND_LPS_2H_STIM_DC_DN	Genes up-regulated in monocyte-derived dendritic cells stimulated with R848 for 2 hours compared to R848 and LPS in vitro stimulated cells.	197	< 1E-06	Peripheral Blood	Human	15995707
19	GSE9988_ANTI_TREM1_VS_ANTI_TREM1_AND_LPS_MONOCYTE_DN	Genes up-regulated in monocyte stimulated in vitro with LPS and anti-TREM1 antibodies compared to anti-TREM1 antibodies alone.	196	< 1E-06	Peripheral Blood	Human	18292579
20	GSE9988_LPS_VS_VEHICLE_TREATED_MONOCYTE_UP	Genes up-regulated in monocyte stimulated in vitro with LPS compared to unstimulated cells.	198	< 1E-06	Peripheral Blood	Human	18292579

Li

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	ACTIVATED_(LPS)_DENDRITIC_CELL_SURFACE_SIGNATURE_(S11)		37	< 1E-06	Peripheral Blood	Human	24336226
2	ENRICHED_IN_ACTIVATED_DENDRITIC_CELLS_(II)_(M165)		35	< 1E-06	Peripheral Blood	Human	24336226
3	CHEMOKINE_CLUSTER_(I)_(M27.0)		26	< 1E-06	Peripheral Blood	Human	24336226
4	ANTIVIRAL_IFN_SIGNATURE_(M75)		22	< 1E-06	Peripheral Blood	Human	24336226
5	CHEMOKINES_AND_INFLAMMATORY_MOLECULES_IN_MYELOID_CELLS_(M86.0)		18	< 1E-06	Peripheral Blood	Human	24336226
6	MYELOID,_DENDRITIC_CELL_ACTIVATION_VIA_NFKB_(I)_(M43.0)		15	< 1E-06	Peripheral Blood	Human	24336226
7	ENRICHED_IN_ACTIVATED_DENDRITIC_CELLS/MONOCYTES_(M64)		17	< 1E-06	Peripheral Blood	Human	24336226
8	CHEMOKINE_CLUSTER_(II)_(M27.1)		15	2.16E-04	Peripheral Blood	Human	24336226
9	T_CELL_ACTIVATION_(II)_(M7.3)		31	7.70E-04	Peripheral Blood	Human	24336226
10	ENRICHED_IN_NK_CELLS_(I)_(M7.2)		47	0.0016381	Peripheral Blood	Human	24336226
11	PUTATIVE_TARGETS_OF_PAX3_(M89.0)		16	0.00166705	Peripheral Blood	Human	24336226
12	T_CELL_ACTIVATION_(I)_(M7.1)		50	0.00165744	Peripheral Blood	Human	24336226
13	VIRAL_SENSING_&_IMMUNITY;_IRF2_TARGETS_NETWORK_(I)_(M111.0)		17	0.0030623	Peripheral Blood	Human	24336226
14	CELL_ADHESION_(M51)		38	0.00490505	Peripheral Blood	Human	24336226
15	AP-1_TRANSCRIPTION_FACTOR_NETWORK_(M20)		15	0.00491378	Peripheral Blood	Human	24336226
16	ENRICHED_IN_T_CELLS_(I)_(M7.0)		59	0.00775815	Peripheral Blood	Human	24336226
17	DC_SURFACE_SIGNATURE_(S5)		82	0.02313339	Peripheral Blood	Human	24336226
18	LEUKOCYTE_DIFFERENTIATION_(M160)		16	0.02258546	Peripheral Blood	Human	24336226
19	*TBA_(M66)*		17	0.03778308	Peripheral Blood	Human	24336226
20	ENRICHED_IN_ANTIGEN_PRESENTATION_(I)_(M71)		18	0.03734423	Peripheral Blood	Human	24336226

Chaussabel

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	M3.4_Interferon		44	< 1E-06	Peripheral Blood	Human	18631455
2	M1.2_Interferon		24	< 1E-06	Peripheral Blood	Human	18631455
3	M5.12_Interferon		45	< 1E-06	Peripheral Blood	Human	18631455
4	*M7.16_Not_Determined*		48	< 1E-06	Peripheral Blood	Human	18631455
5	*M7.35_Undetermined*		21	0.00296325	Peripheral Blood	Human	18631455
6	*M9.35_Undetermined*		24	0.01223695	Peripheral Blood	Human	18631455
7	*M9.24_Undetermined*		41	0.01429375	Peripheral Blood	Human	18631455
8	M4.2_Inflammation		37	0.01591893	Peripheral Blood	Human	18631455
9	*M9.41_Undetermined*		15	0.04131159	Peripheral Blood	Human	18631455
10	M6.13_Cell_Death		40	0.07234576	Peripheral Blood	Human	18631455
11	*M9.39_Undetermined*		18	0.0752909	Peripheral Blood	Human	18631455
12	M3.6_Cytotoxic/NK_Cell		41	0.0724536	Peripheral Blood	Human	18631455
13	*M8.18_Undetermined*		19	0.06779851	Peripheral Blood	Human	18631455
14	*M7.9_Undetermined*		133	0.06402618	Peripheral Blood	Human	18631455
15	*M8.51_Undetermined*		16	0.16875912	Peripheral Blood	Human	18631455
16	*M8.39_Undetermined*		16	0.24413404	Peripheral Blood	Human	18631455
17	M3.2_Inflammation		108	0.2527887	Peripheral Blood	Human	18631455
18	M5.15_Neutrophils		19	0.2618582	Peripheral Blood	Human	18631455
19	M4.13_Inflammation		67	0.26964584	Peripheral Blood	Human	18631455
20	*M7.29_Not_Determined*		47	0.32821018	Peripheral Blood	Human	18631455

Supplemental Table 6.5. Top gene sets from ImmuneSigDB, Li et al. and Chaussabel et al. collections enriched in Treg compared to Tconv CD4^{+} cells.

GSE25087 GSEA of human regulatory T cells (Treg) compared to conventional CD4+ T cells (Tconv) from peripheral blood.

ImmuneSigDB

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	GSE11057_NAIVE_VS_MEMORY_CD4_TCELL_DN	Genes up-regulated in memory CD4+ T cells compared to naive CD4+ T cells.	198	< 1E-06	Peripheral Blood	Human	19568420
2	GSE11057_NAIVE_VS_EFF_MEMORY_CD4_TCELL_DN	Genes up-regulated in effector memory CD4+ T cells compared to naive CD4+ T cells.	197	< 1E-06	Peripheral Blood	Human	19568420
3	GSE11057_NAIVE_VS_CENT_MEMORY_CD4_TCELL_DN	Genes up-regulated in central memory CD4+ T cells compared to naive CD4+ T cells.	199	< 1E-06	Peripheral Blood	Human	19568420
4	GSE13738_RESTING_VS_BYSTANDER_ACTIVATED_CD4_TCELL_DN	Genes up-regulated in bystander activated CD4+ T cells (incubated in the presence of antigen-stimulated T cells) compared to resting CD4+ T cells.	199	< 1E-06	Peripheral Blood	Human	19201849
5	GSE26495_NAIVE_VS_PD1HIGH_CD8_TCELL_DN	Genes up-regulated in PD-1high CD8+ T cells compared to naive CD8+ T cells from healthy donors.	199	< 1E-06	Peripheral Blood	Human	21383243
6	GSE23321_CENTRAL_MEMORY_VS_NAIVE_CD8_TCELL_UP	Genes up-regulated in central memory CD8+ T cells compared to naive CD8+ T cells.	193	< 1E-06	Peripheral Blood	Human	21926977
7	GSE23321_EFFECTOR_MEMORY_VS_NAIVE_CD8_TCELL_UP	Genes up-regulated in effector memory CD8+ T cells compared to naive CD8+ T cells.	195	< 1E-06	Peripheral Blood	Human	21926977
8	GSE26495_NAIVE_VS_PD1LOW_CD8_TCELL_DN	Genes up-regulated in PD-1low CD8+ T cells compared to naive CD8+ T cells from healthy donors.	199	< 1E-06	Peripheral Blood	Human	21383243
9	GSE41978_KLRG1_HIGH_VS_LOW_EFFECTOR_CD8_TCELL_UP	Genes up-regulated in KLRG1 high CD8+ OT-1 T cells compared to KLRG1 low CD8+ OT-1 T cells 6 days following Listeria-OVA infection .	169	< 1E-06	Spleen	Mouse	23325888
10	GSE23321_CD8_STEM_CELL_MEMORY_VS_CENTRAL_MEMORY_CD8_TCELL_DN	Genes up-regulated in central memory CD8+ T cells compared to stem cell memory CD8+ T cells.	193	< 1E-06	Peripheral Blood	Human	21926977
11	GSE23321_CD8_STEM_CELL_MEMORY_VS_NAIVE_CD8_TCELL_UP	Genes up-regulated in stem cell memory CD8+ T cells compared to naive CD8+ T cells.	185	< 1E-06	Peripheral Blood	Human	21926977
12	GSE23321_CD8_STEM_CELL_MEMORY_VS_EFFECTOR_MEMORY_CD8_TCELL_DN	Genes up-regulated in effector memory CD8+ T cells compared to stem cell memory CD8+ T cells.	194	< 1E-06	Peripheral Blood	Human	21926977
13	GSE32901_NAIVE_VS_TH17_ENRICHED_CD4_TCELL_DN	Genes up-regulated in in vitro expanded Th17-enriched effector CD4+ T cells compared to naive CD4+ T cells.	195	< 1E-06	Peripheral Blood	Human	22715389
14	GOLDRATH_NAIVE_VS_EFF_CD8_TCELL_DN	Genes up-regulated in effector CD8+ OT-1 T cells responding in vivo to VV-ova virus compared to naive CD8+ T cells.	199	< 1E-06	Spleen	Mouse	15548615
15	GSE32901_NAIVE_VS_TH17_NEG_CD4_TCELL_DN	Genes up-regulated in in vitro expanded Th17-negative effector CD4+ T cells compared to naive CD4+ T cells.	197	< 1E-06	Peripheral Blood	Human	22715389
16	GSE24634_TEFF_VS_TCONV_DAY7_IN_CULTURE_UP	Genes up-regulated in CD25+ effector CD4+ T cells activated but not incubated in Treg differentiating conditions compared to CD25- T cells incubated with IL-4 for 7 days.	198	< 1E-06	Peripheral Blood	Human	21347372
17	KAECH_NAIVE_VS_DAY8_EFF_CD8_TCELL_DN	Genes up-regulated in effector CD8+ T cells 8 days following LCMV infection compared to naive CD8+ T cells.	196	< 1E-06	Spleen	Mouse	12526810
18	GSE21360_NAIVE_VS_QUATERNARY_MEMORY_CD8_TCELL_DN	Genes up-regulated in quarternary (fourth time infection responders) CD8+ T cells compared to naive CD8+ T cells.	159	< 1E-06	Spleen	Mouse	20619696
19	GSE40685_NAIVE_CD4_TCELL_VS_FOXP3_KO_TREG_PRECURSOR__DN	Genes up-regulated in Treg precursos (FoxP3-KO GFP+) compared to naive CD4+ T cells.	199	< 1E-06	Spleen and Lymph Node	Mouse	23021222
20	GSE41867_NAIVE_VS_DAY8_LCMV_EFFECTOR_CD8_TCELL_DN	Genes up-regulated in effector CD8+ T cells 8 days following LCMV infection compared to naive CD8+ T cells.	173	< 1E-06	Spleen	Mouse	23159438

Li

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	CELL_CYCLE_(I)_(M4.1)		145	< 1E-06	Peripheral Blood	Human	24336226
2	CELL_CYCLE_AND_TRANSCRIPTION_(M4.0)		328	< 1E-06	Peripheral Blood	Human	24336226
3	MITOTIC_CELL_CYCLE_IN_STIMULATED_CD4_T_CELLS_(M4.5)		35	< 1E-06	Peripheral Blood	Human	24336226
4	PLK1_SIGNALING_EVENTS_(M4.2)		34	< 1E-06	Peripheral Blood	Human	24336226
5	MITOTIC_CELL_DIVISION_(M6)		32	< 1E-06	Peripheral Blood	Human	24336226
6	NK_CELL_SURFACE_SIGNATURE_(S1)		48	< 1E-06	Peripheral Blood	Human	24336226
7	TRANSCRIPTION_REGULATION_IN_CELL_DEVELOPMENT_(M49)		47	< 1E-06	Peripheral Blood	Human	24336226
8	CELL_DIVISION_IN_STIMULATED_CD4_T_CELLS_(M4.6)		19	1.80E-04	Peripheral Blood	Human	24336226
9	MHC-TLR7-TLR8_CLUSTER_(M146)		17	3.03E-04	Peripheral Blood	Human	24336226
10	RECEPTORS,_CELL_MIGRATION_(M109)		15	2.73E-04	Peripheral Blood	Human	24336226
11	CELL_CYCLE_(III)_(M103)		51	2.48E-04	Peripheral Blood	Human	24336226
12	ENRICHED_IN_ANTIGEN_PRESENTATION_(I)_(M71)		18	2.27E-04	Peripheral Blood	Human	24336226
13	DC_SURFACE_SIGNATURE_(S5)		82	4.48E-04	Peripheral Blood	Human	24336226
14	CELL_DIVISION_(STIMULATED_CD4+_T_CELLS)_(M46)		28	4.16E-04	Peripheral Blood	Human	24336226
15	MITOTIC_CELL_CYCLE_IN_STIMULATED_CD4_T_CELLS_(M4.9)		16	9.95E-04	Peripheral Blood	Human	24336226
16	RESTING_DENDRITIC_CELL_SURFACE_SIGNATURE_(S10)		75	0.0027663	Peripheral Blood	Human	24336226
17	ENRICHED_IN_NK_CELLS_(I)_(M7.2)		47	0.00269882	Peripheral Blood	Human	24336226
18	REGULATION_OF_ANTIGEN_PRESENTATION_AND_IMMUNE_RESPONSE_(M5.0)		81	0.00254888	Peripheral Blood	Human	24336226
19	ACTIVATED_(LPS)_DENDRITIC_CELL_SURFACE_SIGNATURE_(S11)		37	0.0037057	Peripheral Blood	Human	24336226
20	MYELOID,_DENDRITIC_CELL_ACTIVATION_VIA_NFKB_(I)_(M43.0)		15	0.00497282	Peripheral Blood	Human	24336226

Chaussabel

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	M3.3_Cell_Cycle		44	< 1E-06	Peripheral Blood	Human	18631455
2	*M4.9_Not_Determined*		79	0.0024239	Peripheral Blood	Human	18631455
3	M5.1_Inflammation		207	0.0087603	Peripheral Blood	Human	18631455
4	M3.2_Inflammation		109	0.00753987	Peripheral Blood	Human	18631455
5	M6.16_Cell_Cycle		26	0.00603189	Peripheral Blood	Human	18631455
6	*M9.12_Undetermined*		57	0.00891114	Peripheral Blood	Human	18631455
7	M5.7_Inflammation		117	0.00763812	Peripheral Blood	Human	18631455
8	*M9.35_Undetermined*		24	0.01574711	Peripheral Blood	Human	18631455
9	*M5.2_Not_Determined*		50	0.0358064	Peripheral Blood	Human	18631455
10	M3.4_Interferon		44	0.0367342	Peripheral Blood	Human	18631455
11	M4.1_T_cell		26	0.03388736	Peripheral Blood	Human	18631455
12	M4.2_Inflammation		37	0.03660791	Peripheral Blood	Human	18631455
13	*M9.32_Undetermined*		25	0.04676838	Peripheral Blood	Human	18631455
14	M4.6_Inflammation		97	0.05164539	Peripheral Blood	Human	18631455
15	*M8.32_Undetermined*		18	0.05411964	Peripheral Blood	Human	18631455
16	M2.3_Erythrocytes		58	0.05826901	Peripheral Blood	Human	18631455
17	*M7.7_Undetermined*		95	0.07566726	Peripheral Blood	Human	18631455
18	*M8.37_Undetermined*		18	0.07938689	Peripheral Blood	Human	18631455
19	M4.13_Inflammation		67	0.08084902	Peripheral Blood	Human	18631455
20	M5.12_Interferon		46	0.11432808	Peripheral Blood	Human	18631455

Supplemental Table 6.6. Top gene sets from ImmuneSigDB, Li et al. and Chaussabel et al. collections enriched in plasma cells compared to B cells.

GSE22886 GSEA of human plasma cells compared to naive B cells from peripheral blood.

ImmuneSigDB

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	GSE12845_IGD_POS_VS_NEG_BLOOD_BCELL_DN	Genes up-regulated in blood IgD- B cells compared to IgD+ B cells.	199	< 1E-06	Peripheral Blood	Human	19023113
2	GSE29614_DAY3_VS_DAY7_TIV_FLU_VACCINE_PBMC_DN	Genes up-regulated in blood PBMCs 7 days following trivalent inactivated influenza vaccine (TIV) compared to PBMCs 3 days following vaccination, 2007/08 flu season.	158	< 1E-06	Peripheral Blood	Human	21743478
3	GSE13411_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes up-regulated in plasma cells compared to memory B cells from the spleen.	199	< 1E-06	Spleen	Human	19124732
4	GSE30153_LUPUS_VS_HEALTHY_DONOR_BCELL_UP	Genes up-regulated in B cells from the blood of patients with systemic lupus erythematosus compared to B cells from the blood of healthy donors.	150	< 1E-06	Peripheral Blood	Human	21886837
5	GSE29614_CTRL_VS_DAY7_TIV_FLU_VACCINE_PBMC_DN	Genes up-regulated in blood PBMCs 7 days following trivalent inactivated influenza vaccine (TIV) compared to PBMCs prior to vaccination, 2007/08 flu season.	163	< 1E-06	Peripheral Blood	Human	21743478
6	GSE12845_IGD_NEG_BLOOD_VS_NAIVE_TONSIL_BCELL_UP	Genes up-regulated in blood IgD- B cells compared to naive B cells from the tonsil.	200	< 1E-06	Peripheral blood and tonsil	Human	19023113
7	GSE12366_PLASMA_CELL_VS_MEMORY_BCELL_UP	Genes up-regulated in plasma cells compared to memory B cells from the tonsil.	147	< 1E-06	Tonsil	Human	19023113
8	GSE12366_PLASMA_CELL_VS_NAIVE_BCELL_UP	Genes up-regulated in plasma cells compared to naive B cells from the tonsil.	142	< 1E-06	Tonsil	Human	19023113
9	GSE10325_BCELL_VS_LUPUS_BCELL_DN	Genes up-regulated in B cells from the blood of patients with systemic lupus erythematosus compared to B cells from the blood of healthy donors.	200	< 1E-06	Peripheral Blood	Human	18275831
10	GSE29617_CTRL_VS_DAY7_TIV_FLU_VACCINE_PBMC_2008_DN	Genes up-regulated in blood PBMCs 7 days following trivalent inactivated influenza vaccine (TIV) compared to PBMCs prior to vaccination, 2008/09 flu season.	166	< 1E-06	Peripheral Blood	Human	21743478
11	GSE12366_GC_BCELL_VS_PLASMA_CELL_DN	Genes up-regulated in plasma cells compared to germinal center B cells from the tonsil.	153	< 1E-06	Tonsil	Human	19023113
12	GSE4142_PLASMA_CELL_VS_GC_BCELL_UP	Genes up-regulated in plasma cells compared to germinal center B cells from the spleen.	164	< 1E-06	Spleen	Mouse	16492737
13	GSE42724_B1_BCELL_VS_PLASMABLAST_DN	Genes up-regulated in plasmablast compared to B1 B cells from the blood.	160	< 1E-06	Peripheral Blood	Human	23613519
14	GSE12845_IGD_NEG_BLOOD_VS_PRE_GC_TONSIL_BCELL_UP	Genes up-regulated in blood IgD- B cells compared to pre-germinal center B cells from the tonsil.	198	< 1E-06	Peripheral blood and tonsil	Human	19023113
15	GSE13411_NAIVE_BCELL_VS_PLASMA_CELL_DN	Genes up-regulated in plasma cells compared to naive B cells from the spleen.	199	< 1E-06	Spleen	Human	19124732
16	GSE4142_NAIVE_BCELL_VS_PLASMA_CELL_DN	Genes up-regulated in plasma cells compared to naive B cells from the spleen.	166	< 1E-06	Spleen	Mouse	16492737
17	GSE9601_NFKB_INHIBITOR_VS_PI3K_INHIBITOR_TREATED_HCMV_INF_MONOCYTE_UP	Genes up-regulated in HCMV-infected monocytes treated with NF-kB inhibitor Bay11 compared to HCMV-infected cells treated with PI3K inhibitor LY294002.	199	< 1E-06	Peripheral Blood	Human	18003728
18	GSE13411_IGM_MEMORY_BCELL_VS_PLASMA_CELL_DN	Genes up-regulated in plasma cells compared to IgM+ memory B cells from the spleen.	199	< 1E-06	Spleen	Human	19124732
19	GSE29617_DAY3_VS_DAY7_TIV_FLU_VACCINE_PBMC_2008_DN	Genes up-regulated in blood PBMCs 7 days following trivalent inactivated influenza vaccine (TIV) compared to PBMCs 3 days following vaccination, 2008/09 flu season.	144	< 1E-06	Peripheral Blood	Human	21743478
20	GSE12845_IGD_NEG_BLOOD_VS_DARKZONE_GC_TONSIL_BCELL_UP	Genes up-regulated in blood IgD- B cells compared to B cells from the dark zone of the tonsil.	199	< 1E-06	Peripheral blood and tonsil	Human	19023113

Li

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	PLASMA_CELLS,_IMMUNOGLOBULINS_(M156.1)		25	< 1E-06	Peripheral Blood	Human	24336226
2	RESPIRATORY_ELECTRON_TRANSPORT_CHAIN_(MITOCHONDRION)_(M238)		16	< 1E-06	Peripheral Blood	Human	24336226
3	PLASMA_CELL_SURFACE_SIGNATURE_(S3)		17	< 1E-06	Peripheral Blood	Human	24336226
4	RESPIRATORY_ELECTRON_TRANSPORT_CHAIN_(MITOCHONDRION)_(M219)		18	1.45E-04	Peripheral Blood	Human	24336226
5	PLASMA_CELLS_&_B_CELLS,_IMMUNOGLOBULINS_(M156.0)		34	1.16E-04	Peripheral Blood	Human	24336226
6	*TBA_(M184.0)*		15	0.001479268	Peripheral Blood	Human	24336226
7	*TBA_(M70.0)*		17	0.09481692	Peripheral Blood	Human	24336226
8	*TBA_(M136)*		17	0.12377888	Peripheral Blood	Human	24336226
9	ACTIVATED_(LPS)_DENDRITIC_CELL_SURFACE_SIGNATURE_(S11)		27	0.6037416	Peripheral Blood	Human	24336226
10	ENRICHED_IN_NK_CELLS_(I)_(M7.2)		42	0.76805484	Peripheral Blood	Human	24336226
11	LEUKOCYTE_DIFFERENTIATION_(M160)		16	0.74690443	Peripheral Blood	Human	24336226
12	CHEMOKINES_AND_INFLAMMATORY_MOLECULES_IN_MYELOID_CELLS_(M86.0)		18	0.70106184	Peripheral Blood	Human	24336226
13	CORO1A-DEF6_NETWORK_(I)_(M32.2)		20	0.8959738	Peripheral Blood	Human	24336226
14	ENRICHED_IN_MONOCYTES_(IV)_(M118.0)		47	0.885626	Peripheral Blood	Human	24336226
15	VIRAL_SENSING_&_IMMUNITY:_IRF2_TARGETS_NETWORK_(I)_(M111.0)		15	0.83604485	Peripheral Blood	Human	24336226
16	EXTRACELLULAR_MATRIX_(II)_(M2.1)		44	0.86151814	Peripheral Blood	Human	24336226
17	ENRICHED_IN_MONOCYTES_(II)_(M11.0)		176	0.82706755	Peripheral Blood	Human	24336226
18	CELL_CYCLE,_ATP_BINDING_(M144)		16	1	Peripheral Blood	Human	24336226
19	RESTING_DENDRITIC_CELL_SURFACE_SIGNATURE_(S10)		64	0.94787514	Peripheral Blood	Human	24336226
20	CELL_MOVEMENT,_ADHESION_&_PLATELET_ACTIVATION_(M30)		17	0.91048986	Peripheral Blood	Human	24336226

Chaussabel

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	*M7.7_Undetermined*		78	< 1E-06	Peripheral Blood	Human	18631455
2	*M7.11_Undetermined*		89	< 1E-06	Peripheral Blood	Human	18631455
3	M3.1_Erythrocytes		50	< 1E-06	Peripheral Blood	Human	18631455
4	M6.2_Mitochondrial_Respiration		111	< 1E-06	Peripheral Blood	Human	18631455
5	M5.1_Inflammation		160	3.86E-04	Peripheral Blood	Human	18631455
6	*M7.4_Not_Determined*		91	6.30E-04	Peripheral Blood	Human	18631455
7	M3.5_Cell_Cycle		104	6.83E-04	Peripheral Blood	Human	18631455
8	M5.6_Mitochondrial_Stress_Proteasome		95	0.010274475	Peripheral Blood	Human	18631455
9	*M4.4_Not_Determined*		50	0.022653263	Peripheral Blood	Human	18631455
10	*M6.14_Not_Determined*		28	0.0243098	Peripheral Blood	Human	18631455
11	*M5.13_Not_Determined*		125	0.0224363	Peripheral Blood	Human	18631455
12	*M6.8_Not_Determined*		26	0.02130558	Peripheral Blood	Human	18631455
13	M6.13_Cell_Death		32	0.031561315	Peripheral Blood	Human	18631455
14	*M6.17_Not_Determined*		44	0.03791923	Peripheral Blood	Human	18631455
15	*M7.8_Undetermined*		66	0.039329473	Peripheral Blood	Human	18631455
16	*M5.3_Not_Determined*		66	0.046752	Peripheral Blood	Human	18631455
17	*M7.21_Undetermined*		69	0.051384736	Peripheral Blood	Human	18631455
18	*M7.19_Undetermined*		69	0.0705958	Peripheral Blood	Human	18631455
19	M5.1_Inflammation		184	0.08181434	Peripheral Blood	Human	18631455
20	M3.2_Inflammation		94	0.10395143	Peripheral Blood	Human	18631455

Supplemental Table 6.7. Top gene sets from ImmuneSigDB, Li et al and Chaussabel et al. collections enriched in memory compared to naive B cells.

GSE42724 GSEA of human memory and naive B cells from peripheral blood.

ImmuneSigDB

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	GSE22886_NAIVE_VS_IGG_IGA_MEMORY_BCELL_DN	Genes up-regulated in IgG+/IgA+ memory B cells compared to naive B cells.	185	< 1E-06	Peripheral Blood	Human	15789058
2	GSE22886_NAIVE_VS_IGM_MEMORY_BCELL_DN	Genes up-regulated in IgM+ memory B cells compared to naive B cells.	184	< 1E-06	Peripheral Blood	Human	15789058
3	GSE12366_NAIVE_VS_MEMORY_BCELL_DN	Genes up-regulated in memory B cells compared to naive B cells.	176	< 1E-06	Tonsil	Human	19023113
4	GSE11057_NAIVE_VS_MEMORY_CD4_TCELL_DN	Genes up-regulated in memory (CD45RO+) CD4+ T cells compared to naive (CD45RA+) CD4+ T cells.	186	< 1E-06	Peripheral Blood	Human	19568420
5	GSE17186_MEMORY_VS_CD21HIGH_TRANSITIONAL_BCELL_UP	Genes up-regulated in memory B cells(CD20+CD10-CD27+) compared to CD21high transitional B cells (CD20+CD10+CD27-CD21hi).	170	< 1E-06	Peripheral and cord blood	Human	19965666
6	GSE17186_MEMORY_VS_NAIVE_BCELL_UP	Genes up-regulated in memory B cell (CD20+CD10-CD27+) compared to naive B cell (CD20+CD10-CD27-).	184	< 1E-06	Peripheral Blood	Human	19965666
7	GSE11057_NAIVE_VS_EFF_MEMORY_CD4_TCELL_DN	Genes up-regulated in effector memory (CD45RO+ CCR7- CD62L+/-) CD4+ T cells compared to naive (CD45RA+) CD4+ T cells.	183	< 1E-06	Peripheral Blood	Human	19568420
8	GSE12845_IGD_POS_VS_NEG_BLOOD_BCELL_DN	Genes up-regulated in IgD+ B cells compared to IgD- B cells.	175	< 1E-06	Peripheral Blood	Human	19023113
9	GSE17186_MEMORY_VS_CD21LOW_TRANSITIONAL_BCELL_UP	Genes up-regulated in memory B cells(CD20+CD10-CD27+) compared to CD21low transitional B cells (CD20+CD10+CD27-CD21lo).	176	< 1E-06	Peripheral Blood	Human	19965666
10	GSE32901_NAIVE_VS_TH17_ENRICHED_CD4_TCELL_DN	Genes up-regulated in in vitro expanded Th17-enriched effector CD4+ T cells compared to naive CD4+ T cells.	200	< 1E-06	Peripheral Blood	Human	22715389
11	GSE11057_NAIVE_VS_CENT_MEMORY_CD4_TCELL_DN	Genes up-regulated in central memory (CD45RO+ CCR7+ CD62L+) CD4+ T cells compared to naive (CD45RA+) CD4+ T cells.	190	< 1E-06	Peripheral Blood	Human	19568420
12	GSE13411_NAIVE_VS_IGM_MEMORY_BCELL_DN	Genes up-regulated in IgM+ memory B cells compared to naive B cells.	187	< 1E-06	Spleen	Human	19124732
13	GSE21360_NAIVE_VS_QUATERNARY_MEMORY_CD8_TCELL_DN	Genes up-regulated in quarternary (fourth time infection responders) CD8+ T cells compared to naive CD8+ T cells.	154	< 1E-06	Spleen	Mouse	20619696
14	GSE22886_IGA_VS_IGM_MEMORY_BCELL_UP	Genes up-regulated in IgG+/IgA+ memory B cells compared to IgM+ memory B cells.	182	< 1E-06	Peripheral Blood	Human	15789058
15	GSE41867_NAIVE_VS_EFFECTOR_CD8_TCELL_DN	Genes up-regulated in effector CD8+ T cells responding to LCMV compared to naive CD8+ T cells.	168	< 1E-06	Peripheral Blood	Mouse	23159438
16	GSE11386_NAIVE_VS_MEMORY_BCELL_DN	Genes up-regulated in memory B cells compared to naive B cells.	185	< 1E-06	Spleen	Mouse	18566367
17	GSE40493_BCL6_KO_VS_WT_TREG_UP	Genes up-regulated in Bcl6 knockout Tregs compared to wildtypoe Tregs following 16h anti-CD3/anti-CD28 stimulation.	168	< 1E-06	Spleen and lymph nodes	Mouse	23053511
18	GSE13411_NAIVE_VS_MEMORY_BCELL_DN	Genes up-regulated in memory B cells compared to naive B cells.	188	< 1E-06	Spleen	Human	19124732
19	GSE21360_NAIVE_VS_TERTIARY_MEMORY_CD8_TCELL_DN	Genes up-regulated in tertiary (third time infection responders) CD8+ T cells compared to naive CD8+ T cells.	166	< 1E-06	Spleen	Mouse	20619696
20	GSE32901_NAIVE_VS_TH17_NEG_CD4_TCELL_DN	Genes up-regulated in in vitro expanded Th17-negative effector CD4+ T cells compared to naive CD4+ T cells.	200	< 1E-06	Peripheral Blood	Human	22715389

Li

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	NK_CELL_SURFACE_SIGNATURE_(S1)		35	0.004304194	Peripheral Blood	Human	24336226
2	ENRICHED_IN_MONOCYTES_(II)_(M11.0)		174	0.016894067	Peripheral Blood	Human	24336226
3	ENRICHED_IN_ANTIGEN_PRESENTATION_(II)_(M95.0)		18	0.020241437	Peripheral Blood	Human	24336226
4	PUTATIVE_TARGETS_OF_PAX3_(M89.0)		16	0.097893655	Peripheral Blood	Human	24336226
5	CELL_CYCLE_AND_TRANSCRIPTION_(M4.0)		301	0.10261368	Peripheral Blood	Human	24336226
6	ENRICHED_IN_ACTIVATED_DENDRITIC_CELLS/MONOCYTES_(M64)		16	0.116604	Peripheral Blood	Human	24336226
7	ENRICHED_IN_NK_CELLS_(I)_(M7.2)		46	0.12543195	Peripheral Blood	Human	24336226
8	ENRICHED_IN_ANTIGEN_PRESENTATION_(I)_(M71)		16	0.12386507	Peripheral Blood	Human	24336226
9	*TBA_(M66)*		16	0.12103282	Peripheral Blood	Human	24336226
10	CYTOSKELETON/ACTIN_(SRF_TRANSCRIPTION_TARGETS)_(M145.0)		15	0.1196479	Peripheral Blood	Human	24336226
11	LEUKOCYTE_DIFFERENTIATION_(M160)		15	0.15270436	Peripheral Blood	Human	24336226
12	EXTRACELLULAR_MATRIX_(II)_(M2.1)		43	0.17023705	Peripheral Blood	Human	24336226
13	MEMORY_B_CELL_SURFACE_SIGNATURE_(S9)		34	0.1759184	Peripheral Blood	Human	24336226
14	ENRICHED_IN_MEMBRANE_PROTEINS_(M124)		16	0.19762957	Peripheral Blood	Human	24336226
15	PLATELET_ACTIVATION_(II)_(M32.1)		18	0.1958992	Peripheral Blood	Human	24336226
16	ENRICHED_IN_T_CELLS_(I)_(M7.0)		55	0.22823855	Peripheral Blood	Human	24336226
17	T_CELL_ACTIVATION_(I)_(M7.1)		46	0.27058327	Peripheral Blood	Human	24336226
18	*TBA_(M184.0)*		17	0.2766716	Peripheral Blood	Human	24336226
19	T_CELL_ACTIVATION_(II)_(M7.3)		30	0.28000584	Peripheral Blood	Human	24336226
20	ACTIVATED_(LPS)_DENDRITIC_CELL_SURFACE_SIGNATURE_(S11)		35	0.30645204	Peripheral Blood	Human	24336226

Chaussabel

	Name	Description	# Genes	FDR	Tissue	Organism	PMID
1	*M9.32_Undetermined*		23	0.03650531	Peripheral Blood	Human	18631455
2	*M7.21_Undetermined*		73	0.34071872	Peripheral Blood	Human	18631455
3	M6.13_Cell_Death		38	0.43733603	Peripheral Blood	Human	18631455
4	*M7.15_Undetermined*		71	0.38143057	Peripheral Blood	Human	18631455
5	*M8.18_Undetermined*		18	0.3548228	Peripheral Blood	Human	18631455
6	*M8.41_Undetermined*		15	0.33282864	Peripheral Blood	Human	18631455
7	M4.6_Inflammation		93	0.33361447	Peripheral Blood	Human	18631455
8	M3.1_Erythrocytes		62	0.44778514	Peripheral Blood	Human	18631455
9	*M7.16_Not_Determined*		48	0.42654887	Peripheral Blood	Human	18631455
10	*M8.8_Undetermined*		26	0.3929025	Peripheral Blood	Human	18631455
11	*M9.31_Undetermined*		22	0.40187752	Peripheral Blood	Human	18631455
12	*M6.14_Not_Determined*		30	0.38957456	Peripheral Blood	Human	18631455
13	M3.6_Cytotoxic/NK_Cell		41	0.3637774	Peripheral Blood	Human	18631455
14	*M9.6_Undetermined*		70	0.36457974	Peripheral Blood	Human	18631455
15	M3.2_Inflammation		106	0.524135	Peripheral Blood	Human	18631455
16	*M7.22_Undetermined*		42	0.50671184	Peripheral Blood	Human	18631455
17	*M8.26_Undetermined*		17	0.47988358	Peripheral Blood	Human	18631455
18	*M7.13_Not_Determined*		46	0.46933657	Peripheral Blood	Human	18631455
19	*M5.4_Not_Determined*		81	0.45617348	Peripheral Blood	Human	18631455
20	M7.1_Inflammation		141	0.4750514	Peripheral Blood	Human	18631455

Supplemental Table 6.8. List of ImmuneSigDB gene sets listed in Figure 6.7A.

Rank	Gene Set Name	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	GSE14000_UNSTIM_VS_4H_LPS_DC_DN_HOMO_SAPIENS	200	0.94	3.56	0	0	0	1150	tags=92%, list=6%, signal=96%
2	GSE2706_UNSTIM_VS_2H_LPS_DC_DN_HOMO_SAPIENS	200	0.92	3.54	0	0	0	1207	tags=81%, list=6%, signal=85%
3	GSE2706_UNSTIM_VS_8H_LPS_DC_DN_HOMO_SAPIENS	200	0.91	3.5	0	0	0	1325	tags=80%, list=6%, signal=85%
4	GSE4748_CTRL_VS_LPS_STIM_DC_3H_DN_HOMO_SAPIENS	200	0.9	3.49	0	0	0	1311	tags=71%, list=6%, signal=75%
5	GSE14000_UNSTIM_VS_16H_LPS_DC_DN_HOMO_SAPIENS	200	0.89	3.39	0	0	0	1793	tags=78%, list=9%, signal=84%
6	GSE22886_CTRL_VS_LPS_24H_DC_DN_HOMO_SAPIENS	200	0.87	3.34	0	0	0	1244	tags=63%, list=6%, signal=66%
7	GSE4984_UNTREATED_VS_LPS_TREATED_DC_DN_HOMO_SAPIENS	200	0.85	3.26	0	0	0	1375	tags=58%, list=7%, signal=61%
8	GSE3982_CTRL_VS_LPS_4H_MAC_DN_HOMO_SAPIENS	199	0.83	3.18	0	0	0	1683	tags=59%, list=8%, signal=64%
9	GSE3982_CTRL_VS_LPS_48H_DC_DN_HOMO_SAPIENS	200	0.79	3.04	0	0	0	1343	tags=49%, list=6%, signal=51%
10	GSE14769_UNSTIM_VS_80MIN_LPS_BMDM_DN_MUS_MUSCULUS	194	0.78	2.97	0	0	0	1326	tags=50%, list=6%, signal=53%
11	GSE30971_CTRL_VS_LPS_STIM_MACROPHAGE_WBP7_HET_2H_DN_MUS_MUSCULUS	165	0.78	2.94	0	0	0	1461	tags=51%, list=7%, signal=54%
12	GSE14769_UNSTIM_VS_40MIN_LPS_BMDM_DN_MUS_MUSCULUS	196	0.71	2.74	0	0	0	1344	tags=46%, list=6%, signal=49%
13	GSE14769_UNSTIM_VS_60MIN_LPS_BMDM_DN_MUS_MUSCULUS	194	0.71	2.71	0	0	0	1405	tags=46%, list=7%, signal=49%
14	GSE30971_CTRL_VS_LPS_STIM_MACROPHAGE_WBP7_HET_4H_DN_MUS_MUSCULUS	158	0.7	2.63	0	0	0	1633	tags=42%, list=8%, signal=46%
15	GSE32255_UNSTIM_VS_4H_LPS_STIM_DC_DN_MUS_MUSCULUS	159	0.68	2.6	0	0	0	2104	tags=42%, list=10%, signal=47%

Supplemental Table 6.9. List of ImmuneSigDB gene sets listed in Figure 6.7B.

Rank	Gene Set Name	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	GSE22886_NAIVE_BCELL_VS_BM_PLASMA_CELL_DN_HOMO_SAPIENS	200	0.78	3.79	0	0	0	1241	tags=66%, list=10%, signal=72%
2	GSE13411_NAIVE_BCELL_VS_PLASMA_CELL_DN_HOMO_SAPIENS	200	0.74	3.62	0	0	0	2254	tags=75%, list=18%, signal=89%
3	GSE22886_NAIVE_BCELL_VS_BLOOD_PLASMA_CELL_DN_HOMO_SAPIENS	200	0.69	3.35	0	0	0	894	tags=43%, list=7%, signal=45%
4	GSE12366_PLASMA_CELL_VS_NAIVE_BCELL_UP_HOMO_SAPIENS	143	0.62	2.9	0	0	0	2881	tags=60%, list=22%, signal=77%
5	GSE4142_NAIVE_BCELL_VS_PLASMA_CELL_DN_MUS_MUSCULUS	166	0.55	2.58	0	0	0	2558	tags=43%, list=20%, signal=53%
6	GSE42724_NAIVE_BCELL_VS_PLASMABLAST_DN_HOMO_SAPIENS	167	0.52	2.44	0	0	0	2583	tags=41%, list=20%, signal=50%

Supplemental Table 6.10. List of ImmuneSigDB gene sets listed in Figure 6.7C.

Rank	Gene Set Name	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	GSE25087_TREG_VS_TCONV_ADULT_UP_HOMO_SAPIENS	200	0.88	3.47	0	0	0	1625	tags=79%, list=8%, signal=85%
2	GSE25087_TREG_VS_TCONV_FETUS_UP_HOMO_SAPIENS	200	0.82	3.22	0	0	0	2368	tags=67%, list=12%, signal=75%
3	GSE23321_CENTRAL_MEMORY_VS_NAIVE_CD8_TCELL_UP_HOMO_SAPIENS	193	0.81	3.17	0	0	0	2061	tags=58%, list=10%, signal=63%
4	GSE23321_EFFECTOR_MEMORY_VS_NAIVE_CD8_TCELL_UP_HOMO_SAPIENS	195	0.79	3.09	0	0	0	1666	tags=52%, list=8%, signal=56%
5	GSE7460_TCONV_VS_TREG_THYMUS_DN_MUS_MUSCULUS	200	0.7	2.77	0	0	0	1705	tags=36%, list=8%, signal=38%
6	GSE7852_TREG_VS_TCONV_LN_UP_MUS_MUSCULUS	197	0.7	2.76	0	0	0	1690	tags=37%, list=8%, signal=40%
7	GSE22045_TREG_VS_TCONV_UP_HOMO_SAPIENS	200	0.68	2.7	0	0	0	2021	tags=40%, list=10%, signal=44%
8	GSE7460_TCONV_VS_TREG_LN_DN_MUS_MUSCULUS	198	0.67	2.67	0	0	0	1889	tags=35%, list=9%, signal=38%
9	GSE7852_TREG_VS_TCONV_THYMUS_UP_MUS_MUSCULUS	199	0.68	2.66	0	0	0	1755	tags=35%, list=9%, signal=38%
10	GSE22886_NAIVE_VS_IGG_IGA_MEMORY_BCELL_DN_HOMO_SAPIENS	200	0.61	2.42	0	0	0	2885	tags=35%, list=14%, signal=40%
11	KAECH_NAIVE_VS_MEMORY_CD8_TCELL_DN_MUS_MUSCULUS	198	0.61	2.4	0	0	0	1755	tags=32%, list=9%, signal=35%
12	GSE13738_RESTING_VS_TCR_ACTIVATED_CD4_TCELL_DN_HOMO_SAPIENS	200	0.6	2.35	0	0	0	1893	tags=32%, list=9%, signal=35%
13	GSE12366_NAIVE_VS_MEMORY_BCELL_DN_HOMO_SAPIENS	200	0.6	2.34	0	0	0	1192	tags=26%, list=6%, signal=27%
14	GSE37532_TREG_VS_TCONV_CD4_TCELL_FROM_LN_UP_MUS_MUSCULUS	178	0.61	2.33	0	0	0	2343	tags=37%, list=12%, signal=41%

Supplemental Table 6.11. List of ImmuneSigDB gene sets listed in Figure 6.7D.

Rank	Gene Set Name	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX	LEADING EDGE
1	GSE42724_NAIVE_VS_MEMORY_BCELL_DN_HOMO_SAPIENS	200	0.87	3.6	0	0	0	1704	tags=84%, list=9%, signal=92%
2	GSE22886_NAIVE_VS_IGG_IGA_MEMORY_BCELL_DN_HOMO_SAPIENS	184	0.81	3.32	0	0	0	1372	tags=62%, list=7%, signal=66%
3	GSE22886_NAIVE_VS_IGM_MEMORY_BCELL_DN_HOMO_SAPIENS	183	0.8	3.26	0	0	0	1330	tags=57%, list=7%, signal=61%
4	GSE12366_NAIVE_VS_MEMORY_BCELL_DN_HOMO_SAPIENS	175	0.8	3.19	0	0	0	1787	tags=59%, list=10%, signal=65%
5	GSE17186_MEMORY_VS_NAIVE_BCELL_UP_HOMO_SAPIENS	184	0.67	2.76	0	0	0	1894	tags=38%, list=10%, signal=42%
6	GSE13411_NAIVE_VS_IGM_MEMORY_BCELL_DN_HOMO_SAPIENS	185	0.6	2.42	0	0	0	1650	tags=30%, list=9%, signal=33%
7	GSE13411_NAIVE_VS_MEMORY_BCELL_DN_HOMO_SAPIENS	186	0.56	2.26	0	0	0	2126	tags=28%, list=11%, signal=31%
8	GSE11386_NAIVE_VS_MEMORY_BCELL_DN_MUS_MUSCULUS	185	0.55	2.26	0	0	0	2652	tags=33%, list=14%, signal=38%
9	GSE4142_NAIVE_VS_MEMORY_BCELL_DN_MUS_MUSCULUS	183	0.53	2.15	0	0	0	1831	tags=26%, list=10%, signal=29%
10	GSE13411_NAIVE_VS_SWITCHED_MEMORY_BCELL_DN_HOMO_SAPIENS	191	0.46	1.86	0	0	0.001	2200	tags=21%, list=12%, signal=24%

Supplemental Table 6.12. List of ImmuneSigDB gene sets listed in Figure 6.8B.

Rank	Gene set
1	GSE22886_NAIVE_BCELL_VS_NEUTROPHIL_DN
2	GSE29618_MONOCYTE_VS_PDC_UP
3	GSE6269_E_COLI_VS_STREP_PNEUMO_INF_PBMC_DN
4	GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE_DN
5	GSE34156_UNTREATED_VS_24H_NOD2_AND_TLR1_TLR2_LIGAND_TREATED_MONOCYTE_DN
6	GSE34156_UNTREATED_VS_24H_TLR1_TLR2_LIGAND_TREATED_MONOCYTE_DN
7	GSE6269_E_COLI_VS_STREP_AUREUS_INF_PBMC_DN
8	GSE22886_NAIVE_TCELL_VS_MONOCYTE_DN
9	GSE29618_MONOCYTE_VS_MDC_UP
10	GSE22886_NAIVE_CD8_TCELL_VS_MONOCYTE_DN

Supplemental Table 6.13. List of ImmuneSigDB gene sets listed in Figure 6.9B.

Rank	Gene set
1	GSE6269_E_COLI_VS_STREP_PNEUMO_INF_PBMC_DN
2	GSE22886_NAIVE_BCELL_VS_NEUTROPHIL_DN
3	GSE3982_NEUTROPHIL_VS_NKCELL_UP
4	GSE6269_HEALTHY_VS_STREP_PNEUMO_INF_PBMC_DN
5	GSE34156_NOD2_LIGAND_VS_NOD2_AND_TLR1_TLR2_LIGAND_24H_TREATED_MONOCYTE_DN
6	GSE34156_UNTREATED_VS_24H_NOD2_AND_TLR1_TLR2_LIGAND_TREATED_MONOCYTE_DN
7	GSE6269_E_COLI_VS_STREP_AUREUS_INF_PBMC_DN
8	GSE29618_MONOCYTE_VS_MDC_UP
9	GSE6269_HEALTHY_VS_STREP_AUREUS_INF_PBMC_DN
10	GSE34156_UNTREATED_VS_24H_TLR1_TLR2_LIGAND_TREATED_MONOCYTE_DN

Supplemental Table 6.14. List of ImmuneSigDB gene sets listed in Figure 6.10B.

Rank	Gene set
1	GSE22886_NAIVE_TCELL_VS_MONOCYTE_UP
2	GSE10325_LUPUS_CD4_TCELL_VS_LUPUS_MYELOID_UP
3	GSE22886_NAIVE_CD4_TCELL_VS_MONOCYTE_UP
4	GSE10325_LUPUS_BCELL_VS_LUPUS_MYELOID_UP
5	GSE22886_NAIVE_CD8_TCELL_VS_MONOCYTE_UP
6	GSE22229_RENAL_TRANSPLANT_IMMUNOSUPP_THERAPY_VS_HEALTHY_PBMC_DN
7	GSE10325_BCELL_VS_MYELOID_UP
8	GSE11057_PBMC_VS_MEM_CD4_TCELL_DN
9	GSE22886_NAIVE_TCELL_VS_DC_UP
10	GSE6269_HEALTHY_VS_STAPH_AUREUS_INF_PBMC_UP

Appendix

Appendix A: Godec et al. Immunity 2016

This is a reprint of:

Jernej Godec, Yan Tan, Arthur Liberzon, Pablo Tamayo, Atul J. Butte, Jill P. Mesirov, W. Nicholas Haining. A novel compendium of immune signatures identifies both conserved and species-specific biology in the mouse and human response to inflammation. *Immunity*. 2016 Jan 19;44(1):194-206. (PMID: 26795250)

Immunity
Resource

Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation

Jernej Godec,[1,2] Yan Tan,[3,4] Arthur Liberzon,[3] Pablo Tamayo,[3] Sanchita Bhattacharya,[5] Atul J. Butte,[5] Jill P. Mesirov,[3,4] and W. Nicholas Haining[1,3,6,*]
[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA
[2]Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Bioinformatics Program, Boston University, Boston, MA 02215, USA
[5]Institute for Computational Health Science, University of California, San Francisco, San Francisco, CA 94158, USA
[6]Division of Hematology/Oncology, Children's Hospital, Harvard Medical School, Boston, MA 02115, USA
*Correspondence: nicholas_haining@dfci.harvard.edu
http://dx.doi.org/10.1016/j.immuni.2015.12.006

SUMMARY

Gene-expression profiling has become a mainstay in immunology, but subtle changes in gene networks related to biological processes are hard to discern when comparing various datasets. For instance, conservation of the transcriptional response to sepsis in mouse models and human disease remains controversial. To improve transcriptional analysis in immunology, we created ImmuneSigDB: a manually annotated compendium of ~5,000 gene-sets from diverse cell states, experimental manipulations, and genetic perturbations in immunology. Analysis using ImmuneSigDB identified signatures induced in activated myeloid cells and differentiating lymphocytes that were highly conserved between humans and mice. Sepsis triggered conserved patterns of gene expression in humans and mouse models. However, we also identified species-specific biological processes in the sepsis transcriptional response: although both species upregulated phagocytosis-related genes, a mitosis signature was specific to humans. ImmuneSigDB enables granular analysis of transcriptomic data to improve biological understanding of immune processes of the human and mouse immune systems.

INTRODUCTION

Experiments in both human cells and mouse models have been used to discover many of the mechanisms by which the immune system functions. Identifying aspects of immunobiology that are evolutionarily conserved between humans and mouse models is useful because it can reveal mechanisms of fundamental importance to both species. Moreover, it can provide reassurance that information gleaned from mouse models will be applicable to the human condition. This is crucial, because much of immunobiology cannot be examined physiologically in humans due to inaccessibility of certain tissues or cell types or the difficulty in recapitulating complex biological milieu in vitro. However, considerable controversy exists as to the degree to which mouse models can recapitulate events occurring in immunologic disease states in humans (Davis, 2008; Hackam and Redelmeier, 2006; Rice, 2012; Shay et al., 2014; van der Worp et al., 2010; Warren et al., 2014). These concerns have extended to the analysis of genome-wide analysis of mRNA levels where analyses of the same datasets from mouse and human sepsis reached opposite conclusions regarding the degree of cross-species similarity (Seok et al., 2013; Takao and Miyakawa, 2014). Contradictory findings have also been reported in the comparison of gene expression across a range of human and mouse tissues (Gilad Y, 2015; Lin et al., 2014).

One of the challenges in identifying similarities between gene-expression datasets is that major changes in the cell state can be associated with relatively small alterations in the expression level of a relatively large numbers of genes. Analysis of co-regulated changes in sets of functionally related genes, rather than individual genes, has therefore become an important strategy to identify subtle, but biologically meaningful, differences in gene expression (Haining and Wherry, 2010; Mootha et al., 2003; Subramanian et al., 2005). This is a particularly useful approach when analyzing samples in which experimental variability (such as those collected from heterogeneous human subjects) or evolutionary divergence (such as comparisons between species) add experimental "noise" to gene-expression profiles. Several approaches for testing for the enrichment of gene sets have been developed, including gene set enrichment analysis (GSEA) (Subramanian et al., 2005). GSEA has been made more powerful by the availability of curated collections of gene-expression signatures extracted from a variety of sources including published experimental datasets. The largest of these collections, the Molecular Signatures Database (MSigDB), contains more than 8,000 signatures (Liberzon et al., 2011). However, only a small fraction of these gene sets pertain to immune processes and cell types.

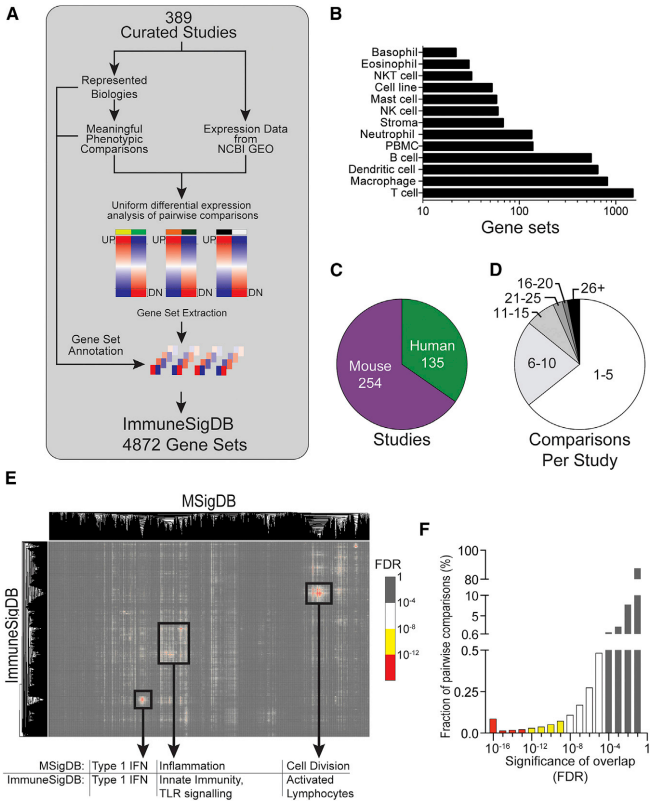

Figure 1. ImmuneSigDB Collection Is Derived from Re-Analysis of Published Data

(A) A schematic of the ImmuneSigDB pipeline.

(B) Number gene sets corresponding to major immune lineages or cell lines and (C) species of origin contained in ImmuneSigDB.

(D) Number of pairwise comparisons made per each study used in ImmuneSigDB.

(E) Overlap in gene set membership in ImmuneSigDB with MSigDB gene sets. Heatmap indicates False Discovery Rate (FDR) values of each pairwise comparison between gene sets. Highlighted are representative biological processes in each of the significantly overlapping clusters of gene sets.

(F) Distribution of the FDR ranges of significance across all pair-wise comparisons of gene set membership. See also Figures S1 and S2.

RESULTS

Generating a Compendium of Gene Signatures Curated from Immune-Expression Profiles

We generated a comprehensive compendium of gene sets pertaining to immune biology. The term "gene-set" in this study refers to groups of genes identified by selecting either up- or downregulated genes in comparisons of gene-expression profiles of interest. We identified and uniformly analyzed 389 published studies in the immunology literature that included genome-wide expression profiling data (outlined in Figure 1A). We selected studies to analyze based on immunological key words in the title or abstract followed by additional manual review. We prioritized studies published in immunology journals of broad interest (Table S1). We identified the corresponding publicly available datasets in the NCBI Gene Expression Omnibus (GEO) and, for uniformity, focused on studies performed on Affymetrix platforms (Table S1) that included three or more biological replicates. Each study was reviewed to identify and annotate the biology represented and to define meaningful pairwise comparisons that would create biologically useful gene sets. For example, an individual study might include a single comparison, such as stimulated versus unstimulated cells, or might have multiple comparisons, as is the case where several cell types were subjected to different culture conditions and analyzed at several time points. In such cases, only meaningful pairwise comparisons, rather than all possible comparisons, were made (Figure S1).

The raw expression data obtained from each GEO study was pre-processed uniformly (see Experimental Procedures). We identified and extracted differentially expressed genes (see Experimental Procedures and Figure 1A), which comprised the gene sets for the ImmuneSigDB collection. These sets represented genes coordinately up- or downregulated in many major

We now report the creation of ImmuneSigDB (http://software.broadinstitute.org/gsea/msigdb/collections.jsp#C7), a compendium of ~5,000 well-annotated signatures generated by analysis of 389 published studies of cell states and perturbations in the mouse and human immune systems. Using this collection of signatures, we demonstrated that signatures of cell differentiation in lymphoid cells and endotoxin stimulation in myeloid cells are highly conserved between humans and mouse models. Moreover, analysis of the transcriptional response to sepsis in human samples and mouse models showed that there was highly significant conservation of gene expression between the species when measured at the gene set level. However, in addition to the conserved transcriptional programs, we also identify species-specific differences in the biological processes associated with sepsis. These findings help interpret contradictory observations regarding the extent of evolutionary conservation in the transcriptional response to sepsis. ImmuneSigDB will enable the detailed analysis of cross-species gene expression that is critical to establishing which biological processes are conserved and which are not, thus allowing mouse models to better inform our understanding of human disease.

immune cell types (Figure 1B) either in their baseline state or following a range of genetic or chemical perturbations. ImmuneSigDB included data from healthy human subjects, patients with immune or non-immune diseases, and mouse models. Mapping orthologous genes to a common identifier allowed us to include both human (n = 135) and mouse (n = 254) studies (Figure 1C). From these studies, we identified 2,436 meaningful comparisons and extracted 4,872 gene sets of up- or downregulated genes comprising the ImmuneSigDB (see Experimental Procedures). The number of gene sets identified per published study ranged from one comparison, (e.g., representing an activated versus unperturbed state or knockout versus wild-type cell) to over 50 (e.g., often representing several cell types cultured in different conditions for varying amounts of time) (Figure 1D). Particular biological conditions over-represented in the literature, such as those related to T cell biology, are correspondingly over-represented in ImmuneSigDB, with slightly fewer gene sets from myeloid cells and B cells (Figure 1B). ImmuneSigDB is publicly available at http://www.msigdb.org.

ImmuneSigDB Expands the Biological Coverage of the MSigDB

We compared the gene sets generated from immune cells (ImmuneSigDB) with those in gene sets in the MSigDB collection. MSigDB is a curated collection of gene sets generated from published gene-expression studies that are generally not from the immunology literature (Liberzon et al., 2011). We measured overlap in constituent genes between each gene set in the ImmuneSigDB and all the other MSigDB collections and found that only a small minority of gene sets significantly overlapped (Figures 1E and 1F), suggesting that ImmunSigDB added a large amount of new transcriptional information. A small subset of gene sets in ImmuneSigDB and MSigDB were highly similar (0.64% of gene sets with $p < 10^{-8}$) and these could be clustered into three groups related to proliferation, inflammation, or type 1 interferon response (Figure 1E). This suggested that with the exception of these core biological processes, gene sets derived from immune cell expression profiles contained genes distinct from non-immune-related gene-expression profiles that previously predominated the MSigDB.

We performed an analogous analysis of pairwise overlaps in gene membership between gene sets within ImmuneSigDB. While most were unique, we found a larger number of gene sets with significant overlap (1.46% with $p < 10^{-8}$) within ImmuneSigDB than between ImmuneSigDB and MSigDB (Figure S2A). These gene sets largely related to lineage-specific signatures shared between datasets generated from similar types of cells. Therefore, ImmuneSigDB has minimal overlap with MSigDB and provides new gene sets describing immune biology.

ImmuneSigDB Provides a Complementary Resource to Existing Immune Module Collections

Several groups have previously created collections of gene modules in the immune system. In studies by Chaussabel et al. (2008) and Li et al. (2014), existing studies of gene-expression profiles in human peripheral blood mononuclear cell (PBMC) or whole blood were analyzed to identify modules of co-regulated genes to aid in the analysis of gene-expression profiles from immune cells. Several features distinguish ImmuneSigDB from either of these collections (summarized in Table S2). First, ImmuneSigDB was generated by direct comparison of the genes that were up- or downregulated in two known sample classes from each study. This allowed the published study to serve as a source of comprehensive annotation of each gene set, in contrast to either of the module collections that were generated by analysis of aggregated pools of samples, limiting the direct experimental annotation of each module. Second, ImmuneSigDB was considerably larger than either module collection (Table S2). Third, ImmuneSigDB included data from both mouse models and humans, and from 13 cell or tissue types, rather than solely from human PBMC and whole blood profiles.

To compare directly the gene-sets in ImmuneSigDB with the module collections of Chaussabel and Li, we measured overlap in constituent genes between each gene set in the ImmuneSigDB and all modules in either the Chaussabel or Li collections (Figures S2B and S2C). We found that only a small fraction of ImmuneSigDB gene sets significantly overlapped with either collection (0.06% and 0.18% with FDR of $< 10^{-8}$ for Chaussabel and Li, respectively), suggesting that the gene-sets within ImmuneSigDB and the modules in the Chaussabel and Li collections were largely distinct. The small number of significantly overlapping gene-sets and/or modules contained genes predominantly related to immune cell lineages (e.g., T cell or myeloid) or to the response to interferon-α (IFN-α) or Toll-like receptor (TLR) ligands. Interestingly, the overlap between modules contained in the Chaussabel and Li collections was similarly limited (Figure S2D), suggesting analysis of immune-expression profiles using each of the three collections could provide complementary information.

Finally, we performed GSEA using four published datasets in human immune cells (LPS stimulated DC, Tregs, plasma B cells, and memory B cells) to compare the results using ImmuneSigDB with the module collections by Chaussabel and Li (Figure S2E). A larger number of ImmuneSigDB gene sets were significantly enriched (even after correcting for multiple hypothesis testing) in each of the four datasets than with either the Chaussabel or Li collections. Moreover, inspection of the top 20 most enriched gene sets from ImmuneSigDB and modules from the Chaussabel or Li collections illustrates the extensive biological annotations (including links to the original studies) available for each ImmuneSigDB gene set (Table S3). Thus analysis with ImmuneSigDB provides a resource for the analysis of gene-expression data in the immune system that is complementary to existing collections.

Enrichment of ImmuneSigDB Gene Sets Recapitulates Known Lineage-Specific Differences in Mouse and Human Hematopoietic Cell Lineages

We next tested whether enrichment analysis of gene expression using ImmuneSigDB could recapitulate known differences in lineage-specific gene expression within the immune system. We analyzed a large, publicly available dataset of gene-expression profiles from the Immunological Genome Project (ImmGen) consisting of immune cell types and cell states in mice (Heng et al., 2008) using a single sample version of GSEA (ssGSEA) (Barbie et al., 2009). In this approach, gene sets are tested for enrichment in the list of genes in a single sample ranked by absolute expression rather than by comparison with another sample. The resulting ssGSEA scores provided an estimate of the

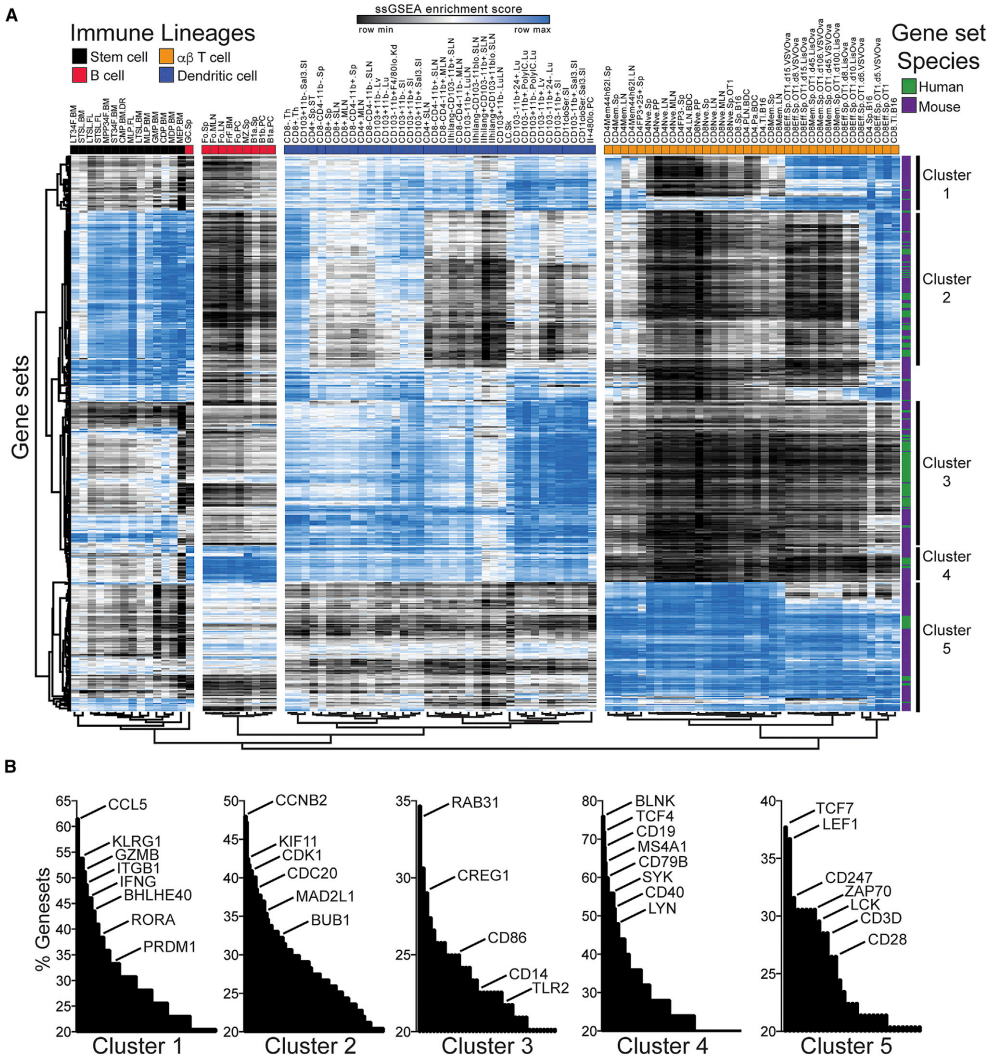

Figure 2. Mouse Immune Lineages Are Accurately Clustered using ImmuneSigDB Enrichments

(A) Unsupervised bi-clustering of ssGSEA values using ImmuneSigDB in samples of four representative mouse immune lineages. Hierarchical clustering of the 10% of gene sets with highest mean absolute deviation is shown. Species of origin of gene sets indicated by green (human) and purple (mouse) bars on the right. Major branches of the gene set dendrogram clusters are indicated by the numbered black bars on the right.

(B) Distribution of genes contained in gene sets in the same gene set dendrogram clusters as indicated in (A). See also Figure S3.

degree of enrichment of each ImmuneSigDB gene set in each individual sample in the dataset. In this way, we generated a dataset containing as rows the profiles of enrichment of each ImmuneSigDB gene set and as columns the samples. Unsupervised hierarchical clustering of samples from four distinct immune cell types—dendritic cells, B cells, αβ T cells, and stem cells—in the space of gene set enrichment scores revealed near-perfect clusters of the respective cell types (Figure 2A).

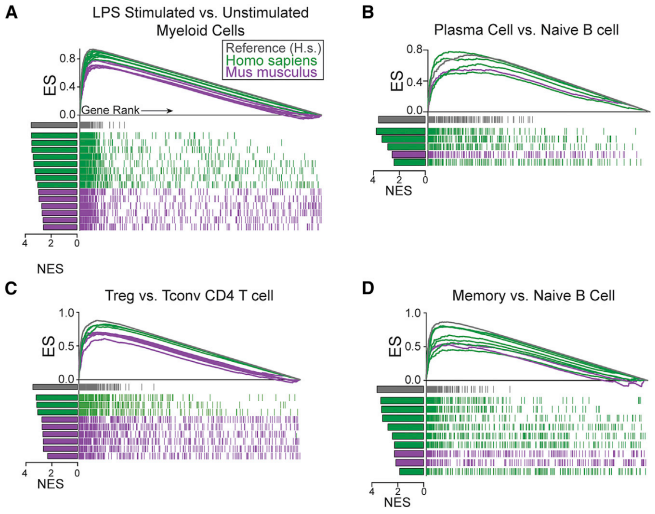

Figure 3. Transcriptional Programs Are Conserved across Mouse and Human Immune Lineages

(A) GSEA of a randomly selected human study comparing LPS-stimulated and unstimulated dendritic cells using ImmuneSigDB gene sets derived from the study itself (gray) or gene sets from other mouse (purple) or human (green) datasets of LPS-stimulated myeloid cells. Mountain plots show all genes ranked by differential expression in sepsis versus control conditions on the x axis, and the curves indicate cumulative enrichment (measured by enrichment score on the y axis). The ticks below the line correspond to the position of genes in each gene set.

(B-D) Analysis as in (A) for three additional cell differentiation states: plasma cells (B), Tregs (C), and memory B cells (D). All gene sets shown are significantly enriched (FDR < 0.001).

Within each lineage, subgroups, such as naive T cells or memory T cells also were clustered accurately together. Similarly, accurate clustering of different lineages was observed when we analyzed human-derived cells in the Differentiation Map (DMAP) (Figure S3) (Novershtern et al., 2011). Hematopoietic stem cells (HSCs) were accurately distinguished from other lineages despite the fact that very few (1.62%) gene sets derived from stem cells were included in the ImmuneSigDB. This suggests that HSCs are characterized by differential expression of gene sets related to biological processes shared with immune cells (Luckey et al., 2006).

We noted that distinct clusters of gene sets showed differential enrichment in specific cell lineages (Figure 2A, clusters 1–5). We characterized these gene set clusters by determining the relative frequencies of genes shared by the gene sets in these clusters (Figure 2B). For example, gene sets in cluster 1, which predominantly distinguished effector and memory T cells from naive T cells, most commonly included genes encoding effector molecules such as granzyme B (*GZMB*), IFN-γ (*IFNG*), as well as Blimp1 (*PRDM1*) and integrin beta 1 (*ITGB1*), and were predominantly derived from expression profiles of effector and memory $CD8^+$ T cells in the context of viral infection and anti-tumoral responses (Table S4, top). Cluster 5, which predominantly distinguished T cells from other cell lineages, included T cell genes such as transcription factors TCF7 and LEF1 as well as components of T cell receptor signaling, CD3ζ (*CD247*), CD3δ (*CD3D*), ZAP70, and Lck and included most gene sets derived from comparing T cells to other immune cell types (Figure 2B and Table S4, bottom). Stem cells showed strong enrichment of gene sets in cluster 2 whose predominant genes play a dominant role in regulating cell cycle (Figure 2B and Table S4, bottom). B cells and dendritic cells were distinguished by a separate cluster of gene sets that included known genes representing those lineages.

We noted that each gene set cluster that distinguished different lineages included both human and mouse gene sets (Figure 2, purple and green bars), suggesting that similar patterns of enrichment were observed using gene sets derived from expression profiles from tissues of both species. Thus, ImmuneSigDB robustly clustered human and mouse immune lineages based on whole transcriptome enrichments of both mouse and human-derived gene sets.

Analogous Cell Types and Contexts in Mice and Humans Show Common Patterns of Gene Expression

We, and others, have previously used GSEA to show that the transcriptional profiles from memory and exhausted $CD8^+$ T cells are highly concordant between mouse and human datasets (Baitsch et al., 2011; Haining et al., 2008; Quigley et al., 2010). To test whether this similarity in gene expression between species is also observed for other cell states we extended this analysis to other cell types and perturbations included in ImmuneSigDB. We focused on four separate biological comparisons (Figure 3) where analysis of gene expression had been made in analogous cell types or perturbations in both human and mouse immune cells. This allowed us to test whether sets of genes differentially expressed in mouse immune cells showed enrichment in profiles from the analogous comparisons in humans and vice versa (Sweet-Cordero et al., 2005).

We first identified 15 studies (6 mouse; 9 human) in which the transcriptional response to lipopolysaccharide (LPS) stimulation had been profiled in myeloid cells; each study had been used to generate a gene set in ImmuneSigDB. We selected one human dataset and generated a ranked list of genes differentially expressed following LPS stimulation. We then performed GSEA using gene sets from the other 14 mouse and human datasets. We found that both human- and mouse-derived gene sets showed highly significant enrichment (FDR < 0.001), suggesting a strong conservation in the transcriptional response to LPS between the two species (Figure 3A). Gene sets derived from studies on

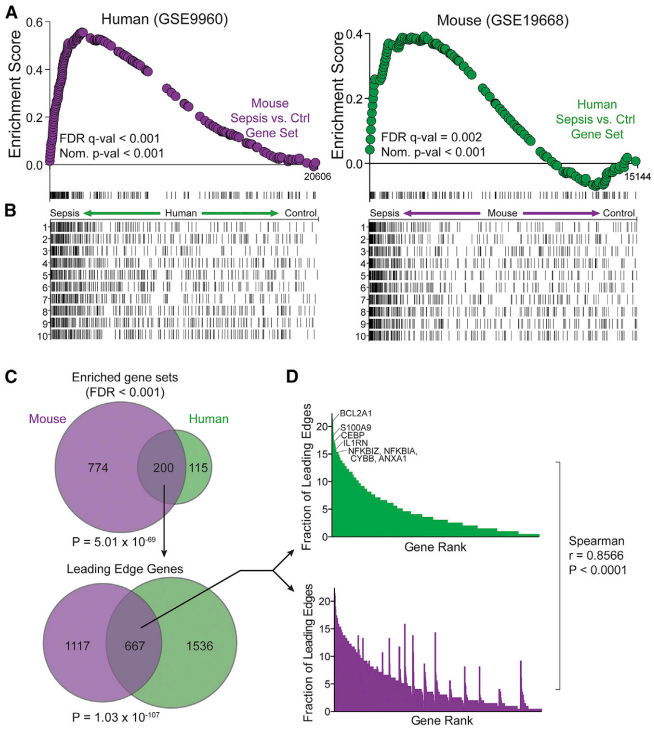

Figure 4. The Transcriptional Response to Sepsis Is Conserved in Humans and Mouse Models

(A and B) GSEA of the set of genes upregulated in mouse sepsis (GSE19668, C57BL/6) in the ranked list of genes upregulated in human sepsis (GSE9960, Gram negative infection) (A, left); and of the corresponding human sepsis gene set enriched in rank ordered list of genes upregulated in mouse sepsis (A, right). Mountain plots indicate cumulative enrichment, and (B) ticks below the line correspond to the position of genes in the ten most enriched gene sets from ImmuneSigDB in the rank order of genes upregulated in sepsis versus control conditions (x axis).

(C) Venn diagram showing overlap in the identity of significantly enriched ImmuneSigDB gene sets in mouse (purple) or human (green) sepsis dataset (top) and the number of shared leading edge genes in the gene sets enriched in both species (bottom). Statistical significance calculated by the hypergeometric test.

(D) Frequency of the genes occurring in the leading edge of a gene sets enriched in human (green) and mouse (purple) sepsis datasets. Statistical significance of the similarity in gene rank is calculated by the Spearman test. See also Figure S4.

human cells tended to show slightly higher enrichment scores than those generated from mouse cells.

We then selected additional cross-species comparisons that were represented by multiple datasets within the ImmuneSigDB. We found cross-species similarity in the gene-expression profiles of comparisons of plasma versus naive B cells, memory B cells versus naive B cells, and regulatory T (Treg) versus conventional T cells (Tconv) (FDR < 0.001, Figure 3B–3D and Table S5). Furthermore, we observed that the biology was not just conserved to the same extent but in some cases mouse-derived gene sets were more strongly enriched in human datasets than other human gene sets, as depicted by the peak height of the respective graphs of their GSEA enrichment scores. These findings indicate that components of the transcriptional signatures of LPS stimulation and some T and B cell differentiation programs are similar in humans and mouse models.

Blood Cells from Human and Mouse Sepsis Share Conserved Biology Reflected in Their Transcriptomes

As we observed common patterns of gene expression in these cross-species comparisons, we next studied more a complex transcriptional dataset from human sepsis and the corresponding mouse models to test whether ImmuneSigDB could resolve similarities or differences between human and mouse transcriptional profiles. Recent studies have analyzed the transcriptional response to sepsis in multiple datasets of gene-expression profiles obtained from human PBMC samples or from mouse models (Seok et al., 2013; Takao and Miyakawa, 2014). However, these studies have differed in their conclusions regarding the degree of similarity between species. We reasoned that analysis with ImmuneSigDB might allow a more detailed analysis of immune signatures elicited by the sepsis response in both species.

We began by using ImmuneSigDB to compare the similarity in gene expression in human and mouse datasets included in the previous studies of the transcriptional responses to sepsis. We selected, at random, a pair of human and mouse studies in which peripheral blood cell gene expression was measured in sepsis versus control conditions (PBMC following sepsis in human [GSE9960] and mouse [GSE19668]) (Ahn et al., 2010; Tang et al., 2009). We first identified genes upregulated following sepsis in each study and tested whether that signature was enriched in the corresponding profile of sepsis versus control in the other species.

We observed strong enrichment of the set of sepsis-induced genes derived from the mouse study in the human dataset (FDR < 0.001, Figure 4A, left). Similarly, we found that a gene set comprising genes upregulated in human PBMC samples in sepsis versus control was strongly enriched in the mouse sepsis gene-expression profile (FDR = 0.002, Figure 4A, right). This internal comparison suggests that there was marked similarity between the genes upregulated by sepsis in humans and in a mouse model.

Next, we identified similarity in gene expression in the sepsis response by testing for enrichment of all gene sets in ImmuneSigDB gene sets in the same pair of human and mouse studies. We compared the ImmuneSigDB gene sets that were significantly enriched in the gene-expression profiles of the human and mouse gram negative and/or positive sepsis response (Figure 4B and Figures S4A–S4D). We observed marked similarity in ImmuneSigDB gene sets that were enriched the sepsis-induced signatures in each species ($p = 5.01 \times 10^{-69}$, Figures 4B and 4C, and Table S6).

To identify which genes in the gene sets that were enriched in both species were "driving" the enrichment of the shared gene sets, we focused on the "leading edge" of enrichment. Leading edge genes in a gene set enrichment analysis are those that contribute most to the enrichment of a particular gene set and include the most significantly upregulated genes in a given gene set. We found that the leading edges of gene sets that were enriched in both species were similar (Spearman $r = 0.857$, $p < 0.0001$; Figure 4D, S4D, and Table S7) indicating that the strong enrichment of shared gene sets is due to the upregulation of similar genes. We found the same results when we performed the same set of analyses using a pair of human and mouse datasets where both were from gram-positive sepsis or when we analyzed gene sets enriched in downregulated genes in sepsis compared to control (Figure S4, Table S7). These data demonstrate a high degree of concordance in gene sets that are enriched following sepsis in humans and mouse models.

Identifying Species-Specific Components of Transcriptional Responses Induced by Sepsis in Human and Mouse

We noted that while many gene sets in ImmuneSigDB were enriched in both species, there were also many gene sets enriched in one species but not the other (Figures 4C and S6). This suggested that in addition to similarities in the sepsis response, there might be species-specific differences in the transcriptional signatures of sepsis. In order to identify the biological basis for these species-specific differences, we devised an analytic approach, termed Leading Edge Metagene (LEM) analysis, to identify main biological "themes" in groups of ImmuneSigDB gene sets enriched in the sepsis datasets. We introduce LEM here and describe it in more detail elsewhere (see Experimental Procedures). LEM analysis is a novel method to identify the groups of co-regulated genes—which we term metagenes—that are highly enriched in multiple gene sets in a comparison of interest (such as sepsis versus control).

For LEM analysis, we first considered all gene sets that were significantly enriched in each dataset of sepsis versus control comparison group (FDR < 0.001). We then filtered the genes in these enriched gene sets to include only leading edge genes (Figure 5A, top and middle). These leading edge genes represented the subset of genes in the group of enriched gene-sets that drive the enrichment score with respect to upregulation in the sepsis phenotype. We then used non-negative matrix factorization (NMF) (Brunet et al., 2004; Lee and Seung, 1999; Lee, 2000; Tamayo et al., 2007) to identify groups of genes that are members of multiple gene sets (Figures S5A and S5B). NMF analysis therefore identifies groups of genes—which we term metagenes—that are members of the leading edge of multiple gene sets that are enriched in the transcriptional response to sepsis (Figure 5A, bottom).

LEM analysis of the gene sets enriched in human sepsis (316 gene sets) and mouse sepsis (974 gene sets) studied in Figure 4 identified three metagenes that were correlated with the sepsis response in each study. Individual metagenes were strongly overrepresented for genes related to distinctive biological processes as annotated by GO terms and Reactome (Ashburner et al., 2000; Croft et al., 2011) (Figure 5B). For instance, in the human sepsis response, we identified a metagene with an overrepresentation of genes involved in mitosis ($p = 4.9 \times 10^{-22}$) such as *CCNA2*, *BUB1*, and *KIF11*. A second metagene was enriched for genes related to phagocytosis ($p = 2.02 \times 10^{-13}$; *LAMP2*, *NCF4*, and *ATPV0B*) and a third metagene was enriched for genes related to inflammation ($p = 3.7 \times 10^{-4}$; *IL1A*, *NFKB1*, and *CCL20*) (Figure 5B). Overlap between the metagene gene memberships and specific GO terms revealed one predominant biological process in each (Figure 5B). However, the while each metagene was significantly enriched for one predominant GO term, only 5%–15% of genes contained in each metagene overlapped with genes in the predominant GO term (Figures S5C–S5F). This suggests that the genes contained in each LEM are related to recognizable biological processes, but that the metagenes represent discrete modules of genes that overlap with but are distinct from GO term categories.

We reasoned that metagenes would provide a more "refined" list of functionally related genes than their parental gene-sets. We therefore tested whether leading edge metagenes were more highly enriched for genes related to biological processes (again as annotated by overlap with GO terms) than their parental gene sets (Figure 5C). We tested the set of three leading edge metagenes for overlap with the collection of GO annotated gene lists, and determined the significance of each GO term's overlap. We compared the p values generated by GO term overlap with the set of genes comprising each metagene with an equivalent number of genes randomly sampled from the original pool of leading edge genes, or from all genes in the genome. We found that the significance of GO term overlap was much higher in the leading edge metagenes than in the original leading edge genes or in a random set of genes. LEM analysis therefore is an effective strategy to both identify major biological processes active in a phenotype of interest and simplify the list of 315 and 974 enriched gene sets in human and mouse, respectively, to a core set of 3 metagenes in each organism that correspond to major biological themes.

We next compared the similarity between metagenes identified in the sepsis response in humans with those in mouse sepsis models. We visualized the pairwise overlap in genes in each metagene using a Circos plot (Figure 5D) and determined the significance of the overlap for each pairwise comparison of mouse and human metagenes (Figures 5E and S6) (Krzywinski et al., 2009). We found striking cross-species similarities for some but not all metagenes. For example, a metagene annotated as "Phagocytic Vesicle" correlated with both the human and mouse sepsis response and contained a very similar set of genes (hypergeometric test $p = 1.09 \times 10^{-31}$, dark blue ribbon, Figure 5D). Similarly there was a highly significant overlap in the metagene annotated as "inflammatory response" in the human dataset and "TRIF-mediated TLR signaling" in the mouse model ($p = 2.79 \times 10^{-23}$).

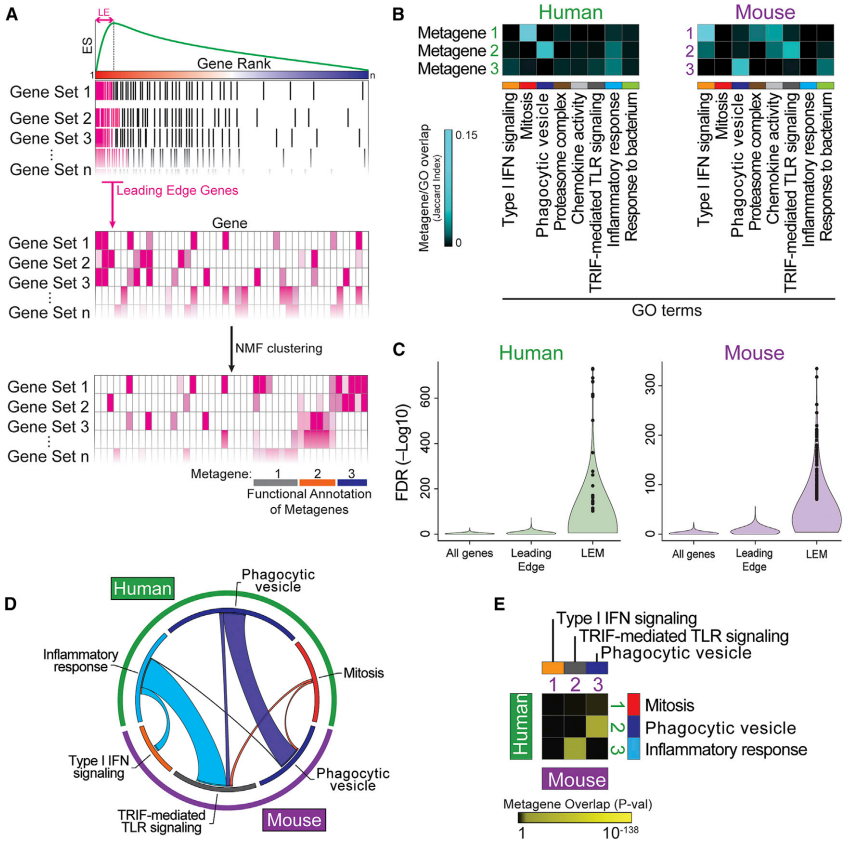

Figure 5. Leading Edge Clustering using Non-Negative Matrix Factorization Identifies Metagenes Representing Distinct Biological Processes

(A) A schematic of the process by which leading edge metagenes are identified.

(B) Biological annotation of metagenes identified in the studies analyzed in Figure 4 generated using GO terms.

(C) Violin plots showing p values of significance of GO Term overlaps with human (left) and mouse (right) sepsis metagenes (LEM), or equivalent-size samples of leading edge genes, or randomly selected genes.

(D) Circos plot of the relative size and overlap of metagenes in mouse (purple, outer segment) and human (green, outer segment) sepsis datasets. Relative number of genes in metagenes is indicated by segment length of the inner circle. Thickness of the ribbon corresponds to the relative number of genes shared between metagenes in the two species.

(E) Heatmap of p values corresponding to significance of overlap in pairwise comparison of metagene gene membership (yellow, highly significant; black, not significant). Statistical significance of the overlap was calculated by hypergeometric test. See also Figures S5 and S6.

However, we also identified metagenes that were not conserved between humans with sepsis and the mouse model. For example, a metagene enriched for genes pertaining to cell cycle ("mitosis" GO term) in humans did not share a corresponding metagene in the mouse model. In the mouse, a type 1 interferon signaling metagene overlapped with very few human metagenes. This analysis approach using ImmuneSigDB suggests that while some biological processes are strongly conserved between these two human and mouse datasets (e.g., phagocytosis, TLR mediated inflammatory response), other biological components are not (e.g., mitosis).

Global Shared and Species-Specific Biological Processes Can Be Identified using ImmuneSigDB and NMF Clustering

We next extended this approach to six datasets of sepsis versus control conditions from three independent studies in humans and from four comparisons in two mouse studies. We identified

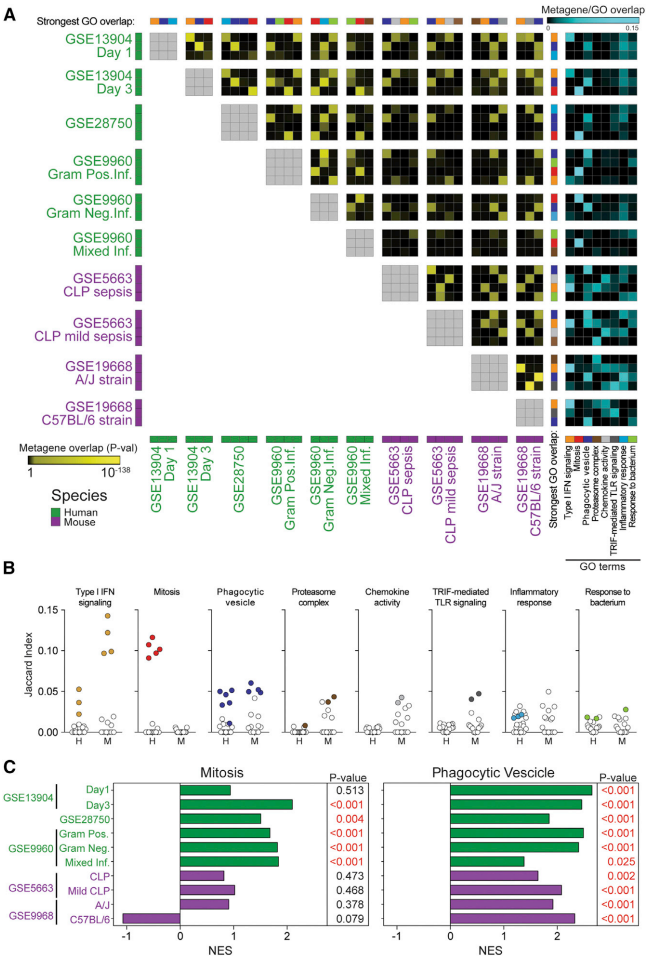

Figure 6. ImmuneSigDB Identifies Shared and Unique Biology in Mouse and Human Sepsis Studies

(A) Pairwise overlaps of all metagenes from mouse (purple bars) and human (green bars) sepsis studies. Heatmap indicates p values corresponding to significance of overlap between each metagene (small squares) in each study (larger squares; yellow, highly significant; black, not significant). The biological annotation of each metagene is based on the significance of enrichment of the GO term indicated (blue, large overlap; black, no overlap) (right). The most significantly enriched GO term annotating each metagene is indicated by the key in lower right.

(B) Jaccard index representing the extent of overlap of metagenes from human (H) and mouse (M) studies. Colored are metagenes that are annotated with the respective biological process as in (A).

(C) Enrichment scores of biological processes that are species-specific (e.g., mitosis, left) or shared (e.g., phagocytic vesicle, right) in the human (green bars) and mouse (purple bars) sepsis datasets. Significance of the enrichment of the named biological process in each dataset is indicated by the p values on the right.

between three or four metagenes in each study providing a total of 35 metagenes present in the collected group of sepsis studies (Figure 6A). We annotated each metagene based on enrichment of GO Terms (Figure 6A, right) and evaluated the significance of pair-wise overlap in the genes present in each metagene.

We found that in almost every study, there was at least one metagene induced by sepsis that showed a highly significant overlap (indicated in yellow in the heatmap, Figure 6A) with metagenes from every other sepsis study, regardless of species of origin. One study that proved an exception was the human study GSE9960, which studied a response to mixed infection, and showed relatively little overlap with any mouse study. However the metagenes identified in that study also showed limited overlap with metagenes from other human studies, suggesting that that transcriptional response contained in that study may represent a different type of biological response to the other human and mouse studies.

In addition to these strongly conserved metagenes, we also found that there were metagenes induced by sepsis that had a striking species-specific distribution. For example, the phagocytic vesicle metagene was either present or strongly overlapped with a metagene present in every dataset, both mouse and human. In contrast, the mitosis metagene was much more specific to human datasets with no significant overlap in mouse studies (Figure 6B). To confirm these results, we tested the significance of enrichment of two GO terms—mitosis and phagocytic vesicle. When looking at the whole-genome transcriptional changes, we indeed observed that mitosis was represented exclusively in human cells in sepsis while phagocytic vesicle process was represented in both species, as we predicted based on the LEM analysis (Figure 6C). These data reveal context-specific transcriptional modules induced by sepsis in humans and mice and also highlight the distinct transcriptional components of the response present in one species but not the other.

DISCUSSION

We analyzed expression profiles from 389 published studies of mouse and human immune cells to generate a collection of curated gene signatures corresponding to cell states and

perturbations in the immune system. This collection of almost 5,000 genes sets contains substantial biological information that was not currently contained in existing collections. We used this new compendium to show that transcriptional signatures induced by LPS stimulation in dendritic cells, and transcriptional programs of T cell and B cell differentiation were highly conserved between humans and mouse models. Moreover, we used ImmuneSigDB to analyze expression profiles from patients and mouse models of sepsis and showed highly significant overlap, suggesting that components of the transcriptional response to sepsis were highly conserved between species. However, we also find that there are substantial species-specific differences, both in enriched gene sets and their component metagenes, in sepsis response signatures, suggesting that not all biological processes induced by sepsis evident at the transcriptional level in humans are present in mouse models and vice versa. These findings suggest that ImmuneSigDB provides a useful tool for detecting subtle patterns of similarity and difference in large-scale datasets of gene expression from cells and tissues in the immune system.

Several studies have directly compared the transcriptional programs in the human and mouse immune systems. We, and others, previously identified conserved patterns of gene expression that change during the differentiation of memory T and B cells, and in exhausted to $CD8^+$ T cells (Baitsch et al., 2011; Haining et al., 2008; Quigley et al., 2010). A recent comparison of gene expression in seven immune cell groups from humans and mice also found a highly significant degree of similarity in global patterns of expression and in the putative transcriptional regulators of these genes (Shay et al., 2013). However, in that study, although the majority of genes showed a pattern of expression that was highly correlated between species, 30%–50% of genes did not show significant correlation between species. Two recent studies of the sepsis datasets analyzed in the present study reached opposite conclusions regarding the degree of similarity between the mouse and human response to sepsis (Seok et al., 2013; Takao and Miyakawa, 2014). Thus, the degree of conservation of transcriptional signatures in the mouse and human immune systems remains controversial (Davis, 2008; Gilad and Mizrahi-Man, 2015; Hackam and Redelmeier, 2006; Lin et al., 2014; Rice, 2012; Shay et al., 2014; van der Worp et al., 2010; Warren et al., 2014).

Our analysis using ImmuneSigDB suggests that there are both conserved and species-specific transcriptional programs induced by sepsis in the immune system. Overall, the transcriptional program shows highly significant similarity between sepsis in the human and mouse. Specifically, analysis of the leading edge metagenes across human (six comparisons) and mouse sepsis datasets (four comparisons) found that in many of the datasets from both species there was coordinate upregulation of metagenes involved in interferon-response and phagocytic processes. This suggests that features of the sepsis response such as interferon release and neutrophilia are shared between species.

However, we also show that many gene sets are enriched in only one species, and metagenes related to mitosis were highly enriched in sepsis-induced profiles in humans but were not significantly enriched in the mouse model. Thus, it is likely that although some components of the sepsis response are highly conserved between species, there is also substantial divergence in the biological processes detected by transcriptional profiling each. Detailed analysis of the transcriptional features of the mouse and human immune systems is therefore required to substantiate conclusions regarding the conservation of a particular biology of interest in two datasets. Whether the differences we observe are due to inherent biological differences between the two species remains unclear. For example, it is possible that the mitotic signature is present in human, but not mouse, because the exact timing of the initiation of activation of immune cells in humans with sepsis is not precisely known and might be more variable compared to tightly controlled, narrow window of induction of sepsis in mouse models.

Our compendium adds to a growing list of collections of transcriptionally co-regulated genes in the immune system. In the human immune system, several studies have identified groups of co-regulated gene modules from expression profiles derived from blood samples representing a range of states of health and disease (Chaussabel et al., 2008; Li et al., 2014). This modular approach to the analysis of gene expression can aid interpretation of gene-expression profiles, increase robustness, and facilitate analyses that span multiple datasets. However, ImmuneSigDB is distinct from those previously described in several respects. Studies by Chaussabel and by Li have focused on identifying collections of genes—termed modules—that tend to vary in expression in a coordinate fashion across a reference set of expression profiles (Chaussabel et al., 2008; Li et al., 2014). Defining modules based on network reconstruction across hundreds or thousands of experimental conditions makes it difficult to associate a particular module with a defined cell state or perturbation that usually results in its up- or downregulation. In contrast, the annotations describing each gene set in the ImmuneSigDB include all the experimental details from a published manuscript, allowing a more transparent connection between gene set and biology. Moreover, ImmuneSigDB was designed for use with GSEA, because each gene set contains either up- or downregulated genes only, rather than a combination of both as can appear in Chaussabel or Li modules, which might limit the use of the latter collections in analyses such as GSEA (Figure S2D, Table S3). Finally, each collection of previously-published modules was defined in a single species (humans), making the generalizability of these compendia to other species hard to predict.

The ImmuneSigDB collection differs in another important respect from previous module collections. The studies by Chaussabel et al. and Li et al. were designed to identify non-overlapping modules of gene expression. However, ImmuneSigDB contains gene sets derived from experimental perturbations that are likely to induce multiple biological processes, each of which might be represented by sub-signatures in a given gene set. Moreover, several gene sets might contain the transcriptional correlate of the same biological processes. For some analytic purposes, it might be useful to have a single gene set that includes the multiple biological processes that are initiated by the complex stimulus such as receptor-ligand engagement or cell differentiation. However, for other applications, such as the analysis which we conducted of the sepsis datasets, a more "atomic" approach might be preferred. We have therefore developed an analytic approach to extract

non-redundant leading edge metagenes from the experimentally derived gene-expression profiles.

Analysis with ImmuneSigDB using GSEA or GSEA combined with a leading-edge metagene analysis might therefore provide the systems immunologist with a useful resource for the analysis of gene expression in the immune system.

EXPERIMENTAL PROCEDURES

ImmuneSigDB Generation

We surveyed the immunology literature and identified published studies that included human or mouse microarray Affymetrix gene expression data in NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). We downloaded the corresponding datasets from GEO (Barrett et al., 2007). When available, raw microarray data in the form of the CEL files were normalized by the Robust Multichip Average (RMA) (Bolstad et al., 2003) using justRMA function from the R Bioconductor package affy (version 1.40.0) (Gautier et al., 2004). When CEL files were absent, we downloaded processed expression data from GEO by means of R GEO query package (version 2.28.0) (Davis and Meltzer, 2007). We mapped Affymetrix probe set identifiers to human gene symbols using the Collapse Dataset tool (max probe algorithm) of the GSEA program (Subramanian et al., 2007). We used ortholog gene assignments from Mouse Genome Informatics. The specific mappings were retrieved from the MGI web site on 14 April 2012 and contained 17,827 human-mouse ortholog gene pairings. Phenotype classes were assigned manually according to the original sample annotations and based on review of meaningful biological comparisons (Figure S1). We implemented a pipeline in R, which combined processed microarray data with the phenotype annotations and produced standard formatted files (.gct and .cls) for each comparison as needed.

For each two-class comparison, the genes were ranked according to an information-based similarity metric (RNMI) (Abazeed et al., 2013) from top upregulated to bottom downregulated genes in the two groups. Gene sets comprised genes differentially expressed with an FDR < 0.02, and a maximum number of genes was set at 200 (i.e., all gene sets had at most 200 differentially expressed genes). This way we generated two gene sets from each assigned biological comparison of two groups—"Group_A_vs_Group_B_UP" and "Group_A_vs_Group_B_DN," for the top upregulated and bottom downregulated genes, respectively, identified for the genes most different in the samples in group A compared to the samples in group B. The resource is accessible as the C7 collection at http://www.msigdb.org.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) was performed as described previously (Mootha et al., 2003; Subramanian et al., 2005). To analyze transcriptional data from Immunological Genome Project (ImmGen) (Heng et al., 2008) and Differentiation Map (DMAP) (Novershtern et al., 2011), we used single sample GSEA (ssGSEA) as described previously (Barbie et al., 2009; Reich et al., 2006), to create a matrix in which columns represented individual samples and rows corresponded to gene sets, and the values represented the single sample ssGSEA score of each gene set in each sample. We averaged the biological replicates and filtered this matrix to include only the top 10% of gene sets based on mean absolute deviation (MAD) across sample types and bi-clustered using 1-Pearson Correlation.

Leading Edge Metagene Analysis

We developed an approach to identify groups of genes—termed leading edge metagenes (LEM)—that are both associated with a phenotype of interest and shared between multiple gene sets enriched in that phenotypic comparison (Y.T., unpublished data). We reasoned that groups of genes that are co-regulated in the phenotype of interest and also present in multiple gene sets are likely to represent the core sub-signatures of genes related to distinct biological processes or pathways. Our approach leverages the notion of the leading edge genes in a GSEA analysis, which are the genes whose expression profile is most highly correlated with the phenotype distinction in a comparison of biological states and thus drives the GSEA enrichment statistic. LEM analysis identifies groups of genes (metagenes) that are common to multiple gene sets returned in a GSEA result, and strongly correlated with the phenotype of interest.

First we perform GSEA using the ImmuneSigDB in a two-class comparison of interest (e.g., sepsis versus control). GSEA yields an enrichment score to quantify the overrepresentation of a gene set (e.g., genes coordinately up- or downregulated in previous experiments) at the top or bottom of a ranked list of genes (e.g., generated by differential expression of in a comparison of interest). The leading edge of each enriched gene set is defined as the subset of genes with positive contribution to the enrichment score before it reaches its peak; i.e., those that are most correlated with the phenotype of interest.

We then consolidate the leading-edges of the m top-scoring gene sets into a sparse n by m matrix M, where the number of rows is the cardinality of the union of genes from all the leading-edges in the m top gene sets, and the columns correspond to the genes in the m enriched gene sets. The value of each entry in the matrix is the signal to noise ratio of the corresponding gene between two conditions in comparison (Equation 1) and 0 if the gene is not in the leading edge of that gene set. A large signal to noise ratio indicates a significant difference in expressions of the corresponding genes between the two conditions.

$$s2n = \frac{\mu_A - \mu_B}{\sigma_A - \sigma_B} \quad (1)$$

To identify clusters in this matrix, we use non-matrix factorization (Brunet et al., 2004; Lee and Seung, 1999; Lee and Seung, 2000; Tamayo et al., 2007) to yield two matrices, W and H. W matrix is a low-dimensional representation of the M matrix and each dimension of W is a linear combination of n genes, called a metagene. The entries in the H matrix represent the quantity of each metagene required to reconstruct each of the M gene sets. The coefficient in W matrix can be viewed as the contribution of each gene to the corresponding metagene. Inspection of the W matrix shows that in each metagene, the coefficients of most genes are usually very small, and only a small number of genes have a coefficient significantly larger than 0. As each metagene is a positive linear combination of all the genes, a small coefficient indicates negligible contribution to the metagene. Thus the next step of our algorithm is to filter out genes with small coefficients in each metagene. To do that, we first assume that the background distribution of coefficients fulfills an exponential distribution. We set a filtering threshold at the 95% quantile of the fitted exponential distribution and set all coefficients below this to zero.

Because each gene can contribute to more than one metagene we next need to assign each gene to a single metagene. The assignment of genes to metagenes uses the following rules: (1) if one gene has no contribution to any of the metagenes, then it will be defined as not in any metagene, and (2) each gene with a coefficient above the threshold (defined above) will be assigned to the metagene where it has the largest coefficient. Each metagene is annotated with a biological "theme" based on the Jaccard overlap of its constituent genes with GO categories (Ashburner et al., 2000).

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.immuni.2015.12.006.

AUTHOR CONTRIBUTIONS

Conceptualization, J.G., J.P.M., and W.N.H.; Methodology, J.G., Y.T., A.L., P.T., J.P.M., and W.N.H.; Software, Y.T. and A.L.; Formal analysis, J.G., A.L., and P.T.; Investigation, J.G., A.L., and P.T.; Data curation, J.G. and A.L.; Writing of original draft, J.G., J.P.M., and W.N.H.; Review and editing, J.G., A.J.B., S.B., J.P.M., and W.N.H.; Visualization, J.G. and W.N.H.; Funding acquisition, J.P.M., A.J.B., and W.N.H.; Project administration, W.N.H.; Supervision, W.N.H. and J.P.M.

ACKNOWLEDGMENTS

We would like to thank Arlene H. Sharpe, William Kim, and Aravind Subramanian for useful discussions. This work was supported by an Infrastructure and Opportunity Grant from the Human Immunology Project Consortium

(to A.J.B., J.P.M., and W.N.H.), by U19 AI090023 to W.N.H.; NIAID Bioinformatics Support Contract HHSN272201200028C to A.J.B.; R01CA154480, R01CA121941, R01GM074024, and U54CA112962 to P.T. and J.P.M.; and the Cancer Research Institute Predoctoral Emphasis Pathway in Tumor Immunology to J.G.

Received: April 13, 2015
Revised: August 2, 2015
Accepted: September 30, 2015
Published: January 12, 2016

REFERENCES

Abazeed, M.E., Adams, D.J., Hurov, K.E., Tamayo, P., Creighton, C.J., Sonkin, D., Giacomelli, A.O., Du, C., Fries, D.F., Wong, K.K., et al. (2013). Integrative radiogenomic profiling of squamous cell lung cancer. Cancer Res. *73*, 6289–6298.

Ahn, S.H., Deshmukh, H., Johnson, N., Cowell, L.G., Rude, T.H., Scott, W.K., Nelson, C.L., Zaas, A.K., Marchuk, D.A., Keum, S., et al. (2010). Two genes on A/J chromosome 18 are associated with susceptibility to Staphylococcus aureus infection by combined microarray and QTL analyses. PLoS Pathog. *6*, e1001088.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

Baitsch, L., Baumgaertner, P., Devêvre, E., Raghav, S.K., Legat, A., Barba, L., Wieckowski, S., Bouzourene, H., Deplancke, B., Romero, P., et al. (2011). Exhaustion of tumor-specific $CD8^+$ T cells in metastases from melanoma patients. J. Clin. Invest. *121*, 2350–2360.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles–database and tools update. Nucleic Acids Res. *35*, D760–D765.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Brunet, J.P., Tamayo, P., Golub, T.R., and Mesirov, J.P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. USA *101*, 4164–4169.

Chaussabel, D., Quinn, C., Shen, J., Patel, P., Glaser, C., Baldwin, N., Stichweh, D., Blankenship, D., Li, L., Munagala, I., et al. (2008). A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity *29*, 150–164.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. *39*, D691–D697.

Davis, M.M. (2008). A prescription for human immunology. Immunity *29*, 835–838.

Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics *23*, 1846–1847.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307–315.

Gilad, Y., and Mizrahi-Man, O. (2015). A reanalysis of mouse ENCODE comparative gene expression data. F1000Research *4*, http://dx.doi.org/10.12688/f1000research.6536.1.

Hackam, D.G., and Redelmeier, D.A. (2006). Translation of research evidence from animals to humans. JAMA *296*, 1731–1732.

Haining, W.N., and Wherry, E.J. (2010). Integrating genomic signatures for immunologic discovery. Immunity *32*, 152–161.

Haining, W.N., Ebert, B.L., Subrmanian, A., Wherry, E.J., Eichbaum, Q., Evans, J.W., Mak, R., Rivoli, S., Pretz, J., Angelosanto, J., et al. (2008). Identification of an evolutionarily conserved transcriptional signature of CD8 memory differentiation that is shared by T and B cells. J. Immunol. *181*, 1859–1868.

Heng, T.S., and Painter, M.W.; Immunological Genome Project Consortium (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. *9*, 1091–1094.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

Lee, D.D., and Seung, H.S. (2000). Algorithms for Non-negative Matrix Factorization. In Proceedings of Neural Information Processing Systems *13*, 556–562.

Lee, D.D., and Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature *401*, 788–791.

Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D.S., Johnson, S.E., Milton, A., Rajam, G., et al. (2014). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. Nat. Immunol. *15*, 195–204.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Lin, S., Lin, Y., Nery, J.R., Urich, M.A., Breschi, A., Davis, C.A., Dobin, A., Zaleski, C., Beer, M.A., Chapman, W.C., et al. (2014). Comparison of the transcriptional landscapes between human and mouse tissues. Proc. Natl. Acad. Sci. USA *111*, 17224–17229.

Luckey, C.J., Bhattacharya, D., Goldrath, A.W., Weissman, I.L., Benoist, C., and Mathis, D. (2006). Memory T and memory B cells share a transcriptional program of self-renewal with long-term hematopoietic stem cells. Proc. Natl. Acad. Sci. USA *103*, 3304–3309.

Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309.

Quigley, M., Pereyra, F., Nilsson, B., Porichis, F., Fonseca, C., Eichbaum, Q., Julg, B., Jesneck, J.L., Brosnahan, K., Imam, S., et al. (2010). Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF. Nat. Med. *16*, 1147–1151.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J.P. (2006). GenePattern 2.0. Nat. Genet. *38*, 500–501.

Rice, J. (2012). Animal models: Not close enough. Nature *484*, S9.

Seok, J., Warren, H.S., Cuenca, A.G., Mindrinos, M.N., Baker, H.V., Xu, W., Richards, D.R., McDonald-Smith, G.P., Gao, H., Hennessy, L., et al.; Inflammation and Host Response to Injury, Large Scale Collaborative Research Program (2013). Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc. Natl. Acad. Sci. USA *110*, 3507–3512.

Shay, T., Jojic, V., Zuk, O., Rothamel, K., Puyraimond-Zemmour, D., Feng, T., Wakamatsu, E., Benoist, C., Koller, D., and Regev, A.; ImmGen Consortium (2013). Conservation and divergence in the transcriptional programs of the human and mouse immune systems. Proc. Natl. Acad. Sci. USA *110*, 2946–2951.

Shay, T., Lederer, J.A., and Benoist, C. (2014). Genomic responses to inflammation in mouse models mimic humans: We concur, apples to oranges comparisons won't do. Proc. Natl. Acad. Sci. USA *112*, E346.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J.P. (2007). GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics *23*, 3251–3253.

Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. (2005). An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. Nat. Genet. *37*, 48–55.

Takao, K., and Miyakawa, T. (2014). Genomic responses in mouse models greatly mimic human inflammatory diseases. Proc. Natl. Acad. Sci. USA *112*, 1167–1172.

Tamayo, P., Scanfeld, D., Ebert, B.L., Gillette, M.A., Roberts, C.W., and Mesirov, J.P. (2007). Metagene projection for cross-platform, cross-species characterization of global transcriptional states. Proc. Natl. Acad. Sci. USA *104*, 5959–5964.

Tang, B.M., McLean, A.S., Dawes, I.W., Huang, S.J., and Lin, R.C. (2009). Gene-expression profiling of peripheral blood mononuclear cells in sepsis. Crit. Care Med. *37*, 882–888.

van der Worp, H.B., Howells, D.W., Sena, E.S., Porritt, M.J., Rewell, S., O'Collins, V., and Macleod, M.R. (2010). Can animal models of disease reliably inform human studies? PLoS Med. *7*, e1000245.

Warren, H.S., Tompkins, R.G., Moldawer, L.L., Seok, J., Xu, W., Mindrinos, M.N., Maier, R.V., Xiao, W., and Davis, R.W. (2014). Mice are not men. Proc. Natl. Acad. Sci. USA *112*, E345.

Appendix B: Gupta, Godec, and Wolski et al. PLOS Pathogens 2015

This is a reprint of:

Prakash K. Gupta*, Jernej Godec*, David Wolski*, Emily Adland, Kathleen Yates, Cormac Cosgrove, Carola Ledderose, Wolfgang G. Junger, Simon C. Robson, E. John Wherry, Galit Alter, Philip J. R. Goulder, Paul Klenerman, Arlene H. Sharpe, Georg M. Lauer, W. Nicholas Haining. CD39 expression identifies terminally exhausted $CD8^+$ T cells. *PLOS Pathogens*. 2015 Oct 20;11(10):e1005177. (PMID: 26485519). *Co-first author.

PLOS | PATHOGENS

RESEARCH ARTICLE

CD39 Expression Identifies Terminally Exhausted CD8$^+$ T Cells

Prakash K. Gupta[1,2☯], Jernej Godec[1,3☯], David Wolski[4☯], Emily Adland[2], Kathleen Yates[1], Kristen E. Pauken[5], Cormac Cosgrove[6], Carola Ledderose[7], Wolfgang G. Junger[7], Simon C. Robson[8], E. John Wherry[5], Galit Alter[6], Philip J. R. Goulder[2], Paul Klenerman[2], Arlene H. Sharpe[3,9], Georg M. Lauer[4], W. Nicholas Haining[1,9,10]*

1 Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, United States of America, **2** Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, United Kingdom, **3** Department of Microbiology and Immunobiology and Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **4** Gastrointestinal Unit, Massachusetts General Hospital and Harvard Medical School, Massachusetts, United States of America, **5** Department of Microbiology and Institute for Immunology, University of Pennsylvania Perelman School Medicine, Philadelphia, Pennsylvania, United States of America, **6** Ragon Institute of Massachusetts General Hospital, Harvard University and Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **7** Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, United States of America, **8** Division of Gastroenterology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard University, Boston, Massachusetts, United States of America, **9** Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **10** Division of Hematology/Oncology, Children's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

☯ These authors contributed equally to this work.
* Nicholas_Haining@dfci.harvard.edu

OPEN ACCESS

Citation: Gupta PK, Godec J, Wolski D, Adland E, Yates K, Pauken KE, et al. (2015) CD39 Expression Identifies Terminally Exhausted CD8$^+$ T Cells. PLoS Pathog 11(10): e1005177. doi:10.1371/journal.ppat.1005177

Editor: Daniel C. Douek, Vaccine Research Center, UNITED STATES

Received: June 15, 2015

Accepted: August 30, 2015

Published: October 20, 2015

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Raw microarray files are publicly available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE72752.

Funding: The author(s) received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Exhausted T cells express multiple co-inhibitory molecules that impair their function and limit immunity to chronic viral infection. Defining novel markers of exhaustion is important both for identifying and potentially reversing T cell exhaustion. Herein, we show that the ectonucleotidse CD39 is a marker of exhausted CD8$^+$ T cells. CD8$^+$ T cells specific for HCV or HIV express high levels of CD39, but those specific for EBV and CMV do not. CD39 expressed by CD8$^+$ T cells in chronic infection is enzymatically active, co-expressed with PD-1, marks cells with a transcriptional signature of T cell exhaustion and correlates with viral load in HIV and HCV. In the mouse model of chronic Lymphocytic Choriomeningitis Virus infection, virus-specific CD8$^+$ T cells contain a population of CD39^{high} CD8$^+$ T cells that is absent in functional memory cells elicited by acute infection. This CD39^{high} CD8$^+$ T cell population is enriched for cells with the phenotypic and functional profile of terminal exhaustion. These findings provide a new marker of T cell exhaustion, and implicate the purinergic pathway in the regulation of T cell exhaustion.

Author Summary

Chronic viral infection induces an acquired state of T cell dysfunction known as exhaustion. Discovering surface markers of exhausted T cells is important for both to identify

exhausted T cells as well as to develop potential therapies. We report that the ectonucleotidase CD39 is expressed by T cells specific for chronic viral infections in humans and a mouse model, but is rare in T cells following clearance of acute infections. In the mouse model of chronic viral infection, CD39 demarcates a subpopulation of dysfunctional, exhausted CD8$^+$ T cells with the phenotype of irreversible exhaustion. CD39 expression therefore identifies terminal CD8$^+$ T cell exhaustion in mice and humans, and implicates the purinergic pathway in the regulation of exhaustion.

Introduction

In acute infections, antigen-specific T cells differentiate into activated effector cells and then into memory T cells which rapidly gain effector functions and re-expand on subsequent encounter with the same pathogen [1]. In contrast, during chronic infections, pathogen-specific T cells gradually lose effector functions, fail to expand, and can eventually become physically deleted [2]. These traits are collectively termed T cell exhaustion, and have been described both in animal models of chronic viral infection as well as in human infections with hepatitis C virus (HCV) and human immunodeficiency virus (HIV) [2–4]. Identifying reversible mechanisms of T cell exhaustion is therefore a major goal in medicine.

Prolonged or high-level expression of multiple inhibitory receptors such as PD-1, Lag3, and CD244 (2B4) is a cardinal feature of exhausted T cells in both animal models and human disease [5–7]. Expression of PD-1 appears to be a particularly important feature of exhausted CD8$^+$ T cells, as the majority of exhausted cells in mouse models of chronic infection express this receptor, and blockade of the PD-1:PD-L1 axis can restore the function of exhausted CD8$^+$ T cells in humans and mouse models [2,6]. However, in humans, many inhibitory receptors also can be expressed by a large fraction of fully functional memory CD8$^+$ T cells. PD-1, for instance, can be expressed by up to 60% of memory CD8$^+$ T cells in healthy individuals, making it challenging to use PD-1 to identify exhausted CD8$^+$ T cells in humans, particularly when the antigen-specificity of potentially exhausted CD8$^+$ T cells is not known [8].

Studies in mice and humans suggest that exhausted CD8$^+$ T cells are not a homogeneous population, but instead include at least two subpopulations of T cells that differentially express the transcription factors T-bet and Eomesodermin (Eomes) [9–11]. T-bet^{high} CD8$^+$ T cells represent a progenitor subset with proliferative potential that give rise to Eomes^{high} CD8$^+$ T cells, which are terminally differentiated and can no longer proliferate in response to antigen or be rescued by PD-1 blockade [9,12]. Both populations express PD-1, but Eomes^{high} exhausted cells express the highest levels of PD-1. However, no specific cell-surface markers of this terminally differentiated population of exhausted cells have thus far been identified.

CD39 (*ENTPD1*) is an ectonucleotidase originally identified as an activation marker on human lymphocytes and as the vascular ecto-ADPase [13], but has subsequently been shown to be a hallmark feature of regulatory T cells [14–16]. CD39 hydrolyzes extracellular ATP and ADP into adenosine monophosphate, which is then processed into adenosine by CD73, an ecto-5'-nucleotidase [17]. Adenosine is a potent immunoregulator that binds to A2A receptors expressed by lymphocytes causing accumulation of intracellular cAMP, preventing T cell activation and NK cytotoxicity [18–20]. Loss of CD39 in Tregs markedly impairs their ability to suppress T cell activation, suggesting that the juxtacrine activity of CD39 serves to negatively regulate T cell function [15]. However, blood CD8$^+$ T cells have generally been reported to be CD39$^-$ [14,21–23], and the expression of this marker on exhausted T cells has not been examined.

In this study, we demonstrate that, in contrast to CD8$^+$ T cells from healthy donors, antigen-specific CD8$^+$ T cells responding to chronic viral infection in humans and a mouse model express high levels of biochemically active CD39. CD39$^+$ CD8$^+$ T cells co-express PD-1 and are enriched for a gene signature of T cell exhaustion. In the mouse model of chronic LCMV infection, high levels of CD39 expression demarcate terminally differentiated virus-specific CD8$^+$ T cells within the pool of exhausted CD8$^+$ T cells. Thus, CD39 provides a specific, pathological marker of exhausted CD8$^+$ T cells in chronic viral infection in humans and mouse models of chronic viral infection.

Results

CD39 is expressed by CD8$^+$ T cells responding to chronic infection

We surveyed the expression of CD39 by CD8$^+$ T cells from healthy adult subjects without chronic viral infection. Consistent with previous reports we found that only a small fraction (mean 6%) of CD8$^+$ T cells in healthy individuals expressed CD39 (Fig 1A and 1B) [14,21–23]. This small population of CD39$^+$ CD8$^+$ T cells in healthy donors was primarily found in the central and effector memory compartments while virtually no naive CD8$^+$ T cells expressed CD39 (S1 Fig). We next focused on CD39 expression by antigen-specific CD8$^+$ T cells specific for latent viruses in healthy subjects and found that only a very small fraction of CMV- or EBV-specific CD8$^+$ T cells expressed CD39 (Fig 1A and 1B) (mean 3% and 7% respectively).

We next measured CD39 expression by T cells specific for the chronic viral pathogens HCV and HIV. We measured CD39 expression in 57 subjects with acute HCV infections (23 with acute resolving infection and 34 with chronically evolving infection), and in 40 subjects with HIV infection (28 chronic progressors and 12 controllers; clinical characteristics of the subjects are summarized in S1 Table). We found a mean of 51% of HCV-specific CD8$^+$ T cells and 31% of HIV-specific CD8$^+$ T cells expressed CD39, a number significantly higher than CD8$^+$ T cells specific for EBV or CMV, or in total CD8$^+$ T cell populations from healthy individuals (Fig 1A and 1B). A slightly greater fraction of virus-specific CD8$^+$ T cells from HCV-infected subjects expressed CD39 than did those from HIV-infected subjects.

In subjects with chronic infection, the frequency of CD39-expressing cells in the virus-specific (tetramer$^+$) CD8$^+$ T cell population was significantly higher than in the total CD8$^+$ T cell population (Fig 1C and 1D). However the fraction of total CD8$^+$ T cells expressing CD39 in the CD8$^+$ T cell compartment of individuals with HCV or HIV infection was slightly increased compared to healthy controls (Fig 1E), consistent with the presence of other, unmeasured virus-specific CD8$^+$ T cells that were also CD39$^+$ in the tetramer$^-$fraction of CD8$^+$ T cells. Thus CD39 is expressed infrequently by CD8$^+$ T cells in healthy donors, but marks a large fraction of pathogen-specific cells CD8$^+$ T cells in patients with chronic infection.

CD39 expressed by CD8$^+$ T cells hydrolyzes ATP

CD39 expressed by regulatory T cells catalyzes the hydrolysis of ADP to 5'-AMP [14–16] but its enzymatic activity can be regulated by a range of post-transcriptional mechanisms [PMID. We therefore tested CD39 expressed by CD8$^+$ T cells from patients infected with chronic HCV was functional using ATP hydrolysis as a surrogate marker of CD39 activity [24–26]. We sorted CD39$^-$ and CD39$^+$ CD8$^+$ T cells from six HCV-infected individuals (four with chronic infection and two with resolved infection) and incubated equal numbers of cells in the presence of extracellular ATP (eATP). Remaining levels of eATP were measured in the supernatant by HPLC. As a control, we assessed ATP hydrolysis by CD4$^+$ CD25$^+$ CD39$^+$ regulatory T cells (Tregs) sorted from the same individuals (Fig 2A).

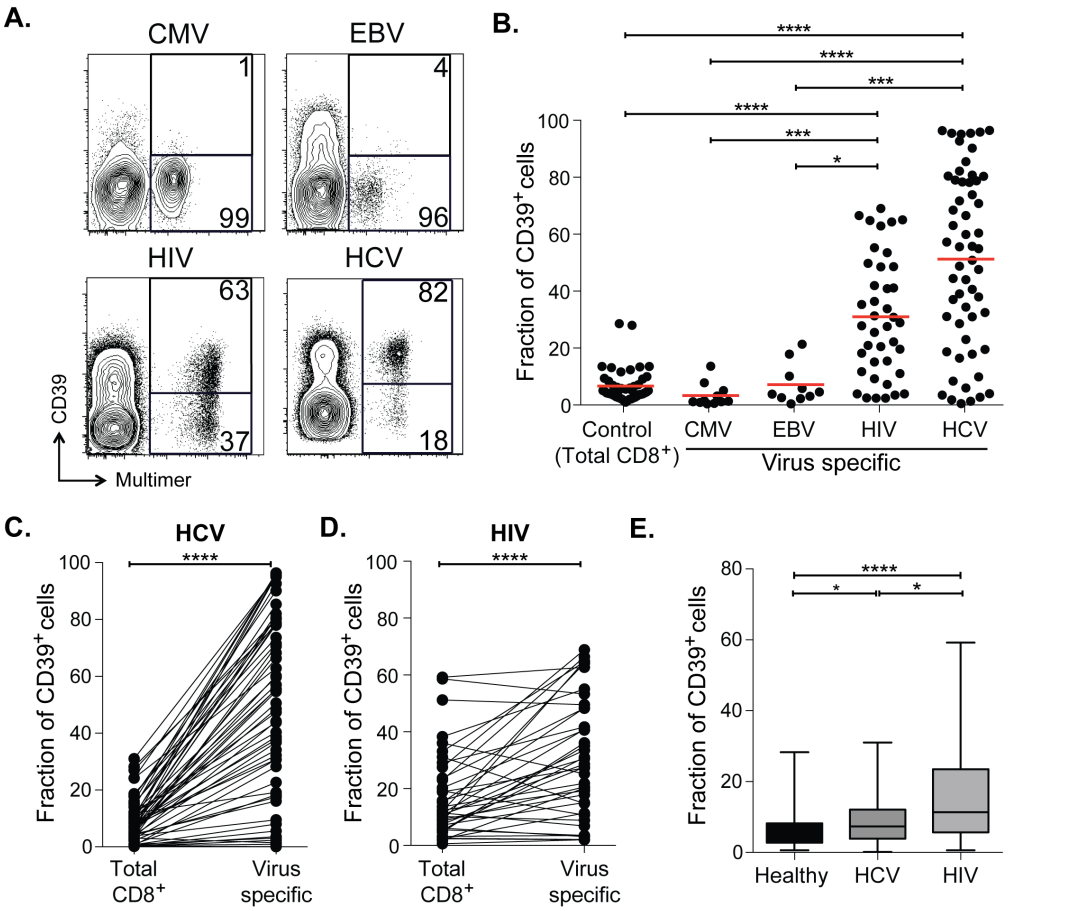

Fig 1. CD39 is highly expressed by virus-specific CD8+ T cells in chronic viral infection. (A) Expression of CD39 by virus-specific CD8+ T cells. Plots are gated on CD8+. **(B)** Fraction of total or antigen-specific CD8+ T cells expressing CD39. **(C, D)** Comparison of CD39 expression by total CD8+ T cells with virus-specific CD8+ T cells from patients with HCV (C) and HIV (D) infections. **(E)** Fraction of total CD8+ T cells expressing CD39 in healthy, HIV or HCV infected donors. Error bars represent SEM. Statistical significance was assessed by Kruskal-Wallis test (B, E), or Wilcoxon test (C, D). **P* <0.05, ****P* <0.001, *****P* <0.0001.

doi:10.1371/journal.ppat.1005177.g001

Within the CD39+ CD8+ T cell population the level of CD39 expression was lower than in Tregs (Fig 2B). Consistent with reduced CD39 expression relative to Tregs, ATP hydrolysis by CD39+ CD8+ T cells was less than that by Tregs (Fig 2C). However ATP hydrolysis by CD39+ CD8+ T cells was significantly greater than that of CD39− cells (Fig 2C). Thus CD39 expressed by CD8+ T cells in HCV infection is enzymatically active and capable of hydrolyzing ATP.

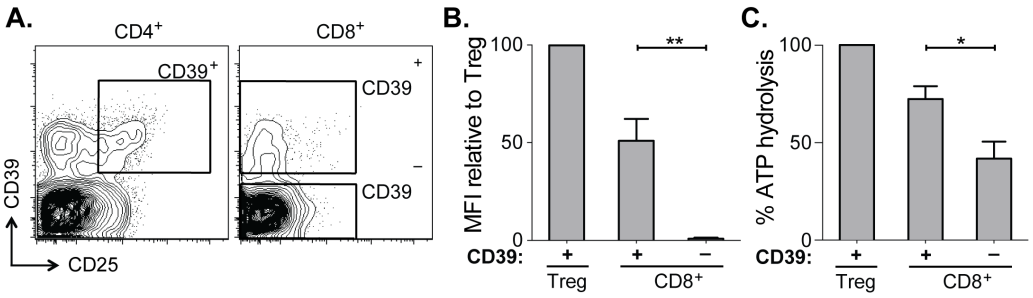

Fig 2. CD39 expressed by CD8+ T cells in HCV infection is enzymatically active. (A) Flow cytometry sorting gates of CD39+ and CD39− CD8+ T cells and CD39+ CD25+ CD4+ Tregs used for rpHPLC analysis of CD39 activity. **(B)** Summary of CD39 expression level relative to Tregs in the same subjects. **(C)** ATP hydrolysis by CD8+ T cell populations relative to Tregs. Data represent 6 patients with chronic HCV infection. Error bars represent SEM. Statistical significance was assessed by paired Student's t-test (B, C). *P <0.05, **P <0.01.

doi:10.1371/journal.ppat.1005177.g002

CD39 is co-expressed with PD-1 on virus-specific CD8+ T cells and correlates with viral load in both HCV and HIV infection

CD8+ T cells specific for chronic viruses such as HCV and HIV express increased levels of PD-1 [3,27]. We therefore examined the relationship between CD39 and PD-1 expression by virus-specific CD8+ T cells in 54 patients with HCV (23 chronically infected and 31 resolvers) and 40 patients infected with HIV (28 chronic progressors, 7 viremic controllers and 5 elite controllers). In both diseases we found a significant association between the level of expression (mean fluorescence intensity, MFI) of CD39 and PD-1 on antigen-specific CD8+ T cells in subjects with HCV and with HIV (r = 0.70, P <0.0001 and r = 0.54, P<0.05, respectively) (Fig 3A and 3B).

We next examined the relationship between CD39 and PD-1 expression and viral load in HCV and HIV infection. We found that in both the HCV and HIV infection there was a modest but significant correlation between viral load and the level of CD39 expression on virus-specific CD8+ T cells measured by MFI (Fig 3C). The fraction of CD39+, virus-specific CD8+ T cells was significantly higher in HIV progressors compared with those from HIV controllers (S2 Fig). A similar, but non-significant, trend was seen comparing CD39 expression in HCV-specific CD8+ T cells in patients with chronic versus resolved disease. However, in HCV, a significantly higher fraction of virus-specific CD8+ T cells co-expressed both CD39 and PD-1 in patients with chronic versus resolved disease (S2 Fig). Consistent with these findings, there was a significant correlation between viral load and the fraction of virus-specific CD8+ T cells that were CD39+ PD-1+ double positive in both HCV and HIV infection (S2 Fig). PD-1 expression was also modestly correlated with the viral load in HCV and in HIV-infected patients (Fig 3D) [3,27]. Thus CD39 expression by virus-specific CD8+ T cells is greatest in setting of high antigen burden.

Transcriptional analysis of CD39+ CD8+ T cells in HCV infection

In order to characterize more broadly the phenotype of CD39+ CD8+ T cells from individuals with chronic infection, we compared the global gene expression profiles of sorted CD39+ and CD39− CD8+ T cells from 8 HCV-infected subjects (3 with acute resolving infection and 5 with chronically evolving infection; S4 Table). Limited numbers of cells precluded the comparison of CD39+ and CD39− CD8+ T cells within HCV-specific cells, leading us to focus on the total

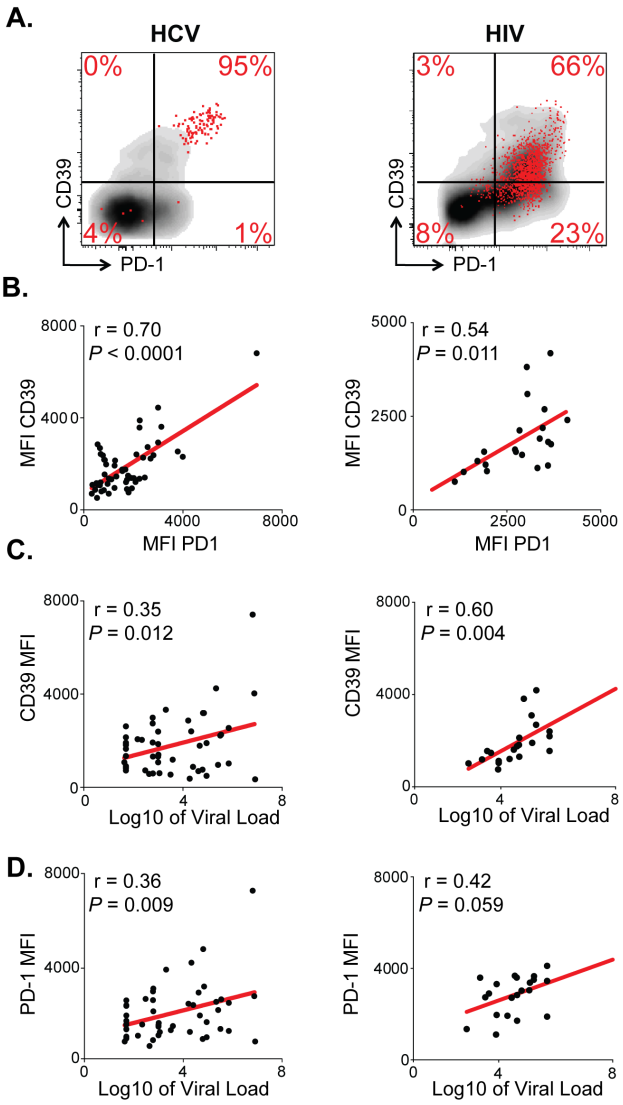

Fig 3. CD39 expression correlates with PD-1 expression and viral load in chronic viral infection. (A) CD39 and PD-1 expression in chronic HCV (left) or HIV infection (right). Representative plots demonstrate total (gray) and virus-specific (red) $CD8^+$ T cells. **(B)** Correlation between CD39 and PD-1 expression by HCV- (left) and HIV-specific (right) $CD8^+$ T cells. **(C)** Correlation between CD39 expression by virus-specific $CD8^+$ T cells and viral load count in HCV (left) or HIV (right) infection. **(D)** Correlation between PD-1 expression by virus-specific $CD8^+$ T cells and viral load in HCV (left) or HIV (right) infection. Correlation was assessed by Pearson correlation coefficient (B, C, D). MFI; mean fluorescence intensity.

doi:10.1371/journal.ppat.1005177.g003

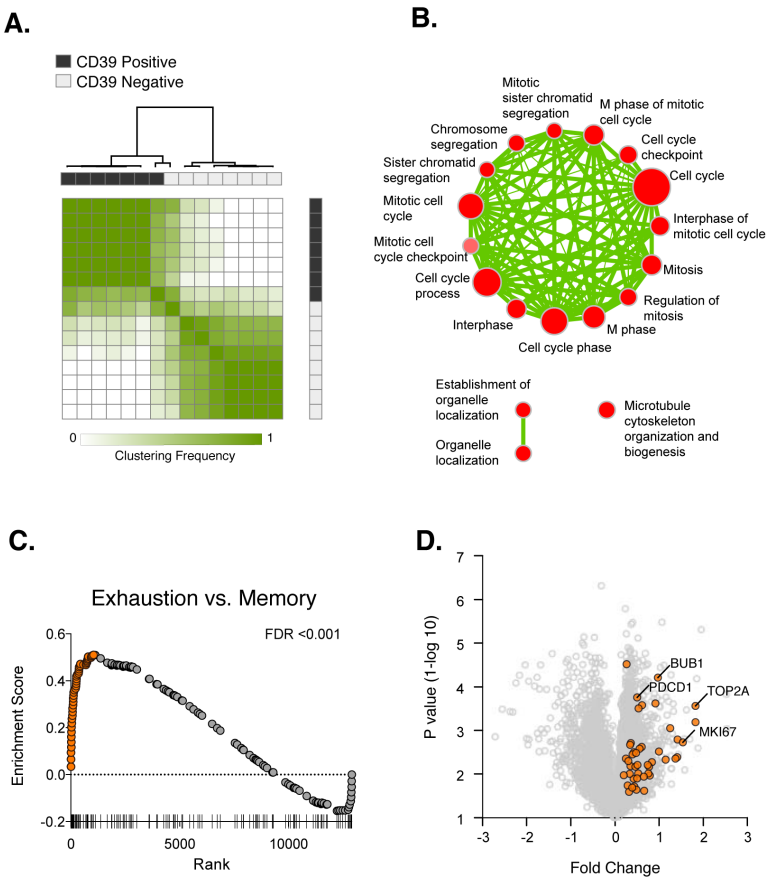

Fig 4. Transcriptional analysis of CD39$^+$ and CD39$^-$ CD8$^+$ T cells in HCV infection. (A) Consensus hierarchical clustering of expression profiles from CD39$^+$ (black) and CD39$^-$ (grey) CD8$^+$ T cells from 8 HCV infected patients. Clustering is based on the top 10% of genes by variance across the dataset. Sample similarity (1-Pearson correlation coefficient) is annotated with color from low (white) to high (green). **(B)** Gene set enrichment map displaying Gene Ontology gene sets enriched (FDR < 0.1) in CD39$^+$ CD8$^+$ T cells from (A). Nodes (in red) are sized in proportion to gene set size; connecting line thickness represents extent of gene member overlap between gene sets. **(C)** Gene set enrichment analysis of a signature of 200 genes up-regulated in exhausted CD8$^+$ T cells from the mouse model of chronic viral infection versus acute infection (day 30 post infection) in the ranked list of genes differentially expressed by CD39$^+$ vs. CD39$^-$ CD8$^+$ T cells. Leading edge genes are indicated by orange symbols. **(D)** Volcano plot of all genes (grey) or exhausted leading edge genes (orange) from (C).

doi:10.1371/journal.ppat.1005177.g004

CD8$^+$ population of antigen-experienced CD8$^+$ T cells (S4 Table). Because naive CD8$^+$ T cells express little CD39 (S1 Fig), we excluded this population from the sorted cells (S3 Fig) to enable direct comparison of antigen-experienced CD39$^+$ and CD39$^-$ CD8$^+$ T cells.

We first used unbiased clustering approaches to identify whether CD39$^+$ and CD39$^-$ CD8$^+$ T cells showed distinct patterns of gene expression. Analysis of gene expression profiles using consensus hierarchical clustering (Fig 4A) showed two distinct clusters of samples that

corresponded almost exactly to CD39$^+$ and CD39$^-$ populations, suggesting that that in both acute and chronic infection, CD39 expression demarcates two types of CD8$^+$ T cells with markedly different patterns of gene expression. Supervised analysis of differential gene expression identified 619 genes differentially expressed (FDR<0.15) between CD39$^+$ and CD39$^-$ CD8$^+$ T cells (S4 Table). Inspection of the list of differentially expressed genes revealed many with known roles in CD8$^+$ T cell biology including increased expression of the inhibitory receptors PD-1 and CTLA-4 in CD39$^+$ CD8$^+$ T cells.

To identify biological processes that were differentially active in CD39$^+$ vs. CD39$^-$ cells, we performed gene set enrichment analysis using the Gene Ontology collection of gene sets [28]. We found no significant enrichment of GO terms in the CD39$^-$ CD8$^+$ subset. In contrast, 21 gene sets significantly enriched (FDR<0.1) in CD39$^+$ population, almost all of which were related to mitosis and cell-cycle associated genes or cytoskeleton organization (Fig 4B). This suggests that CD39$^+$ CD8$^+$ T cells in chronic viral infection show coordinate up-regulation of genes related to proliferation.

The expression of CD39 by CD8$^+$ T cells in chronic but not acute/latent infection, suggests that it may be a marker of T cell exhaustion. We therefore tested whether the profile of CD39$^+$ CD8$^+$ T cells was enriched for genes expressed by exhausted CD8$^+$ cells. Previous studies of gene expression in CD8$^+$ T cells in the mouse model of chronic viral infection with the Clone 13 strain of LCMV have identified signatures of T cell exhaustion that are also enriched in exhausted CD8$^+$ T cells in humans [29–31]. We therefore curated a signature of 200 genes up-regulated by exhausted CD8$^+$ T cells responding to chronic infection relative to functional memory CD8$^+$ T cells generated by acute infection (LCMV Armstrong strain). We found that the exhausted CD8$^+$ T cell signature from LCMV model was significantly enriched in CD39$^+$ vs. CD39$^-$ CD8$^+$ T cells in subjects with HCV infection (Fig 4C). We focused on the "leading edge" genes contributing most to the enrichment [32], which correspond to genes up-regulated both in the mouse exhausted signature and in the human CD39$^+$ profile. As expected, the leading edge genes included PD-1 (*PDCD1*), a feature of both human CD39$^+$ CD8$^+$ T cells and of exhausted CD8$^+$ T cells in the mouse model (Fig 4D). In addition we found that up-regulation of many genes associated with proliferation including *BUB1*, *TOP2A* and *MKI67* was common to mouse exhausted CD8$^+$ T cells and human CD39$^+$ CD8$^+$ T cells. Thus CD39$^+$ CD8$^+$ T cells in HCV infection and exhausted CD8$^+$ T cells in a mouse model of chronic infection share transcriptional features that include genes related to proliferation.

CD39 is increased in exhausted CD8$^+$ T cells in the mouse model of chronic LCMV infection

Because the mouse signature of CD8$^+$ T cell exhaustion was significantly enriched in the transcriptional profile of CD39$^+$ CD8$^+$ T cells in HCV-infected patients, we next asked if CD39 was up-regulated by CD8$^+$ T cells in the mouse model of chronic viral infection. To address this question we compared two well-described mouse models of viral infection using two strains of LCMV: LCMV Armstrong that causes an acute infection that is resolved in up to 8 days; and LCMV Clone 13 that persists in mice for up to 3 months and leads to T cell exhaustion [5,6].

We measured CD39 expression and compared it to PD-1 expression in CD8$^+$ T cells responding to each infection. While naive CD8$^+$ T cells expressed neither CD39 nor PD-1 (Fig 5A), both were rapidly and coordinately up-regulated by antigen-experienced cells following either infection (day 7 post infection [d7 p.i.], Fig 5B). However, in acute infection, the fraction of CD39 bright PD-1$^+$ population decreased with time. In contrast, high expression of CD39 and PD-1 was maintained in Clone 13 infection. The accumulation of CD39 bright PD-1$^+$ cells

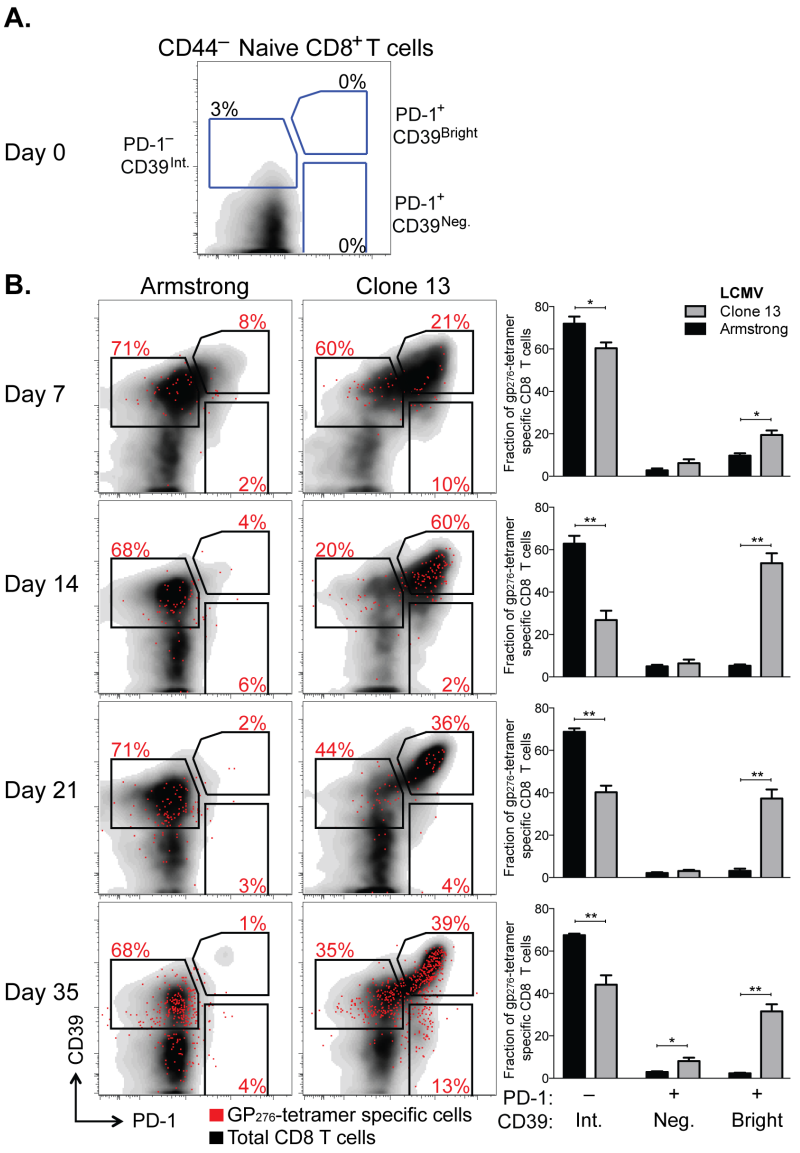

Fig 5. CD39 is highly up-regulated by exhausted CD8+ T cells in a mouse model of chronic infection. (**A, B**) Expression of CD39 and PD-1 by CD44− naive mouse CD8+ T cells (A) and in CD8+ T cells at indicated times following LCMV Armstrong (acute) or Clone 13 (chronic) infection (B). Representative plots show total (black) and H-2Db $GP_{276-286}$ tetramer-specific CD8+ T cells (red). Summary of results in 5 mice per group is shown in bar-graphs on the right. Statistical significance was assessed with Mann-Whitney test. *$P < 0.5$, **$P < 0.01$.

doi:10.1371/journal.ppat.1005177.g005

among the total CD8$^+$ population was most apparent in the H-2D^b GP$_{276-286}$ tetramer-specific CD8$^+$ T cells (Fig 5B).

Thus after chronic viral infection, antigen-specific CD8$^+$ T cells can be identified by high expression of both CD39 and PD-1. This difference in expression of both markers between chronic and acute infection is noticeable as early as d7 p.i. but becomes more pronounced with time after infection.

CD39 expression correlates with a terminally exhausted phenotype in virus-specific CD8$^+$ T cells in chronic infection

Having determined that high, persistent expression of CD39 is a feature of LCMV-specific CD8$^+$ T cells during chronic LCMV infection, we next sought to further characterize the phenotype of CD39$^+$ CD8$^+$ T cells during Clone 13 infection. We analyzed CD39 expression in antigen-experienced, CD44$^+$ CD8$^+$ T cells and found that mice infected with Clone 13 developed a population of cells with particularly high expression of CD39 (CD39^{high}). This population was entirely absent in mice infected with the acute LCMV Armstrong strain at d35 p.i., which only exhibited the presence of intermediate levels of CD39 staining (CD39^{int}) (Fig 6A). Further characterization of the two sub-populations in Clone 13 infected mice revealed that the CD39^{high} cells showed more down-regulation of CD127 (Fig 6B) and higher expression of PD-1 (Fig 6C) than did the CD39^{int} population.

Because the highest levels of PD-1 are characteristic of terminally exhausted CD8$^+$ T cells in chronic infection [12,33], we tested whether CD39^{high} T cells in chronic infection showed other phenotypic characteristics of terminal exhaustion. Analysis of expression of two additional co-inhibitory receptors, CD244 (2B4) and Lag3, showed that a significantly higher fraction of CD39^{high} cells co-expressed multiple receptors, consistent with terminal exhaustion. In contrast, CD39^{int} CD8$^+$ T cells were generally negative for all three receptors analyzed (Fig 6D and 6E). We next examined the expression of the transcription factors T-bet and Eomes. We found that the CD39^{high} subset of CD8$^+$ T cells was comprised primarily of Eomes^{high} T-bet^{low} terminally exhausted phenotype, while the CD39^{int} CD8$^+$ T cells showed a comparable distribution of both (Fig 6F). Similarly, we found that in CD8$^+$ T cells from subjects with either HCV or HIV infection, the CD39$^+$ CD8$^+$ T cell compartment contained a significantly higher ratio of Eomes^{high} T-bet^{low}: Eomes^{low} T-bet^{high} relative to CD39$^-$ CD8$^+$ T cells (S4 Fig). Thus in both humans and mice with chronic viral infection, CD39$^+$ CD8$^+$ T cells show a phenotype consistent with previous descriptions of terminal exhaustion [9].

CD39 correlates with reduced functionality in virus-specific CD8$^+$ T cells in chronic infection

We next examined the functional properties of CD39^{high} and CD39^{int} CD8$^+$ T cells from mice with chronic LCMV infection. Co-production of cytokines IFN-γ and TNFα is a feature of virus-specific T cells responding to acute infection and in the early stages of chronic infection but is progressively lost as exhaustion evolves [2]. To compare the functionality of CD39^{high} and CD39^{int} virus-specific CD8$^+$ T cells, we isolated CD8$^+$ T cells from mice with chronic infection at d35 post-infection and stained for IFN-γ and TNFα following in vitro stimulation with GP$_{33-41}$ peptide. We found a significantly smaller fraction of antigen-specific coproduced IFN-γ and TNFα in CD39^{high} CD8$^+$ T cells compared to CD39^{int} CD8$^+$ T cells (Fig 7A and 7B).

To confirm this finding, we analyzed the function of transferred P14 CD8$^+$ T cells in chronic infection. The P14 TCR transgene recognizes the GP$_{33-41}$ peptide of LCMV presented on H-2D^b. We found that both the frequency of IFN-γ-producing and IFN-γ/TNFα co-producing

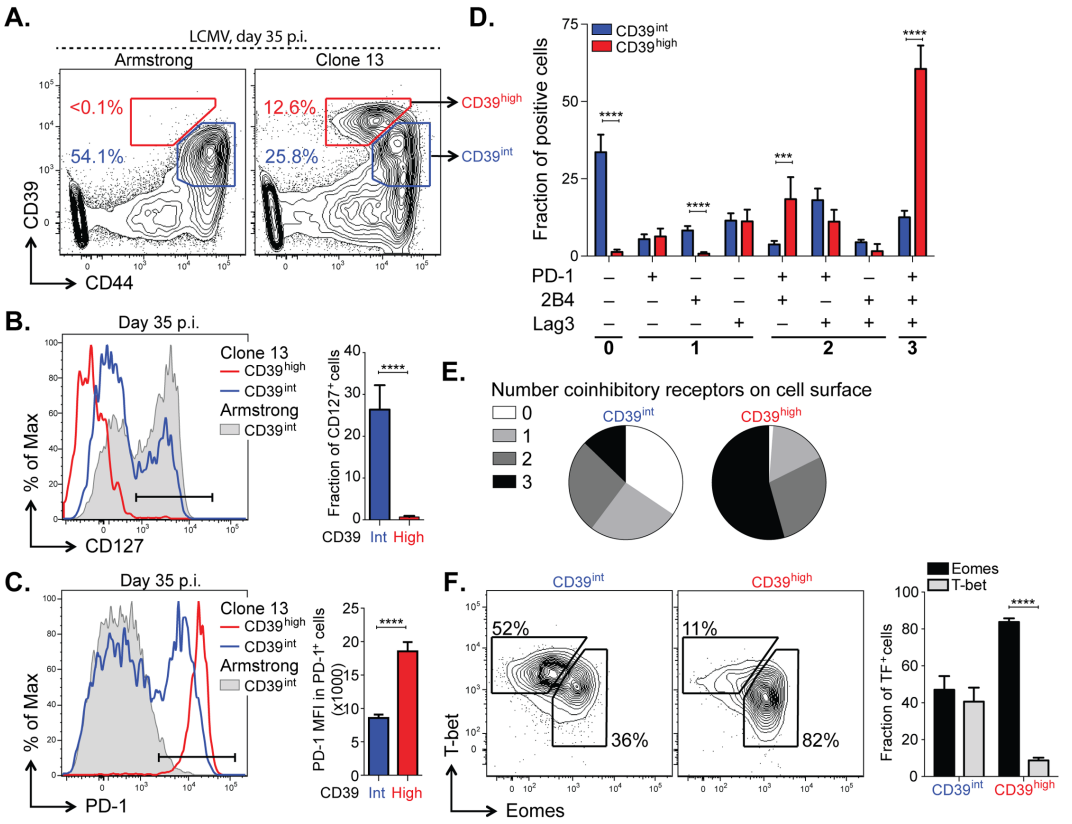

Fig 6. CD39 identifies terminally exhausted CD8⁺ T cells in mice with chronic LCMV infection. (**A**) Expression of CD39 and CD44⁺ by mouse CD8⁺ T cells 30–35 days following LCMV Armstrong (left) or Clone 13 (right) infection. (**B**, **C**) Representative histograms (left) of CD127 (B) and PD-1 (C) expression by CD39^{high} and CD39^{int} CD8⁺ T cells from Clone 13 (red and blue, respectively) and CD39^{int} from Armstrong (filled gray) infected mice on d35 p.i. (left). Fraction of CD127⁺ (B) and MFI of PD-1 in PD-1⁺ cells (C) is shown on the right. Results are from 5 mice. (**D**) Fraction of CD39^{high} and CD39^{int} CD44⁺ CD8⁺ T cells expressing different combinations of co-inhibitory receptors PD-1, 2B4, and Lag3. (**E**) Average number of co-inhibitory receptors expressed by CD39^{int} (left) or CD39^{high} (right) CD8⁺ T cells at d35 p.i. following LCMV Clone 13 infection. (**F**) Representative plots of T-bet and Eomes expression in CD39^{int} (left) and CD39^{high} (right) cells as in (A). Summary of results is shown on the right. Data are representative of three experiments of 5 mice per group. Statistical significance was assessed with Student's t-test (B, C, F) with Holm-Sidak multiple comparison correction (D). $^{**}P < 0.01$, $^{****}P < 0.0001$.

doi:10.1371/journal.ppat.1005177.g006

P14 T cells was significantly lower in CD39^{high} CD8⁺ T cells compared to CD39^{int} CD8⁺ T cells (Fig 7C and 7D). The defect in cytokine secretion was not only observed in terms of the frequency of cytokine-secreting cells, but also in the amount of cytokine detected per cell. Even among cells that did secrete IFN-γ, we found the MFI of expression to be significantly lower in CD39^{high} CD8⁺ T cells compared to CD39^{int} CD8⁺ T cells (Fig 7E and 7F). Thus high levels of CD39 expression demarcate a population of exhausted cells with the poorest function in chronic infection.

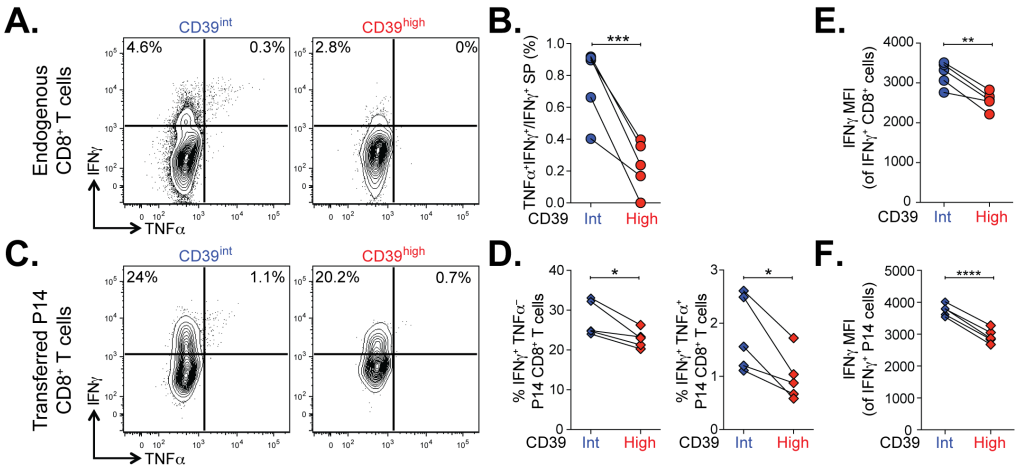

Fig 7. Terminally exhausted CD8+ T cells marked by high levels of CD39 are most impaired in their effector function. (**A**) Representative plots showing the production of IFN-γ and TNFα in CD39^{int} or CD39^{high} CD8+ T cells 36 days following LCMV Clone 13 infection. (**B**) Quantification of cells in (A) that produce both TNFα and IFN-γ relative to IFN-γ only. (**C, D**) Cytokine production by P14 cells (C) gated from an infection as in (A) and summary of IFN-γ and TNFα producing cells. (**E, F**) Mean fluorescence intensity (MFI) of IFN-γ in IFN-γ positive endogenous (E) and transferred P14 cells (F). Statistical significance was assessed with paired Student's t-test. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.

doi:10.1371/journal.ppat.1005177.g007

Discussion

The state of CD8+ T cell exhaustion is characterized by widespread changes in gene expression relative to functional memory CD8+ T cells [5]. However, in humans, identification of specific T cell exhaustion markers that are not shared by more functional CD8+ T cell populations has been challenging [8]. We show that high-level expression of the ectonucleotidase CD39 is characteristic of CD8+ T cells specific for chronic viral infections in humans and mice, but is otherwise rare in the CD8+ T cell compartment of healthy donors. Persistent, high-level expression is also seen in the LCMV mouse model of chronic viral infection, suggesting that CD39 expression is a phenotypic marker of CD8+ T cell exhaustion. Moreover, within the exhausted population in the mouse model, CD39^{high} CD8+ T cells express the highest levels of PD-1, co-express multiple inhibitory receptors and have profoundly impaired function. We found that in both mice and humans, CD39 is expressed preferentially by CD8+ T cells that are T-bet^{low}/Eomes^{high}. These data suggest that CD39 expression by CD8+ T cells is a pathological finding and demarcates the population of CD8+ T cells previously identify as being terminally exhausted [9].

The fact that peripheral blood CD8+ T cells in humans can express CD39 is surprising. Previous data have shown that CD39 expression is restricted to CD4+ regulatory T cells, Th17 cells, and small populations of regulatory-like CD8+ T cells [14,21–23]. Indeed, we find that in the bulk population of CD8+ T cells in healthy donors only a small minority of CD8+ T cells expresses CD39. However, CD39 is abundantly expressed by virus-specific CD8+ T cells in two human chronic infections (HIV and HCV). This helps explain why CD39+ CD8+ T cells have not been appreciated in earlier studies that have focused on healthy individuals, and suggests that, in steady-state conditions, the expression of CD39 by CD8+ T cells is a pathological occurrence that is related to the development of T cell exhaustion. Whether the small fraction of

CD8$^+$ T cells expressing CD39 in healthy donors represents acutely activated CD8$^+$ T cells, or those exhausted by asymptomatic chronic pathogens or inflammatory signals is an important question for future studies.

Several features of CD39-expressing CD8$^+$ T cells suggest that CD39 is a diagnostically valuable marker of T cell exhaustion. First, in both human and mouse CD8$^+$ T cells responding to chronic infection, CD39 is co-expressed with PD-1, an inhibitory receptor expressed by the majority of exhausted T cells [5,6]. Second, CD39 expression correlates with viral load in subjects with HIV and HCV infection suggesting that the conditions of high levels of inflammation and antigen load that lead to exhaustion also increase CD39 expression in the virus-specific pool of CD8$^+$ T cells, as has been observed for PD-1 [3,34]. Third, gene signatures characteristic of exhausted mouse CD8$^+$ T cells are enriched in CD39$^+$ cells relative to CD39$^-$ CD8$^+$ T cells in subjects with HCV infection, underscoring the association between CD39 expression and T cell exhaustion. Finally, chronic LCMV infection in the mouse model increases CD39 expression by exhausted virus-specific CD8$^+$ T cells, and elicits a population of CD39^{high} cells that are absent in functional memory cells. Previous studies show that CD39, like PD-1, is transiently up-regulated by acute T cell activation [14,35]. Additional studies will therefore be required to determine the extent to which T cell activation (rather than exhaustion *per se*) contributes to the up-regulation of CD39 and PD-1 in chronic infection. However, the strong association between CD39 expression and the hallmark phenotypic features of T cell exhaustion in humans and a mouse model suggests that it can serve as a valuable marker of the exhausted T cells state.

The expression of molecules, such as PD-1, that inhibit T cell function has been used to identify exhausted CD8$^+$ T cells in several studies of human chronic infection and cancer [2]. However, there are important distinctions between the pattern of CD39 expression and that of inhibitory receptors. Many inhibitory receptors, such as PD-1 [3,8,36] and CD244 [37,38] are also expressed by a substantial fraction of CD8$^+$ T cells in healthy donors that are not exhausted. In contrast, CD39 expression is found in only a very small minority of CD8$^+$ T cells from healthy donors. This expression pattern suggests that CD39 expression, particularly in combination with PD-1, may be useful as a more specific phenotype of exhausted CD8$^+$ T cells, at least in HCV and HIV infection. In addition, CD39 may provide a useful marker to isolate exhausted CD8$^+$ T cells in settings such as tumor-specific responses where very few reagents are available to identify antigen-specific T cells. Importantly, while CD39 is rare in the CD8$^+$ compartment in healthy donors, it is expressed by CD4$^+$ Tregs–as is PD-1 –making it difficult to distinguish between exhausted CD4$^+$ T cells and Tregs by CD39 expression alone.

Analysis of global expression profiles of CD39$^+$ *versus* CD39$^-$ CD8$^+$ T cells in HCV-infected subjects showed that the CD39$^+$ fraction was strongly enriched for genes related to proliferation. This may at first seem counterintuitive, given the functional defects that have been described in exhausted CD8$^+$ T cells [2,5]. However, data from the mouse model of chronic infection suggest that, unlike memory CD8$^+$ T cells, exhausted CD8$^+$ T cells are dependent on continuous exposure to viral antigen to ensure their survival and undergo extensive cell division at a rate higher than that seen in physiological homeostatic proliferation of the memory CD8$^+$ T cell pool [39]. Exhausted CD8$^+$ T cells therefore have a paradoxical increase in their proliferation *in vivo* despite reduced proliferative potential *in vitro* [40], explaining the increased expression of proliferation-associated genes in CD39$^+$ CD8$^+$ T cells in HCV infection and in mouse exhausted CD8$^+$ T cells [9,41].

Recent studies of exhausted CD8$^+$ T cells have revealed that two distinct states of virus-specific CD8$^+$ T cells exist in chronically infected mice and humans [9,10]. Differential expression of the T-box transcription factors T-bet and Eomes characterize two populations, which form a progenitor-progeny relationship. T-bet^{high} cells display low intrinsic turnover but are capable

of proliferation in response to persisting antigen, giving rise to Eomes^{high} terminal progeny. In contrast, Eomes^{high} CD8$^+$ T cells responding to chronic infection had reduced capacity to undergo additional proliferation *in vivo*. The T-bet^{low} /Eomes^{high} exhausted subset of CD8$^+$ T cells correspond to the PD-1 bright population that has also been shown to be unresponsive to PD-1:PD-L1 blockade. These data suggest that the differential expression of these transcription factors identifies subpopulations of exhausted CD8$^+$ T cells with fundamentally different fates and functional profiles. Our data show that in the LCMV mouse model of chronic infection and in HIV infection, the CD39^{high} subset of CD8$^+$ T cells demarcates T-bet^{low} /Eomes^{high} cells. Consistent with this, CD39$^+$ CD8$^+$ T cells in the mouse model express the highest levels of PD-1, co-express multiple inhibitory receptors and show marked functional defects. These findings suggest that CD39 may be a marker not only of the exhausted state, but specifically of the most terminally exhausted cells, at least in the mouse model. Additional studies of the fate of transferred CD39$^+$ vs. CD39$^-$ exhausted CD8$^+$ T cells in the mouse model, and broader surveys of CD39 expression in human chronic infections will be required to determine whether this marker can be used as a surrogate for terminal exhaustion. However, the strong association between CD39 expression and the key features of terminal exhaustion suggests that it may prove a useful marker to help distinguish between "reversible" and "irreversible" T cell exhaustion. Moroever, the fact that isolating CD39$^+$ cells does not require intracellular staining (as is required for T-bet and Eomes) makes this marker useful for studying the function of this terminally exhausted cells *ex vivo*.

The fact that CD39 is expressed by a slightly larger fraction of HCV-specific CD8$^+$ T cells than HIV-specific CD8$^+$ T cells may be related to differences in the timing of blood sampling during the course of infection, or may be due to differences in the extent of antigen-load and inflammation in the two infections. Alternatively, it may be consistent with a model in which HCV-specific CD8$^+$ T cells are in a more “terminal” state of exhaustion than CD8$^+$ T cells specific for HIV. This possibility is supported by profound loss of HCV-specific CD8$^+$ T cells over the course of chronic infection [42] that is not seen in the HIV-specific CD8$^+$ T cell pool, consistent with the clonal deletion seen in mouse models of extreme CD8$^+$ T cell exhaustion [43,44]

It is tempting to speculate that expression of CD39 contributes to the dysfunction of exhausted T cells [45]. For instance, the expression of CD39 might enable CD8$^+$ T cells to provide negative regulation in an autocrine or juxtacrine fashion via adenosine [18–20] in the same manner as Tregs [15,35]. The fact that CD39 requires both a substrate (ATP) and a downstream enzyme (CD73) to generate adenosine could provide a mechanism to ensure that this negative signaling occurred only in certain contexts such as in inflamed, damaged tissues where the extracellular concentrations of ATP are high and CD73-expressing cells are present [46]. Moreover, CD39-expressing CD8$^+$ T cells may contribute to the general inhibitory milieu by contributing to the inhibition of activated T cells that express the adenosine receptor but are not yet exhausted. It will therefore be important to determine whether inhibition of CD39 activity could provide an additional therapeutic strategy to rescue the function of exhausted T cells.

Materials and Methods

Human Subjects

Healthy human donors were recruited at the Kraft family Blood Donor Center, Dana-Farber Cancer Institute. All human subjects with HCV infection were recruited at the Gastrointestinal Unit and the Department of Surgery of the Massachusetts General Hospital (Boston, MA) (S1 Table).

Individuals with chronic HCV infection (n = 82) were defined by positive anti-HCV antibody and detectable viral load. Patients with spontaneous clearance of HCV, termed resolvers (n = 30), were defined by positive anti-HCV antibody but an undetectable viral load for at least 6 months. The estimated time of infection was calculated either using the exposure date or the time of onset of symptoms and peak ALT (which are assumed to be 7 weeks post infection). All HCV patients were treatment naive and studied at 5.9 and 219.7 weeks post infection. HCV RNA levels were determined using the VERSANT HCV RNA 3.0 (bDNA 3.0) assay (Bayer Diagnostics).

All HIV infected subjects (n = 40) were recruited at the Ragon Institute at the Massachusetts General Hospital (Boston, USA) or the Peter Medawar Building for Pathogen Research (Oxford, UK) (S2 Table). HIV controllers included elite controllers (n = 5) defined as having HIV RNA below the level of detection (<75 viral copies per ml) and viremic controllers (n = 7) with HIV RNA levels < 2,000 viral copies per ml. HIV chronic progressors (n = 28) were defined as having > 2,000 viral copies per ml. All subjects were off therapy. Viral load during chronic infection was measured using the Roche Amplicor version 1.5 assay.

MHC Class I Tetramers

Major histocompatibility complex (MHC) class I HIV Gag-specific tetramers were produced as previously described [47] or obtained from Proimmune. CMV- and EBV-specific MHC class I dextramers conjugated with FITC and APC were purchased from Immudex. Mouse MHC class I tetramers of H-2D^b complexed with LCMV GP$_{276-286}$ were produced as previously described [48,49]. Biotinylated complexes were tetramerized using allophycocyanin-conjugated streptavidin (Molecular Probes). The complete list of multimers can be found in supplemental materials (S3 Table).

Antibodies and flow cytometry

The following anti-human (hu) and anti-mouse (m) fluorochrome-conjugated antibodies were used for flow cytometry: huCD8α (RPA-T8), huCD4 (OKT4), huCD3 (OKT3), huCD39 (A1), huPD-1 (EG12.2H7), huCD25 (BC96), huCCR7 (G043H7), huCD45RA (HI100), huT-bet (4B10), mCD8α (53–6.7), mCD4 (GK1.5), mCD3 (145-2C11), mCD244.2 (m2B4 (B6)458.1), mPD-1 (RMP1-30), mLag3 (C9B7W), mCD44 (IM7), mCD127 (A7R34), mTNFα (MP6X T22) (all from Biolegend), mT-bet (04–46; BD Pharmingen), mCD39 (24DMS1), mIFN-γ (XMG1.2), huEomes (WD1928) and mEomes (Dan11mag) (eBioscience). Intracellular staining was performed following surface staining and fixed and permeabilized using the FoxP3/Transcription Factor Staining Buffer Set (eBioscience). Cells were sorted by BD FACS ARIA II and all other analyses were performed on BD LSR II and BD LSR Fortessa flow cytometers equipped with FACSDiva v6.1. Gates were set using Full Minus One (FMO) controls. Data were analyzed using FlowJo software v9.8 (Treestar).

For intracellular cytokine analysis of mouse T cells, 2x10^6 splenocytes were cultured in the presence of GP$_{33-41}$ peptide (0.2 μg/ml) (sequence KAVYNFATM), brefeldin A (BD), and monensin (BD) for 4.5 hours at 37°C. Following staining for surface antigens, cells were permeabilized and stained for intracellular cytokines with the Cytofix/Cytoperm kit according to manufacturer's instructions (BD Biosciences).

Mice and infections

Wild-type C57BL/6J mice were purchased from The Jackson Laboratory. Female mice (6–8 weeks old) were infected with 2 x 10^5 plaque forming units (p.f.u.) of LCMV-Armstrong intraperitoneally or 4 x 10^6 p.f.u. of LCMV-Clone 13 intravenously and analyzed at indicated time

points by homogenizing the spleen into a single-cell suspension, Ammonium-Chloride-Potassium lysis of red blood cells, followed by antibody staining. For experiments involving P14 cell transfers, Ly5.1+ P14s were isolated from peripheral blood, and 500 P14 cells were transferred i.v. into 5–6 week old wild-type female mice one day prior to infection. Viruses were propagated as described previously [48–50].

HPLC analysis of ATP levels

The concentration of ATP hydrolyzed by $CD8^+$ T cells from subjects with HCV infection (n = 6) was assessed by high performance liquid chromatography (HPLC) as previously described [51]. Briefly, 10,000 $CD39^+$ $CD8^+$ T cells were sorted and placed on ice to minimize ATP production by cells. 20 μM of ATP was added and incubated for 1 h at 37°C in 5% CO_2 to allow for cellular activity to increase and CD39-mediated ATP hydrolysis to occur. Samples were then placed in an ice bath for 10 min to halt enzymatic activity, collected, and centrifuged for 10 min at 380 x g and 0°C. Cells were discarded and supernatant centrifuged again to remove remaining cells (2350 x g, 5 min, 0°C). The resulting RPMI samples (160 μl) were treated with 10 μl of an 8 M perchloric acid solution (Sigma-Aldrich) and centrifuged at 15,900 x g for 10 min at 0°C to precipitate proteins. In order to neutralize the pH of the resulting solutions and to remove lipids, supernatants (80 μl) were treated with 4 M K_2HPO_4 (8 μl) and tri-N-octylamine (50 μl). These samples were mixed with 50 μl of 1,1,2-trichloro-trifluoroethane and centrifuged (15,900 x g, 10 min, 0°C) and this last lipid extraction step was repeated once. The resulting supernatants were subjected to the following procedure to generate fluorescent etheno-adenine products: 150 μl supernatant (or nucleotide standard solution) was incubated at 72°C for 30 min with 250 mM Na_2HPO_4 (20 μl) and 1 M chloroacetaldehyde (30 μl; Sigma-Aldrich) in a final reaction volume of 200 μl, resulting in the formation of 1, N6-etheno derivatives as previously described [51]. Samples were placed on ice, alkalinized with 0.5 M NH_4HCO_3 (50 μl), filtered with a 1 ml syringe and 0.45 μM filter and analyzed using a Waters HPLC system and Supelcosil 3 μM LC-18T reverse phase column (Sigma), consisting of a gradient system described previously, a Waters autosampler, and a Waters 474 fluorescence detector [52]. Empower2 software was used for the analysis of data and all samples were compared with water and ATP standard controls as well as a sample with no cells to determine background degradation of ATP.

Microarray data acquisition

$CD8^+$ T cells from subjects with HCV infection were sorted and pelleted and re-suspended in TRIzol (Invitrogen). RNA extraction was performed using the RNAdvance Tissue Isolation kit (Agencourt). Concentrations of total RNA were determined with a Nanodrop spectrophotometer or Ribogreen RNA quantification kits (Molecular Probes/Invitrogen). RNA purity was determined by Bioanalyzer 2100 traces (Agilent Technologies). Total RNA was amplified with the WT-Ovation Pico RNA Amplification system (NuGEN) according to the manufacturer's instructions. After fragmentation and biotinylation, cDNA was hybridized to HG-U133A 2.0 microarrays (Affymetrix). Data have been deposited in Gene Expression Omnibus with accession code GSE72752.

Statistics

Prior to analysis, microarray data were pre-processed and normalized using robust multichip averaging, as previously described [53]. Differentially gene expression and consensus clustering [54] were performed using Gene-E software (www.broadinstitute.org/cancer/software/GENE-E/), and gene set enrichment analysis was performed as described previously using gene sets from MSigDB [55] or published resources [29,32].

Consensus hierarchical clustering was performed using the top 10% of genes that varied across the dataset, without reference to sample identity. Consensus cluster assesses the "stability" of the clusters discovered using unbiased methods such as hierarchical clustering i.e. the robustness of the putative clusters to sampling variability. The basic assumption is that if the data represent a sample of items drawn from distinct sub-populations, a different sample drawn from the same sub-populations, would result in cluster composition and number should not be radically different. Therefore, the more the attained clusters are robust to sampling variability, the greater the likelihood that the observed clusters represent real structure. The result of consensus clustering is a matrix that shows, for each pair of samples, the proportion of clustering runs on sub-sampled data in which those two items cluster together (shown on a scale of 0 to 1).

Enrichment Map analysis of GSEA results was performed as described [56]. The gene signature of exhaustion was generated by identifying the top 200 genes upregulated in CD8$^+$ T cells responding to chronic vs. acute LCMV infection in microarray data from a previously published study [29].

Ethics statement

All human subjects were recruited with recruited with written informed consent in accordance with Dana-Farber Cancer Institute IRB approval DFCI 00–159, Partners IRB approvals 2010P002121, 2010P002463, 1999P004983, and Oxford Research Ethics Committee approval 06/Q1604/12. The mouse work was performed under a protocol 01214 approved by the HMA Institutional Animal Care and Use Committee (IACUC), in strict accordance with the recommendations in the Guide for the care and use of Laboratory Animals of the National Institutes of Health. The Harvard Medical School animal management program is accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care International (AAALAC).

Supporting Information

S1 Fig. CD39 is expressed by few CD8$^+$ T cells in health donors. Fraction of CD39$^+$ cells in naïve CD8$^+$ T and central memory (CM), effector memory (EM) and effector memory RA$^+$ (EMRA) subpopulations of CD8$^+$ T cells based on CD45RA and CCR7 staining from 18 healthy human donors. Error bars represent SEM. Statistical significance was assessed by Friedman test. $^{**}P < 0.01$, $^{***}P < 0.001$.
(TIF)

S2 Fig. CD39 and PD-1 co-expression in HCV and HIV. (A, B) Fraction of HCV-specific (A) and HIV-specific (B) CD8$^+$ T cells expressing PD-1, CD39, or both in patients with persistent high viral load (black) or patients controlling the disease (grey). Correlation of the fraction of PD-1 and CD39 double positive virus specific CD8$^+$ T cells with the viral load in the blood in HCV (C) and HIV (D) infected patients. Statistical significance was assessed by Mann-Whitney test with Bonferroni correction (A, B). $^{*}P < 0.05$. Correlation was assessed by Pearson correlation coefficient (C, D). MFI; mean fluorescence intensity.
(TIF)

S3 Fig. Cell sorting strategy for microarray analysis. Gating strategy for CD39$^+$ and CD39$^-$ live non-naive CD8$^+$ T cells from HCV-infected patients.
(TIF)

S4 Fig. Comparison of T-bet and Eomes expression by CD39$^+$ and CD39$^-$ CD8$^+$ T cells in patients with chronic viral infection. (**A, D**) Expression of CD39 in CD8$^+$ T cells in patients

infected with HCV (A) and HIV (D). (**B, E**) Expression of transcription factors T-bet and Eomes on CD39$^-$ and CD39$^+$ populations identified in (A) and (D). (**C, F**) Summary of the ratio of terminally exhausted Eomes^{high}/T-bet^{low} CD8$^+$ T cells in CD39$^-$ and CD39$^+$ subsets in HCV (C) and HIV (F) infection. Statistical significance was assessed with paired Student's t-test. $^*P < 0.05$, $^{***}P < 0.001$.
(TIF)

S1 Table. Clinical characteristics of the subjects with HCV infection.
(XLSX)

S2 Table. Clinical characteristics of the subjects with HIV infection.
(XLSX)

S3 Table. The complete list of MHC-peptide multimers used in the study.
(XLSX)

S4 Table. List of genes differentially expressed in CD39$^+$ vs CD39$^-$ CD8$^+$ T cells in HCV infected subjects (FDR<0.15).
(XLSX)

Author Contributions

Conceived and designed the experiments: PKG JG DW EJW GML PK AHS WNH. Performed the experiments: PKG JG DW KY KEP CC EA CL. Analyzed the data: PKG JG DW KY KEP EA. Contributed reagents/materials/analysis tools: WGJ SCR GA PJRG. Wrote the paper: PKG JG DW PK AHS GML WNH.

References

1. Kaech SM, Cui W (2012) Transcriptional control of effector and memory CD8+ T cell differentiation. Nat Rev Immunol 12: 749–761. doi: 10.1038/nri3307 PMID: 23080391
2. Wherry EJ (2011) T cell exhaustion. Nat Immunol 12: 492–499. PMID: 21739672
3. Day CL, Kaufmann DE, Kiepiela P, Brown JA, Moodley ES, et al. (2006) PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. Nature 443: 350–354. PMID: 16921384
4. Lechner F, Wong DK, Dunbar PR, Chapman R, Chung RT, et al. (2000) Analysis of successful immune responses in persons infected with hepatitis C virus. J Exp Med 191: 1499–1512. PMID: 10790425
5. Wherry EJ, Ha SJ, Kaech SM, Haining WN, Sarkar S, et al. (2007) Molecular signature of CD8+ T cell exhaustion during chronic viral infection. Immunity 27: 670–684. PMID: 17950003
6. Barber DL, Wherry EJ, Masopust D, Zhu B, Allison JP, et al. (2006) Restoring function in exhausted CD8 T cells during chronic viral infection. Nature 439: 682–687. PMID: 16382236
7. Kroy DC, Ciuffreda D, Cooperrider JH, Tomlinson M, Hauck GD, et al. (2014) Liver environment and HCV replication affect human T-cell phenotype and expression of inhibitory receptors. Gastroenterology 146: 550–561. doi: 10.1053/j.gastro.2013.10.022 PMID: 24148617
8. Duraiswamy J, Ibegbu CC, Masopust D, Miller JD, Araki K, et al. (2011) Phenotype, function, and gene expression profiles of programmed death-1(hi) CD8 T cells in healthy human adults. J Immunol 186: 4200–4212. doi: 10.4049/jimmunol.1001783 PMID: 21383243
9. Paley MA, Kroy DC, Odorizzi PM, Johnnidis JB, Dolfi DV, et al. (2012) Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. Science 338: 1220–1225. doi: 10.1126/science.1229620 PMID: 23197535
10. Buggert M, Tauriainen J, Yamamoto T, Frederiksen J, Ivarsson MA, et al. (2014) T-bet and Eomes are differentially linked to the exhausted phenotype of CD8+ T cells in HIV infection. PLoS Pathog 10: e1004251. doi: 10.1371/journal.ppat.1004251 PMID: 25032686
11. Kurktschiev PD, Raziorrouh B, Schraut W, Backmund M, Wachtler M, et al. (2014) Dysfunctional CD8+ T cells in hepatitis B and C are characterized by a lack of antigen-specific T-bet induction. J Exp Med 211: 2047–2059. doi: 10.1084/jem.20131333 PMID: 25225458

12. Blackburn SD, Shin H, Freeman GJ, Wherry EJ (2008) Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. Proc Natl Acad Sci U S A 105: 15016–15021. doi: 10.1073/pnas.0801497105 PMID: 18809920

13. Kaczmarek E, Koziak K, Sevigny J, Siegel JB, Anrather J, et al. (1996) Identification and characterization of CD39/vascular ATP diphosphohydrolase. J Biol Chem 271: 33116–33122. PMID: 8955160

14. Kansas GS, Wood GS, Tedder TF (1991) Expression, distribution, and biochemistry of human CD39. Role in activation-associated homotypic adhesion of lymphocytes. J Immunol 146: 2235–2244. PMID: 1672348

15. Deaglio S, Dwyer KM, Gao W, Friedman D, Usheva A, et al. (2007) Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. J Exp Med 204: 1257–1265. PMID: 17502665

16. Borsellino G, Kleinewietfeld M, Di Mitri D, Sternjak A, Diamantini A, et al. (2007) Expression of ectonucleotidase CD39 by Foxp3+ Treg cells: hydrolysis of extracellular ATP and immune suppression. Blood 110: 1225–1232. PMID: 17449799

17. Junger WG (2011) Immune cell regulation by autocrine purinergic signalling. Nature Publishing Group: Nature Publishing Group. pp. 201–212. doi: 10.1038/nri2938 PMID: 21331080

18. Zarek PE, Huang CT, Lutz ER, Kowalski J, Horton MR, et al. (2008) A2A receptor signaling promotes peripheral tolerance by inducing T-cell anergy and the generation of adaptive regulatory T cells. Blood 111: 251–259. PMID: 17909080

19. Huang S, Apasov S, Koshiba M, Sitkovsky M (1997) Role of A2a extracellular adenosine receptor-mediated signaling in adenosine-mediated inhibition of T-cell activation and expansion. Blood 90: 1600–1610. PMID: 9269779

20. Lokshin A, Raskovalova T, Huang X, Zacharia LC, Jackson EK, et al. (2006) Adenosine-mediated inhibition of the cytotoxic activity and cytokine production by activated natural killer cells. Cancer Res 66: 7758–7765. PMID: 16885379

21. Moncrieffe H, Nistala K, Kamhieh Y, Evans J, Eddaoudi A, et al. (2010) High Expression of the Ectonucleotidase CD39 on T Cells from the Inflamed Site Identifies Two Distinct Populations, One Regulatory and One Memory T Cell Population. The Journal of Immunology. pp. 134–143. doi: 10.4049/jimmunol.0803474 PMID: 20498355

22. Pulte D, Furman RR, Broekman MJ, Drosopoulos JH, Ballard HS, et al. (2011) CD39 expression on T lymphocytes correlates with severity of disease in patients with chronic lymphocytic leukemia. Clin Lymphoma Myeloma Leuk 11: 367–372. doi: 10.1016/j.clml.2011.06.005 PMID: 21816376

23. Boer MC, van Meijgaarden KE, Bastid J, Ottenhoff TH, Joosten SA (2013) CD39 is involved in mediating suppression by Mycobacterium bovis BCG-activated human CD8(+) CD39(+) regulatory T cells. Eur J Immunol 43: 1925–1932. doi: 10.1002/eji.201243286 PMID: 23606272

24. Robson SC, Kaczmarek E, Siegel JB, Candinas D, Koziak K, et al. (1997) Loss of ATP diphosphohydrolase activity with endothelial cell activation. J Exp Med 185: 153–163. PMID: 8996251

25. Papanikolaou A, Papafotika A, Murphy C, Papamarcaki T, Tsolas O, et al. (2005) Cholesterol-dependent lipid assemblies regulate the activity of the ecto-nucleotidase CD39. J Biol Chem 280: 26406–26414. PMID: 15890655

26. Wu Y, Sun X, Kaczmarek E, Dwyer KM, Bianchi E, et al. (2006) RanBPM associates with CD39 and modulates ecto-nucleotidase activity. Biochem J 396: 23–30. PMID: 16478441

27. Kasprowicz V, Schulze Zur Wiesch J, Kuntzen T, Nolan BE, Longworth S, et al. (2008) High level of PD-1 expression on hepatitis C virus (HCV)-specific CD8+ and CD4+ T cells during acute HCV infection, irrespective of clinical outcome. J Virol 82: 3154–3160. PMID: 18160439

28. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25–29. PMID: 10802651

29. Doering TA, Crawford A, Angelosanto JM, Paley MA, Ziegler CG, et al. (2012) Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory. Immunity 37: 1130–1144. doi: 10.1016/j.immuni.2012.08.021 PMID: 23159438

30. Quigley M, Pereyra F, Nilsson B, Porichis F, Fonseca C, et al. (2010) Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF. Nat Med 16: 1147–1151. doi: 10.1038/nm.2232 PMID: 20890291

31. Baitsch L, Baumgaertner P, Devevre E, Raghav SK, Legat A, et al. (2011) Exhaustion of tumor-specific CD8(+) T cells in metastases from melanoma patients. J Clin Invest 121: 2350–2360. doi: 10.1172/JCI46102 PMID: 21555851

32. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102: 15545–15550. PMID: 16199517

33. Blackburn SD, Shin H, Haining WN, Zou T, Workman CJ, et al. (2009) Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. Nat Immunol 10: 29–37. doi: 10.1038/ni.1679 PMID: 19043418

34. Trautmann L, Janbazian L, Chomont N, Said EA, Gimmig S, et al. (2006) Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. Nat Med 12: 1198–1202. PMID: 16917489

35. Jenabian MA, Seddiki N, Yatim A, Carriere M, Hulin A, et al. (2013) Regulatory T cells negatively affect IL-2 production of effector T cells through CD39/adenosine pathway in HIV infection. PLoS Pathog 9: e1003319. doi: 10.1371/journal.ppat.1003319 PMID: 23658513

36. Petrovas C, Casazza JP, Brenchley JM, Price DA, Gostick E, et al. (2006) PD-1 is a regulator of virus-specific CD8+ T cell survival in HIV infection. J Exp Med 203: 2281–2292. PMID: 16954372

37. Pita-Lopez ML, Gayoso I, DelaRosa O, Casado JG, Alonso C, et al. (2009) Effect of ageing on CMV-specific CD8 T cells from CMV seropositive healthy donors. Immun Ageing 6: 11. doi: 10.1186/1742-4933-6-11 PMID: 19715573

38. Rey J, Giustiniani J, Mallet F, Schiavon V, Boumsell L, et al. (2006) The co-expression of 2B4 (CD244) and CD160 delineates a subpopulation of human CD8+ T cells with a potent CD160-mediated cytolytic effector function. Eur J Immunol 36: 2359–2366. PMID: 16917959

39. Shin H, Blackburn SD, Intlekofer AM, Kao C, Angelosanto JM, et al. (2009) A role for the transcriptional repressor Blimp-1 in CD8(+) T cell exhaustion during chronic viral infection. Immunity 31: 309–320. doi: 10.1016/j.immuni.2009.06.019 PMID: 19664943

40. Migueles SA, Laborico AC, Shupert WL, Sabbaghian MS, Rabin R, et al. (2002) HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors. Nat Immunol 3: 1061–1068. PMID: 12368910

41. Shin H, Blackburn SD, Blattman JN, Wherry EJ (2007) Viral antigen and extensive division maintain virus-specific CD8 T cells during chronic infection. J Exp Med 204: 941–949. PMID: 17420267

42. Cox AL, Mosbruger T, Lauer GM, Pardoll D, Thomas DL, et al. (2005) Comprehensive analyses of CD8 + T cell responses during longitudinal study of acute human hepatitis C. Hepatology 42: 104–112. PMID: 15962289

43. Zajac AJ, Blattman JN, Murali-Krishna K, Sourdive DJ, Suresh M, et al. (1998) Viral immune evasion due to persistence of activated T cells without effector function. J Exp Med 188: 2205–2213. PMID: 9858507

44. Moskophidis D, Lechner F, Pircher H, Zinkernagel RM (1993) Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. Nature 362: 758–761. PMID: 8469287

45. Seddiki N, Brezar V, Draenert R (2014) Cell exhaustion in HIV-1 infection: role of suppressor cells. Curr Opin HIV AIDS 9: 452–458. doi: 10.1097/COH.0000000000000087 PMID: 25010895

46. Toth I, Le AQ, Hartjen P, Thomssen A, Matzat V, et al. (2013) Decreased frequency of CD73+CD8+ T cells of HIV-infected patients correlates with immune activation and T cell exhaustion. J Leukoc Biol 94: 551–561. doi: 10.1189/jlb.0113018 PMID: 23709688

47. Leisner C, Loeth N, Lamberth K, Justesen S, Sylvester-Hvid C, et al. (2008) One-pot, mix-and-read peptide-MHC tetramers. PLoS One 3: e1678. doi: 10.1371/journal.pone.0001678 PMID: 18301755

48. Wherry EJ, Blattman JN, Murali-Krishna K, van der Most R, Ahmed R (2003) Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. J Virol 77: 4911–4927. PMID: 12663797

49. Murali-Krishna K, Altman JD, Suresh M, Sourdive DJ, Zajac AJ, et al. (1998) Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. Immunity 8: 177–187. PMID: 9491999

50. Ahmed R, Salmi A, Butler LD, Chiller JM, Oldstone MB (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. J Exp Med 160: 521–540. PMID: 6332167

51. Lazarowski ER, Tarran R, Grubb BR, van Heusden CA, Okada S, et al. (2004) Nucleotide release provides a mechanism for airway surface liquid homeostasis. J Biol Chem 279: 36855–36864. PMID: 15210701

52. Chen Y, Corriden R, Inoue Y, Yip L, Hashiguchi N, et al. (2006) ATP release guides neutrophil chemotaxis via P2Y2 and A3 receptors. Science 314: 1792–1795. PMID: 17170310

53. Haining WN, Ebert BL, Subrmanian A, Wherry EJ, Eichbaum Q, et al. (2008) Identification of an evolutionarily conserved transcriptional signature of CD8 memory differentiation that is shared by T and B cells. J Immunol 181: 1859–1868. PMID: 18641323

54. Monti S TP, Mesirov J, Golub T (2003) Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning 52: 91–118.

55. Liberzon A (2014) A description of the Molecular Signatures Database (MSigDB) Web site. Methods Mol Biol 1150: 153–160. doi: 10.1007/978-1-4939-0512-6_9 PMID: 24743996

56. Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5: e13984. doi: 10.1371/journal.pone.0013984 PMID: 21085593

Appendix C: Godec et al. PNAS 2015

This is a reprint of:

Jernej Godec, Glenn S. Cowley, R. Anthony Barnitz, Ozan Alkan, David E. Root, Arlene H. Sharpe, W. Nicholas Haining. Inducible RNAi *in vivo* reveals that BATF is required to initiate but not to maintain CD8+ T cell effector differentiation. *Proceedings of the National Academy of Sciences USA*. 2015 Jan 13;112(2):512-7. (PMID: 25548173)

Inducible RNAi in vivo reveals that the transcription factor BATF is required to initiate but not maintain CD8$^+$ T-cell effector differentiation

Jernej Godec[a,b], Glenn S. Cowley[c], R. Anthony Barnitz[a], Ozan Alkan[c], David E. Root[c], Arlene H. Sharpe[b,c,d,1], and W. Nicholas Haining[a,c,e,1]

[a]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115; [b]Department of Microbiology and Immunobiology, Harvard Medical School, and Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA 02115; [c]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [d]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115; and [e]Division of Hematology/Oncology, Children's Hospital, Harvard Medical School, Boston, MA 02115

Edited by Rafi Ahmed, Emory University, Atlanta, GA, and approved December 3, 2014 (received for review July 14, 2014)

The differentiation of effector CD8$^+$ T cells is critical for the development of protective responses to pathogens and for effective vaccines. In the first few hours after activation, naive CD8$^+$ T cells initiate a transcriptional program that leads to the formation of effector and memory T cells, but the regulation of this process is poorly understood. Investigating the role of specific transcription factors (TFs) in determining CD8$^+$ effector T-cell fate by gene knockdown with RNAi is challenging because naive T cells are refractory to transduction with viral vectors without extensive ex vivo stimulation, which obscures the earliest events in effector differentiation. To overcome this obstacle, we developed a novel strategy to test the function of genes in naive CD8$^+$ T cells in vivo by creating bone marrow chimera from hematopoietic progenitors transduced with an inducible shRNA construct. Following hematopoietic reconstitution, this approach allowed inducible in vivo gene knockdown in any cell type that developed from this transduced progenitor pool. We demonstrated that lentivirus-transduced progenitor cells could reconstitute normal hematopoiesis and develop into naive CD8$^+$ T cells that were indistinguishable from wild-type naive T cells. This experimental system enabled induction of efficient gene knockdown in vivo without subsequent manipulation. We applied this strategy to show that the TF BATF is essential for initial commitment of naive CD8$^+$ T cells to effector development but becomes dispensable by 72h. This approach makes possible the study of gene function in vivo in unperturbed cells of hematopoietic origin that are refractory to viral transduction.

CD8 T cell | RNAi | transcription factor | BATF

Following activation by antigen, costimulation, and inflammation, naive CD8$^+$ T cells initiate a differentiation program resulting in massive changes in gene expression and cell function, which leads to the formation of effector and memory T cells (1). This differentiation program is critical for the development of effective tumor immunity (2) and the control of pathogens (3). Although the development of effector CD8$^+$ T cells occurs over a period of days (4), early events in the life history of CD8$^+$ T cells are critical in determining their fate (5–10), suggesting that investigating the events that occur in the hours following initial antigen encounter will be essential for defining the mechanisms that regulate the fate of effector CD8$^+$ T cells.

The AP-1 family transcription factor (TF) BATF is absolutely required for effector CD8$^+$ T-cell differentiation and coordinates the program of gene expression essential for this process (11). The role of specific TFs in regulating CD8$^+$ T-cell effector differentiation has generally been investigated using germ-line or conditional KOs. However, these approaches are restricted to studying a small number of candidate genes (12). In contrast, perturbing genes with RNAi could permit the study of many more candidate regulators in parallel (13), but techniques to deliver shRNAs to T cells are limited by the need to stimulate cells to divide using T-cell receptor (TCR) cross-linking (14), infection (15, 16), or cytokine stimulation (17) to achieve meaningful transduction frequencies with viral vectors encoding shRNA constructs. The need to activate T cells for delivery of shRNAs raises concerns about whether this activation alters these T cells at a critical phase of time when even subtle perturbations of TFs can profoundly influence T-cell fate (10).

To address these limitations, we have developed an experimental system to knock down gene expression in T cells in vivo using shRNA without the need to transduce T cells directly. We generated bone marrow (BM) chimera from hematopoietic stem cells (HSCs) transduced with an inducible shRNA vector. Following hematopoietic reconstitution, this strategy allows inducible gene knockdown in any cell type that developed from this transduced progenitor pool, including resting naive CD8$^+$ T cells in vivo. We have applied this system to show that BATF is essential for initial commitment of naive CD8$^+$ T cells to effector cell development, but becomes dispensable after 72 h.

Results

Lentivirus-Transduced Stem Cells Reconstitute Blood Immune Lineages and Give Rise to Effector CD8$^+$ T Cells with Unaltered Functionality. Resting T cells are refractory to lentiviral transduction, but HSCs are more readily transduced. We therefore generated bone marrow chimeric animals using lentivirus-transduced

Significance

Effector CD8$^+$ T-cell differentiation is essential for protective immunity. Investigating specific genes in this process by knockdown with RNAi is challenging because naive T cells are refractory to viral transduction. To overcome this obstacle, we developed a novel strategy to knock down genes in naive CD8$^+$ T cells by creating bone marrow chimera from hematopoietic progenitors transduced with an inducible shRNA. This approach enabled inducible in vivo gene knockdown in any cell type developed from this progenitor pool. We applied this strategy to show that the transcription factor BATF is essential for initial commitment to effector differentiation but becomes dispensable by 72 h. This approach now enables the study of gene function in vivo in cells of hematopoietic origin otherwise refractory to viral transduction.

Author contributions: J.G., A.H.S., and W.N.H. designed research; J.G. and R.A.B. performed research; G.S.C., O.A., and D.E.R. contributed new reagents/analytic tools; J.G., A.H.S., and W.N.H. analyzed data; and J.G., A.H.S., and W.N.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: nicholas_haining@dfci.harvard.edu or arlene_sharpe@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1413291112/-/DCSupplemental.

hematopoietic progenitor cells in which hematopoietic lineages (including T cells) are reconstituted with transduced cells (Fig. 1*A*). We isolated lineage$^-$/sca-1$^+$/c-kit$^+$ (LSK) cells (which include HSCs and multipotent progenitors), from the bone marrow of P14 TCR transgenic (Tg) mice in which most CD8$^+$ T cells express a TCR specific for lymphocytic choriomeningitis virus (LCMV) glycoprotein $(GP)_{33–41}$ peptide presented on H-2D^b (Fig. 1*A*), and transduced them with a lentivirus carrying a GFP expression cassette so that the fate of transduced cells could be tracked. We used congenic markers to distinguish transplanted cells from recipient cells in bone marrow chimeras.

To test whether lentivirus-transduced LSK cells could be used to generate fully functional CD8$^+$ T cells, we first transduced the LSK cells with lentivirus encoding only GFP under a human phosphoglycerate kinase (PGK) promoter (PGK-eGFP) and transplanted them into lethally irradiated animals (50,000 cells per animal). Following reconstitution (8–12 wk later), analyses of major lineages in the immune system showed that the frequencies of GFP$^+$ B cells (B220$^+$), CD4$^+$ and CD8$^+$ T cells, dendritic cells (CD11c$^+$), and monocytes (CD11b$^+$) were similar to that of the LSK inoculum (Fig. 1*B*), suggesting efficient engraftment of transduced cells. CD4$^+$ and CD8$^+$ T, total T, B, and myeloid lineages developed from transplanted GFP$^+$ (transduced) and GFP$^-$ (untransduced) LSK with equal efficiency (Fig. 1*C* and Fig. S1). We next compared effector CD8$^+$ T-cell differentiation of naive CD8$^+$ T cells derived from transplanted PGK-eGFP transduced (GFP$^+$) LSK cells with differentiation of naive CD8$^+$ T cells derived from transplanted but untransduced (GFP$^-$) LSK cells. We transferred equal ratios of GFP$^+$ and GFP$^-$ naive P14 CD8$^+$ T cells to naive wild-type recipients (10,000 cells per animal) and infected them with H1N1 influenza PR8 engineered to express GP33 (PR8-GP33) (Fig. 1*D*). We found equal expansion and persistence of GFP$^+$ and GFP$^-$ effector CD8$^+$ T cells at 10 d postinfection (p.i.).

We next compared the phenotype and function of effector CD8$^+$ T cells arising from naive CD8$^+$ T cells that developed from transduced LSKs with effector CD8$^+$ T cells differentiating from untransduced naive CD8$^+$ T cells. We analyzed the proliferative capacity, expression of cell surface molecules, key transcription factors, and production of cytokines upon restimulation and found no difference between untransduced and transduced effector CD8$^+$ T cells at d 8 postinfection (p.i.) (Fig. S2). Thus, lentiviral transduction of LSK neither impairs the development of lymphoid and myeloid lineages following transplantation nor alters effector CD8$^+$ T-cell generation, proliferative capacity, or survival following transfer of naive CD8$^+$ T cells.

To compare the persistence and phenotype of effector CD8$^+$ T cells derived from untransduced naive CD8$^+$ T cells, or naive CD8$^+$ T cells generated from bone marrow chimeras, we transferred either unmodified naive P14 CD8$^+$ T cells or naive P14 CD8$^+$ T cells carrying the 1xLacO–shLacZ construct into congenically distinct LCMV-infected wild-type recipients. We analyzed the fraction of transferred cells at d28 p.i. and observed no difference in the frequency of unmodified CD8$^+$ T cells and those carrying the 1xLacO–shLacZ construct (Fig. S3 *A* and *B*). Additionally, 1xLacO and WT P14 memory cells were indistinguishable in their expression of cell surface molecules and production of cytokines upon restimulation (Fig. S3 *C* and *D*). This finding suggests that the persistence of CD8$^+$ T cells following effector differentiation is not altered by the presence of the lentiviral vector.

Naive T Cells That Develop from Transduced LSK Are Indistinguishable from Wild-Type Naive T Cells. We next examined whether naive CD8$^+$ T cells that developed from transduced LSK cells show any alterations of surface phenotype, proliferative status, or gene expression that might obscure analysis of early differentiation events. We compared naive CD8$^+$ T cells that were derived from transduced LSK cells with control wild-type P14 CD8$^+$ T cells. We also studied P14 CD8$^+$ T cells cultured in conditions used in

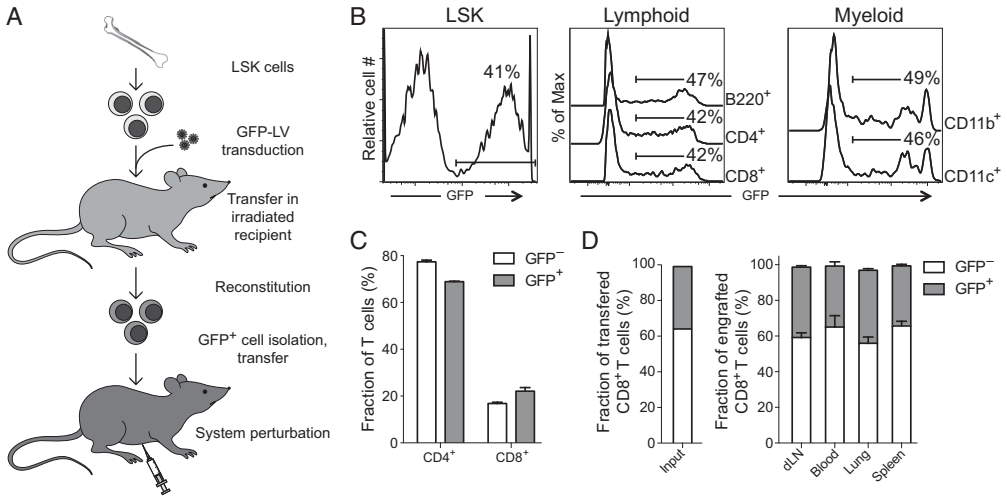

Fig. 1. Transduced bone marrow progenitor populations efficiently reconstitute myeloid and lymphoid compartments and develop normally into functional CD8$^+$ T cells. (*A*) Schematic diagram of transduction strategy. (*B*) Fraction of LSK cells (*Left*) transduced with GFP-expressing lentivirus at the time of transplant, and in lymphoid (*Middle*) and myeloid (*Right*) cell populations following engraftment. (*C*) Quantitation of fractions of transduced (GFP$^+$) and untransduced (GFP$^-$), donor-derived T cells following engraftment as in *B*. (*D*) Fraction of transduced (GFP$^+$) and untransduced (GFP$^-$), donor-derived naive P14 CD8$^+$ T cells (*Left*) before adoptive transfer, and of effector (*Right*) P14 CD8$^+$ T cells in tissues indicated 10 d following transfer and subsequent host infection with PR8-GP33 influenza virus. Representative data are shown from two to three experiments with four to five mice per group.

previous studies to facilitate direct viral transduction of T cells (14–18) to compare $CD8^+$ T cells generated by our approach and previously reported methods: (*i*) activation of $CD8^+$ T cells in vivo by infecting P14 mice with 2×10^5 p.f.u. LCMV Armstrong, (*ii*) activation in vitro by stimulation with anti-CD3 and anti-CD28, or (*iii*) incubation in vitro with a combination of IL-7 and IL-15 cytokines.

The proportions of naive ($CD62L^+$ $CD44^-$), central memory ($CD62L^+$ $CD44^+$), and effector memory ($CD62L^-$ $CD44^+$) cells were similar in the GFP^+ naive $CD8^+$ T cells from the transduced BM chimeras and in naive $CD8^+$ T cells from wild-type mice, but were markedly altered by the other stimulation conditions, particularly with cytokine treatment (Fig. 2*A*). The expression of cytokine receptors including CD25, CD127, and CD122 was not different in GFP^+ naive and wild-type naive $CD8^+$ T cells, but was altered in naive $CD8^+$ T cells treated with anti-CD3/CD28 or cytokines (Fig. 2*B*). The GFP^+ naive $CD8^+$ T cells also showed a low rate of homeostatic turnover that was similar to wild-type naive $CD8^+$ T cells (Fig. 2*C*). In contrast, all of the other stimulation conditions induced varying degrees of cell proliferation (Fig. 2*C*).

We measured transcript abundance for TFs and effector molecules that change during $CD8^+$ effector T-cell differentiation. Important regulators of effector differentiation such as T-bet (*Tbx21*), Eomesodermin (*Eomes*), and Blimp1 (*Prdm1*), as well as effector molecules including granzyme A and B, perforin1, and IFNγ and TNFα were unchanged in GFP^+ naive $CD8^+$ T cells relative to wild-type naive $CD8^+$ T cells, but were up-regulated in the other stimulation conditions (Fig. 2*D*). Thus, GFP^+ naive $CD8^+$ T cells that had developed from transduced LSK cells were indistinguishable from untransduced naive $CD8^+$ T cells. In contrast, existing protocols used to achieve viral transduction of naive $CD8^+$ T cells were associated with marked perturbation of the T-cell state.

Lac Operon-Regulated shRNA Allows Inducible, Efficient, and Transient Gene Knockdown in Vivo at Low Concentrations of Isopropyl β-D-1-Thiogalactopyranoside. Because constitutive gene knockdown in LSK could compromise the development of immune lineages, we used an inducible shRNA expression vector that uses the Lac operon system to regulate the shRNA promoter following addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) (Fig. S4*A*). We confirmed the inducibility of gene knockdown by targeting a control gene in a Jurkat cell line. Target gene (GFP) expression was only minimally affected in the uninduced state (Fig. 3*A*). However, gene knockdown following IPTG induction of shRNA expression was as efficient as that achieved by a constitutive shRNA expressing vector (Fig. 3*A*) even at low concentrations of IPTG (Fig. S4*B*).

To test knockdown efficiency in primary $CD8^+$ T cells, we generated bone marrow chimeras with an IPTG-inducible vector encoding an shRNA targeting BATF (shBATF) and a GFP expression cassette to create GFP^+ naive T cells that carried the inducible shRNA vector (hereafter "shBATF–naive T cells"). We first tested inducible knockdown in vitro by stimulating the cells with anti-CD3/CD28 and assessing the *Batf* transcript levels 3 d following activation. IPTG was administered to the bone marrow chimeras 3 d before activation (d −3) or 1 d following activation (d +1). Decreased target gene expression was apparent in both transcript and protein abundance as early as 2 d following IPTG addition in vitro (Fig. 3 *B* and *C*) and was comparable to knockdown with the constitutive vector (mean = 80.4%, SD = 7.8%). To test inducible knockdown in vivo, we transferred shBATF–naive P14 $CD8^+$ T cells into mice that were at the same time infected with LCMV-infected mice treated with IPTG, and measured BATF expression after 3 d. Initiating IPTG induction 1 d following cell activation resulted in modest (18%) gene knockdown but treating bone marrow chimera 3 d before

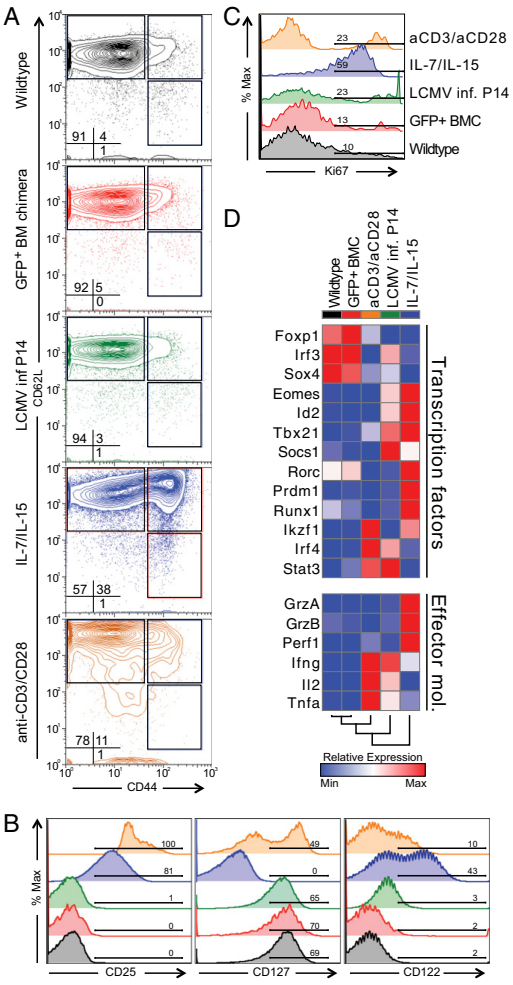

Fig. 2. $CD8^+$ T cells derived from transduced LSK cells are indistinguishable from untransduced naive $CD8^+$ T cells. Expression of (*A*) naive surface markers, (*B*) cytokine receptors, and (*C*) Ki-67 by wild-type P14 $CD8^+$ T cells (black); P14 $CD8^+$ T cells derived from transduced LSKs (red); P14 $CD8^+$ T cells stimulated by LCMV infection (green); by the cytokines IL-7 and IL-15 (blue); or by anti-CD3 and anti-CD28 antibodies (orange). (*D*) Relative transcript abundance of transcriptional regulators and effector molecules changes measured by quantitative RT-PCR in examined groups.

transfer resulted in a significantly greater degree of gene silencing (68% knockdown) when it was measured 3 d following transfer and infection (Fig. 3*C*). Thus, efficient, inducible gene knockdown can be achieved in following activation of shBATF–naive $CD8^+$ T cells in vivo.

BATF Knockdown Impairs the Development of $CD8^+$ Effector T Cells Following Acute Viral Infection. We have recently shown that $Batf^{-/-}$ $CD8^+$ T cells show profoundly impaired effector $CD8^+$ T-cell differentiation (11). To test whether BATF knockdown in wild-type

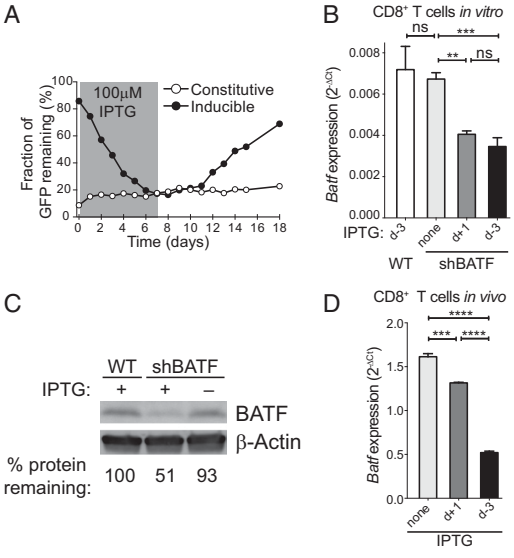

Fig. 3. Novel shRNA vector enables efficient, inducible, and transient gene knockdown in vitro and in vivo. (*A*) Fraction of GFP-expressing Jurkat cells transduced with lentivirus expressing an shRNA targeting GFP under constitutive (white symbols) or inducible (black) promoters, cultured with 100 μM IPTG (gray box) for times indicated. BATF transcript levels (*B*) in anti-CD3/CD28-stimulated shBATF-naive CD8$^+$ T cells cultured in vitro with (dark gray or black bars) or without (light gray) IPTG starting at the day indicated. Cells were continuously exposed to IPTG by in vivo exposure in bone marrow chimeric mice 3 d before T-cell sort (d −3) or 1 d following activation (d +1) and for the remainder of the experiment. (*C*) BATF protein abundance in anti-CD3/CD28-stimulated wild-type and shBATF–naive CD8$^+$ T cells exposed to IPTG d −3 or incubated in medium alone. Numbers represent BATF densiometry values normalized to β-actin in shBATF relative to the wild-type cells. (*D*) Cells treated as in *B* were transferred into recipient mice that were also infected with LCMV and IPTG exposure was maintained by treating mice with 20 mM IPTG in drinking water starting 3 d prior to transfer (in bone marrow chimeras) or 1 d following transfer until 3 d following transfer. *Batf* mRNA level was normalized to *Hprt* and $2^{-\Delta Ct}$ values reported. Significance was assessed with one-way ANOVA; $*P < 0.05$, $***P < 0.001$, $****P < 0.0001$. Representative data are shown from two experiments.

CD8$^+$ T cells also impaired CD8$^+$ effector T-cell development, we adoptively transferred naive P14 CD8$^+$ T cells from bone marrow chimeras transduced with either an inducible shBATF vector or a control shRNA vector targeting LacZ in a 1:1 ratio with naive P14 CD8$^+$ T cells from a bone marrow chimera transduced with a second control shRNA (shRFP) into wild-type recipients (Fig. S5*A*). Endogenous, shBATF– (or control shLacZ–) and shRFP–naive CD8$^+$ T cells were distinguished by the use of congenic markers. Comparison of the ratios of numbers shBATF– or shLacZ–effector T cells to shRFP–effector T cells was used to determine the effect of BATF knockdown, while controlling for any effect of shRNA expression on differentiation. We found markedly reduced numbers of P14 shBATF–effector CD8$^+$ T cells at days 7–9 p.i. relative to shRFP P14 CD8$^+$ T cells, when the cells were exposed to IPTG from d −3 until the end of the experiment. In contrast, the ratio of shLacZ–effectors to shRFP–effectors remained constant (Fig. 4 *A* and *B*). This reduction in shBATF–effector cell numbers was seen with three different BATF shRNAs designed with different seed regions, making this unlikely to be due to off-target effects (Fig. 4*B*) (19).

To identify the reason for the reduced population size of effector CD8$^+$ T cells following BATF knockdown, we measured both cell death and proliferation in shBATF–effector CD8$^+$ T cells at d 5 p.i. Analysis of active caspase abundance showed significantly higher apoptosis in shBATF–effector CD8$^+$ T cells (Fig. 4*C*). In addition, there was a modest increase in the proliferation of the fraction of remaining shBATF–effector CD8$^+$ T cells compared with shLacZ–effector CD8$^+$ T cells (Fig. S5*B*). Thus, knockdown of BATF impairs the development of an effector CD8$^+$ T-cell response primarily by increasing cell death during early differentiation. These findings are consistent with previous studies using germ-line deletion of BATF, which have demonstrated that naive *Batf*$^{-/-}$ T cells undergo massive cell death at 72–96 h after stimulation (11).

BATF Is Required to Initiate but Not Maintain Effector CD8$^+$ T-Cell Development. Because previous studies of the role of BATF in effector CD8$^+$ T-cell differentiation have been carried out using T cells with constitutive germ-line deletion, it is not known whether BATF is required only to initiate the development of CD8$^+$ effector T cells (i.e., at the time of initial antigen encounter) or whether BATF is also needed to maintain CD8$^+$ effector T-cell development once underway. To address this question, we adoptively transferred 1:1 mixtures of congenically distinguishable P14 shBATF– and shLacZ–CD8$^+$ T cells into recipient wild-type animals, which were then infected with LCMV Armstrong. IPTG was administered to induce BATF knockdown either before infection, at the time of infection, or 72 h p.i. (Fig. 5*A*). We assessed BATF knockdown at d 8 p.i. and found that BATF transcript abundance was significantly reduced in shBATF compared with shRFP–effector CD8$^+$ T cells, regardless of when IPTG was initiated (Fig. 5*B*), and was not significantly different between any of the shBATF–effector cell conditions.

We observed profound differences in the ratio of shBATF: shLacZ–CD8$^+$ T cells at d 8 p.i., depending on the time at which BATF knockdown had been initiated. BATF knockdown initiated 3 d before infection or at the time of infection was associated with a significant reduction in the numbers of d 8 p.i. effector CD8$^+$ T cells compared with controls with no IPTG induction. In contrast, inducing BATF knockdown 72 h postinfection did not significantly change the numbers of effector CD8$^+$ T cells d 8 p.i. (Fig. 5*C*). These finding show that, whereas BATF is required for effector CD8$^+$ T-cell development at the time of initial antigen encounter, by 72 h p.i., BATF becomes largely dispensable, at least through d 8 of CD8$^+$ effector T-cell differentiation.

Discussion

We have developed a strategy to inducibly silence gene expression in unperturbed hematopoietic cells in vivo using RNAi. We used this system to show that BATF is initially required for the development of effector CD8$^+$ T cells, but becomes dispensable after 72 h. Our findings suggest that this experimental approach can be used to accelerate the understanding of how effector and memory T-cell responses are regulated.

Viral vectors expressing shRNA molecules have been used to investigate gene function in the immune system (13, 15–17, 20). However, this approach is limited by the inability to deliver viral vectors to quiescent cells. To study gene function in T cells, most approaches have used T-cell activation in vitro (14, 17) or in vivo (15, 16) to achieve efficient transduction. We show that both of these approaches profoundly alter the underlying transcriptional and functional state of naive T cells, suggesting that many aspects of effector differentiation are initiated by these manipulations before gene knockdown can occur. Our experimental system permits the inducible knockdown of genes in quiescent, unperturbed naive T cells, allowing the events that occur during

cytolytic function in naive CD8$^+$ T cells (5, 7, 8). These studies indicate that CD8$^+$ T cells encounter an irreversible decision point within hours of antigen encounter (5). Our findings suggest that transcriptional regulation by BATF may be one component of that decision point. BATF may launch differentiation by irreversibly engaging the effector transcriptional program within the first 24 h of stimulation.

Although we have used this experimental approach to investigate the role of BATF in the early commitment events in effector CD8$^+$ T-cell differentiation, this strategy also could be used to discover genes that regulate differentiation and longevity of memory CD8$^+$ T cells and the mechanisms leading to CD8$^+$ T-cell exhaustion. We also anticipate that pooled in vivo screens of gene function will be possible using this approach. Moreover, the use of a bone marrow chimeric system results in transduction of all hematopoietic-derived lineages with the inducible shRNA vector. This system, therefore, provides a feasible strategy for analyzing gene function in other cell types that are refractory to viral transduction, such as naive CD4$^+$ T cells or B cells.

Methods

Mice. Wild-type C57BL/6J, Ly5.1 (CD45.1), and Thy1.1 mice were purchased from The Jackson Laboratory. The P14 TCR Tg mice were previously described (22). All mice were used according to the Harvard Medical School Standing Committee on Animals and National Institutes of Animal Healthcare Guidelines.

Generation of Bone Marrow Chimeras. LSK cells from bone marrow were enriched using anti-CD117 microbeads (Miltenyi Biotech) and then sorted using a BD FACSAria cytometer. Sorted cells were plated overnight in StemSpan SFEM (Stemcell Technologies) with 100 μg/mL recombinant stem cell factor, thrombopoietin, IL-7, and Flt3-ligand (PeproTech). Cells were then spin infected with lentiviral supernatants at 650 × *g* for 90 min at 37 °C on 100 μg/mL RetroNectin (Takara Bio)-coated plates. Fresh medium was added after 1 h. The following day, the cells were washed in PBS (Gibco) and 50,000 cells were injected i.v. into recipient mice that had been irradiated with two doses of 600 cGy, 3 h apart.

Lentivirus Production. The 293T cells were seeded in DMEM with 10% (vol/vol) FBS. The following day, the cells were transfected with shRNA construct pLKO.1 or 1xLacO (now available from Sigma-Aldrich under the name "MISSION 1X LacO Inducible") and the packaging plasmids Pax2 (gag, pol) and VSV-G using TransIT-LT1 (Mirus Bio) or ExGen 500 transfection reagents (Thermo Scientific Fermentas). Viral supernatants were collected 48–72 h later.

In Vitro Knockdown of GFP. A stable GFP-expressing Jurkat cell line was constructed using PGK-eGFP lentivirus. GFP-Jurkat cells were lentivirus-transduced with shRNA targeting GFP under constitutive (pLKO.1, with puromycin resistance) or inducible (1xLacO, with Thy1.1 reporter) promoters. Varying doses of dioxane-free IPTG (Promega) were added at the indicated concentrations and durations. GFP expression was assessed with Accuri C6 flow cytometer (BD Biosciences).

T-Cell Transfers and Infections. CD8$^+$ T cells were magnetically separated using the CD8a$^+$ T Cell Isolation Kit II (Miltenyi Biotech) and then GFP$^+$ CD8$^+$ congenic cells were sorted using a BD FACSAria cytometer. P14 CD8$^+$ T cells (10^4 to 10^6 cells per animal) were injected into recipient mice i.v. Mice were subsequently infected intraperitoneally with 2×10^5 p.f.u. LCMV Armstrong or with influenza intranasally. For influenza virus infection, the mice were anesthetized with 2.5% avertin and infected with 0.5 LD_{50} H1N1 influenza virus (PR8), engineered to express GP_{33-41} peptide of LCMV (PR8-GP33) (23). Both viruses were a generous gift from E. John Wherry (University of Pennsylvania School of Medicine, Philadelphia).

Flow Cytometry and Cell Sorting. Spleen or bone marrow tissue was homogenized into single cell suspension and resuspended in staining buffer (2 mM EDTA and 1% FBS in PBS; Gibco) together with combinations of the indicated antibodies. Data were acquired using LSR II or Accuri C6 (BD Biosciences) cytometers and analyzed with FlowJo software (v9.7.2; TreeStar).

shRNA Construct Generation. Target sequences of the shRNA used are: shBatf 1 (CCGCAAAGAGATCAAACAGCT), shBatf 2 (CTGGACAAGTATTGAACACAA), shBatf 3 (GAGCTCAAGTACTTCACATCA), shLacZ (CCGTCATAGCGATAACGAGTT), shRFP (GCTTCAAGTGGGAGCGCGTGA), and shGFP (ACAACAGCCACAACGTCTATA). Cloning methods can be found at www.broadinstitute.org/rnai/public/.

Additional Methods. Additional descriptions for all methods are available in *SI Methods*, including antibodies used, infections, immunoblotting, and RT-QPCR.

ACKNOWLEDGMENTS. We thank John G. Doench, E. John Wherry, and the members of the W.N.H. and A.H.S. laboratories for meaningful discussions. This work was supported by the National Institutes of Health Grants AI091493, AI057266, AI082630 (to W.N.H.), AI38310 (to A.H.S.), and Cancer Research Institute Predoctoral Emphasis Pathway in Tumor Immunology (J.G.).

1. Doering TA, et al. (2012) Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory. *Immunity* 37(6):1130–1144.
2. Galon J, et al. (2006) Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 313(5795):1960–1964.
3. Wong P, Pamer EG (2003) CD8 T cell responses to infectious pathogens. *Annu Rev Immunol* 21:29–70.
4. Kaech SM, Hemby S, Kersh E, Ahmed R (2002) Molecular and functional profiling of memory CD8 T cell differentiation. *Cell* 111(6):837–851.
5. Kaech SM, Ahmed R (2001) Memory CD8+ T cell differentiation: Initial antigen encounter triggers a developmental program in naïve cells. *Nat Immunol* 2(5):415–422.
6. Mercado R, et al. (2000) Early programming of T cell populations responding to bacterial infection. *J Immunol* 165(12):6833–6839.
7. van Stipdonk MJ, Lemmens EE, Schoenberger SP (2001) Naïve CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nat Immunol* 2(5):423–429.
8. van Stipdonk MJ, et al. (2003) Dynamic programming of CD8+ T lymphocyte responses. *Nat Immunol* 4(4):361–365.
9. Wong P, Pamer EG (2001) Cutting edge: Antigen-independent CD8 T cell proliferation. *J Immunol* 166(10):5864–5868.
10. Chang JT, et al. (2007) Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science* 315(5819):1687–1691.
11. Kurachi M, et al. (2014) The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. *Nat Immunol* 15(4):373–383.
12. Kaech SM, Cui W (2012) Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol* 12(11):749–761.
13. Amit I, et al. (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326(5950):257–263.
14. Yang L, et al. (2012) miR-146a controls the resolution of T cell responses in mice. *J Exp Med* 209(9):1655–1670.
15. Joshi NS, et al. (2007) Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. *Immunity* 27(2):281–295.
16. Araki K, et al. (2009) mTOR regulates memory CD8 T-cell differentiation. *Nature* 460(7251):108–112.
17. Zhou P, et al. (2014) In vivo discovery of immunotherapy targets in the tumour microenvironment. *Nature* 506(7486):52–57.
18. Chen R, et al. (2014) In vivo RNA interference screens identify regulators of antiviral CD4(+) and CD8(+) T cell differentiation. *Immunity* 41(2):325–338.
19. Jackson AL, et al. (2006) Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* 12(7):1179–1187.
20. Chevrier N, et al. (2011) Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* 147(4):853–867.
21. Ciofani M, et al. (2012) A validated regulatory network for Th17 cell specification. *Cell* 151(2):289–303.
22. Pircher H, Bürki K, Lang R, Hengartner H, Zinkernagel RM (1989) Tolerance induction in double specific T-cell receptor transgenic mice varies with antigen. *Nature* 342(6249):559–561.
23. Mueller SN, et al. (2010) Qualitatively different memory CD8+ T cells are generated after lymphocytic choriomeningitis virus and influenza virus infections. *J Immunol* 185(4):2182–2190.