Systems Immunology Characterization of the Interaction of Immunological Set Points and Disease Outcomes

McCarthy, Elizabeth Edna

2022

Systems Immunology Characterization of the Interaction of Immunological Set Points and Disease Outcomes

by
Elizabeth McCarthy

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biological and Medical Informatics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Matthew Spitzer
4793AE6232BE479...
Matthew Spitzer
Chair

DocuSigned by:
Jimmie Ye
[illegible]
Jimmie Ye

DocuSigned by:
K. Mark Ansel
[illegible]
K. Mark Ansel

DocuSigned by:
Lawrence Fong
91A2D66309BE4D7...
Lawrence Fong

Committee Members

Acknowledgements

The work in this thesis would not have been possible without the support and mentorship of many individuals. I would like to thank my thesis advisors, Jimmie Ye and Matthew Spitzer, for their mentorship and support throughout my PhD studies.

Jimmie welcomed me into his lab as a rotation student with no experience in computational biology. His guidance and the instruction from my rotation mentor, Meena Subramaniam, were essential in the beginning and continuation of my scientific development as a computational biologist. Jimmie has been a champion for team science throughout my graduate career and his instruction in single-cell RNA sequencing data generation and analysis and statistical methods were invaluable.

Matt set me up with a rotation project that sparked my interest in longitudinal human immunology and ultimately led to my first, first-author paper of graduate school! As a mentor, he formalized a system for career check-ins and goal setting through our quarterly meetings which helped to streamline and focus my efforts during my graduate career. His encyclopedic knowledge of immunology fueled my development and confidence in the “immunology” portion of my computational immunology studies.

Both Matt and Jimmie have created supportive, friendly, and collaborative research environments and I am grateful to be scientific colleagues and friends with the members of both labs. I would like to specifically thank the “Ye lab babies”, Cody, Gracie, George, Matt, and Min Cheol, fellow graduate students and now friends who all joined the Ye lab with me. Similarly, I would like to thank the first graduate students in the Ye (Meena and Rachel) and Spitzer (Bree and Kamir) labs whose efforts and mentorship were essential to my success. I would also like to thank the graduate students in the Spitzer lab, Casey and Rachel, who also joined close to when I

did and have both been near-peer mentors in their leadership efforts around diversity and inclusion in IgEquity and ImmunoDiverse. I also want to thank the students who I mentored in their rotations (Tianna and Sophia) - mentors have been and continue to be the driving force behind my accomplishments and I really enjoyed being able to pay it forward by mentoring you during your rotations! And finally thank you to every current and past member of the Spitzer and Ye labs for creating such dynamic and enjoyable research environments.

I would also like to thank the leadership and administrative staff of the UCSF Medical Scientist Training Program and Biological and Medical Informatics Program for their support. I am also deeply grateful to the UCSF Discovery Fellows Program and the NIH Ruth L. Kirschstein NRSA Institutional Predoctoral Fellowship for their financial support. Finally, I would like to thank my thesis committee members, Mark Ansel and Larry Fong, for their sponsorship and scientific and career feedback and mentorship throughout my PhD. I am also so thankful for the team of female science and life mentors I serendipitously assembled, including Rachel Rutishauser, Judy Ashouri, Pam Odirozzi, and Bridget Keenan, who have been and continue to be incredible role-models as scientists/physician-scientists, career mentors, and friends.

They say "It takes a village to raise a child" and it also takes one to “raise” a graduate student. My village rests on my family, starting with my parents, Cathy and Frank, who I thank for instilling a sense of curiosity and academic rigor that has propelled me to completing this thesis. I would like to thank my grandparents, Grandma Jean, and Grandpa Jim and MJ for loving me and cheerleading for me in my academic pursuits from preschool to now (grade 22!). I would also like to thank all of my siblings and their partners, Frank and Leela, Amy and Michael, Dannielle and Spencer, Esther and Will, and Sarah and Gus. You all have been so supportive and loving throughout my life and especially during the ups and downs in my PhD. I have to give a special shout-out to Dannielle, of course, who is my #littlesisterbutbigmentor due to the support and feedback that she has given me during my entire PhD but especially during the last intensive writing portion!

My fiancé, Jonathan, is my source of strength, support, and humor. Whether he is listening to a practice talk for the second or third time or making me dinner while I am working late, he always works to be supportive even though he has his own busy career as a psychiatry resident. I am so thankful to you, Jonathan, for being my partner and I am so excited about the life we are building together. Jane, Mitch, and Nina and Adam, you are like a second nuclear family to me and I am so grateful for how you have welcomed me and also supported me and celebrated with me throughout my PhD!

I also have to give a big-shout out to my “niblings”, Sophia, Isaac, and Zoe - your joy, wonder, and fun are infectious!

My village extends to friends in lab and also outside of lab named here that I met during college (Nat, Neil, Lukas, Hannah Laura, Eva, AJ, Esther, and Alyssa), medical and graduate school (Bill, Carson, Dillon, McKenna, Cody, Kate, Hannah, Paul, and Camille) and too many others to name but who are all so loved and appreciated in my heart.

I also want to ask for a moment of silence for my loved ones who are no longer with us but whose life, and now memory, are my guiding forces: my late Grandmother, Milele Chikasa Anana, and my late friends, David Cline, Griffin Madden, and Max Feinstein.

Finally, I want to thank my graduate school communities of my graduate school classmates in BE-STEM, my MSTP cohort, and my iPQB graduate cohort - the people who are walking next to you understand your challenges and triumphs better than anyone and I am so grateful to have been on this academic and life journey with such an amazing group of individuals.

Contributions

The chapter entitled “A cytotoxic-skewed immune set point predicts low neutralizing antibody levels after Zika virus infection” was published in Cell Reports (PMID: 35584677, doi:10.1016/j.celrep.2022.110815).

Beauty isn’t all about just nice, loveliness like. Beauty is about a more rounded substantial becoming. Beauty in that sense is about an emerging fullness, a greater sense of grace and elegance, a deeper sense of depth, and also a kind of homecoming for the enriched memory of your unfolding life.

- John O’Donohue

Systems Immunology Characterization of the Interaction of Immunological Set Points and Disease Outcomes

Elizabeth McCarthy

Abstract

Systems immunology characterizes how immune cell types interact within different contexts to lead to immunological outcomes. Several of the tools of systems immunology are drawn from the single-cell -omics world, fittingly since the first single-cell technology, flow cytometry, was and is primarily used in immunology to phenotype diverse immune cell types. Here, we present the use of single-cell transcriptomics and proteomics to take a systems immunology approach to study how immune set points, created by genetic and environmental exposures, affect individual immune responses in three major classes of disease, autoimmune (rheumatoid arthritis), infection (Zika virus), and cancer (metastatic prostate cancer).

How pathogenic CD4 T cells develop to cause autoimmunity remains unknown. In the SKG autoimmune arthritis mouse model, we profiled arthritogenic naïve CD4 T cells by bulk and single cell RNA and T cell antigen receptor (TCR) sequencing prior to arthritis onset. Our analyses reveal that despite impaired proximal TCR signaling, a subset of SKG naïve CD4 T cells that have most recently encountered antigen more highly express gene programs associated with positive regulation of T cell activation and cytokine signaling compared to wild type cells. These cells also induce genes associated with negative regulation of T cell activation but for a

subset of tolerogenic markers (e.g., *Izumo1r, Tnfrsf9, Bach2, Eomes, Tigit, Tox, Tox2*) do so at lesser amounts than wild type cells. Furthermore, their TCR sequences exhibit a previously unrecognized bias towards Vbs that recognize superantigen from endogenous retrovirus (ERV) mouse mammary tumor virus (MMTV). In arthritic joints, these biased Vbs are further expanded and ERVs are readily detected. Inhibition of viral reverse transcription significantly reduced SKG arthritis development. Together, our results suggest that endogenous viral products promote autoreactive naïve CD4 T cells which recognize endogenous viral superantigens to break tolerance via changes to their transcriptome and activation state.

Although generating high neutralizing antibody levels is a key component of protective immunity after acute viral infection or vaccination, little is known about why some individuals generate high versus low neutralizing antibody titers. Here, we leverage the high-dimensional single-cell profiling capacity of mass cytometry to characterize the longitudinal cellular immune response to Zika virus (ZIKV) infection in viremic human blood donors in Puerto Rico. During acute ZIKV infection, we identify widely coordinated responses across innate and adaptive immune cell lineages. High frequencies of multiple activated cell types during acute infection are associated with high titers of ZIKV neutralizing antibodies 6 months post-infection, while stable immune features suggesting a cytotoxic-skewed immune set point are associated with low titers. Our study offers insight into the coordination of immune responses and identifies candidate cellular biomarkers that may offer predictive value in vaccine efficacy trials aimed at inducing high levels of antiviral neutralizing antibodies.

Cancer immunotherapy has been a revolutionary anti-tumor treatment, but in prostate cancers, and other solid tumors, the response has been limited. Sipuleucel-T, an autologous antigen-presenting cell vaccine involving *ex vivo* peptide stimulation, is the only approved

immunotherapy for advanced prostate cancer. Understanding the immunogenic and tolerogenic myeloid cell states in prostate cancer could improve the limited immunotherapy response. We used a genetic multiplexing strategy to simultaneously profile gene and protein expression on single cells from ~400,000 peripheral blood mononuclear cells (PBMCs) from longitudinal sampling of a metastatic castration-resistant prostate cancer (mCRPC) human cohort receiving combined immunotherapy (sipuleucel-T and ipilimumab). We identified co-expressed chronic interferon and complement gene signatures in the peripheral CD14+ myeloid compartment which predicted immunotherapy resistance. In contrast, the responders had higher frequency of progenitor exhausted CD8+ T cells (T_{pex}) suggesting a permissive myeloid environment requires a productive T cell response for anti-tumor activity. Future trials could use low co-expression of the CD14+ myeloid chronic interferon and complement gene signatures as a biomarker to select participants who are more likely to respond to therapy and could specifically target this myeloid state to potentially improve response to immunotherapy.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
APCs	antigen presenting cells
ASC	antibody secreting cells
AUC	area under the curve
CITE-seq	cellular indexing of transcriptomes and epitopes by sequencing
CyTOF	mass cytometry by time-of-flight
DEGs	differentially expressed genes
DENV	dengue virus
ERV	endogenous retrovirus
FR4	folate receptor 4
GM-CSF	granulocyte macrophage colony-stimulating factor
GSEA	gene set enrichment analysis
HVGs	highly variable genes
ICIs	immune checkpoint receptor inhibitors
IFN	interferon
IL-1	interleukin 1
IL-17	interleukin 17
IL-6	interleukin 6
IRDS	interferon-related DNA damage response
ISGs	interferon signaling genes
M-MDSC	monocytic-MDSC
mCRPC	metastatic castration-resistant prostate cancer

MDSCs	myeloid derived suppressor cells
MMRD	mismatch repair deficiency
MMTV	mouse mammary tumor virus
MSI	microsatellite instability
OR	odds ratio
PAP	prostatic acid phosphatase
PBMCs	peripheral blood mononuclear cells
PCI	prostate cancer immunotherapy
PSA	prostate-specific antigen
RA	rheumatoid arthritis
RBC	red blood cell
RNA	ribonucleic acid
ROC	receiver operating characteristic
Sag	superantigen
scRNA-seq	single-cell RNA sequencing
SOCS3	suppressor of cytokine signaling 3
TCR	T cell receptor
Tpex	progenitor exhausted CD8+ T cells
TRBV	TCR Beta Variable Chain
UMAP	uniform manifold approximation and projection
Vb	variable beta
WT	wild-type
ZIKV	Zika virus

Chapter 1: Introduction

Systems immunology utilizes a variety of technologies to generate large-scale datasets to gain insights about the cellular states and interactions within the immune system using a combination of bioinformatics tools and immunological expertise (1). While the cells that make up the immune system have complex individual phenotypes and functions, the cell-cell interactions and distribution of the total immune cell milieu across individual states, which can be captured by a systems immunology approach, are critical to understanding immune system function and dysfunction. Single-cell -omics are part of the bedrock of systems immunology tools.

Many of the datasets generated by single-cell -omics technologies use a similar analytical pipeline. Here we focus on technologies used to generate single cell proteomics (mass cytometry by time-of-flight (CyTOF) (2)), single-cell transcriptomics (single-cell RNA sequencing (scRNA-seq) (3) and single-cell T cell receptor sequencing (scTCR-seq) (4)), and their combination (cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)) (5). Each experimental technology requires data quality control, normalization, and other pre-processing steps (6,7). These single-cell data can be visualized with numerous dimensionality reduction algorithms (8,9), and clustering algorithms (10,11) are commonly applied to partition cells that share similar multidimensional expression profiles. Statistical methods for differential abundances (12,13) are commonly applied to identify biological differences across experimental conditions or groups of samples, as are methods to identify differential expression (14) of the quantified cellular feature (e.g., protein, RNA, etc.) within clusters of interest. Trajectory inference algorithms (15) can reconstruct differentiation or activation processes, leveraging the single-cell nature of these data. For scTCR-seq, analytical techniques (16) which can capture unequal distribution of clonotypes across cells such as Gini coefficient or Shannon entropy are used to measure clonal restriction across cells within a sample.

A fundamental question in immunology is what features and interactions within the immune system determine a successful or unsuccessful response to an immunological perturbation (e.g., infection, malignancy, etc.). In many contexts, immune set points have emerged as key factors that can determine immunological responses before a perturbation. Immune set points are determined by the interaction of intrinsic host factors (e.g., microbiome, genetics, etc.) and the extrinsic environmental factors (e.g., previous infections, carcinogen exposure, etc.). For example, the level of regulatory T cells was shown to be a tunable immune set point that could be used to decrease the risk of the development of autoimmune disease in mouse models (17). Additionally, the neonatal microbiome composition has been shown to alter the risk of developing childhood asthma (18). Immune set points can also affect the response to immune-mediated treatments. Indeed, a higher risk for Th17-mediated skin autoimmune disease is associated with response to checkpoint receptor inhibitors in bladder cancer (19). Characterizing how immune set points can affect the development of disease and the response to treatment can reveal important biomarkers to predict disease risk or to guide treatment selection and can also suggest potential mechanisms for disease pathogenesis.

In chapter 2, we investigate how genetic factors, namely a hypomorphic mutation in Zap70 which impairs proximal TCR signaling in the SKG mouse, create an arthritogenic immune set point. We used bulk RNA sequencing and single-cell RNA and TCR sequencing to investigate the naïve CD4 T cell compartment before arthritis induction. We found a defect in the induction of peripheral tolerance and a TCR Beta Variable chain (TRBV) bias towards an autoreactive TCR repertoire in the SKG mice compared to wild-type (WT) mice.

In chapters 3 and 4, we shift focus from mouse models to human immunology. In chapter 3, we explored how immune set points affect the maintained levels of long-term neutralizing antibodies after a viral infection. We profiled peripheral blood mononuclear cells (PBMCs) with CyTOF that were longitudinally sampled from participants who were infected with Zika virus (ZIKV). A cytotoxic-skewed immune set point expressed during acute infection and after the resolution of infection that predicted lower levels of ZIKV neutralizing antibodies 6 months post-infection.

In chapter 4, we asked if an immune set point could predict resistance to immunotherapy in metastatic castration-resistant prostate cancer (mCRPC). We used multiplexed CITE-seq to profile longitudinal samples from clinical trial participants receiving combined immunotherapy of sipuleucel-T, an antigen presenting vaccine, and ipilimumab, a checkpoint receptor inhibitor, for mCRPC. We found a co-expressed chronic interferon and complement signature in the peripheral CD14+ myeloid compartment which predicted resistance to immunotherapy which could potentially inhibit the progenitor exhausted (T_{pex})-like CD8+ T cell population which was enriched in responders.

References

1. Davis MM, Tato CM, Furman D. Systems immunology: just getting started. Nature Immunology. 2017;18(7):725–32.
2. Spitzer MH, Nolan GP. Mass Cytometry: Single Cells, Many Features. Cell. 2016;165(4):780–91.
3. Haque A, Engel J, Teichmann SA, Lönnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. 2017;9(1):75.
4. Pai JA, Satpathy AT. High-throughput and single-cell T cell receptor sequencing technologies. Nat Methods. 2021;18(8):881–92.
5. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017;14(9):865–8.
6. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15(6):e8746.
7. Rybakowska P, Alarcón-Riquelme ME, Marañón C. Key steps and methods in the experimental design and data analysis of highly multi-parametric flow and mass cytometry. Comput Struct Biotechnol J. 2020;18:874–86.
8. Wang K, Yang Y, Wu F, Song B, Wang X, Wang T. Comparative Analysis of Dimension Reduction Methods for Cytometry by Time-of-Flight Data [Internet]. bioRxiv [Preprint]. 2022 [cited 2022 Dec 06]: 26 p. Available from: https://doi.org/10.1101/2022.04.26.489549
9. Xiang R, Wang W, Yang L, Wang S, Xu C, Chen X. A Comparison for Dimensionality Reduction Methods of Single-Cell RNA-seq Data. Front Genet. 2021;12:646936.

10. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. Cytometry A. 2016;89(12):1084–96.
11. Yu L, Cao Y, Yang JYH, Yang P. Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. Genome Biol. 2022;23(1):49.
12. Dann E, Henderson NC, Teichmann SA, Morgan MD, Marioni JC. Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat Biotechnol. 2022;40(2):245–53.
13. Weber LM, Nowicka M, Soneson C, Robinson MD. diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. Commun Biol. 2019;2:183.
14. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021 Sep 28;12(1):5692.
15. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019;37(5):547–54.
16. Chiffelle J, Genolet R, Perez MA, Coukos G, Zoete V, Harari A. T-cell repertoire analysis and metrics of diversity and clonality. Curr Opin Biotechnol. 2020;65:284–95.
17. Hsieh CS, Bautista JL. Sliding set-points of immune responses for therapy of autoimmunity. J Exp Med. 2010;207(9):1819–23.
18. Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosh D, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. Nat Med. 2016;22(10):1187–91.

19. Khan Z, Di Nucci F, Kwan A, Hammer C, Mariathasan S, Rouilly V, et al. Polygenic risk for skin autoimmunity impacts immune checkpoint blockade in bladder cancer. PNAS. 2020;117(22):12288–94.

Chapter 2: Naïve Arthritogenic SKG T Cells have a Defect in Peripheral Tolerance and a Repertoire Pruned by Superantigen

Judith F. Ashouri[1]†*, Elizabeth E. McCarthy[1,2,3]†, Steven Yu[1], Noah Perlmutter[1], Charles Lin[1,4], Joe DeRisi[5,6,7], Chun Jimmie Ye[1,3,6,7,]*, Arthur Weiss[1]*

[1]Rosalind Russell and Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, UCSF; San Francisco, CA.

[2]Department of Epidemiology and Biostatistics, UCSF; San Francisco, CA.

[3]Institute for Human Genetics, UCSF; San Francisco, CA.

[4]Stanford University; Palo Alto, CA.

[5]Department of Biochemistry and Biophysics, UCSF; San Francisco, CA.

[6]Chan Zuckerberg Biohub; San Francisco, CA.

[7]Bakar Computational Health Sciences Institute, UCSF; San Francisco, CA.

†Contributed equally

*Co-corresponding authors

Introduction

Activation of conventional CD4 T cells that recognize specific self-antigen(s) is thought to be necessary for rheumatoid arthritis (RA) onset (1). A paradoxical observation from patients with RA is that CD4 T cells can differentiate into pathogenic effector cells despite impaired TCR signaling (2-8). Yet, how these T cells subvert tolerance to cause disease remains incompletely understood. T cell-intrinsic mechanisms that operate during thymic development (negative selection of self-reactive cells) and in peripheral T cells (functional unresponsiveness or ‘anergy’) are essential to maintain tolerance to self.

The SKG mouse model derived from the BALB/c genetic background is a powerful tool to define how defects in tolerance contribute to arthritis. Due to a hypomorphic mutation in ZAP70, a cytoplasmic tyrosine kinase critical for proximal TCR signaling, SKG mice exhibit impaired thymocyte negative selection resulting in a break in central tolerance and escape of self-reactive CD4 T cells into the periphery (9-12). In response to an innate immune stimulus, arthritogenic CD4 T cells that are otherwise dormant become activated resulting in erosive inflammatory arthritis that resembles RA (9, 13). SKG CD4 T cells are sufficient and necessary to cause arthritis (9), and we have shown that adoptive transfer of even naïve SKG CD4 T cells (into immunodeficient hosts) are sufficient to trigger disease (10). While the SKG mice have a known defect in central tolerance, it is unclear how SKG T cells subvert peripheral tolerance to differentiate into pathogenic effector cells that cause frank disease in the setting of severely impaired TCR signaling.

To address this question, we previously developed the SKGNur mouse which combines the SKG model with a reporter of TCR signaling, Nur77/*Nr4a1*-eGFP, that tethers GFP expression to the regulatory region of *Nr4a1* (encoding the orphan nuclear hormone receptor

Nur77). Because NR4A1 is rapidly and selectively upregulated in response to antigen but not inflammatory stimuli (14, 15), the SKGNur mouse allows us to identify antigen-activated T cells before and during disease. We previously demonstrated that high Nur77-eGFP expression in SKGNur mice identifies self-reactive naïve CD4 T cells before disease onset with greater potential to cause arthritis, in part, because of abnormally heightened responses to interleukin 6 (IL-6) (10) (**Supplementary Fig. 2.1A**). We proposed that chronic antigen stimulation of peripheral SKG CD4 T cells results in lower levels of suppressor of cytokine signaling 3 (SOCS3)—a key negative regulator of IL-6 signaling, and showed that this mechanism may operate in patients with RA (10). This led us to hypothesize that SKG CD4 T cells might exhibit dysregulated expression of a broader gene program of negative immune regulators rendering them more susceptible to stimuli that could lead to a breach in peripheral tolerance.

To test this hypothesis, we studied the TCR repertoire and transcriptome of arthritogenic naïve CD4 T cells in SKG mice by performing both bulk and single cell RNA sequencing. We capitalized on the SKGNur model in order to capture arthritogenic (SKGNur GFP^{hi}) cells before disease onset (akin to the pre-RA phase of disease (16)). We reasoned this could reveal early events in pathogenesis and identify novel targets to preserve tolerance and prevent disease. Within arthritogenic naïve CD4 T cells, we found a cluster of cells marked by high *Nr4a1* expression that upregulate gene programs associated with TCR signaling in response to antigen-encounter. Though *Nr4a1* high expressing SKG CD4 T cells also upregulate tolerogenic programs, we identified a defect in the extent to which they do so relative to WT CD4 T cells. Furthermore, TCR sequencing of the arthritogenic CD4 T cells revealed an enrichment of variable beta (Vb) chains that recognize superantigen (Sag) from mouse mammary tumor virus (MMTV), an endogenous retrovirus (ERV), in BALB/c mice. We had previously shown that these Vbs escape negative selection in SKG mice (12). Here we find peripheral T cells bearing these Vbs are strongly associated with an activated TCR signaling program. We confirmed enrichment of these *TRBV*s in arthritogenic SKG naïve CD4 T cells by TCR Vb protein expression. Moreover, the frequency of CD4 T cells bearing Vbs that recognize MMTV Sag are expanded in the arthritic joints of SKG mice and may contribute to development and/or severity of arthritis. Indeed, we find that antiretrovirals shown to suppress ERV reverse transcriptase and MMTV viral load (17, 18) significantly ameliorate and delay SKG arthritis onset. Our results reveal that self-reactive SKG T cells which escape negative selection harbor an independent defect in peripheral tolerance which, together with chronic antigen stimulation, sets the stage for disease. Moreover, we propose a novel role for endogenous MMTV Sag in promoting arthritogenic T cell responses.

Results

Arthritogenic SKG Naïve CD4 T Cells Display a Signature of TCR Activation

We recently demonstrated that in the SKGNur model, it is possible to identify arthritogenic naïve CD4 T cells prior to disease onset on the basis of Nur77-eGFP expression (10). To characterize the transcriptional program of the arthritogenic CD4 T cells (the SKGNur GFP^{hi} population) prior to disease onset, we performed bulk RNA-sequencing on naïve (CD62L^{hi}CD44^{lo}CD25^{-}) CD4 T cells with the highest (GFP^{hi}) or lowest GFP expression (GFP^{lo}) from SKG and wild-type (SKGNur and WTNur) mice (**Fig. 2.1A, Supplementary Fig. 2.1B**). Principal component (PC) analysis revealed all four subgroups are transcriptionally distinct with the component that explains the largest amount (57%) of variance separating samples by GFP expression followed by the second component, which explains 22% of variance, separating samples by genotype (**Fig. 2.1B**). Hierarchical clustering of the 991 differentially expressed genes (DEGs) between subgroups identified six gene modules (**Fig. 2.1C**). Gene ontology analysis (19) revealed functional heterogeneity between, and in some cases within, these modules (**Fig. 2.1D**). To further dissect the transcriptomic differences between WTNur GFP^{hi} and SKGNur GFP^{hi} cells, we focused on the 260 DEGs between these two subgroups (**Supplemental Fig. 2.1C-D**). Cell cycle genes (e.g., *Cdca3, Cdk2nc*, *Mki67*, primarily represented in module 1) were more highly expressed in the SKGNur GFP^{hi} cells. SKGNur GFP^{hi} cells were also found to have increased expression of genes (e.g., *Socs1, Tnfsf14, Il2ra, H2-Aa, H2-Ab1* represented in module 6) and pathways associated with cytokine signaling and antigen processing compared to WTNur GFP^{hi} cells (**Fig. 2.1C-E**). This signature appears to be independent of GFP expression in SKG cells (**Fig. 2.1C**).

Despite their hypomorphic *Zap70* allele and impaired proximal TCR signaling (9, 10, 12), SKGNur GFP^{hi} cells upregulate TCR signaling response genes, both positive (*Egr1*, *Id3, Icos, Irf4, Tnfrsf9, Tnfrsf4, Myb*) and negative regulators (*Nr4a1, Nr4a3, Cd5, Folr4/Izumo1r, Tigit, Tox, Pdcd1, Lag3, Ctla4, Birc5, Nrp1*) of TCR signaling (found in modules 1 and 2, **Fig. 2.1C-D**) compared to GFP^{lo} cells. Paradoxically, SKGNur GFP^{hi} CD4 T cells have higher expression of pathways associated with T cell activation and signaling compared to WTNur GFP^{hi} cells (**Fig. 2.1E**). Additionally, we found that genes associated with tolerogenic programs, and induced by TCR signaling, are broadly upregulated in SKGNur and WTNur GFP^{hi} cells, though a subset of tolerogenic genes are not as highly expressed in SKGNur GFP^{hi} T cells compared to WT (e.g., *Folr4/Izumo1r, Tigit, Tox, Lag3*). The lower expression of *Folr4*—which is also known as *Izumo1r and* encodes Folate Receptor 4 (FR4), a specific marker of anergic cells—suggests that perhaps SKGNur GFP^{hi} CD4 T cells may sub optimally induce anergy and/or other tolerogenic programs. (**Fig. 2.1C**). Our results indicate SKGNur GFP^{hi} CD4 T cells upregulate TCR activation and cytokine signaling signatures likely in response to chronic antigen encounter but may inefficiently induce tolerogenic programs to restrain these responses.

An Enhanced TCR Signaling Program Defines a Subset of SKG GFP^{hi} Naïve CD4 T Cells

The long half-life of eGFP (20) compared to the more dynamic turnover of NUR77/*Nr4a1* protein and transcript (21-23) (**Supplementary Fig. 2.2A-C**) means GFP^{hi} cells likely consist of mixed populations of more and less recently stimulated cells. To overcome this limitation and to understand whether the TCR signaling and effector cytokine gene modules we identified were uniformly or heterogeneously activated in GFP^{hi} cells, we performed paired single-cell RNA and TCR sequencing (scRNA- and scTCR-seq) on GFP^{hi} and GFP^{lo} naïve CD4 T cells from SKGNur and WTNur mice (**Fig. 2.2A**). In our scRNA-seq dataset, we identify eight distinct clusters (**Fig. 2.2B-C**) which recapitulate and further refine our bulk RNA-seq gene signatures (**Fig. 2.1**) and demonstrate pronounced heterogeneity within the GFP^{hi} population (**Fig. 2.2D**). While GFP^{hi} CD4 T cells from both SKGNur and WTNur mice were present across all eight clusters, there was clear enrichment in the T.N4$_{Nr4a1}$ cluster compared to GFP^{lo} CD4 T cells by a mean of > 4-fold (**Supplementary Fig. 2.2E-F**). Cluster T.4N$_{Nr4a1}$ had the highest expression of *Nr4a1* and *eGfp* (**Fig. 2.2C, Supplementary Fig. 2.2F**). GFP^{hi} T cells were also enriched in the T.N4$_{Izumo1r\_Id2}$ and, albeit to a lesser extent, the Cyto$_{Nkg7}$ clusters (**Fig. 2.2D, Supplementary Fig. 2.2D-F**).

Given the specificity of NR4A1 (NUR77) as a reporter of TCR signaling (24), the high expression of *Nr4a1* in the T.N4$_{Nr4a1}$ cluster signifies that these cells most recently encountered endogenous antigen(s) (25, 26). Indeed, T.N4$_{Nr4a1}$ cells overwhelmingly up-regulate genes associated with TCR signal transduction *(*e.g., *Nr4a1, Nr4a3, Egr1-3, Tnfrsf9, Tnfrsf4, Ifr4*, *Cd69*, **Fig. 2.2E**) which most closely mirror TCR signaling genes found in module 2 of our bulk

RNA-seq analysis (**Supplementary Fig. 2.2G**). In fact, several of the TCR response genes highly expressed in cluster T.N4$_{Nr4a1}$ have been identified as targets of the LAT-PLCg-HDAC7 pathway (e.g., *Nr4a1, Egr1-3, Irf4*) and correlate with strength of tonic signaling (25, 26). These TCR signaling signatures are more highly expressed in SKGNur GFP^{hi} CD4 T cells in the T.N4$_{Nr4a1}$ cluster compared to those of WT (**Fig. 2.2F-G**), as observed in our bulk RNA-seq dataset. This suggests the signaling impaired SKGNur GFP^{hi} cells in cluster T.N4$_{Nr4a1}$ have, nonetheless, responded more strongly to encountered antigen than WTNur GFP^{hi} cells.

Antigen-activated T cells Upregulate a Tolerogenic Transcriptional Program

We next investigated additional T cell transcriptomic signatures that could further illuminate how SKGNur GFP^{hi} CD4 T cells in the setting of chronic antigen encounter may be more able to escape tolerance and differentiate into pathogenic effector cells. Therefore, we examined the expression of candidate genes associated with tolerance programs (27-32) within cells that have most recently encountered antigen. We found GFP^{hi} CD4 T cells in cluster T.N4$_{Nr4a1}$, and to a lesser extent in the overall dataset, upregulated genes associated with tolerogenic programs in both WTNur and SKGNur GFP^{hi} subgroups (including *Egr2, Izumo1r, Pdcd1lg2, Pdcd1, Lag3, Tigit, Tox*) (**Fig. 2.2H, Supplementary Fig. 2.2H**). This likely reflects the triggering of a negative regulatory program in naïve CD4 T cells in response to tonic TCR signaling (10, 33, 34) driven in part by NR4A family members that have been shown to play negative regulatory roles in peripheral T cells (22, 32, 35, 36).

However, several of these TCR negative regulators are less efficiently induced in SKGNur GFP^{hi} cells compared to WTNur GFP^{hi} cells (e.g., *Tigit, Izumo1r, Eomes, Tox, Tnfrsf9, Tox2, Bach2*) (**Fig. 2.2H**). Therefore, in addition to a known loss in central tolerance, SKG mice likely have an independent defect in mechanisms maintaining peripheral tolerance. This defect is likely derived from their impaired proximal TCR signaling capacity and may explain the reduced frequency of anergic peripheral CD4 T cells we previously observed in SKGNur mice (10).

SKG's Hyperresponsiveness to IL-6 is Pre-programmed Transcriptionally

IL-6 production in SKG mice is indispensable for SKG arthritis development (37, 38). Recognition of major histocompatibility complex/self-peptide complexes, stimulate antigen-presenting cells (APCs) to secrete IL-6 (38). We previously found that SKGNur GFP^{hi} T cells were more responsive to IL-6 and more readily produced IL-17 in the most autoreactive T cells, in part due to lower levels of SOCS3 (suppressor of cytokine signaling 3) – a critical negative regulator of IL-6 (10) (**Supplementary Fig. 2.1A**). In our current study, we found that genes associated with IL-6 signaling machinery and the Th17 pathway were uniquely enriched in SKGNur GFP^{hi} T cells (39) in the T.N4$_{Nr4a1}$ cluster (**Supplementary Fig. 2.2I**).

SOCS3 is downregulated in naïve CD4 T cells in response to antigen (40) and in patients with RA (10, 41). Its expression has a strong inverse correlation with murine arthritis severity (42-44). Therefore, we examined the expression of SOCS family members in our single cell dataset. Of these family members, *Socs3* was specifically downregulated in SKGNur GFP^{hi} cells within the T.N4$_{Nr4a1}$ cluster (**Supplementary Fig. 2.2I**). Moreover, we found a striking inverse correlation between the expression of *Nr4a1* and *Socs3* (**Fig. 2.2I**), corresponding to a published report (40). The inverse correlation between *Nr4a1* and *Socs3* expression in SKGNur GFP^{hi} T cells within the T.N4$_{Nr4a1}$ cluster provides orthogonal validation of our previous results. It highlights the interdependence between signaling via the TCR and heightened sensitivity to cytokines such as IL-6 (10).

T.N4$_{Nr4a1}$ Cells Segregate into Two Distinct TCR Signaling Modules

To explore heterogeneity within cluster T.N4$_{Nr4a1}$, we performed co-expression analysis with the highly variable genes (HVGs) and *Nr4a1*. We identified three distinct modules of HVGs that positively correlated with *Nr4a1* (**Fig. 2.3A**). Genes from two of these modules, *Egr* family members (immediate early gene transcription factors) and *Tnfrsf9* (4-1BB – the TCR inducible co-stimulatory receptor), uniquely identified distinct subsets of cells within the T.N4$_{Nr4a1}$ cluster (**Fig. 2.3B, Supplementary Fig. 2.3A**). The *Egr* module contained additional immediate early genes or markers of early T cell activation (e.g., *Egr1, Egr2, Cd69, Ier2, Egr3, Nfkbid, Junb, Fos, Myc, Cd40lg*), whereas the *Tnfrsf9* module correlated with markers that upregulate in response to prolonged TCR signaling (e.g., *Pou2f2, Myb, Tnfrsf4, Lag3*) (**Fig. 2.3C, Supplementary Fig. 2.3B**).

Cells that expressed *Egr2* had increased expression of genes enriched in pathways induced early after TCR stimulation (0.5h and 1h), whereas*Tnfrsf9* expressing cells had increased expression of genes enriched in pathways upregulated after prolonged TCR stimulation (24h and 72h, **Fig. 2.3D**). Our findings suggest the T.N4$_{Nr4a1}$ cluster contains subclusters driven by their TCR signaling kinetics (early vs prolonged stimulation). Interestingly, the tolerogenic genes that are inefficiently induced in SKGNur GFP^{hi} T cells within the T.N4$_{Nr4a1}$ cluster compared to those of WT are also lower in the subset of the *Tnfrsf9* high SKGNur GFP^{hi} T cells, which seem to have undergone prolonged stimulation, compared to the corresponding WT subset (**Fig. 2.3E**). This further suggested a defect in the upregulation of tolerogenic gene programs in SKGNur GFP^{hi} CD4 T cells in response to prolonged TCR stimulation rather than simply a compositional shift of SKGNur GFP^{hi} T cells away from the prolonged TCR stimulation state.

To examine other covariates which could lead to segregation of early versus prolonged antigen-activated T cells, we performed cell-cycle analysis on our dataset. Though cell-cycle appeared to contribute somewhat to the heterogeneity among the T.N4$_{Nr4a1}$ cluster, as *Tnfrsf9* expressing cells were more likely to be in S phase (**Supplementary Fig. 2.3C**), it did not fully account for the division between *Egr* family member and *Tnfrsf9* expression.

Cell States and Trajectories of T.4N$_{Nr4a1}$ Cells have a Distinct Distribution in SKGNur GFP^{hi} Subset

We next asked if the early versus prolonged TCR signaling states in the T.N4$_{Nr4a1}$ cluster represented endpoints of a trajectory. We discovered a continuum of cell states in the T.N4$_{Nr4a1}$ cluster using latent time based on RNA velocity (45) (**Fig. 2.4A**). The expression of the *Egr* family peaks in earlier latent time cells, while the expression of *Tnfrsf9* and associated genes peak in later latent time cells (**Fig. 2.4B-C**).

We separated the cells into four cell states labeled “Stage 1” to “Stage 4” from earlier to later latent time (**Fig. 2.4D, Supplementary Fig. 2.3D-E**). The RNA velocity vector field (**Supplementary Fig. 2.3F**) and trajectory inference analysis (46) supported a trajectory from Stage 1 to Stage 4 (**Fig. 2.4E**). The expression of *Egr2* and *Nr4a1* peak within cells from Stage 1 while the expression of *Tnfrsf9* peaks within cells from Stage 4 (**Fig. 2.4F**), and the genes overexpressed in Stage 1 and Stage 4 cells are enriched for early or prolonged TCR stimulation pathways, respectively (**Supplementary Fig. 2.3G**). Thus, these cell states seem to be the endpoints of a trajectory of cell states from early to prolonged TCR stimulation.

Cells from SKGNur GFP^{hi} and WTNur GFP^{hi} groups had significantly different distributions across latent time with a higher density at earlier latent time for the SKGNur GFP^{hi} cells, which also had an increased odds of being in Stage 1 versus Stage 4 compared to WTNur GFP^{hi} cells

(OR = 1.25, p = 0.02). No significant difference was observed between the GFP^{lo} subgroups (**Fig. 2.4G-H**).

We hypothesized this imbalance may result from slower progression of SKG cells through the stages or higher input into, due to higher proliferation, Stage 1 in the SKG pool. The cells from SKGNur GFP^{hi} and WTNur GFP^{hi} within stage 1 did not have a significant difference in their distribution across cell cycle (p = 0.24) suggesting the SKGNur GFP^{hi} cells in stage 1 do not have higher proliferation. While this result favors our slower progression hypothesis, the two hypotheses are not mutually exclusive. Slower progression of the SKGNur GFP^{hi} cells would suggest that SKG CD4 T cells have a defect in peripheral tolerance induction—a program which is upregulated as the cells progress through the stages—and could explain the reduced frequency of anergic cells observed in SKG mice (10).

Naïve SKGNur GFP^{hi} CD4 T Cells Demonstrate a Biased TCR Beta Variable Gene (TRBV) Repertoire

We have previously shown that the SKGNur GFP^{hi} cells have increased self-reactivity and ability to proliferate in response to an undefined endogenous antigen(s) (10). Thus, we asked how their TCR repertoire might contribute to their activation in the periphery. We examined the TCR repertoire using scTCR-seq (**Fig. 2.2A**) and detected paired TCR α (TRA) and TCR β (TRB) genes in 86% of cells (**Supplementary Fig. 2.4A**). Using Gini coefficient analysis, we did not find oligoclonal expansion in the naïve T cells in any of the samples, including all the SKGNur samples. Instead, we found that SKGNur GFP^{hi} cells T cells demonstrate a biased TCR variable β gene (*TRBV*) usage, but not TCR variable α gene (*TRAV*) usage (**Fig. 2.5A-C, Supplementary Fig. 2.4B**). In SKGNur GFP^{hi} CD4 T cells compared to the paired SKGNur GFP^{lo} samples, we found significantly higher (FDR < 0.1) usage of *TRBV26* (corresponding to TCR variable beta 3, Vb3, protein), *TRBV12-1* (Vb5), *TRBV15* (Vb12), *TRBV16* (Vb11), *TRBV3*, and *TRBV29* (Vb7*)*. Each of these *TRBV* genes also had a higher mean frequency in SKGNur GFP^{hi} cells compared to WTNur GFP^{hi} cells (**Fig. 2.5A,C-D**).

Polyclonal Vb expansion occurs in the presence of superantigen (Sag) in both humans and mice (47, 48). The *TRBV* genes enriched in SKGNur GFP^{hi} T cells mark Vbs that recognize endogenous retroviral (ERV) Sag from mouse mammary tumor virus (MMTV) (**Fig. 2.5E**) (49-51). We confirmed that our SKG colony harbors all three endogenous MMTV proviruses (*Mtv-6*, *Mtv-8*, *Mtv-9*) known to be present in BALB/c mice (49, 50, 52, 53) (**Supplementary Fig. 2.4C**). Exogenous MMTV infection can stimulate cell proliferation and facilitate infection by increasing the number of cell targets, but Sag expression from endogenous Mtv leads to clonal T-cell deletion in the thymus and resistance to infection owing to the absence of Vb targets (54). However, due to impairment in SKG TCR signaling, thymic clonal T-cell deletion in response to endogenous Mtv Sag is incomplete (12) allowing for partial escape of these Vb targets into the periphery. In contrast to the *TRBV* genes uniquely enriched in SKGNur GFP^{hi} cells, *TRBV* genes for Vbs that are not MMTV Sag targets in BALB/c (e.g., *TRBV19*/Vb6, *TRBV13-2*/Vb8, *TRBV31*/Vb14) are not enriched in SKGNur GFP^{hi} T cells (**Fig. 2.5C, Supplemental Fig. 2.4D**). These results reinforce not only that negative selection is defective in SKG mice, but also that encounter with endogenous MMTV Sag in the periphery further biases the *TRBV* repertoire in SKGNur GFP^{hi} CD4 T cells since the SKGNur GFP^{lo} cells did not show a bias of the MMTV Sag reactive TRBV repertoire.

SKGNur GFP^{hi} cells in the T.N4$_{Nr4a1}$ cluster also demonstrate enrichment of several of the MMTV Sag Vb targets (e.g., *TRBV15* (Vb12), *TRBV16* (Vb11) **Supplementary Fig. 2.5A-D**). Interestingly, MMTV Sag in the periphery fails to induce the peripheral deletion observed with

MMTV reactive cells that are not deleted in the thymus (55). Instead, the peripheral encounter with MMTV Sag likely drives *Nr4a1* expression and the paradoxically activated transcriptional states for the cells with Vbs that recognize MMTV Sag within the $T.N4_{Nr4a1}$ cluster (**Supplementary Fig. 2.5A-D**).

SKGNur GFP^{hi} CD4 T Cells are Enriched for Vbs Driven by MMTV Sag(s)

To validate our scTCR-seq results, we examined the distribution of TCR Vb protein levels in SKGNur and WTNur peripheral CD4 T cells prior to arthritis initiation using commercially available antibodies against a subset of candidate Vbs (gating strategy **Supplementary Fig. 2.6A**). We found that the Vb protein abundances paralleled the transcript abundances observed in our scTCR-seq dataset. T cells expressing Vb3, Vb5, Vb11 (corresponding to *TRBV26*, *-12*, *-16* respectively) are significantly enriched in SKGNur GFP^{hi} peripheral naïve CD4 T cells from lymph node (LN) (**Fig. 2.5F-G**) and spleen, whereas Vbs that are not MMTV Sag targets (e.g., Vb6, Vb8, Vb14 corresponding to *TRBV19*, *-13*, *-31* respectively) are not enriched in SKGNur GFP^{hi} cells (**Supplementary Fig. 2.6B-C**).

Vb enrichment in SKGNur GFP^{hi} T cells subset may be driven by Sag encounter in the periphery and even the joints. Indeed, we found that BALB/c specific *Mtv* proviruses are detectable in SKG joints (**Supplemental Fig. 2.7A**). Therefore, it is feasible that intra-articular MMTV Sag expression could engage and enrich for SKG T cells uniquely expressing these MMTV target Vbs (Vb3, Vb5, and Vb11) during arthritis. To investigate this hypothesis, we induced moderate to severe inflammatory arthritis in SKG mice (**Supplementary Fig. 2.7B**) and examined Vb usage in CD4 T cells harvested from regional joint draining LN (dLN) and arthritic joints. We found an increased frequency of MMTV Sag targets Vb3, Vb5, and Vb11 in the arthritic joints compared to the periphery (**Fig. 2.6A-B**), but not of the Vbs unresponsive to

BALB/c MMTV Sag (**Supplementary Fig. 2.7C**). The CD4 T cells bearing Vbs known to respond to MMTV Sag had significantly higher NUR77-eGFP mean fluorescence intensity (MFI) compared to CD4 T cells with Vbs that do not respond to MMTV Sag (p = 0.002, **Fig. 2.6C**). Furthermore, we found a significantly higher frequency of Vb3, Vb5, and Vb11 in SKGNur GFP^{hi} T cells infiltrating the arthritic joints compared to GFP^{lo} cells in the joint and to GFP^{hi} cells from the dLN (**Fig. 2.6D-E, Supplemental Fig. 2.7D**). This further enrichment suggests SKG CD4 T cells with these particular Vbs encounter intra-articular antigen (10). This enrichment in the joint was not observed in SKGNur GFP^{hi} T cells expressing Vb6, Vb8, and Vb14 (**Supplementary Fig. 2.7C, E-F**).

Antiretroviral therapy ameliorates SKG arthritis: We next assessed whether cDNA synthesis of endogenous MMTV retroelements impact SKG arthritis development. Mice were treated with a combination of antiretrovirals, emtricitabine and tenofovir (Truvada), which are nucleotide and nucleoside reverse transcriptase inhibitors, respectively, or vehicle control prior to and during arthritis development (**Fig. 2.6F**). MMTV reverse transcriptase is sensitive to these commercially available compounds which prematurely terminate nascent cDNA synthesis during reverse transcription (17, 18) and may result in decreased viral protein expression. Antiretroviral treatment significantly reduces arthritis severity (p = 0.0037) and delays disease onset (p = 0.0016) in SKG mice treated with zymosan (**Fig. 2.6G-I**). These data support the idea that MMTV Sag expression contributes to arthritis disease activity in the SKG mouse model.

Discussion

In this study, we directly probed the gene expression and TCR repertoire in enriched naïve arthritogenic SKG CD4 T (i.e., SKGNur GFP^{hi} T cells) cells before arthritis onset. We identified a subset of these cells (T.4N$_{Nr4a1}$) that upregulate TCR signaling gene programs compared to WT cells despite their hypomorphic *Zap70* allele and impaired proximal TCR signaling. This TCR signaling gene signature is likely driven by chronic endogenous antigen encounter marked by upregulation of LAT-PLCg-HDAC7 pathway targets (e.g., *Nr4a1, Egr1-3, Irf4*) (25, 26). How these SKG cells then break tolerance and produce a focused autoimmune phenotype has been a long-standing puzzle. Results presented in this study address this fundamental question.

Our findings provide multiple lines of evidence that arthritogenic SKG CD4 T cells harbor a defect in peripheral tolerance, independent from their break in central tolerance. Induction of negative regulators by TCR activation fine-tunes and restrains T cell responses in normal T cells, enforcing peripheral tolerance, and limiting immunopathology (56-59). However, we found that arthritogenic SKGNur GFP^{hi} CD4 T cells that have severely defective TCR signaling, despite evidence of chronic endogenous antigen encounter, inefficiently and/or incompletely upregulate genes within the tolerogenic program in response to TCR signaling compared to WT cells (e.g., *Eomes, Tnfrsf9, Izumo1r, Bach2, Tigit, Tox, Tox2*). This aberrant transcriptional program may permit their activation and differentiation into pathogenic effector cells.

Second, using RNA velocity analysis to create a latent time axis for cells that have most recently encountered antigen (T.4N$_{Nr4a1}$), we found arthritogenic SKG CD4 T cells are overrepresented in cells expressing early TCR gene signatures (Stage 1) and may not efficiently transition along the trajectory of cell states to express genes associated with prolonged, and perhaps more robust, TCR signaling (Stage 4). Cells in the latter stage appear to more efficiently upregulate genes associated with tolerogenic programs (e.g., *Izumo1r, Bach2, Tigit, Tox, Tox2*).

Third, endogenous MMTV Sag Vb targets in SKG not only escape negative selection in the thymus (central tolerance), but also avoid expected peripheral deletion by endogenous

MMTV Sag expression (55). Previous examinations of the peripheral TCR Vb repertoire in SKG mice had not identified oligoclonal expansion of a particular Vb before or after arthritis development (60). Here we describe a previously unknown TCR Vb bias in the peripheral naïve SKG repertoire. The Vbs enriched in SKGNur GFP^{hi} T cells suggest their repertoire is further influenced in the periphery by Sag encounter and are further expanded in arthritic SKG joints. The Sag target Vbs that infiltrate arthritic joints significantly upregulate Nur77-eGFP compared to CD4 T cells bearing Vbs that do not recognize MMTV Sag. The expansion of Sag target Vbs in the SKG arthritic joint and their upregulation of Nur77-eGFP are both likely driven by intra-articular MMTV Sag encounter. This begs the question whether MMTV Sag contributes to arthritis development in SKG mice. We found that inhibition of ERV reverse transcriptase (with Truvada) in arthritic SKG mice significantly ameliorates and delays onset of SKG arthritis. Together our data suggests a previously unrecognized role for endogenous Sag in SKG arthritis development.

We propose a model drawn from these and previous results (9, 10, 12, 60) in which inefficient negative selection results in the escape of a portion of the self-reactive biased repertoire (marked by Vbs responsive to MMTV Sags) that is further enriched in the periphery through chronic (super)antigen encounter. Although naïve SKG CD4 T cells can upregulate negative regulators of TCR signaling in response to endogenous antigen encounter, due to weak proximal TCR signaling they are unable to efficiently and fully upregulate these programs resulting in an independent defect in peripheral tolerance and reduced numbers of anergic cells (10). Furthermore, Treg peripheral tolerance mechanisms are also severely impaired due to their attenuated TCR signaling and their altered Treg repertoire (11), releasing another checkpoint on arthritogenic T cells. We propose that costimulatory molecules such as IL-6, and perhaps other cytokines, promote T cell survival and lower the TCR signaling threshold required for peripheral activation and differentiation, thereby allowing the activation of naïve SKG T cells that failed to upregulate TCR signaling-induced tolerogenic programs (61-63). Therefore, the biased self-reactive TCR Vb repertoire unique to SKGNur GFP^{hi} CD4 T cells, together with their activated transcriptional state, allow for an innate immune stimulus to trigger these cells to become potential initiators, or amplifiers, of disease. Future studies will determine whether these Vbs are sufficient, and/or necessary, to initiate or exacerbate SKG arthritis and the extent of the role of endogenous MMTV Sags in SKG arthritis development. These findings have relevant implications in human autoimmune disease, where endogenous or foreign antigens can also prime 'dormant' autoreactive T cells and trigger disease, and may provide new insights in human RA (64-66), and other autoimmune arthritides (67-69), where T cells bearing particular Vbs have been reported to be expanded and retained in the synovial microenvironment.

Figures

Figure 1

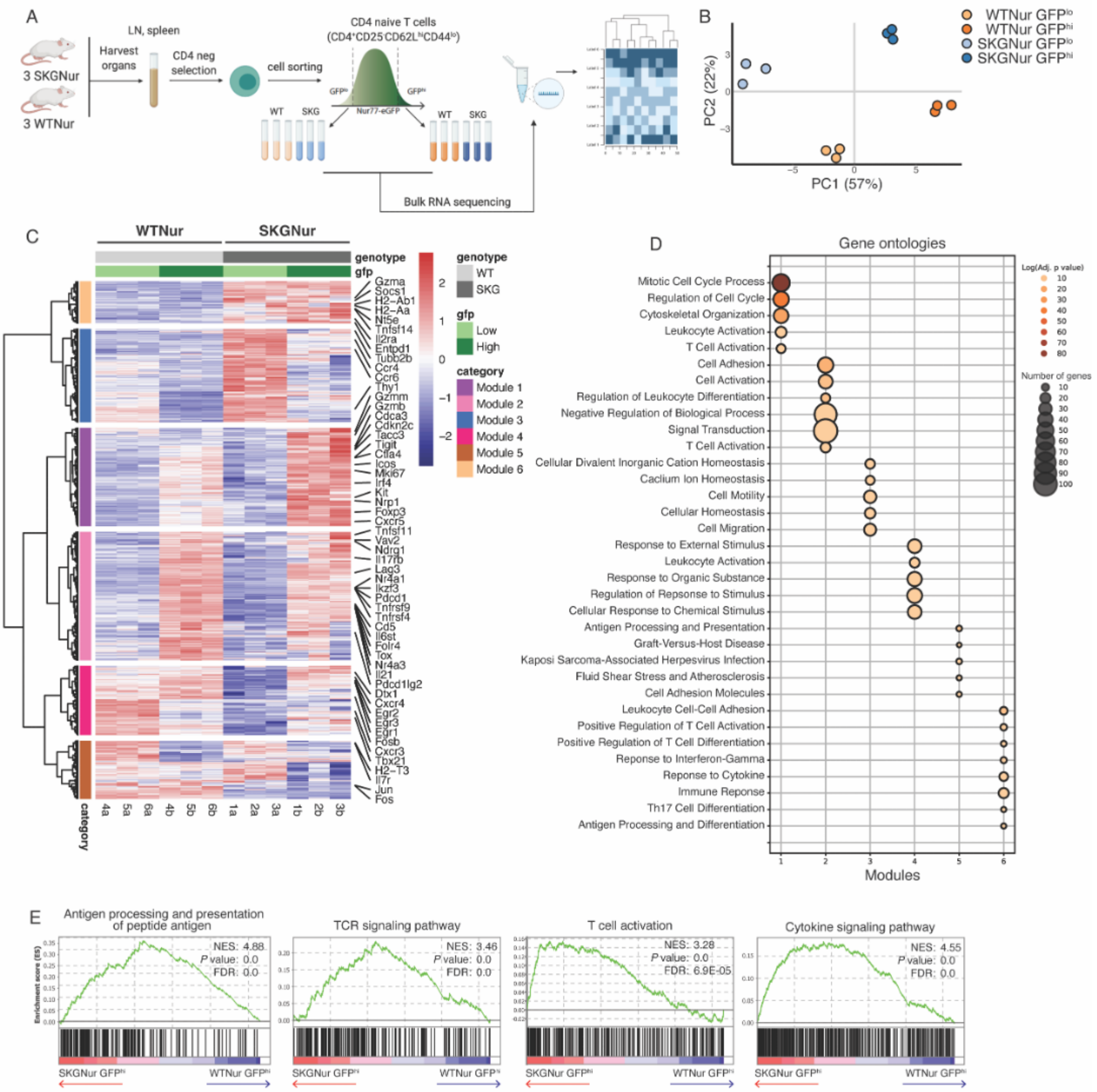
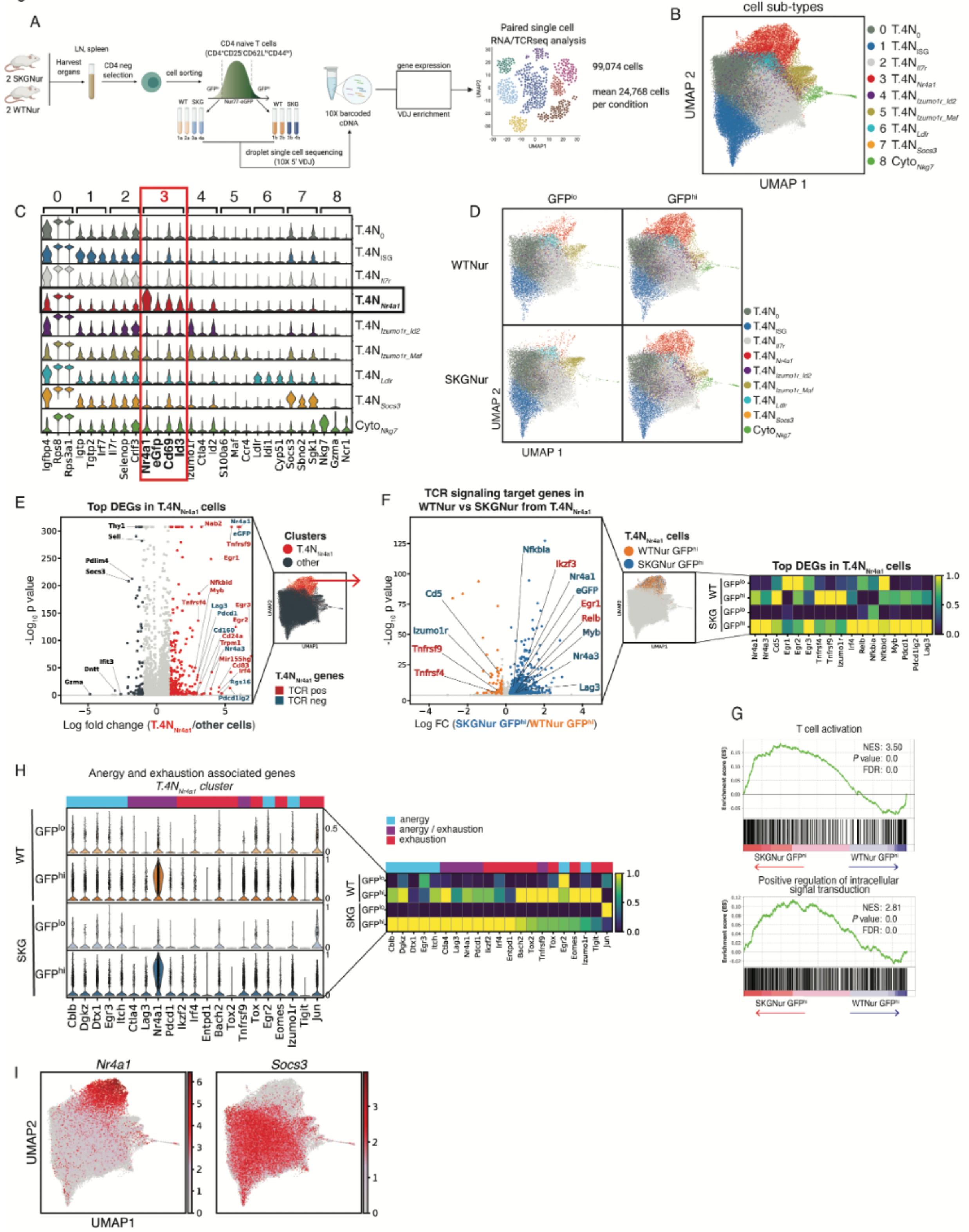


Figure 2.1 Pre-arthritic naïve SKG T cells demonstrate enhanced T cell activation
(A) Experimental scheme of bulk RNA-seq analysis. (B) Principal component analysis (PCA) based on transcriptomic data from bulk RNA-seq data shows distribution of SKGNur and WTNur GFP^{lo} and GFP^{hi} CD4 naïve T cell subsets as shown in (A) (n=3 per subgroup). (C) Heatmap showing expression of 991 significantly differentially expressed genes (DEGs, absolute value(log2(fold-change)) > 1, adjusted *P value* < 0.05) from pairwise comparisons for all samples grouped by subgroup. Hierarchical clustering was used to group DEGs into 6 modules (indicated by dendrogram and row annotation color bar on left). (D) Dot plot of select pathways from gene ontology analysis for each gene module from (C) with dot color indicating adjusted *P* value and dot size proportional to number of genes in overlap between pathway genes and module genes. (E) Enrichment plots of TCR signaling and cytokine pathways from GSEA analysis of all GO:BP pathways for ranked genes from SKGNur GFP^{hi} and WTNur GFP^{hi} differential expression analysis. FDR, false discovery rate. NES, normalized enrichment score.

Figure 2



A
LN, spleen
Harvest organs
CD4 neg selection
cell sorting
CD4 naive T cells (CD4+CD25-CD62L+CD44lo)
2 SKGNur
2 WTNur
WT SKG
Nur77-eGFP
GFPlo
GFPhi
10X barcoded cDNA
gene expression
VDJ enrichment
droplet single cell sequencing (10X 5' VDJ)
Paired single cell RNA/TCRseq analysis
99,074 cells
mean 24,768 cells per condition
UMAP1
UMAP2
B
cell sub-types
0 T.4N0
1 T.4NISG
2 T.4NIl7r
3 T.4NNr4a1
4 T.4NIzumo1r_Id2
5 T.4NIzumo1r_Maf
6 T.4NLdlr
7 T.4NSocs3
8 CytoNkg7
UMAP 1
UMAP 2
C
0 1 2 3 4 5 6 7 8
T.4N0
T.4NISG
T.4NIl7r
T.4NNr4a1
T.4NIzumo1r_Id2
T.4NIzumo1r_Maf
T.4NLdlr
T.4NSocs3
CytoNkg7
Igfbp4 Rps8 Rps3a1 Igtp Tgtp2 Irf7 Il7r Selenop Crlf3 Nr4a1 eGfp Cd69 Id3 Izumo1r Ctla4 Id2 S100a6 Maf Ccr4 Ldlr Idi1 Cyp51 Socs3 Sbno2 Sgk1 Nkg7 Gzma Ncr1
D
GFPlo
GFPhi
WTNur
SKGNur
UMAP 1
UMAP 2
T.4N0
T.4NISG
T.4NIl7r
T.4NNr4a1
T.4NIzumo1r_Id2
T.4NIzumo1r_Maf
T.4NLdlr
T.4NSocs3
CytoNkg7
E
Top DEGs in T.4NNr4a1 cells
Thy1
Sell
Pdlim4
Socs3
Ifit3
Dntt
Gzma
Nab2
Nr4a1
eGFP
Tnfrsf9
Egr1
Nfkbid
Myb
Tnfrsf4
Lag3
Pdcd1
Egr3
Egr2
Cd160
Cd24a
Trpm1
Nr4a3
Mir155hg
Cd83
Irf4
Rgs16
Pdcd1lg2
-Log10 p value
Log fold change (T.4NNr4a1/other cells)
Clusters
T.4NNr4a1
other
T.4NNr4a1 genes
TCR pos
TCR neg
F
TCR signaling target genes in WTNur vs SKGNur from T.4NNr4a1
Nfkbia
Ikzf3
Nr4a1
eGFP
Egr1
Relb
Myb
Nr4a3
Lag3
Cd5
Izumo1r
Tnfrsf9
Tnfrsf4
-Log10 p value
Log FC (SKGNur GFPhi/WTNur GFPhi)
T.4NNr4a1 cells
WTNur GFPhi
SKGNur GFPhi
Top DEGs in T.4NNr4a1 cells
WT
SKG
GFPlo
GFPhi
Nr4a1 Nr4a3 Cd5 Egr1 Egr2 Egr3 Tnfrsf4 Tnfrsf9 Izumo1r Irf4 Relb Nfkbia Nfkbid Myb Pdcd1 Pdcd1lg2 Lag3
1.0
0.5
0.0
G
T cell activation
NES: 3.50
P value: 0.0
FDR: 0.0
Enrichment score (ES)
SKGNur GFPhi
WTNur GFPhi
Positive regulation of intracellular signal transduction
NES: 2.81
P value: 0.0
FDR: 0.0
H
Anergy and exhaustion associated genes
T.4NNr4a1 cluster
WT
SKG
GFPlo
GFPhi
Cblb Dgkz Dtx1 Egr3 Itch Ctla4 Lag3 Nr4a1 Pdcd1 Ikzf2 Irf4 Entpd1 Bach2 Tox2 Tnfrsf9 Tox Egr2 Eomes Izumo1r Tigit Jun
anergy
anergy / exhaustion
exhaustion
1.0
0.5
0.0
I
Nr4a1
Socs3
UMAP1
UMAP2

Figure 2.2 Single-cell RNA sequencing unveils heterogeneity among naïve CD4 T cells with a subset marked by genes associated with TCR signaling

(A) Experimental scheme of paired scRNA- and TCR-seq of sorted GFP^{hi} and GFP^{lo} naïve (CD62L^{hi}CD44^{lo}CD25-) CD4 T cells. (B) Uniform manifold approximation and projection (UMAP) of 99,074 naïve T cells derived from 8 samples (2 replicates each for GFP^{lo} and GFP^{hi} subsets from WTNur and SKGNur). Cells are colored and annotated by merged leiden clusters assignment. (C) Stacked violin plot with a kernel bandwidth of 0.5 of log normalized expression of marker genes for each annotated cluster. Vertical axis scale is from 0 to 5 for each cluster. Black box highlights T.4N$_{Nr4a1}$ cluster and red box highlights genes uniquely expressed in T.4N$_{Nr4a1}$ cluster. (D) UMAP of cells separated by subgroup (GFP^{lo} and GFP^{hi} subgroups from WTNur and SKGNur). (E) Volcano plot of DEGs from cells in T.4N$_{Nr4a1}$ cluster versus other cells. Dots are colored by significant overexpression (absolute value(log2(fold-change)) > 1, adjusted P value < 0.05) in T.4N$_{Nr4a1}$ cluster (red), other cells (dark gray), or no significant difference (light gray). Labeled genes are colored by their role in regulation of TCR signaling either positive (red) or negative (blue). (F) Volcano plot of DEGs from SKGNur GFP^{hi} versus WTNur GFP^{hi} in T.4N$_{Nr4a1}$ cluster. Dots are colored as significantly overexpressed (absolute value(log2(fold-change)) > 0.2, adjusted P value < 0.05) in WTNur GFP^{hi} (orange), SKGNur GFP^{hi} (blue), or not significantly different between groups (gray). Labeled genes involved in TCR signaling are colored as indicated in (**E**). Heatmap shows average expression of the labeled genes by subgroup normalized by standard scale (subtract minimum and divide by maximum) for each gene. (G) Enrichment plots of TCR activation and signaling pathways from GSEA analysis of all GO:BP pathways for ranked genes from differential gene expression analysis of SKGNur GFP^{hi} versus WTNur GFP^{hi} cells from T.4N$_{Nr4a1}$ cluster. FDR, false discovery rate. NES, normalized enrichment score. (H) Stacked violin plot of expression of candidate anergy and exhaustion associated genes in WTNur and SKGNur GFP^{lo} and GFP^{hi} CD4 naïve cells from T.4N$_{Nr4a1}$ cluster normalized by standard scale for each gene. Color bar indicates annotation [anergy (blue), anergy/exhaustion (purple), exhaustion (red)] for each gene. Heatmap shows average expression of the same genes by subgroup also normalized by standard scale for each gene. (I) UMAP of all cells colored by expression of the indicated genes. Scale is for the log-normalized gene expression.

Figure 3

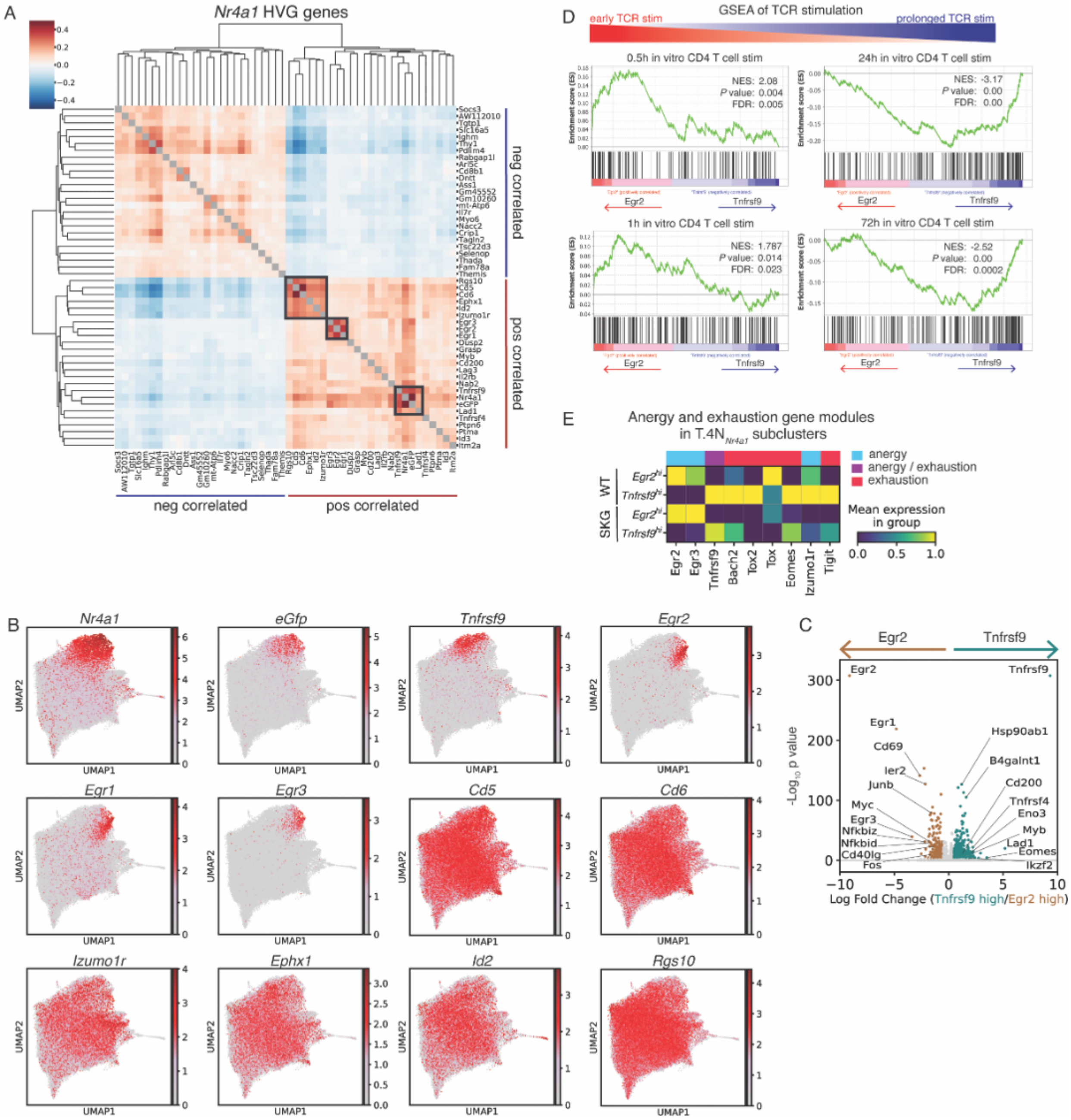

A
Nr4a1 HVG genes
0.4
0.2
0.0
−0.2
−0.4
Socs3
AW112010
Tgtp1
Slc16a5
Ighm
Thy1
Pdlim4
Rabgap1l
Arl5c
Cd8b1
Dntt
Ass1
Gm45552
Gm10260
mt-Atp6
Il7r
Myo6
Nacc2
Crip1
Tagln2
Tsc22d3
Selenop
Thada
Fam78a
Themis
Rgs10
Cd5
Cd6
Ephx1
Id2
Izumo1r
Egr3
Egr2
Egr1
Dusp2
Grasp
Myb
Cd200
Lag3
Il2rb
Nab2
Tnfrsf9
Nr4a1
eGFP
Lad1
Tnfrsf4
Ptpn6
Ptma
Id3
Itm2a
neg correlated
pos correlated
D
GSEA of TCR stimulation
early TCR stim
prolonged TCR stim
0.5h in vitro CD4 T cell stim
NES: 2.08
P value: 0.004
FDR: 0.005
24h in vitro CD4 T cell stim
NES: -3.17
P value: 0.00
FDR: 0.00
1h in vitro CD4 T cell stim
NES: 1.787
P value: 0.014
FDR: 0.023
72h in vitro CD4 T cell stim
NES: -2.52
P value: 0.00
FDR: 0.0002
Enrichment score (ES)
Egr2
Tnfrsf9
E
Anergy and exhaustion gene modules in T.4N$_{Nr4a1}$ subclusters
anergy
anergy / exhaustion
exhaustion
Mean expression in group
0.0
0.5
1.0
WT
SKG
Egr2^{hi}
Tnfrsf9^{hi}
Egr2
Egr3
Tnfrsf9
Bach2
Tox2
Tox
Eomes
Izumo1r
Tigit
B
Nr4a1
eGfp
Tnfrsf9
Egr2
Egr1
Egr3
Cd5
Cd6
Izumo1r
Ephx1
Id2
Rgs10
UMAP1
UMAP2
C
Egr2
Tnfrsf9
Egr1
Cd69
Ier2
Junb
Myc
Egr3
Nfkbiz
Nfkbid
Cd40lg
Fos
Hsp90ab1
B4galnt1
Cd200
Tnfrsf4
Eno3
Myb
Lad1
Eomes
Ikzf2
300
200
100
0
−10
−5
0
5
10
$-\text{Log}_{10}$ p value
Log Fold Change (Tnfrsf9 high/Egr2 high)

Figure 2.3 T.N4$_{Nr4a1}$ cells segregate into two distinct TCR signaling modules that segregate acute from chronic antigen-activated T cells

(A) Correlation matrix shows hierarchical clustering of Spearman's correlation of top 25 highly variable genes (HVG) that positively and negatively correlate with *Nr4a1* expression across all cells. Diagonal gray boxes represent correlation of 1. Dark gray boxes mark distinct gene modules from genes that positively correlate with *Nr4a1* expression. (B) UMAP plots show the expression levels of indicated marker genes positively correlating with *Nr4a1* as identified in (**A**). Scale represents the log-transformed normalized counts of genes. (C) Volcano plot shows DEGs for cells in T.4N$_{Nr4a1}$ cluster that expressed (log normalized expression > 1) *Egr2* or *Tnfrsf9* with dots colored by significant overexpression (absolute value(log2(fold-change)) > 0.5, adjusted P value < 0.05) in *Egr2* (tan) or *Tnfrsf9* (teal) expressing cells. (D) Enrichment plots from GSEA analysis of study GSE17974 pathways of time course *in vitro* activation of CD4+ T cells with aCD3 + CD28 for ranked genes from differential gene expression analysis of cells in T.4N$_{Nr4a1}$ cluster that express *Egr2* versus *Tnfrsf9*. FDR, false discovery rate. NES, normalized enrichment score. (E) Heatmap of average expression of exhaustion and anergy genes from WTNur and SKGNur GFP^{hi} cells expressing *Egr2* or Tnfrsf9 in T.4N$_{Nr4a1}$ cluster normalized by standard scale for each gene. Color bar indicates annotation [anergy (blue), anergy/exhaustion (purple), exhaustion (red)] for each gene.

Figure 4

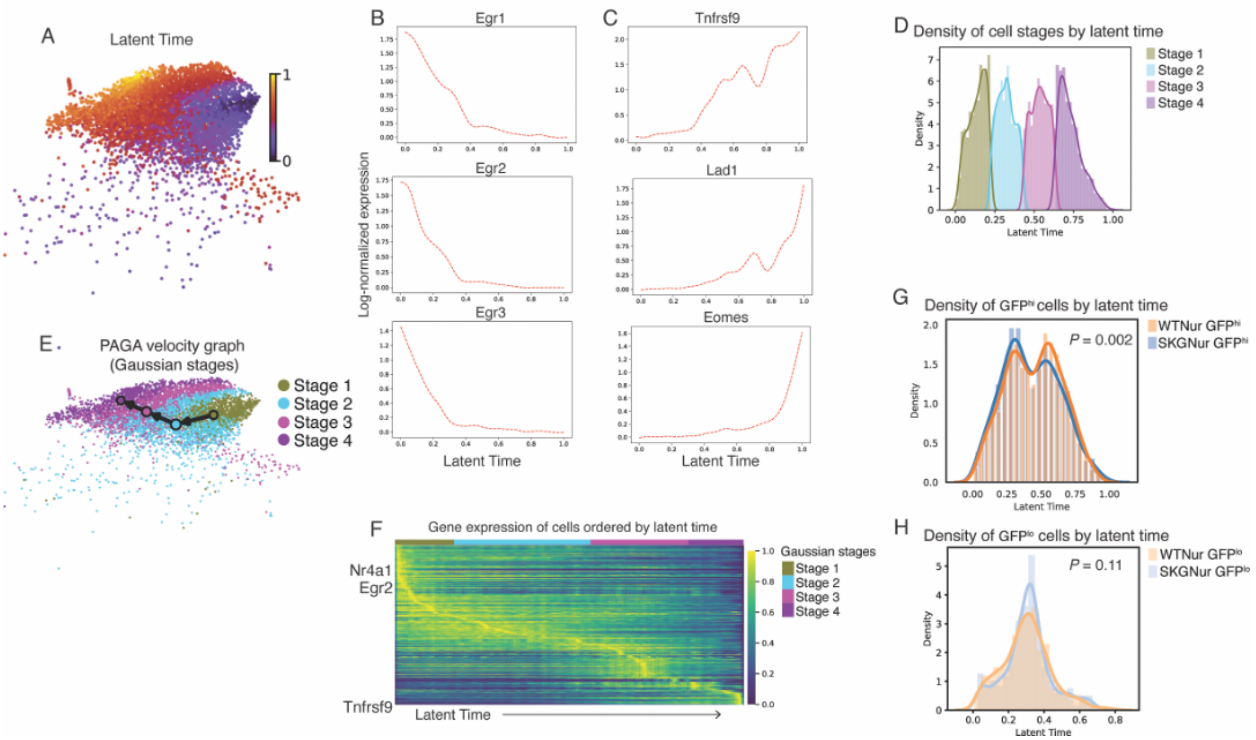
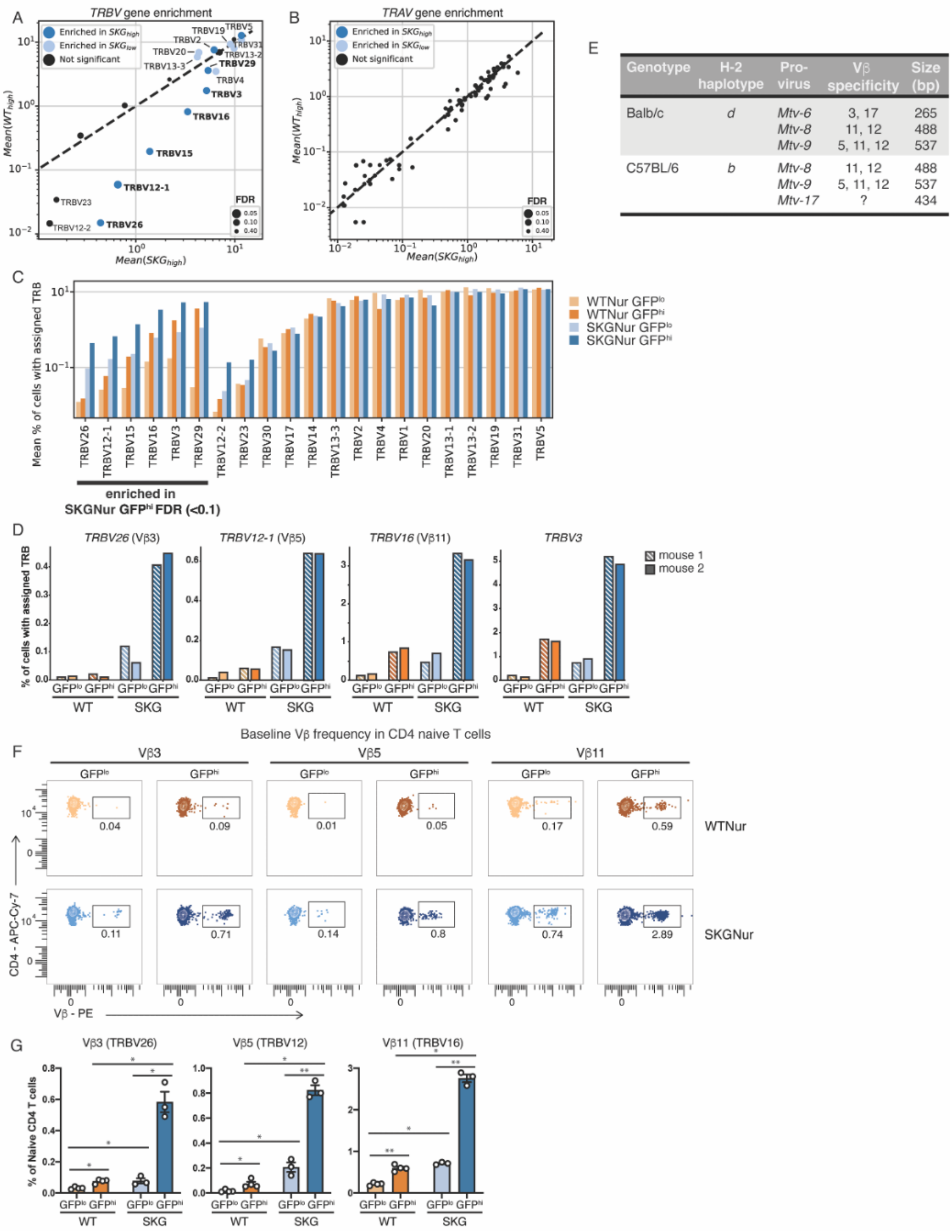

A
Latent Time
1
0
B
Egr1
Egr2
Egr3
Log-normalized expression
Latent Time
C
Tnfrsf9
Lad1
Eomes
Latent Time
D
Density of cell stages by latent time
Stage 1
Stage 2
Stage 3
Stage 4
Density
Latent Time
E
PAGA velocity graph
(Gaussian stages)
Stage 1
Stage 2
Stage 3
Stage 4
F
Gene expression of cells ordered by latent time
Nr4a1
Egr2
Tnfrsf9
Latent Time
Gaussian stages
Stage 1
Stage 2
Stage 3
Stage 4
G
Density of GFPhi cells by latent time
WTNur GFPhi
SKGNur GFPhi
P = 0.002
Density
Latent Time
H
Density of GFPlo cells by latent time
WTNur GFPlo
SKGNur GFPlo
P = 0.11
Density
Latent Time

Figure 2.4 Trajectory analysis of T.4N$_{\mathrm{Nr4a1}}$ cells orthogonally uncovers acute versus chronic antigen-activated T cell states with a distinct distribution in the SKGNur GFP$^{\mathrm{hi}}$ subset
(A) UMAP of cells from T.4N$_{Nr4a1}$ cluster colored by latent time assigned by the scvelo dynamical modeling algorithm. (B-C) Smoothed gene expression from cells in T.4N$_{\mathrm{Nr4a1}}$ cluster of selected genes with highest expression earlier (**B**) or later (**C**) along latent time axis. (D) Probability densities of latent time distribution of cells from T.4N$_{Nr4a1}$ cluster assigned to 4 distinct clusters (labeled “Stage 1” – “Stage 4”) by a Gaussian mixture model. (E) Predicted transitions from PAGA algorithm between cells from stages indicated in (**D**). (F) Heatmap of single cell standard scale normalized expression of genes ordered top to bottom by peak expression at earlier to later latent time. Chosen genes are the top 300 highest confidence genes used in modeling of latent time. Column annotation bar indicates stage assignment of the cell in each column. (G-H) Probability densities of latent time distribution for GFP$^{\mathrm{hi}}$ (**G**) and GFP$^{\mathrm{lo}}$ (**H**) cells from WTNur and SKGNur mice ($P = 0.002$ and $P = 0.11$, respectively, Kolmogorov-Smirnov test).

Figure 5



Genotype	H-2 haplotype	Pro-virus	Vβ specificity	Size (bp)
Balb/c	*d*	*Mtv-6*	3, 17	265
		Mtv-8	11, 12	488
		Mtv-9	5, 11, 12	537
C57BL/6	*b*	*Mtv-8*	11, 12	488
		Mtv-9	5, 11, 12	537
		Mtv-17	?	434

Figure 2.5 SKG CD4 T cells harbor a biased TCR variable beta gene repertoire
(A-B) Scatter plot of mean frequency of cells expressing each TRBV (**A**) or TRAV (**B**) gene for the SKGNur GFP^{hi} samples versus the WTNur GFP^{hi} samples. Dots for each TRBV and TRAV genes are sized according to the FDR from a one-sided paired t-test comparing frequency in SKGNur GFP^{hi} versus SKGNur GFP^{lo}. Dots are colored as either significantly enriched (FDR < 0.1) in SKGNur GFP^{hi} (dark blue), significantly enriched in SKGNur GFP^{lo} (light blue), or not significantly enriched in either subgroup (black). Dots for significant TRBV genes are labeled with the TRBV gene name. Labels for TRBV genes that were significantly enriched in SKGNur GFP^{hi} and were also more highly expressed in SKGNur GFP^{hi} samples versus WTNur GFP^{hi} samples are bolded. (C) Bar plot of mean value of cells expressing each TRBV gene as a percentage of all cells in each sample with an assigned TRBV. Bars are colored according to subgroup and ordered with the TRBV genes enriched in SKGNur GFP^{hi} from (**A**) followed by the other TRBV genes ordered by increasing overall frequency. (D) Bar plots of frequency of cells expressing the indicated TRBV genes significantly enriched in SKGNur GFP^{hi} for the two replicate mice in each subgroup. (E) Table depicting H-2 haplotype, expected Mtv pro-virus, its Vb specificity and base pair (bp) size on gel for Balb/c and C57BL/6 mice. (F-G) Representative FACS plots (**F**) of naïve peripheral CD4 T cells with indicated TCR Vb protein usage determined by flow cytometry in GFP^{lo} and GFP^{hi} T cells from LN of WTNur and SKGNur mice prior to arthritis induction and quantified in (**G**) where bar graphs depict mean frequency (± SEM), n = 3-4 mice per group, experiment repeated at least 3 times. Significance indicated by asterisk for FDR (paired t-test) or *P* value (exact permutation test) < 0.05 (*), < 0.1 (**).

Figure 6

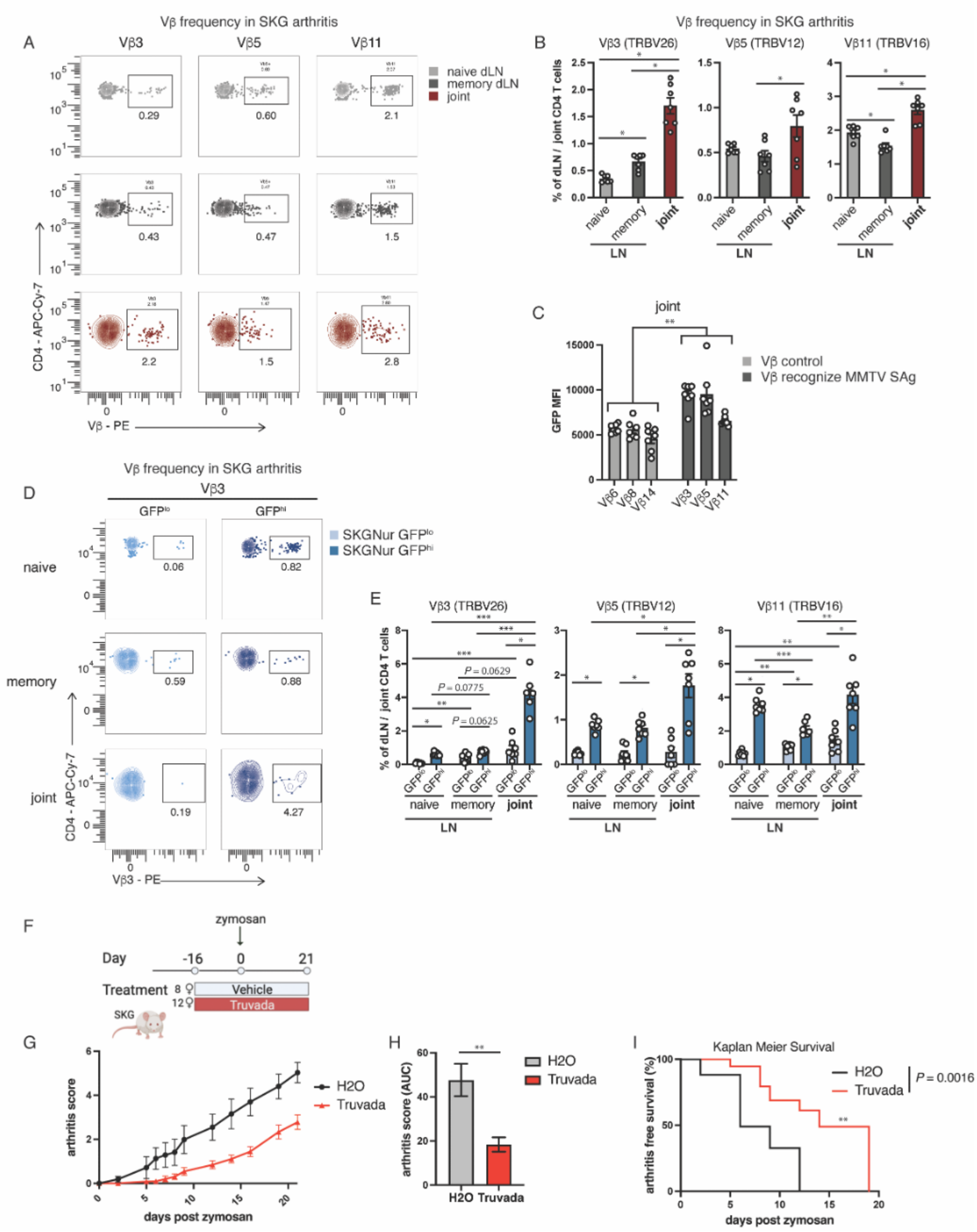
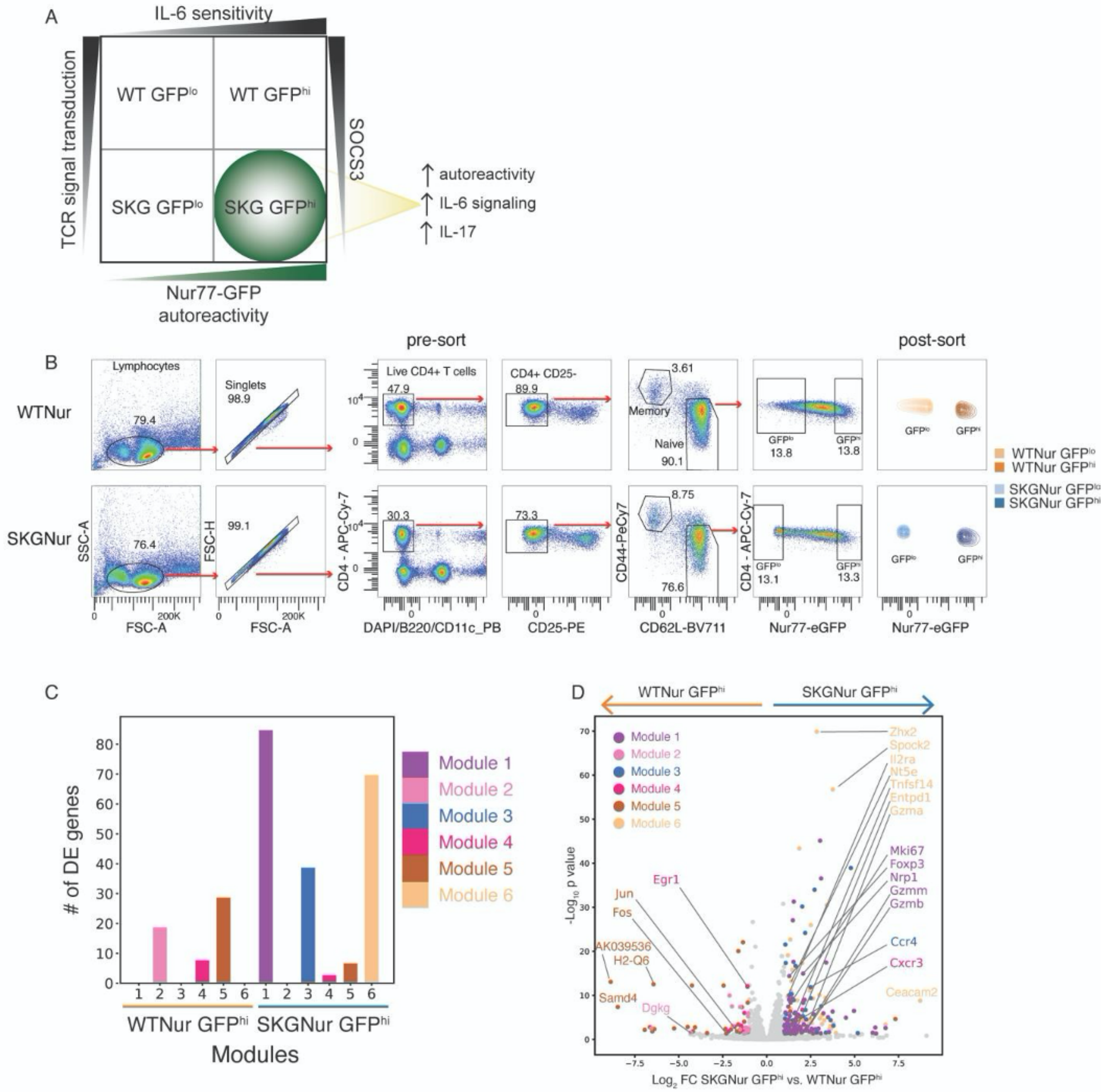

Vβ frequency in SKG arthritis
A
Vβ3
Vβ5
Vβ11
0.29
0.60
2.1
0.43
0.47
1.5
2.2
1.5
2.8
CD4 - APC-Cy-7
Vβ - PE
naive dLN
memory dLN
joint
B
Vβ3 (TRBV26)
Vβ5 (TRBV12)
Vβ11 (TRBV16)
% of dLN / joint CD4 T cells
naive
memory
joint
LN
C
joint
GFP MFI
Vβ control
Vβ recognize MMTV SAg
Vβ6
Vβ8
Vβ14
Vβ3
Vβ5
Vβ11
D
Vβ frequency in SKG arthritis
Vβ3
GFPlo
GFPhi
naive
memory
joint
0.06
0.82
0.59
0.88
0.19
4.27
CD4 - APC-Cy-7
Vβ3 - PE
SKGNur GFPlo
SKGNur GFPhi
E
Vβ3 (TRBV26)
Vβ5 (TRBV12)
Vβ11 (TRBV16)
% of dLN / joint CD4 T cells
P = 0.0629
P = 0.0775
P = 0.0625
GFPlo
GFPhi
naive
memory
joint
LN
F
zymosan
Day
-16
0
21
Treatment
8 ♀
12 ♀
Vehicle
Truvada
SKG
G
arthritis score
days post zymosan
H2O
Truvada
H
arthritis score (AUC)
H2O
Truvada
I
Kaplan Meier Survival
arthritis free survival (%)
days post zymosan
H2O
Truvada
P = 0.0016

Figure 2.6 Arthritogenic CD4 T cells are enriched for TCR Vbs likely driven by endogenous superantigen(s)
(A-B) Representative FACS plots (**A**) of peripheral naïve or memory, or joint CD4 T cells with indicated TCR Vb protein usage determined by flow cytometry in CD4 T cells from draining LN or joints of SKGNur mice 2.5 weeks after arthritis induction with zymosan (as seen in **Supplementary Fig. 2.7B**) and quantified in (**B**) where bar graphs depict mean frequency (± SEM). (C) Bar graphs of GFP mean fluorescence intensity (MFI) of CD4 T cells bearing indicated Vbs from arthritic joints of SKG mice, n = 7 mice pooled from 2 experiments. (D-E) Representative FACS plots (**D**) of peripheral naïve or memory, or joint CD4 T cells with indicated TCR Vb protein usage determined by flow cytometry in GFP^{lo} (light blue) and GFP^{hi} (dark blue) T cells from LN or joints of SKGNur mice 2.5 weeks after arthritis induction with zymosan and quantified in (**E**) where bar graphs depict mean frequency (± SEM), n = 7 mice per group pooled from 2 experiments. (F) Experimental set-up: SKG mice were treated with Truvada (n = 12) or vehicle control (n = 8) at day -16 prior to arthritis induction with i.p. zymosan on day 0. (G) Arthritis score in SKG mice from experimental set-up in (**F**). (H) Bar graph depicting area under the curve (AUC) of arthritis scores from (**G**). (I) Arthritis free survival plotted as Kaplan Meier Curve from experimental set-up in (**F**). Significance indicated by asterisk [< 0.05 (*), < 0.1 (**), or < 0.001 (***)] for FDR (paired t-test) or *P* value from exact permutation test (**B** and **E**), linear mixed effect model (**C**), unpaired t-test with Welch's correction (**H**), or log-rank Mantel-Cox test (**I**).

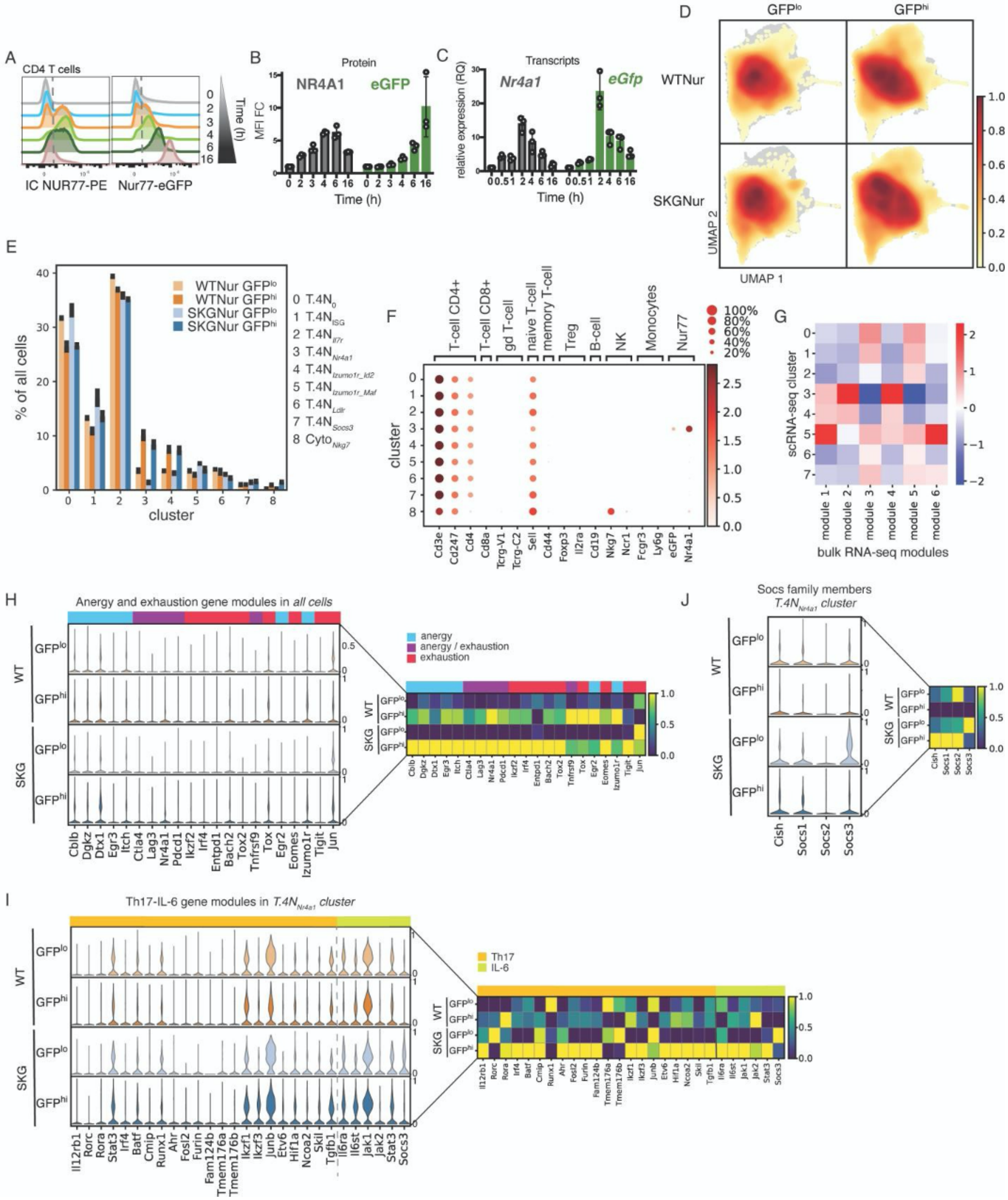
Supplemental Figure 1



Supplementary Figure 2.1 SKGNur GFP^{hi} CD4 T cells readily differentiate into pathogenic effector cells

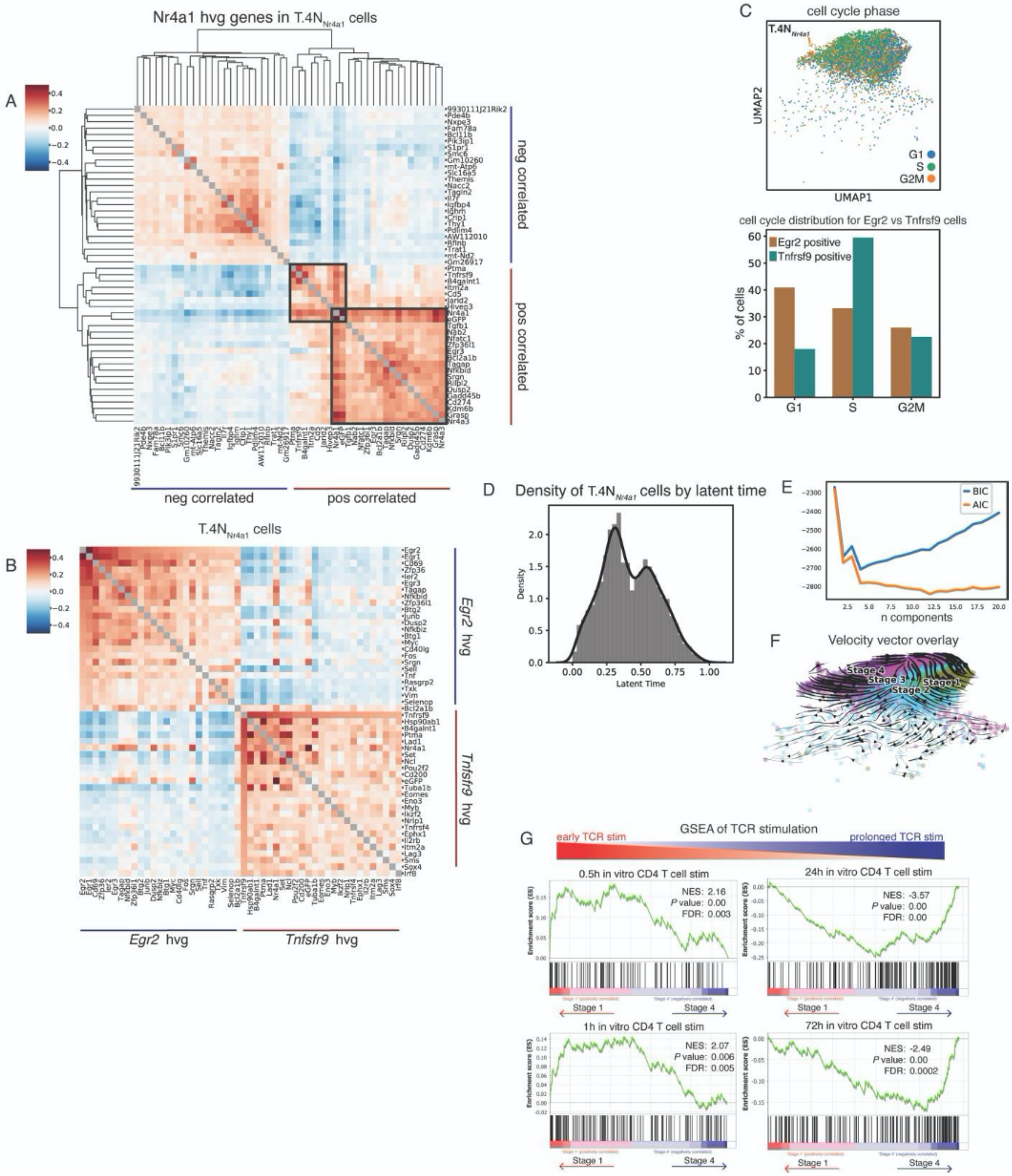
(A) 2 x 2 matrix demonstrates how impaired TCR signaling observed in SKG mice (left y-axis, due to the hypomorphic Zap70 allele), in addition to chronic antigen stimulation (x-axis, resulting in higher levels of Nur77eGFP demarcated by GFP^{hi}) confer heightened sensitivity to IL-6 cytokine signaling, in part due to decreased levels of SOCS3. This contributes to the increased arthritogenicity observed in the autoreactive T cell clones that more readily differentiate into IL-17 producing CD4 T cells in SKG mice. (B) Gating for bulk RNAseq sorting of WTNur and SKGNur lymphocytes. (C) Bar plot of number of DEGs from WTNur GFP^{hi} and SKGNur GFP^{hi} cells contained in each gene module from **Fig. 2.1C**. (D) Volcano plot of DEGs for SKGNur GFP^{hi} versus WTNur GFP^{hi}. DEGs are colored by module membership from gene modules in **Fig. 2.1C**.

Supplemental Figure 2



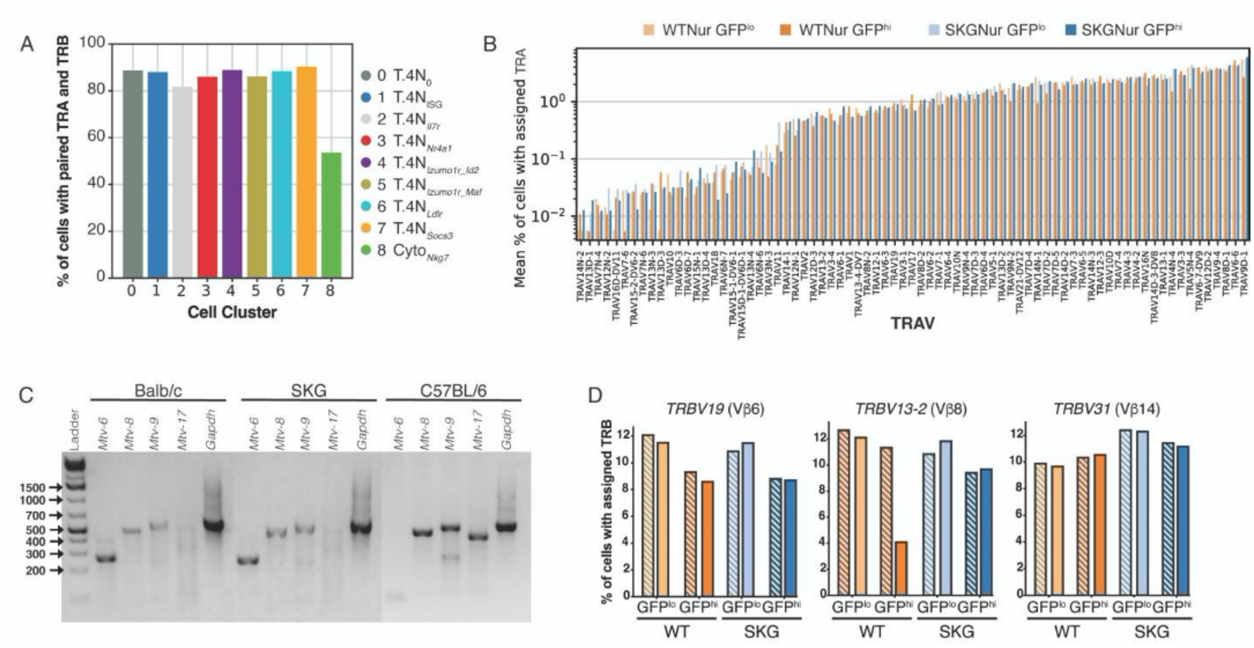
Supplementary Figure 2.2 NUR77/*Nr4a1* identifies naïve CD4 T cells that have recently encountered endogenous antigen resulting in a unique transcriptional program
(A-C) CD4 T cells were stimulated ± plate bound aCD3e + CD28 at indicated times. Levels of NUR77 and eGFP are depicted in representative histogram from 2 experiments (**A**) and MFI fold change (FC) quantified in (**B**) from 3 biological replicates. (**C**) Real-time RT-PCR measuring *Nr4a1* and *eGfp* mRNA levels in stimulated CD4 T cells from 3 biological replicates, from 2 independent experiments. (D) UMAP colored by density of cells for each of the four subgroups (each subgroup contains samples from 2 mice). (E) Bar plot of mean frequency for each subgroup of cells within each cluster. Black bars indicate the difference between mouse 1 and mouse 2 for each subgroup. (F) Expression of labeled genes for each cluster is shown by percentage of cells with expression greater than zero (dot size) and mean expression for cells with nonzero expression (color). (G) Heatmap normalized by standard scale (subtract minimum and divide by maximum) by column of average single cell gene set scores for each cluster (excluding cluster 8 – $\mathrm{Cyto}_{\mathrm{Nkg7}}$) for the gene sets defined by the modules from **Fig. 2.1C**. (H-J) Stacked violin plot demonstrates standard scale normalized expression of candidate anergy and exhaustion associated genes (**H**), Th-17 and IL-6 associated genes (**I**), and *Socs* family members (**J**) in WTNur and SKGNur $\mathrm{GFP}^{\mathrm{lo}}$ and $\mathrm{GFP}^{\mathrm{hi}}$ CD4 naïve T cells in all cells (**H**) or in $\mathrm{T.4N}_{Nr4a1}$ cells (**I-J**). Heatmaps on the right for each panel show mean expression of the indicated genes across subgroup normalized by standard scale for each gene.

Supplemental Figure 3



Supplementary Figure 2.3 Highly variable genes that positively and negatively correlate with *Nr4a1* in T.4N$_{Nr4a1}$ cluster and trajectory analysis of their underlying states
(A) Hierarchical clustering of correlation matrix of top 25 HVGs that positively and negatively correlate with *Nr4a1* expression in T.4N$_{Nr4a1}$ cells using Spearman's correlation. Diagonal gray colored boxes represent correlation of 1. Dark gray boxes mark modules of HVGs that highly correlate with *Nr4a1* expression. (B) Correlation matrix of HVGs that positively correlate with *Egr2* and *Tnfrsf9* expression in T.4N$_{Nr4a1}$ cells using Spearman's correlation. Diagonal gray boxes represent correlation of 1. (C) UMAP of cells from the T.4N$_{Nr4a1}$ cluster colored by cell cycle phase assignment. Bar plot of % of cells in each cell cycle stage for cells expressing *Egr2* or *Tnfrsf9* (log-normalized expression > 1). (D) Probability density of latent time distribution of all cells in T.4N$_{Nr4a1}$ cluster. (E) Line plots for the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) for the Gaussian mixture model deconvolution versus number of underlying distributions or clusters. (F) UMAP colored by cell stage as defined in **Fig. 2.4D** with an overlay of RNA velocity vectors for cell transitions as determined by the scvelo dynamical model. (G) Enrichment plots from GSEA analysis of study GSE17974 pathways of time course *in vitro* activation of CD4+ T cells with CD3 + CD28 for ranked genes from differential gene expression analysis of T.4N$_{Nr4a1}$ cluster cells in Stage 1 versus Stage 4. FDR, false discovery rate. NES, normalized enrichment score

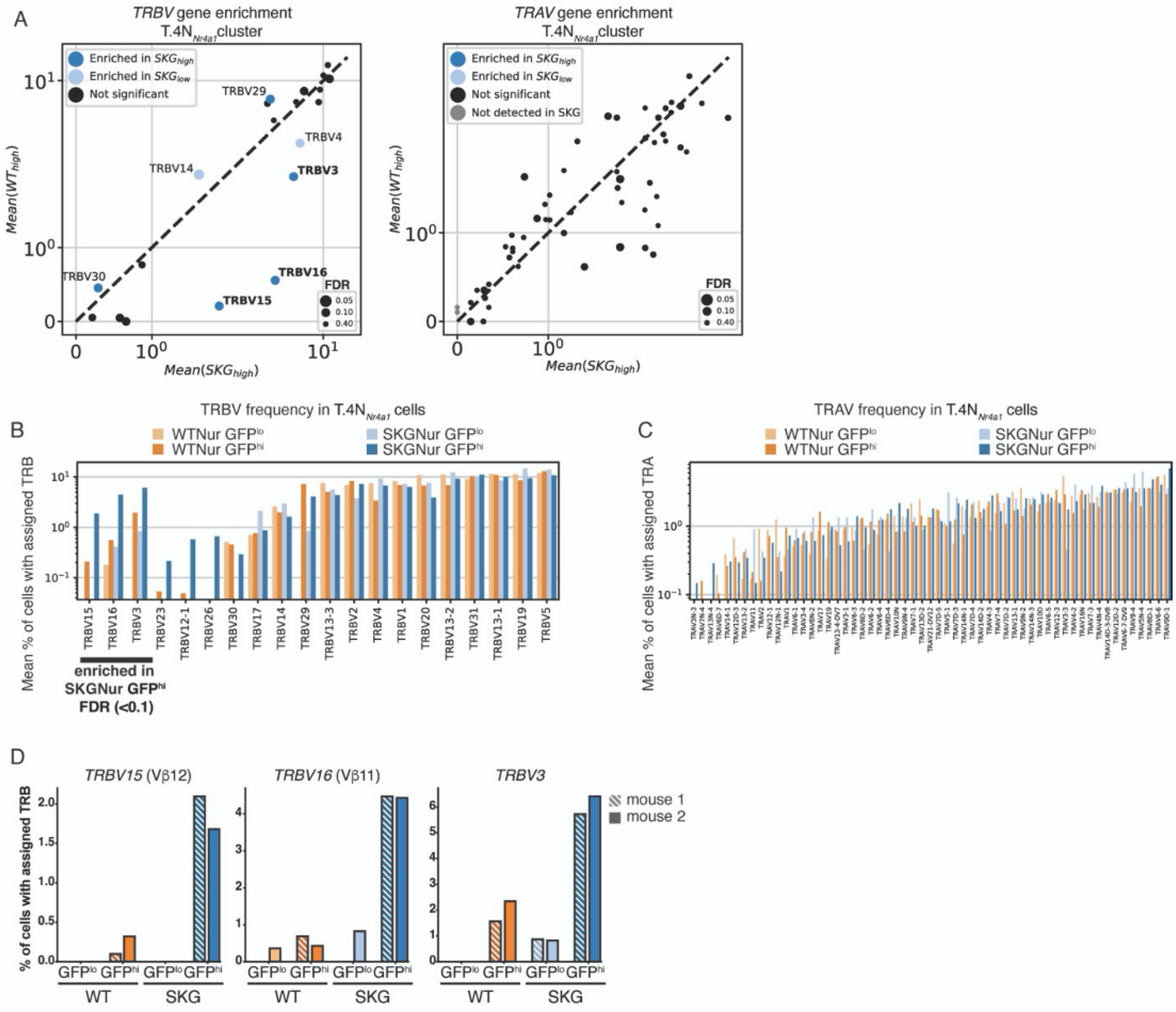
Supplemental Figure 4



Supplementary Figure 2.4 SKGNur mice express superantigens involved in selection of TCR variable beta repertoire

(A) Bar plot with percent of cells with paired TRA and TRB detection by cluster. (B) Bar plot of mean frequency of cells expressing each TRAV gene as a percentage of all cells in each sample with an assigned TRAV. Bars are colored according to subgroup and ordered by increasing overall frequency. (C) BALB/c and SKG tail DNA used in PCR reactions containing primers specific for the indicated Mtv pro-viruses. (D) Bar plots of frequency of cells expressing the indicated TRBV control genes not uniquely enriched in SKGNur GFP^{hi} cells for the two replicate mice in each subgroup.

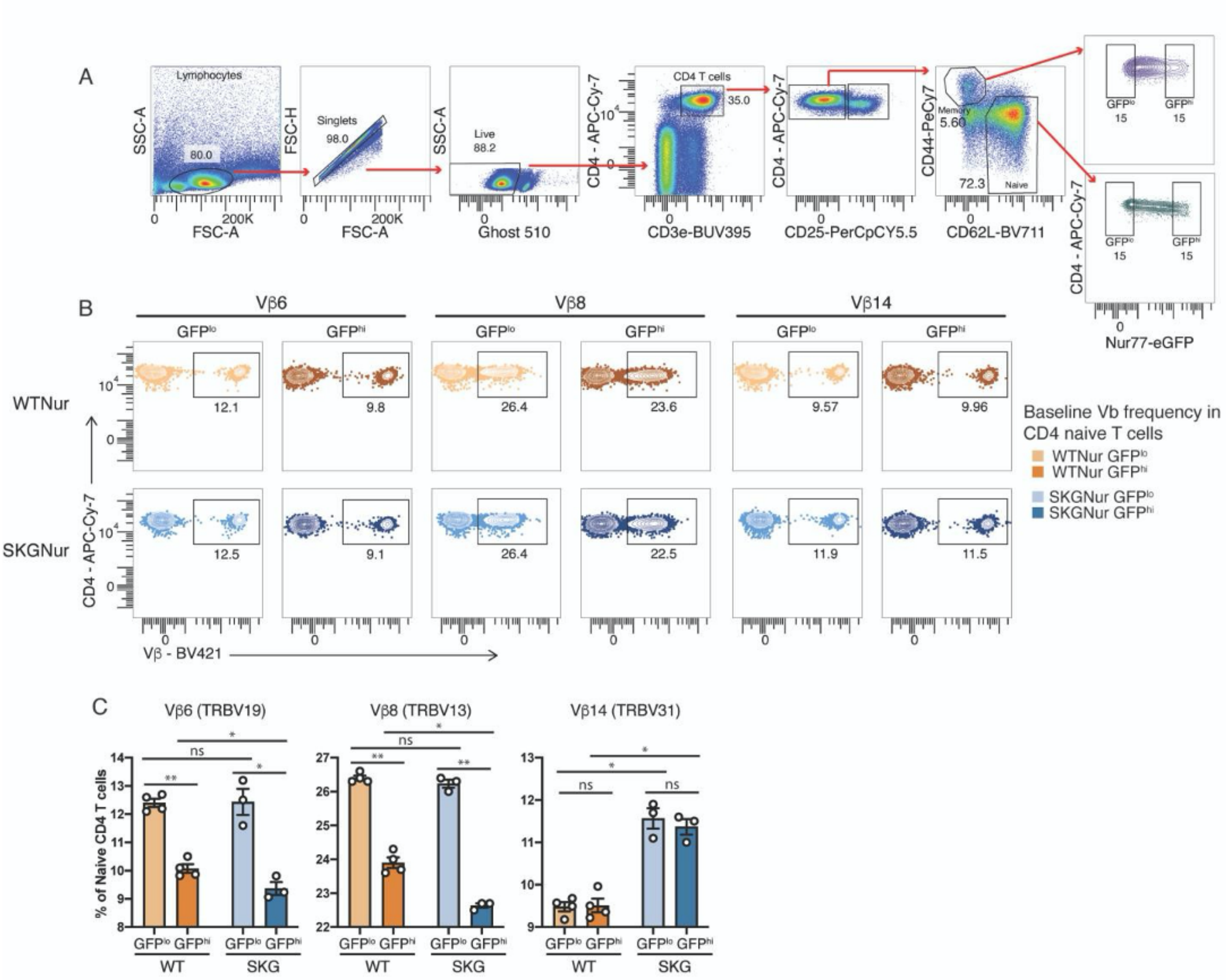
Supplemental Figure 5



Supplementary Figure 2.5 Further enrichment of biased TRBV in SKGNur GFP^{hi} T.4N$_{Nr4a1}$ cells

(A) Scatter plot of mean frequency of cells expressing each TRBV (left) or TRAV (right) gene for the SKGNur GFP^{hi} versus the WTNur GFP^{hi} T.4N$_{Nr4a1}$ cells. Dots for each TRV gene are sized according to the false discovery rate (FDR) from a paired one-sided t-test comparing frequency in SKGNur GFP^{hi} versus SKGNur GFP^{lo}. Dots are colored as either significantly enriched (FDR < 0.1) in SKGNur GFP^{hi} (dark blue), significantly enriched in SKGNur GFP^{lo} (light blue), or not significantly enriched in either subgroup (black). TRBV genes that were significantly enriched in SKGNur GFP^{hi} and were also more highly expressed in SKGNur GFP^{hi} versus WTNur GFP^{hi} T.4N$_{Nr4a1}$ cells are bolded. (B) Bar plot of mean value of T.4N$_{Nr4a1}$ cells expressing each TRBV gene as a percentage of all T.4N$_{Nr4a1}$ cells in each sample with an assigned TRBV. Bars are colored according to subgroup and are ordered with the TRBV genes enriched in SKGNur GFP^{hi} T.4N$_{Nr4a1}$ cells (see **A**) followed by the other TRBV genes ordered by increasing overall frequency. (C) Bar plot of mean value of T.4N$_{Nr4a1}$ cells expressing each TRAV gene as a percentage of all T.4N$_{Nr4a1}$ cells in each sample with an assigned TRAV. Bars are colored according to subgroup and are ordered by increasing overall frequency. (D) Bar plot of frequency of cells expressing indicated TRBV genes significantly enriched in SKGNur GFP^{hi} T.4N$_{Nr4a1}$ cells (see **A**) for the two replicate mice in each subgroup.

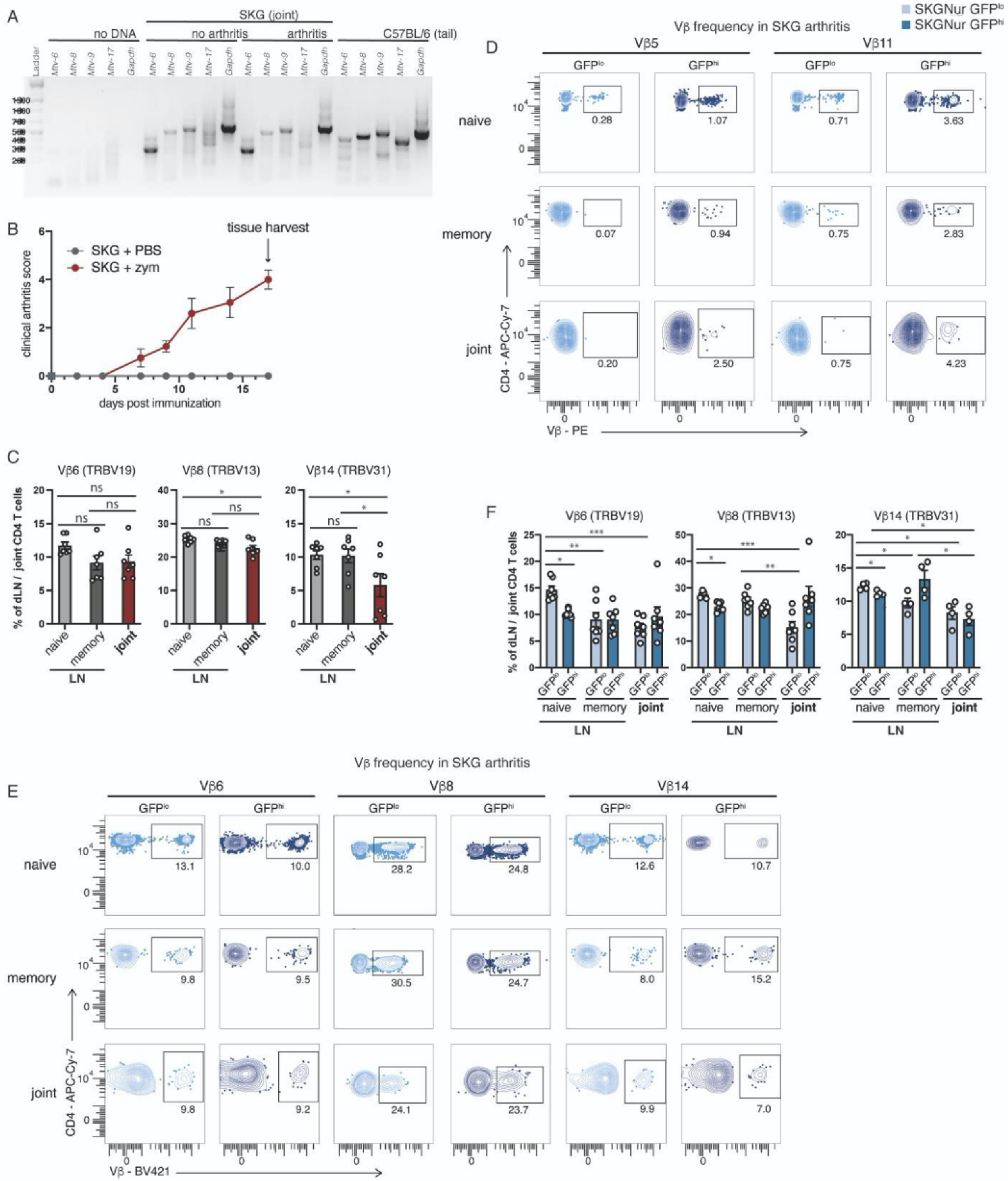
Supplemental Figure 6



Supplementary Figure 2.6 TCR Vbs unresponsive to BALB/c MMTV superantigen do not expand in peripheral T cells marked by TCR signaling reporter

(A) Flow cytometry gating used to identify GFP^{hi} and GFP^{lo} populations in naïve ($CD62L^{hi}CD44l^{o}$) and memory ($CD44^{hi}CD62L^{lo}$) CD4+CD25- T cells for Vb identification in WTNur and SKGNur lymphocytes. (B-C) Representative FACS plots (**B**) of naïve peripheral CD4 T cells with indicated TCR Vb protein usage determined by flow cytometry in GFP^{lo} and GFP^{hi} T cells from LN of WTNur and SKGNur mice prior to arthritis induction and quantified in (**C**) where bar graphs depict mean frequency (± SEM), n = 3-4 mice per group, experiment repeated at least 3 times. Significance indicated by asterisk for FDR (paired t-test) or *P* value (exact permutation test) < 0.05 (*), < 0.1 (**), or < 0.001 (***).

Supplemental Figure 7



Supplementary Figure 2.7 TCR Vbs unresponsive to BALB/c MMTV superantigen do not expand in SKG CD4 T cells after arthritis induction

(A) DNA was used from SKG joints ± arthritis in PCR reactions containing primers specific for the indicated Mtv pro-viruses. Lanes 26 show PCR mixtures lacking template DNA. C57BL/6 tail DNA was used as a positive control for *Mtv-8, -9,* -17 and a negative control for *Mtv*-6. Molecular size markers are shown in lane 1. Each gel is representative of at least 3-4 biological replicates per condition and genotype. (B) Arthritis score in SKGNur mice ± i.p. zymosan (red) or PBS (gray), n=4 mice in each group, representative of at least 3 experiments. (C) Bar graph depicts mean frequency (± SEM) of peripheral naïve or memory, or joint CD4 T cells with indicated TCR Vb protein usage determined by flow cytometry in CD4 T cells from draining LN or joints of SKGNur mice 2.5 weeks after arthritis induction with zymosan (as seen in **B**). (D-F) Representative FACS plots of peripheral naïve or memory, or joint CD4 T cells with indicated TCR Vb protein usage determined by flow cytometry in GFP^{lo} (light blue) and GFP^{hi} (dark blue) T cells from LN or joints of SKGNur mice 2.5 weeks after arthritis induction with zymosan and quantified in **Fig. 2.6E** or (**F**), respectively, where bar graphs depict mean frequency (± SEM), n = 7 mice per group pooled from 2 experiments. (**C**, **F**) Significance indicated by asterisk for *P* value (exact permutation test) or FDR (paired t-test) < 0.05 (*), < 0.1 (**), or < 0.001 (***)**.**

Materials and Methods

Antibodies and Reagents. Ghost Dye Violet 510 (Tonbo Biosciences: 13-0870-T100) was used for live/dead staining. The following antibodies were used for staining as indicated: CD3e-BUV395 (BD Bioscience: 563565, clone: 145-2C11), CD4-APCeFluor 780 (eBioscience: 47-0042-82, clone: RM4-5), CD25-PerCPCy5.5 (Tonbo 65-0251-U100, clone: PC61.5), CD44-PE-Cy7 (BioLegend: 103030, clone: IM7), CD62L-BV711 (BioLegend: 104445, clone: MEL-14), TCR Vβ3-PE (BD Bioscience: 553209, clone: KJ25), TCR Vβ5.1/5.2-PE (BD Bioscience: 562088, clone: MR9-4), TCR Vβ6-BV421 (BD Bioscience: 744590, clone: RR4-7), TCR Vβ8-BV421 (BD Bioscience: 742376, clone: F23), TCR Vβ11-PE (BD Bioscience: 553198, clone: RR3-15), TCR Vβ14-Biotin (BD Bioscience: 553257, clone: 14-2), Streptavidin-BV421 (BioLegend: 405226), FOXP3-eFluor 660 (eBioscience: 50-5773-82, clone: FJK-16s).

Mice. BALB/c and C57BL/6J mice were purchased from Jackson laboratory, and BALB/cNur77-eGFP and SKGNur77-eGFP mice were bred in our facility (University of California, San Francisco) as previously described (10). All mice were housed and bred in specific pathogen-free conditions in the Animal Barrier Facility at UCSF according to the University Animal Care Committee and NIH guidelines. All animal experiments were approved by the UCSF Institutional Animal Care and Use Committee.

Flow Cytometry and Cell Sorting. Cells were stained with antibodies of the indicated specificities and analyzed on a BD LSR Fortessa flow cytometer. Flow cytometry plots and analyses were performed using FlowJo v.10.8.0 (Tree Star). Cells were sorted to >95% purity using a MoFlo XDP (Beckman Coulter).

Murine Synovial Tissue Preparation. Synovial tissues from ankle joints were digested with 1 mg/mL Collagenase IV (Worthington: LS004188) and DNase I (Sigma: 4536282001) in RPMI 1640 medium for 2 h at 37 °C on a rotator then quenched with 10% fetal bovine serum in RPMI 1640 medium; digested cells were filtered through a 70 µm nylon mesh to prepare single cell suspensions.

Surface and Intracellular Staining. After live/dead staining with Ghost Dye Violet 510 as per manufacturer's instructions, cells were stained for surface markers, washed, and then fixed for 10 min with 4% (vol/vol) fresh paraformaldehyde at room temperature protected from light. Cells were then permeabilized using the Mouse Regulatory T-Cell Staining kit 1 (eBioscience: 00-5521-00) per manufacturer's instruction and then stained with FoxP3 e660.

***In vivo* Treatments.** Arthritis: Zymosan A (Sigma-Aldrich) suspended in saline at 10 mg/mL was kept in boiling water for 10 min. Zymosan A solution (2 mg) or saline was intraperitoneally injected into 8–12-week-old mice. Antiretroviral therapy: 5–7-week-old SKG mice were administered Truvada combination therapy with emtricitabine (Sigma-Aldrich) and tenofovir disoproxil fumarate (Acros organics) in a 1:1 ratio (0.5 mg/mL in diH2O for each drug) or vehicle control. Solution was added to the drinking water supply and changed once per week. Mice were also given an intraperitoneal bolus injection x1 of Truvada (~160 mg/kg) or vehicle control in 200 ul PBS at start of treatment. Drinking water dosage with Truvada or diH2O continued throughout the arthritis course.

Statistics. Flow cytometry data were analyzed by comparison of means using paired or unpaired 2-tailed Student's *t*-tests using Prism v.9.2.0 or v.9.3.1 for Mac (GraphPad Software). Unpaired t-test with Welch's correction was used to calculate differences in arthritis scores and log-rank Mantel-Cox test used to calculate differences in Kaplan Meier Survival. Data in all figures represent mean ± SEM unless otherwise indicated. Differences were considered significant at $P < 0.05$: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$.

PCR and RT-PCR. BALB/cJ, C57BL/6J, and SKG tail DNA was typed for *Mtv-6, -8, -9,* and *-17*. Standard PCR protocols were used for preparing PCR mixtures. Primer pairs for the detection of MMTV proviruses were previously described (55). GAPDH primers used: (5' CATGTTTGTGATGGGTGTGAACCA 3') and (5' GTTGCTGTAGCCGTATTCATTGTC 3'). PCR mixtures for *Mtv-6, -8, and -9* were incubated at 94°C for 5 min, then denatured for 44 cycles at 94°C for 1 min, annealed at 46°C for 1 min, polymerized at 72°C for 1 min, and then incubated at 72°C for 5 min. PCRs for *Mtv-17* were conducted similarly except for an annealing temperature of 50°C. Samples were run on 2% agarose gel.

RT-PCR with Joints. Single cell suspensions of synovial tissues from SKG ankle joints were spun down at 1500 RPM at 4C. Cell pellets were flash frozen using dry ice in ethyl alcohol. Frozen cell pellets were used with the RNeasy Mini Kit (Qiagen: 74106) for RNA purification. The qScript cDNA Synthesis Kit (Quantabio: 95047-100) was used for cDNA library synthesis from purified total RNA. RT-PCR was conducted as described previously for PCR.

Bulk RNA Sequencing. Negatively selected CD4 T cells from the lymph node were sorted for $CD62L^{hi}CD44^{lo}CD25^{-}$ and the 10% highest (GFP^{hi}) or lowest (GFP^{lo}) expressing T cells. Cells were washed, pelleted and immediately flash frozen using dry ice in ethyl alcohol. Samples were processed for bulk RNA-sequencing by Q2 solutions using the TruSeq Stranded mRNA kit (Illumina: RS-122-2103) for library preparation. The resulting libraries pool into three batches and sequenced on a Illumina HiSeq 2500 sequencer over three lanes.

Alignment and Initial Processing of Bulk RNA Sequencing Data. The raw fastq files were clipped and filtered using fastq-mcf v.1.04.636 to remove low quality reads and bases, homopolymers, and adapter sequences. The filtered reads were aligned using the STAR v.2.4 (70) with the default settings to the mm10 transcriptome and the resulting bam files were converted to count matrices for each sample with RSEM v.1.2.14. Genes with less than 10 counts across all the samples were filtered out. Raw counts were normalized and transformed by the variance stabilizing transformation (VST) function from DESeq2 v.1.22.2 (71).

Principal Component Analysis. The VST normalized features were used for principal component analysis with the function plotPCA from DESeq2.

Bulk RNA Sequencing Differential Expression. Differential gene expression for the bulk RNA sequencing samples was performed with the raw counts from the filtered gene list for the indicated samples as the inputs. The analysis was run using a negative binomial model with multiple testing correction with Benjamini-Hochberg implemented via the DESeq function which includes an internal normalization from DESeq2. For differential gene expression between samples within the same genotype, mouse identity was included as a covariate.

Functional Enrichment Analysis. The collection of 991 significantly differentially expressed genes (log2FC > 1 and adjusted p value < 0.05) from the four comparisons [SKGNur GFP^{hi} versus SKGNur GFP^{lo} , WTNur GFP^{hi} versus WTNur GFP^{lo}, SKGNur GFP^{hi} versus WTNur GFP^{hi}, SKGNur GFP^{lo} versus WTNur GFP^{lo}] were hierarchically clustered using the Ward linkage (“ward.D2”) with the R package pheatmap v.1.0.12. The resulting dendrogram was used to partition the differentially expressed gene list into six gene modules. The gene lists for each gene module were analyzed using the functional profiling g:GOSt tool from g:Profiler (version e102_eg49_p15_e7ff1c9) with g:SCS multiple testing correction method applying significance threshold of 0.05. Select significantly enriched pathways from the GO:BP or KEGG collections were reported.

Gene Set Enrichment Analysis. For the bulk RNA differential expression, the differential gene list was filtered to remove genes with NA for the adjusted p value or log fold change. For the single-cell RNA differential expression, the differential gene list was filtered to only include genes which were expressed in at least 1% of cells in the T.4N$_{Nr4a1}$ cluster. These filtered gene list were used to create ranked gene lists with the sign(log fold change) times the -log10(raw p value) as the ranking metric. The ranked list was used as input to look for gene set enrichment in the indicated collection of pathways in the 'classic' mode with the GSEAPreranked tool from GSEA v.4.1.0 with the default settings. For pathway collections of human genes, the 'Mouse_Gene_Symbol_Remapping_Human_Orthologs_MSigDB' chip file was used to map mouse genes from the ranked gene list to the human orthologs. Mouse gene symbols that mapped to the same human symbol were collapsed based on the max rank.

Single-cell RNA and TCR Sequencing. Negatively selected CD4 T cells from the lymph node and spleen were sorted for CD62L^{hi}CD44^{lo}CD25^{-} and the 10% highest (GFP^{hi}) or lowest (GFP^{lo}) expressing T cells. Droplet-based paired single-cell RNA and TCR sequencing was performed using the 10x single-cell 5'+V(D)J v.1 kit per manufacturer's instructions. The resulting cDNA libraries were sequenced on four lanes of an Illumina Novaseq 6000 sequencer to yield gene expression (GEX) and T cell receptor (TCR) fastqs.

Alignment and Initial Processing of Single-cell Sequencing Data. The raw fastq files were aligned using CellRanger v3.0.1 and 3.0.2 software with the default settings to the mm10 transcriptome with the addition of the sequence for the eGFP transcript and the vdj GRCm38 v 3.1.0 reference for the GEX and TCR fastqs, respectively.

***eGFP* Transcript Sequence.**

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG

GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC
CACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCC
CTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCC
CGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAG
GAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGT
CTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCA
CAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCA
TCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCC
TGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

Cell Type Classification and Clustering. We filtered out 721 cells with less than 100 or more than 3000 genes detected and filtered out 14,388 genes detected in less than 3 cells. We also filtered out 1,066 cells with more than 10% of total counts (UMIs) mapping to mitochondrial genes and 1008 cells determined to be contaminating B cells based on CD19 expression. The raw counts were normalized to 10,000 counts per cell and log(count + 1) transformed. For technical and batch correction, we regressed out total UMI counts and % counts mapping to mitochondrial genes and used combat for batch correction with each sample as a batch. We identified 1119 highly variable genes (excluding all Trav and Trbv genes to avoid clustering cells based on expression of those genes) which were scaled and used with the default settings in scanpy v.1.4.3 (72) for PCA analysis followed by leiden clustering after nearest neighbor detection and UMAP projection. This analysis identified 13 clusters which we collapsed into 9 cell subtypes based on differential gene analysis.

Single-cell Differential Expression Analysis. Single-cell differential expression was performed using the log-normalized gene counts with the rank_genes_groups function from scanpy with the Wilcoxon rank-sum method and multiple testing correction with Benjamini-Hochberg. Additionally, the adjusted p values that were equal to 0 were updated to the minimum representable positive normalized float (2.2250738585072014e-308).

Cell Cycle Phase Assignment and Module Scoring. To assign cells to the cell cycle phases, the log-normalized scaled gene counts were used with the score_genes_cell_cycle function from the scanpy v.1.5.1 package with the *Mus musculus* G1/S DNA Damage Checkpoints and G2/M Checkpoints gene lists from the REACTOME database being used for the genes associated to the S phase and genes associated to the G2M phase (73, 74), respectively. For the single cell scoring of the bulk RNA sequencing gene modules, the log-normalized scaled gene counts were used with the score_genes function from scanpy.

RNA Velocity Analysis. For each 10x well, we used velocyto v.0.17.17 (75) to create a loom file with the spliced, unspliced, and ambiguous counts with the Dec. 2011 GRCm38/mm10 repeat masking gtf file from the UCSC genome browser (76, 77). The loom files across all wells were merged and then subsetted to all cells in the $\mathrm{T.4N}_{\mathit{Nr4a1}}$ cluster. The resulting object was used to determine the RNA velocity and to predict the latent time for each cell using the 1119 HVGs with the dynamical model from scvelo v.0.2.1.

We used we used a Gaussian mixture model with the GaussianMixture tool from sklearn v.0.23.1 (78) to deconvolute the underlying individual Gaussian distributions from the latent time distribution for cells from the $\mathrm{T.4N}_{\mathit{Nr4a1}}$ cluster. This separated the cells into an optimal number of

4 distributions or clusters as determined by the elbow of the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) plots.

The smoothed gene expression versus latent time was modeled using a linear generalized additive model using default settings with the LinearGAM function from pygam v.0.8.0 (79). For trajectory inference between the four clusters ("Stage 1" – "Stage 4"), we used the graph-based tool PAGA within scvelo to predict velocity-inferred transitions among the clusters. The latent time distributions from different subgroups were compared using the Kolmogorov-Smirnov test. The cell cycle distributions between subgroups within stage 1 were compared using Pearson's chi-squared test.

TCR Analysis. Cells with <=2 TRA chains and <=1 TRB chains were used in the TCR clonotype analyses (47). Cells with two TRA chains were removed for the TRBV and TRAV analyses since the highest frequency for any dual TRA was 0.09% in any one sample (~1 cell). This removed 10,598 cells or 13.6% of all cells which is consistent with the expected dual TRA frequency. *TRBV* and *TRAV* genes which were not present in at least two mice from the same subgroup (i.e., SKGNur GFP^{hi}, WTNur GFP^{hi}, SKGNur GFP^{lo}, and WTNur GFP^{lo}) were removed from the downstream *TRBV* and *TRAV* analyses.

Significant differences in the TRBV frequencies between subgroups was determined by exact permutation test for unpaired samples and exact permutation test (N > 5 paired samples) (80) or paired t-test (N <= 5 paired samples) using scipy v.1.4.1 followed by Benjamini-Hochberg correction with statsmodels v.0.11.1 for paired samples.

Significant difference in GFP mean fluorescence intensity (MFI) for cells assigned TRBVs in the enriched or not-enriched groups was determined with a linear mixed effect model GFP MFI ~

TRBV group (enriched or not-enriched) + mouse id with a random intercept for each TRBV protein.

References

1. Cope AP. T cells in rheumatoid arthritis. Arthritis research & therapy. 2008;10 Suppl 1(S1).
2. Schmidt D, Goronzy JJ, Weyand CM. CD4+ CD7- CD28- T cells are expanded in rheumatoid arthritis and are characterized by autoreactivity. The Journal of clinical investigation. 1996;97(9):2027-37.
3. Thomas R, McIlraith M, Davis LS, Lipsky PE. Rheumatoid synovium is enriched in CD45RBdim mature memory T cells that are potent helpers for B cell differentiation. Arthritis and rheumatism. 1992;35(12):1455-65.
4. Weyand CM, Fujii H, Shao L, Goronzy JJ. Rejuvenating the immune system in rheumatoid arthritis. Nature reviews Rheumatology. 2009;5(10):583-8.
5. Zhang Z, Gorman CL, Vermi AC, Monaco C, Foey A, Owen S, et al. TCRzetadim lymphocytes define populations of circulating effector cells that migrate to inflamed tissues. Blood. 2007;109(10):4328-35.
6. Gringhuis SI, Papendrecht-van der Voort EA, Leow A, Nivine Levarht EW, Breedveld FC, Verweij CL. Effect of redox balance alterations on cellular localization of LAT and downstream T-cell receptor signaling pathways. Molecular and cellular biology. 2002;22(2):400-11.
7. Maurice MM, Lankester AC, Bezemer AC, Geertsma MF, Tak PP, Breedveld FC, et al. Defective TCR-mediated signaling in synovial T cells in rheumatoid arthritis. Journal of immunology (Baltimore, Md : 1950). 1997;159(6):2973-8.
8. Romagnoli P, Strahan D, Pelosi M, Cantagrel A, van Meerwijk JP. A potential role for protein tyrosine kinase p56(lck) in rheumatoid arthritis synovial fluid T lymphocyte hyporesponsiveness. International immunology. 2001;13(3):305-12.

9. Sakaguchi N, Takahashi T, Hata H, Nomura T, Tagami T, Yamazaki S, et al. Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. Nature. 2003;426(6965):454-60.
10. Ashouri JF, Hsu L-Y, Yu S, Rychkov D, Chen Y, Cheng DA, et al. Reporters of TCR signaling identify arthritogenic T cells in murine and human autoimmune arthritis. Proceedings of the National Academy of Sciences. 2019;116(37):18517-27.
11. Tanaka S, Maeda S, Hashimoto M, Fujimori C, Ito Y, Teradaira S, et al. Graded attenuation of TCR signaling elicits distinct autoimmune diseases by altering thymic T cell selection and regulatory T cell function. Journal of immunology (Baltimore, Md : 1950). 2010;185(4):2295-305.
12. Hsu LY, Tan YX, Xiao Z, Malissen M, Weiss A. A hypomorphic allele of ZAP-70 reveals a distinct thymic threshold for autoimmune disease versus autoimmune reactivity. J Exp Med. 2009;206(11):2527-41.
13. Yoshitomi H, Sakaguchi N, Kobayashi K, Brown GD, Tagami T, Sakihama T, et al. A role for fungal {beta}-glucans and their receptor Dectin-1 in the induction of autoimmune arthritis in genetically susceptible mice. J Exp Med. 2005;201(6):949-60.
14. Evans RM. The steroid and thyroid hormone receptor superfamily. Science (New York, NY). 1988;240(4854):889-95.
15. Olefsky JM. Nuclear receptor minireview series. The Journal of biological chemistry. 2001;276(40):36863-4.
16. Deane KD, Holers VM. The Natural History of Rheumatoid Arthritis. Clin Ther. 2019;41(7):1256-69.

17. Lima-Junior DS, Krishnamurthy SR, Bouladoux N, Collins N, Han SJ, Chen EY, et al. Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. Cell. 2021;184(14):3794-811.e19.
18. Sharon D, Chen M, Zhang G, Girgis S, Sis B, Graham D, et al. Impact of combination antiretroviral therapy in the NOD.c3c4 mouse model of autoimmune biliary disease. Liver Int. 2015;35(4):1442-50.
19. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic acids research. 2019;47(W1):W191-w8.
20. Ashouri JF, Weiss A. Endogenous Nur77 Is a Specific Indicator of Antigen Receptor Signaling in Human T and B Cells. Journal of immunology (Baltimore, Md : 1950). 2017;198(2):657-68.
21. Huang B, Pei HZ, Chang HW, Baek SH. The E3 ubiquitin ligase Trim13 regulates Nur77 stability via casein kinase 2alpha. Scientific reports. 2018;8(1):13895.
22. Tan C, Hiwa R, Mueller JL, Vykunta V, Hibiya K, Noviski M, et al. NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. Nature immunology. 2020;21(10):1267-79.
23. Zhang L, Xie F, Zhang J, Dijke PT, Zhou F. SUMO-triggered ubiquitination of NR4A1 controls macrophage cell death. Cell Death Differ. 2017;24(9):1530-9.
24. Jennings E, Elliot TAE, Thawait N, Kanabar S, Yam-Puc JC, Ono M, et al. Nr4a1 and Nr4a3 Reporter Mice Are Differentially Sensitive to T Cell Receptor Signal Strength and Duration. Cell reports. 2020;33(5):108328.

25. Myers DR, Lau T, Markegard E, Lim HW, Kasler H, Zhu M, et al. Tonic LAT-HDAC7 Signals Sustain Nur77 and Irf4 Expression to Tune Naive CD4 T Cells. Cell reports. 2017;19(8):1558-71.
26. Myers DR, Zikherman J, Roose JP. Tonic Signals: Why Do Lymphocytes Bother? Trends in immunology. 2017;38(11):844-57.
27. Kalekar LA, Schmiel SE, Nandiwada SL, Lam WY, Barsness LO, Zhang N, et al. CD4(+) T cell anergy prevents autoimmunity and generates regulatory T cell precursors. Nature immunology. 2016;17(3):304-14.
28. Mueller DL. E3 ubiquitin ligases as T cell anergy factors. Nature immunology. 2004;5(9):883-90.
29. Zheng Y, Zha Y, Driessens G, Locke F, Gajewski TF. Transcriptional regulator early growth response gene 2 (Egr2) is required for T cell anergy in vitro and in vivo. J Exp Med. 2012;209(12):2157-63.
30. Crawford A, Angelosanto JM, Kao C, Doering TA, Odorizzi PM, Barnett BE, et al. Molecular and transcriptional basis of CD4(+) T cell dysfunction during chronic infection. Immunity. 2014;40(2):289-302.
31. Trefzer A, Kadam P, Wang SH, Pennavaria S, Lober B, Akcabozan B, et al. Dynamic adoption of anergy by antigen-exhausted CD4(+) T cells. Cell reports. 2021;34(6):108748.
32. Chen J, Lopez-Moyado IF, Seo H, Lio CJ, Hempleman LJ, Sekiya T, et al. NR4A transcription factors limit CAR T cell function in solid tumours. Nature. 2019;567(7749):530-4.

33. Zinzow-Kramer WM, Weiss A, Au-Yeung BB. Adaptation by naïve CD4(+) T cells to self-antigen-dependent TCR signaling induces functional heterogeneity and tolerance. Proc Natl Acad Sci U S A. 2019;116(30):15160-9.

34. Nguyen TTT, Wang ZE, Shen L, Schroeder A, Eckalbar W, Weiss A. Cbl-b deficiency prevents functional but not phenotypic T cell anergy. J Exp Med. 2021;218(7):e20202477.

35. Liebmann M, Hucke S, Koch K, Eschborn M, Ghelman J, Chasan AI, et al. Nur77 serves as a molecular brake of the metabolic switch during T cell activation to restrict autoimmunity. Proc Natl Acad Sci U S A. 2018;115(34):E8017-e26.

36. Liu X, Wang Y, Lu H, Li J, Yan X, Xiao M, et al. Genome-wide analysis identifies NR4A1 as a key mediator of T cell dysfunction. Nature. 2019;567(7749):525-9.

37. Hata H, Sakaguchi N, Yoshitomi H, Iwakura Y, Sekikawa K, Azuma Y, et al. Distinct contribution of IL-6, TNF-alpha, IL-1, and IL-10 to T cell-mediated spontaneous autoimmune arthritis in mice. The Journal of clinical investigation. 2004;114(4):582-8.

38. Hirota K, Hashimoto M, Yoshitomi H, Tanaka S, Nomura T, Yamaguchi T, et al. T cell self-reactivity forms a cytokine milieu for spontaneous development of IL-17+ Th cells that cause autoimmune arthritis. J Exp Med. 2007;204(1):41-7.

39. Ciofani M, Madar A, Galan C, Sellars M, Mace K, Pauli F, et al. A validated regulatory network for Th17 cell specification. Cell. 2012;151(2):289-303.

40. Yu CR, Mahdi RM, Ebong S, Vistica BP, Gery I, Egwuagu CE. Suppressor of cytokine signaling 3 regulates proliferation and activation of T-helper cells. The Journal of biological chemistry. 2003;278(32):29752-9.

41. Ye H, Zhang J, Wang J, Gao Y, Du Y, Li C, et al. CD4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: evidence from a case-control study. Arthritis research & therapy. 2015;17:76.
42. Atsumi T, Ishihara K, Kamimura D, Ikushima H, Ohtani T, Hirota S, et al. A point mutation of Tyr-759 in interleukin 6 family cytokine receptor subunit gp130 causes autoimmune arthritis. J Exp Med. 2002;196(7):979-90.
43. Shouda T, Yoshida T, Hanada T, Wakioka T, Oishi M, Miyoshi K, et al. Induction of the cytokine signal regulator SOCS3/CIS3 as a therapeutic strategy for treating inflammatory arthritis. The Journal of clinical investigation. 2001;108(12):1781-8.
44. Wong PK, Egan PJ, Croker BA, O'Donnell K, Sims NA, Drake S, et al. SOCS-3 negatively regulates innate and adaptive immune mechanisms in acute IL-1-dependent inflammatory arthritis. The Journal of clinical investigation. 2006;116(6):1571-81.
45. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nature biotechnology. 2020;38(12):1408-14.
46. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome biology. 2019;20(1):59.
47. Abdelnour A, Bremell T, Holmdahl R, Tarkowski A. Clonal expansion of T lymphocytes causes arthritis and mortality in mice infected with toxic shock syndrome toxin-1-producing staphylococci. European journal of immunology. 1994;24(5):1161-6.
48. Kappler J, Kotzin B, Herron L, Gelfand EW, Bigler RD, Boylston A, et al. V beta-specific stimulation of human T cells by staphylococcal toxins. Science (New York, NY). 1989;244(4906):811-3.

49. Hodes RJ, Abe R. Mouse endogenous superantigens: Ms and Mls-like determinants encoded by mouse retroviruses. Current protocols in immunology. 2001;Appendix 1:Appendix 1F.
50. Matsutani T, Ohmori T, Ogata M, Soga H, Yoshioka T, Suzuki R, et al. Alteration of T-cell receptor repertoires during thymic T-cell development. Scandinavian journal of immunology. 2006;64(1):53-60.
51. Czarneski J, Rassa JC, Ross SR. Mouse mammary tumor virus and the immune system. Immunol Res. 2003;27(2-3):469-80.
52. Acha-Orbea H, Shakhov AN, Scarpellino L, Kolb E, Müller V, Vessaz-Shaw A, et al. Clonal deletion of V beta 14-bearing T cells in mice transgenic for mammary tumour virus. Nature. 1991;350(6315):207-11.
53. Choi Y, Kappler JW, Marrack P. A superantigen encoded in the open reading frame of the 3' long terminal repeat of mouse mammary tumour virus. Nature. 1991;350(6315):203-7.
54. Frankel WN, Rudy C, Coffin JM, Huber BT. Linkage of Mls genes to endogenous mammary tumour viruses of inbred mice. Nature. 1991;349(6309):526-8.
55. Barnett A, Mustafa F, Wrona TJ, Lozano M, Dudley JP. Expression of mouse mammary tumor virus superantigen mRNA in the thymus correlates with kinetics of self-reactive T-cell loss. Journal of virology. 1999;73(8):6634-45.
56. Bour-Jordan H, Esensten JH, Martinez-Llordella M, Penaranda C, Stumpf M, Bluestone JA. Intrinsic and extrinsic control of peripheral T-cell tolerance by costimulatory molecules of the CD28/B7 family. Immunological reviews. 2011;241(1):180-205.
57. Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. Nature reviews Immunology. 2013;13(4):227-42.

58. Fife BT, Bluestone JA. Control of peripheral T-cell tolerance and autoimmunity via the CTLA-4 and PD-1 pathways. Immunological reviews. 2008;224:166-82.
59. Lee J, Su EW, Zhu C, Hainline S, Phuah J, Moroco JA, et al. Phosphotyrosine-dependent coupling of Tim-3 to T-cell receptor signaling pathways. Molecular and cellular biology. 2011;31(19):3963-74.
60. Ito Y, Hashimoto M, Hirota K, Ohkura N, Morikawa H, Nishikawa H, et al. Detection of T cell responses to a ubiquitous cellular protein in autoimmune disease. Science (New York, NY). 2014;346(6207):363-8.
61. Lotz M, Jirik F, Kabouridis P, Tsoukas C, Hirano T, Kishimoto T, et al. B cell stimulating factor 2/interleukin 6 is a costimulant for human thymocytes and T lymphocytes. J Exp Med. 1988;167(3):1253-8.
62. Rochman I, Paul WE, Ben-Sasson SZ. IL-6 increases primed cell expansion and survival. Journal of immunology (Baltimore, Md : 1950). 2005;174(8):4761-7.
63. Holsti MA, McArthur J, Allison JP, Raulet DH. Role of IL-6, IL-1, and CD28 signaling in responses of mouse CD4+ T cells to immobilized anti-TCR monoclonal antibody. Journal of immunology (Baltimore, Md : 1950). 1994;152(4):1618-28.
64. Howell MD, Diveley JP, Lundeen KA, Esty A, Winters ST, Carlo DJ, et al. Limited T-cell receptor beta-chain heterogeneity among interleukin 2 receptor-positive synovial T cells suggests a role for superantigen in rheumatoid arthritis. Proc Natl Acad Sci U S A. 1991;88(23):10921-5.
65. Paliard X, West SG, Lafferty JA, Clements JR, Kappler JW, Marrack P, et al. Evidence for the effects of a superantigen in rheumatoid arthritis. Science (New York, NY). 1991;253(5017):325-9.

66. Jenkins RN, Nikaein A, Zimmermann A, Meek K, Lipsky PE. T cell receptor V beta gene bias in rheumatoid arthritis. The Journal of clinical investigation. 1993;92(6):2688-701.
67. Bhardwaj N, Hodtsev AS, Nisanian A, Kabak S, Friedman SM, Cole BC, et al. Human T-cell responses to Mycoplasma arthritidis-derived superantigen. Infection and immunity. 1994;62(1):135-44.
68. Grom AA, Thompson SD, Luyrink L, Passo M, Choi E, Glass DN. Dominant T-cell-receptor beta chain variable region V beta 14+ clones in juvenile rheumatoid arthritis. Proc Natl Acad Sci U S A. 1993;90(23):11104-8.
69. Zhao YX, Brunsberg U, Holmdahl R, Tarkowski A. V beta 11+ T-lymphocyte expansion by toxic shock syndrome toxin-1 differs in mice bearing H-2q versus H-2b haplotypes. Immunology. 1998;94(1):1-4.
70. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15-21.
71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.
72. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome biology. 2018;19(1):15.
73. G1/S DNA Damage Checkpoints. Reactome, v.75, https://reactome.org/content/detail/R-MMU-69615 [accessed January 16, 2021]. [Internet]. 2021 [cited January 16, 2021]. Available from: https://reactome.org/content/detail/R-MMU-69615.

74. G2/M Checkpoints. Reactome, v.75, https://reactome.org/content/detail/R-MMU-69473 [accessed January 16, 2021]. [Internet]. 2021 [cited January 16, 2021]. Available from: https://reactome.org/content/detail/R-MMU-69473.

75. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;560(7719):494-8.

76. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic acids research. 2004;32(Database issue):D493-6.

77. University of California Santa Cruz. Genome Browser 2000. updated July 20, 2021. Available from: http://genome.ucsc.edu

78. Pedregosa et al. Scikit-learn: Machine Learning in Python. JMLR 12, pp. 2825-2830 2011. Available from: https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

79. Servén D. BC. dswah/pyGAM: v0.8.0. Zenodo. 2018. Available from: https://doi.org/10.5281/zenodo.1476122.

80. Thompson WL, White GC, Gowan C. Detection of a Trend in Population Estimates. In: Thompson WL, White GC, Gowan C, editors. Monitoring Vertebrate Populations: Academic Press; 1998. p. 145-69.

Chapter 3: A Cytotoxic-skewed Immune Set Point Predicts Low Neutralizing Antibody Levels after Zika Virus Infection

Elizabeth E. McCarthy[1], Pamela M. Odorizzi[2,‡], Emma Lutz[2], Carolyn P. Smullin[2], Iliana Tenvooren[1], Mars Stone[3,4], Graham Simmons[3,4], Peter W. Hunt[2], Margaret E. Feeney[2,5], Philip J. Norris[2,3,4], Michael P. Busch[3,4], Matthew H. Spitzer[1,6-8]†, Rachel L. Rutishauser[2,6]†

[1]Departments of Otolaryngology-Head and Neck Surgery and of Microbiology and Immunology, University of California San Francisco; San Francisco, CA, 94143, USA.

[2]Department of Medicine, Zuckerberg San Francisco General Hospital, University of California San Francisco; San Francisco, CA, 94110, USA.

[3]Vitalant Research Institute; San Francisco, CA, 94104, USA.

[4]Department of Laboratory Medicine, University of California San Francisco; San Francisco, CA, 94143, USA.

[5]Department of Pediatrics, University of California San Francisco; San Francisco, CA, 94110, USA.

[6]Gladstone-UCSF Institute for Genomic Immunology, San Francisco, CA 94158, USA.

[7]Parker Institute for Cancer Immunotherapy; San Francisco, CA, 94143, USA.

[8]Chan Zuckerberg Biohub; San Francisco, CA, 94158, USA

‡Current address: Gilead Sciences, Inc.; Foster City, CA, 94404, USA.

†Co-corresponding authors

Introduction

Infection of pregnant individuals with Zika virus (ZIKV), a flavivirus primarily transmitted to humans via the bite of an infected mosquito, can lead to persistent viral replication in the placenta and fetal brain that is associated with devastating fetal neurologic outcomes (1–4). In contrast, for the majority of non-pregnant immunocompetent adults, ZIKV virus is rapidly cleared from the plasma (5–9), and infection is accompanied by mild symptoms such as fever, rash, and joint pain or can be asymptomatic (10,11). Since the recent 2015-2016 epidemic in the Americas, there has been a considerable effort towards the development of a ZIKV vaccine, particularly for the prevention of mother-to-child transmission of infection (12–14). The majority of ZIKV vaccine candidates aim to induce durable, high-titer neutralizing antibody responses, which confer protection in animal models (15,16).

Natural infection with ZIKV in humans generates robust ZIKV-specific antibody responses (11,17); however, there is wide inter-individual variation in the levels of ZIKV-specific antibodies that persist in the serum (11,18). Immunity to subsequent infection with ZIKV is likely to be influenced by the magnitude and durability of the ZIKV neutralizing antibody response (17,19,20), but little is known about the factors that contribute to inter-individual variation in antibody responses. There is substantial cross-reactivity between virus-specific antibodies (18,21,22) and T cell responses (23–25) generated after infection with ZIKV and those from the closely-related and often co-circulating dengue virus (DENV). However, prior DENV exposure alone does not appear to explain the wide range of ZIKV antibody titers observed after natural infection (18).

For other pathogens, baseline immune characteristics and/or signatures of early immune responses acutely after infection or vaccination have been shown to correlate with the magnitude of pathogen-specific antibody titers (26–33). Some aspects of the innate cytokine and cellular immune responses to ZIKV infection have been described in humans (34–40). However, the relationship between the acute-phase immune response and the generation of ZIKV-specific antibodies has not been characterized. This is in part due to the inherent challenges in identifying and establishing longitudinal cohorts of individuals identified during the earliest days of the acute phase of a natural infection.

Here, we used high-dimensional single-cell profiling with mass cytometry (CyTOF) to deeply characterize the cellular innate and adaptive immune response during acute and convalescent ZIKV infection. We evaluated longitudinal peripheral blood samples collected from 25 individuals in a natural history cohort of healthy, non-pregnant adults from Puerto Rico who were found to be viremic with ZIKV at the time of blood donation during the recent ZIKV epidemic of 2015-2016 (9,41,42). We found broadly coordinated cellular responses across immune cell lineages during acute ZIKV infection and identified distinct cellular immune signatures during acute ZIKV infection that were associated with the development and persistence of low versus high neutralizing antibody titers. In addition, we identified stable immune features that comprise a cytotoxic immune set point associated with low neutralizing antibody titers. Future vaccine efficacy trials for ZIKV and other acute viral infections may benefit from the inclusion of these candidate cellular biomarkers to aid in the prediction of neutralizing antibody titers, and additional strategies may be required to elicit stronger antibody responses in individuals with cytotoxic-skewed baseline immune set points.

Results

Identifying Immune Cell Populations that Respond to Acute ZIKV Infection

To characterize the cellular immune response to acute ZIKV infection, we designed two CyTOF antibody panels to phenotype innate immune and B cell features (panel 1) and T cell features (panel 2; see **Table 3.1**). We used these panels to analyze peripheral blood mononuclear cells (PBMCs) collected longitudinally at up to three timepoints during acute, early, and late convalescent phases of infection from 25 otherwise healthy blood donors in Puerto Rico who were found to be viremic for ZIKV at the time of blood donation ("index visit"; study participants are part of a larger REDS-III cohort; **Fig. 3.1A, Table 3.2**). 28% (7 of 25) of the participants were female, and the median age was 45 years (range 21-71). All participants mounted a detectable ZIKV IgM, IgG, and neutralizing antibody response (reported as the 80% neutralization titers: NT_{80}; **Fig. 3.1B**). Although all participants were viremic at the index visit, 68% (17 of 25) had not yet formed ZIKV-specific IgM responses. Of the participants with a collection visit at the first ("acute") PBMC collection timepoint (median 8 days after index), 100% had formed IgM antibodies and 22% (5 of 23) had residual detectable plasma viremia. There was substantial variation in both peak neutralizing antibody titers (ZIKV NT_{80} titers: 84-37,872) and follow-up titers 6 months after the index visit (0-6,286).

We first characterized how acute ZIKV infection perturbs the frequency and activation of different immune cell types in peripheral blood. We manually gated major landmark immune cell populations defined by standard lineage markers (e.g., classical [CD14+] monocytes, non-classical [CD14-CD16+] monocytes, plasmacytoid dendritic cells [pDCs], classical DCs [cDCs], $CD56^{bright/dim}$ NK cells, CD4+ T cells, CD8+ T cells, B cells, etc.) and classically-defined adaptive immune cell subsets (see **Supplementary Fig. 3.1** for gating strategy and **Table 3.1** for mass cytometry antibodies). We first evaluated the relative abundance of 40 cell types (landmark populations and adaptive immune subsets). We then evaluated the Boolean expression of 30 different phenotypic surface and intracellular proteins on these parent cell types, which yielded a total of 286 unique phenotypic features (see **Table 3.3** for phenotypic features).

To broadly determine how the immune state is perturbed in the context of ZIKV infection, we performed principal component analysis (PCA) on the manually-gated CyTOF features (adjusted for age and sex). We mapped the trajectories across the three timepoints in PCA space for each individual ZIKV-infected participant (**Fig. 3.1C**) as well as 8 control ZIKV-uninfected blood donors (black triangles). While there was variation between individuals, most participants followed a similar general trajectory from right-to-left along PC1 as they progressed from acute to convalescent ZIKV infection (not observed across longitudinal sampling of 6 separate ZIKV-uninfected individuals; **Supplementary Fig. 3.1D**). The number of days between the index and the acute timepoint negatively correlated with the total distance traveled in PCA space across the top five PCs as well as the value of PC1 at the acute timepoint (**Fig. 3.1D**). These correlations suggest that both the PC1 coordinate and the distance traveled correspond to movement in virtual infection space as participants resolve their ZIKV infection. To understand which cellular features contributed to this coordinated movement over time, we used linear mixed effect modelling on the age- and sex-adjusted feature abundances. While the frequency of most major immune cell types did not change significantly across the three sampled timepoints (**Supplementary Fig. 3.2A-B**), 128 of the 286 phenotypic features did change significantly across the three sampled timepoints ($p\_adj<0.05$; **Fig. 3.1E**). The vast majority (95%) of these changing features were elevated at the acute timepoint and decreased in abundance by the late convalescent timepoint. A subset of these features initially remained elevated at the early convalescent timepoint, while others decreased sharply between the acute and early convalescence stage (e.g., most populations expressing Ki-67 and CD71).

To leverage the richness of our high dimensional single cell dataset, we performed unsupervised clustering using the SCAFFoLD algorithm that we have described previously, which associates cell clusters with user-defined landmark populations (43,44). We observed high concordance in the frequency of the pre-defined landmark immune cell populations between our manual gating and SCAFFoLD approaches (**Supplementary Fig. 3.2C-D**). Linear mixed effect modeling demonstrated that 15 of 34 clusters assigned to innate immune cell types and 23 of 56 clusters assigned to adaptive immune cell types (innate immune and B cell clusters from CyTOF Panel 1, T cell clusters from Panel 2) changed significantly in abundance as a percent of their parent landmark population over time (**Fig. 3.1F-G**). We again observed diversity in the direction and speed with which clusters changed in abundance over the three timepoints.

Innate Immune Cell Activation in Acute ZIKV Infection

Little is known about the innate immune response to acute ZIKV infection in humans. Intermediate (CD14+CD16+) monocytes have been shown to increase in the peripheral blood of children with acute infection and are themselves a major target for ZIKV infection (37,45). In adults, we also observed a transiently elevated level of intermediate monocytes during acute ZIKV infection (**Fig. 3.2A**; see **Supplementary Fig. 3.1** for gating). Intermediate monocytes in acute infection expressed higher levels of activation markers (**Fig. 3.2A**). Manual gating and unsupervised clustering analyses revealed that acute infection was also associated with activation in the broader classical (CD14+) monocyte population (which includes CD14+CD16+ intermediate monocytes; **Fig. 3.2B**) as well as non-classical (CD14-CD16+) monocytes (**Supplementary Fig. 3.3A**).

To understand how the expression of activation markers was coordinated on monocyte populations transiently increased in acute infection, we used our clustering analysis to investigate co-expression on individual cells contained within the classical CD14+ monocyte cluster (cluster 49) with the greatest median relative change in frequency (-75%) from the acute to late convalescent timepoints (**Fig. 3.2C**). This revealed three modules of markers with coordinated expression patterns: (1) a proliferative module (Ki-67, CD71, and CD38), (2) an early activation module (HLA-DR, CD86, PD-1, and CD69), and (3) a monocyte maturation/differentiation module (CD16, CD11c, CD40, and CD4; **Fig. 3.2C**). Thus, with unsupervised analysis we identified distinct modules, representing activation/differentiation states, of transiently expanded monocytes.

The proportion of activated cDCs and pDCs was also increased in acute infection, and several activation markers were co-expressed on the cDC cluster with the greatest relative decrease in abundance as infection resolved (**Supplementary Fig. 3.3B**). Amongst NK cells, acute infection was associated with increased proliferation and activation in both CD56^{bright} and the more cytotoxic CD56^{dim} NK cells which resolved during convalescence (**Supplementary Fig. 3.3C**). Collectively, these data demonstrate that acute ZIKV infection is characterized by the activation and differentiation of diverse innate immune cells.

Accumulation of Activated T and B Cells in Acute ZIKV Infection

The population of HLA-DR+CD38+ CD8+ T cells has been found to be enriched for antigen-specific CD8+ T cells in other acute infections (46–48). Acute ZIKV infection was accompanied by a profound accumulation of cycling, activated non-naïve CD8+ T cells co-expressing HLA-DR and CD38 (**Fig. 3.2D-E**). Indeed, our clustering analysis revealed that the expression of multiple activation markers on CD8+ T cells in acute ZIKV infection was tightly co-regulated on small sub-populations. The majority of HLA-DR+CD38+ CD8+ T cells were contained within three clusters of CD8+ T cells that expressed the highest levels of other activation markers (e.g., Ki-67, ICOS, CTLA-4, TIGIT, CD25) and were only transiently increased in acute infection (summed median frequency of c75, 66, 34 across time: 7.1% => 3.1% => 1.6%; **Fig. 3.2F**). Acute ZIKV infection was also associated with a transient increase in the abundance of small sub-populations of cycling and activated non-naïve CD4+ T cell subsets and cytotoxic-skewed γδ T cells (**Supplementary Fig. 3.4**).

Using two established and correlated (**Supplementary Fig. 3.5A**) methods for identifying B cell populations that are actively secreting antibodies (CD38hiCD20neg plasmablasts and CD71hiCD20neg “Antibody Secreting Cells” [ASCs] ((49)), we noted a significant decrease in the frequency of these cells between acute infection and early/late convalescence (**Fig. 3.2G-H**). Phenotypically, a larger proportion of ASCs and other B cell subsets expressed the transcription factor T-bet, which has been associated with B cell responses to viral infections (50), during acute infection compared to early and late convalescence (p_adj=0.02; **Fig. 3.2H, Supplementary Fig. 3.5C**).

An increased frequency of CD20hiCD71hi “Activated B Cells” (ABCs) has also been described in other acute infections in humans (49,51). We observed a significant decrease over time in the proportion of ABCs expressing Ki-67, FCRL5, and CD40 (**Fig. 3.2H**). Overall, multiple subsets of B cells transiently expressed several activation markers in acute ZIKV infection (**Supplementary Fig. 3.5B-C**). Finally, the expression of CD21 was lower on the two IgD- memory B cell populations during acute ZIKV infection (**Supplementary Fig. 3.5C**), which may identify cells recently exited from a germinal center (52).

Coordinated Activation of Innate and Adaptive Immune Cells in Acute ZIKV Infection

We next asked how these immune parameters during acute infection were coordinated to understand the intercellular dynamics that mediate the observed active immune response. We focused on the acute timepoint from 17 individuals without detectable anti-ZIKV IgM at their index visit to limit the variation in the data collected from participants, who may have been sampled at a different number of days following infection. We first interrogated features enriched for antigen-specific cells and found the frequencies of ASC B cells and CD38+HLA-DR+ non-naïve CD8+ T cells were positively correlated (Spearman's r=0.61, p=0.01; **Fig. 3.3A**).

We broadly characterized the relationships of the acute-phase cellular features by computing correlations between all cellular features at the acute timepoint, revealing 279 feature pairs that were positively correlated and 66 that were negatively correlated during acute ZIKV infection (p_adj<0.05). To focus on the correlations that were exclusive to acute ZIKV infection, we separated the feature pairs that were uniquely correlated during acute ZIKV infection, designated as "unique" (n=169), from the remainder – labeled as "shared" – which were correlated both during acute infection as well as in the uninfected samples (n=176; **Fig. 3.3B**). Compared to the shared correlations, the unique correlations during acute ZIKV infection were more likely to be between features from different major landmark populations (odds of correlations being between features belonging to different/the same landmark populations amongst unique [105/64] versus shared [68/108] correlations; odds ratio [OR] 2.60 [95% CI: 1.65-4.12]; **Fig. 3.3C**). The unique correlations were also more likely to be between (rather than within) adaptive and innate immune features (OR 2.77 [1.54-5.10]). Thus, during acute ZIKV infection, there was more coordination across arms of the immune system (e.g., significant unique correlation between CD38+ pDCs and CD38+ Th1 CD4+ T cells (r=0.79, p_adj=0.03;

Fig. 3.3D). We also found that within the positive correlations, the unique correlations were more likely to be between different markers (OR 2.27 [1.31-3.98]) (e.g., CD40+ cDCs and Ki-67+ double negative (DN; CD27-IgD-) B cells (r=0.78, p_adj=0.03; **Fig. 3.3D**)). Together these findings suggest that during acute ZIKV infection, there is broad coordination of the expression of a diversity of activation markers across adaptive and innate immune cell types.

Individuals Exhibit Inversely Correlated Cellular Immune Signatures During Acute ZIKV Infection

We next asked if there were inter-individual differences between study participants that may help to explain the variable outcomes of acute infection, such as the large differences in neutralizing antibody titers. Indeed, the feature pairs uniquely correlated during acute infection were more likely to be negatively correlated across study participants compared to those shared with the uninfected state (OR 3.80 [2.03-7.42]; **Fig. 3.3B-C**). For example, uniquely in acute ZIKV infection, we observed negative correlations between the frequency of activated B cells and CD4+ Tregs (r=-0.86, p_adj=0.002) and between CD69+ CD56^{dim} NK cells and Helios+ Vδ2- γδ T cells (r=-0.77, p_adj=0.03; **Fig. 3.3E**). We hypothesized that these negatively correlated features unique to acute infection reflected inter-individual variability in the acute-phase immune response. To investigate, we performed hierarchical clustering of the acute infection feature correlation matrix (**Fig. 3.4A**), which revealed the presence of two modules (module 3 and module 5) that contained sets of features which were inversely correlated with one another (average correlation: r = -0.79). While the module 5 immune signature was enriched for features that represent activated innate and adaptive immune cell types, 54% of which were transiently elevated in acute infection, the module 3 signature was enriched for features that reflect more cytotoxic-differentiated cell types, 91% of which were “stable,” meaning they did not change in abundance in the context of acute ZIKV infection. During acute infection, some individuals had a higher module 5 score, reflecting dynamic immune activation during acute ZIKV infection, while others had a higher module 3 score, suggesting a more cytotoxic-differentiated immune state (**Fig. 3.4B**). To determine the clinical significance of these acute-phase signatures, we next asked if these distinct acute-phase cellular immune signatures could predict the magnitude of ZIKV neutralizing antibody responses after infection.

Transient Expansion of Activated Cell Types in Acute Infection Predicts High Neutralizing Antibody Titers after ZIKV Infection

We observed a large range in the titers of ZIKV neutralizing antibodies (NT_{80}) that persisted several months after the resolution of acute infection (**Fig. 3.1B**). In order to identify the cellular immune features during acute ZIKV infection that associated with the development of a high versus low ZIKV NT_{80} titer 6 months after infection, we again focused our analysis on individuals sampled as early as possible in the course of infection (i.e., who were ZIKV IgM- at the index visit). Since prior exposure to DENV is associated with significantly higher long-term ZIKV NT_{80} antibody titers ((11) and **Supplementary Fig. 3.5D**), we also only examined individuals with serologic evidence of prior DENV infection (final n=14). These individuals were separated into “high” or “low” 6-month ZIKV NT_{80} titer groupings based on the tertiles of the 6-month ZIKV NT_{80} titers from the whole DENV-exposed REDS-III cohort (low: n=6, <230, mid: n=3, 230-1240 or high: n=5, >1240), which were measured a median of 181 days after index visit [range 160-196 days]; **Fig. 3.5A**). Of note, there was no significant difference in the age (p=0.31) or sex distribution (p=0.53) between the tertiles.

Using a receiver operating characteristic (ROC) analysis, we found that a module 5-skewed score during acute infection was predictive of a high 6-month ZIKV NT_{80} titer, while a module 3-skewed score during acute infection was predictive of a low 6-month titer (area under the curve [AUC] = 0.800; **Fig. 3.5B**). To investigate which individual features were predictive of high versus low 6-month titers, we returned to an unbiased analysis with the full set of phenotypic features, examining features with significantly different frequencies at the acute timepoint between the high and low titer individuals. While we did not observe an association between the frequencies of antigen specific populations (e.g., ASCs or HLA-DR+CD38+ CD8+ T cells) at the acute timepoint and the level of ZIKV NT_{80} at 6 months post-infection (**Supplementary Fig. 3.5E-F**), we did find unique sets of features associated with high versus low levels of ZIKV titers.

We found that high levels of ZIKV neutralizing antibody titers 6 months post-infection were associated with a significantly higher frequency of 11 cellular features during acute infection (e.g., CD86+ CD14-CD16+ monocytes and pDCs, CD40+ CD14+ monocytes and cDCs, CD69+ NK cells, CD38+ Th1 and Tfh CD4+ T cells, and CD86+ as well as Ki-67+ DN B cells; **Fig. 3.5C-D**). These included multiple activated cell types, eight of which were contained within module 5. Six of the 11 features associated with the high titer group were specifically expanded in acute infection (indicated as “Changing”). These features also tended to be low in frequency in uninfected individuals (see lighter green colors in “Uninfected [UI] Mean” column). Together, these data suggest that high 6-month ZIKV NT_{80} titers are associated with robust but transient expansion of specific, diverse activated cellular features during the acute phase of infection.

A Cytotoxic Immune Set Point Predicts Low Neutralizing Antibody Titers after ZIKV Infection

In contrast, individuals with low titers of neutralizing ZIKV antibodies 6 months after infection had an acute infection immune signature defined by higher frequencies of cytotoxic T cell features. These included higher Granzyme B expression in Tctl CD4+ T cells, a larger TEMRA population in CD8+ T cells, higher Eomesodermin expression in non-naïve CD8+ T cells, a higher overall frequency of non-naïve Vδ2- γδ T cells, and a higher frequency of non-naïve Vδ2- γδ T cells that express Granzyme B, T-bet and Helios (**Fig. 3.5C,E**). Most (8 of 9) of the low titer-associated features were contained within the stable, cytotoxic-skewed module 3. Unlike the cellular features associated with high 6-month NT_{80} titers, most of the features associated with low 6-month NT_{80} titers were present at high baseline abundance in uninfected individuals (see darker green/blue colors in the "Uninfected Mean" column), and most (8 of 9) were not dynamically regulated over the course of ZIKV infection. This supported the notion that the cytotoxic-skewed immune signature associated with the development of low neutralizing antibody titers represents a distinct and stable immunologic set point. A higher frequency of cytotoxic-differentiated T cells can relate to a history of infection with other viruses, in particular cytomegalovirus (CMV), and a positive CMV serostatus can be associated with impaired response to vaccination (53–55). However, CMV seropositivity was not significantly associated with the development of low ZIKV NT_{80} titers in our cohort (p=0.29).

To determine the predictive power of the high and low titer-associated features, we again performed an ROC analysis and found that all of the acute infection cellular immune features associated with high or low antibody titers also reliably predicted these two outcomes in this cohort (minimum AUC=0.833; **Fig. 3.5F**). Interestingly, several of the low titer-associated features at the late convalescent timepoint, after the resolution of infection, remained associated with and were predictive of low 6-month ZIKV NT_{80} titers (**Fig. 3.5C**, black in "Late Convalescence" column, and **Fig. 3.5G**). Collectively, our data suggest that high 6-month ZIKV NT_{80} titers are predicted by an immune state of transiently expanded, highly activated immune cell features during acute infection. In contrast, low 6-month ZIKV NT_{80} titers are instead predicted by a distinct “immune set point” characterized by a stable, high frequency of cytotoxic-differentiated T cell populations that are not dynamically regulated during acute ZIKV infection (**Fig. 3.6**).

Discussion

We present here a deep characterization by mass cytometry of dynamic cellular immune responses to acute ZIKV infection in human adults. Leveraging a well-characterized longitudinal cohort of individuals with viremic ZIKV infection, we found that acute ZIKV infection did not impact the frequency of most major cellular immune populations. However, small populations of highly activated innate and adaptive immune cells were coordinately and transiently expanded during acute infection, and distinct acute-phase immune signatures predicted the persistence of high versus low ZIKV neutralizing antibody titers six months after the resolution of infection. Our findings build upon a small but growing literature describing cellular immune responses in acute viral infection in humans (31,46,47,49,56–66), and they suggest immunologic states to target in order to enhance the efficacy of antiviral vaccines.

Our analysis of cellular activation states enabled us to precisely delineate and characterize the coordination between innate and adaptive immune cell populations that respond to acute ZIKV infection. In prior studies, acute ZIKV infection has been associated with activation of some innate immune cell types (38) and, in children, an increase in the frequency of monocyte populations that are also a target for viral infection *in vivo* (37,45). In our study in adults, we observed not only a similar expansion of intermediate CD14+CD16+ monocytes during acute infection, but also a transient increase in a suite of activation markers on this cell type in the acute phase. Additionally, we identified a cluster of CD14+ monocytes that were present at a higher frequency during acute infection (c49) with distinct co-regulated markers denoting proliferation, activation or differentiation states. Amongst adaptive immune cells, HLA-DR+CD38+ non-naïve CD8+ T cells were expanded at the acute time point, consistent with other acute viral infections (31,46,47). We found that these cells were contained in three distinct clusters of cells that co-express different combinations of activation markers. Acute ZIKV infection was also associated with activation of Th1 and Tctl T cell CD4+ T cell subsets. Finally, using gating strategies to identify populations of B cells enriched for antigen-specific cells in other infections (49,67), we identified an expansion of Tbet+ ASCs during acute ZIKV infection.

Our study describes the diverse and coordinated activation of cellular immune responses during acute ZIKV infection in human adults. Compared to the baseline correlations that exist in the uninfected state, we found that acute ZIKV infection drove new coordination between different immune cell types and across the innate and adaptive immune system. The correlations unique to acute ZIKV infection (e.g., positive correlations in the proportion of CD38+ pDCs and CD38+ Th1 CD4+ T cells, or between CD40+ cDCs and Ki-67+ DN B cells) may reflect interactions that are essential to mount a productive antiviral immune response. Further exploration of the correlated features in acute infection revealed two distinct modules that were inversely correlated: one (module 5) contained features reflecting transiently elevated activated cell populations while the second (module 3) contained features reflecting stable/unchanging cytotoxic cell populations. Remarkably, these two acute infection immune signatures, which we identified using an unbiased analysis approach, appear to truly reflect distinct immune states that differentially impact and predict the development and maintenance of high neutralizing antibody responses.

High titers of ZIKV neutralizing antibodies are likely critical for protective immunity in humans, and they are a key target for ZIKV vaccines (68). Six months following infection, participants across our cohort had a greater than 100-fold difference in ZIKV neutralizing antibody titers. Other than a positive association with prior DENV serostatus observed here and in other studies (11), little is known about what parameters predispose some individuals to maintain higher versus lower ZIKV neutralizing antibody titers. Interestingly, as has been observed in SARS-CoV-2 infection (31), the frequency of antibody secreting cells during acute ZIKV infection did not correlate with antibody levels in convalescence. We did, however, find several other acute-phase cellular features that were associated with and predictive of high versus low neutralizing antibody titers, many of which have plausible roles in augmenting a productive B cell response. For example, CD86 expression on pDCs and monocytes and CD40 expression on cDCs and monocytes can mediate enhanced antigen presentation to and priming of helper CD4+ T cells, IFNγ produced by activated Th1 cells or NK cells can promote B cell activation, and activated Tfh CD4+ T cells can provide direct help to differentiating B cells. Further investigation could elucidate whether robust induction of these same activated cell populations also predicts the long-term immunogenicity of vaccines for ZIKV and other viral infections.

In contrast to the dynamically regulated acute-phase cellular immune features associated with high ZIKV neutralizing antibody titers, a higher frequency of T cells with cytotoxic differentiation features were associated with low 6-month ZIKV neutralizing antibody titers and predictive of levels of 6-month ZIKV neutralizing antibody titers. Most of these features did not dynamically change over the course of infection and were themselves inversely correlated with the cellular immune features associated with high 6-month ZIKV neutralizing antibody titers (similar to the inverse correlation between modules 3 and 5 in the correlation matrix). Several of the low titer-associated cytotoxic features were also present at higher levels in the low titer individuals 3-6 months after resolution of the infection, suggesting that they may represent a stable biological state that is likely reflective of their history of prior antigen encounters. This state is distinct from the small populations of virus-specific (e.g., HLA-DR+CD38+) T cells that are transiently expanded in acute infection (31,46,47). The stability of these features suggests that a cytotoxic immune "set point" may identify individuals predisposed to have a blunted activation response to acute infection that then leads to impaired neutralizing antibody responses. In general, a more cytotoxic-skewed T cell compartment can be a sign of immune senescence, which can in turn be associated with a reduced capacity to generate functional antigen-specific responses after vaccination (53,55). Thus, in addition to identifying candidate biomarkers of a "responsive" immune signature that may be useful for predicting the formation of a robust neutralizing antibody response to other infections or vaccination, our study also provides insight into potential markers of an immune state that impairs the formation of protective immunity after acute viral infection. Future studies should explore the generalizability of our findings to other infections and vaccination and the underlying causes of these distinct immune signatures.

Our study provides a first in-depth characterization of the cellular immune response to acute ZIKV infection in human adults and relates distinct acute-phase cellular immune signatures to the development of high- or low-titers of durable neutralizing antibodies. Our approach offers a powerful tool to test whether these features also predict immunogenicity of vaccines for ZIKV and other viral infections, such as SARS-CoV-2, for which neutralizing antibodies play a major role in protection. Our findings suggest that targeted therapeutic approaches in individuals predicted to have poor neutralizing antibody responses to vaccination (e.g., different adjuvants or a higher dose of vaccine) might increase acute-phase immune activation and subsequently promote enhanced long-term protective antiviral immunity.

Our study has some important limitations. Although we have made an effort to control for the variance introduced by sampling time, it was not possible to align participants according to the exact date they were infected. Our study included only otherwise healthy individuals who presented for volunteer blood donation and does not include pregnant individuals or infants, who are key populations affected by this infection. Finally, while it is likely that neutralizing antibodies play a key role in immunologic protection from ZIKV (69), a titer that correlates with protection in humans has not yet been identified (14) and other antibody functions (70) and/or other types of immune responses (17,19,71), may also be critical for robust protection.

Figures

Figure 1

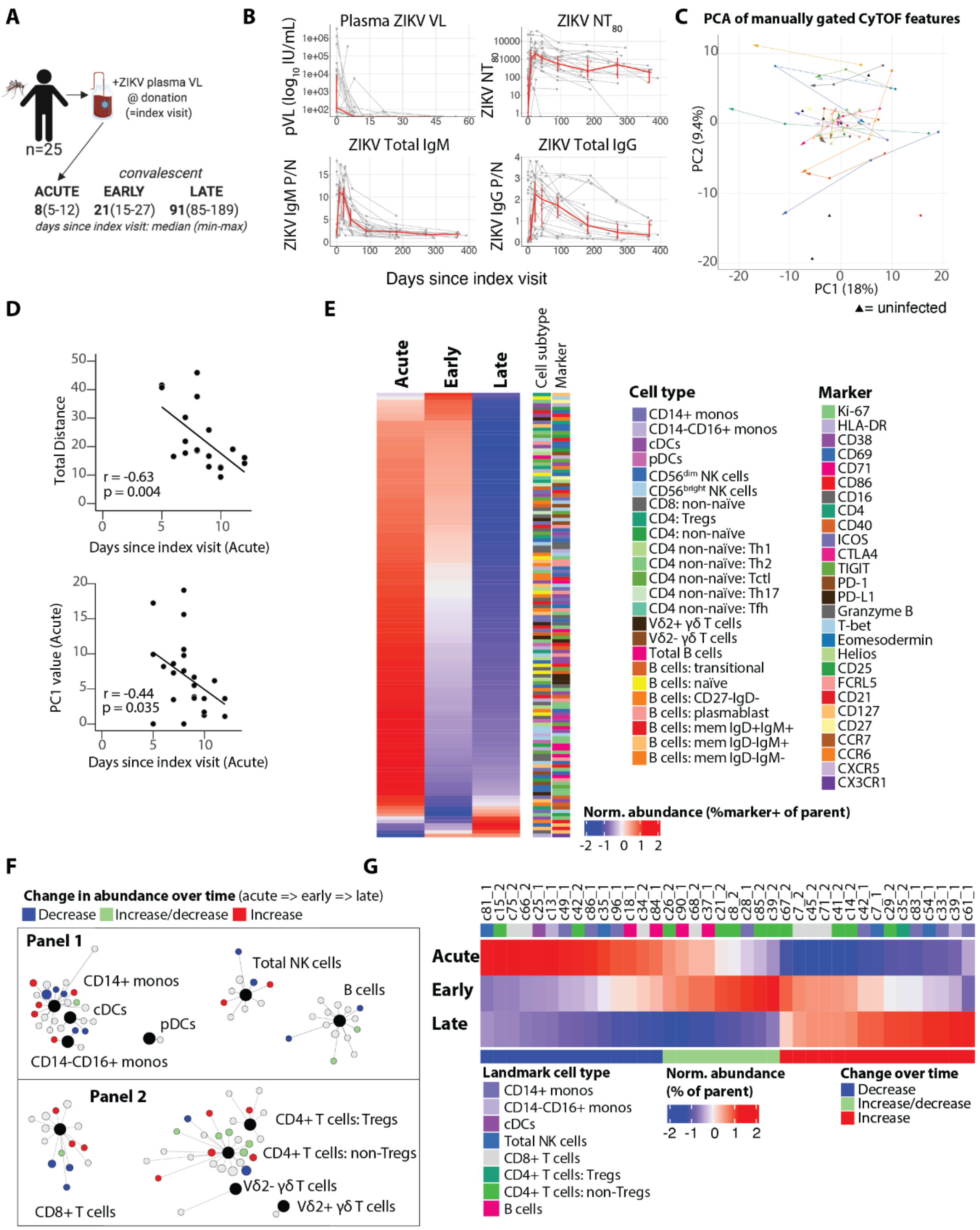
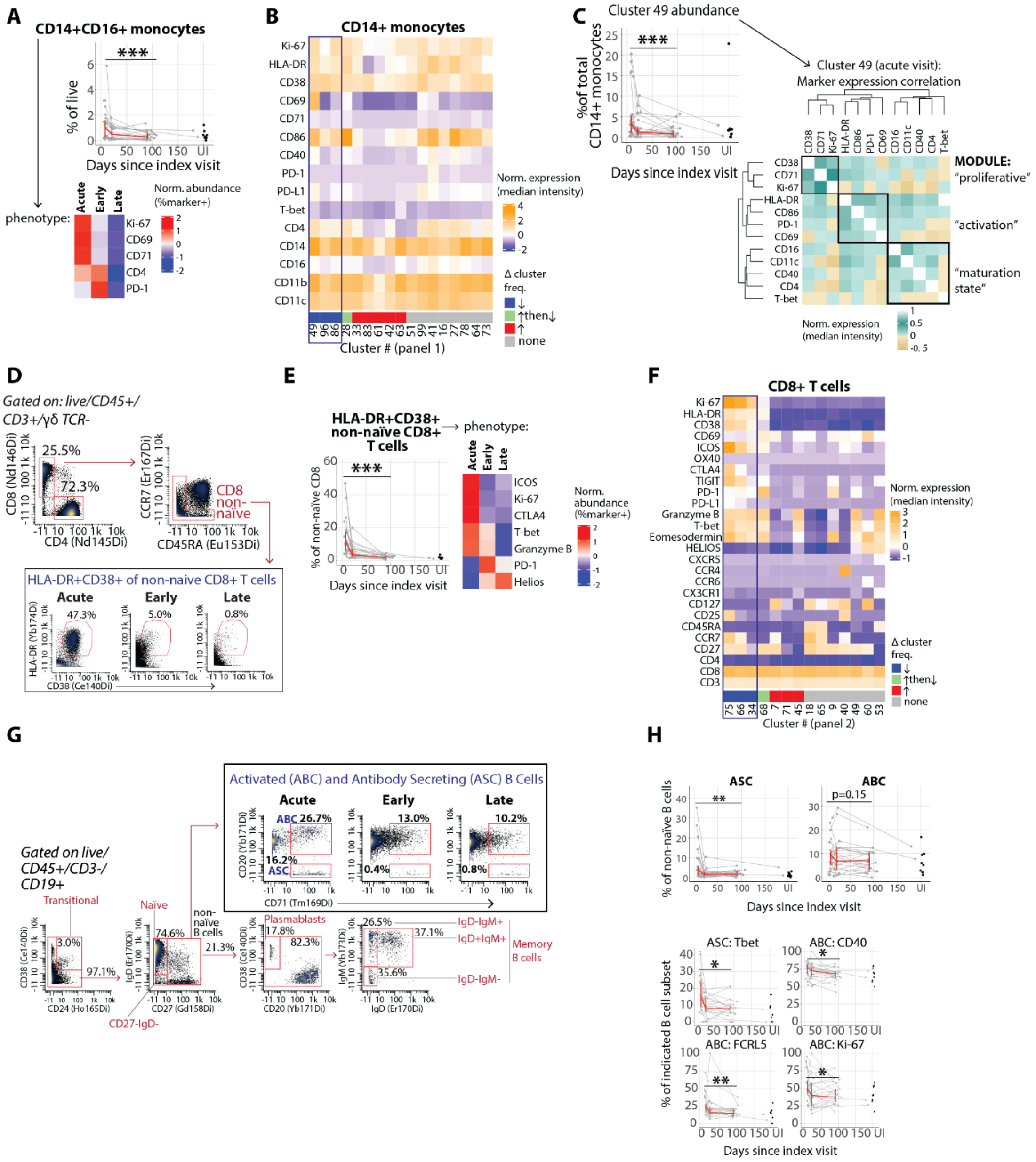


Figure 3.1 Acute infection with ZIKV elicits profound phenotypic changes across peripheral blood cellular immune populations

(A) 25 adults viremic with acute ZIKV infection at the time of blood donation ("index visit") had peripheral blood sampling at up to three timepoints: acute phase of infection, early and/or late convalescence (see Table 3.2 for clinical characteristics). (B) Plasma ZIKV viral load (VL), neutralizing antibody titers (NT_{80}), and total IgG and IgM levels of study cohort participants. Red line connects median values at each sampling timepoint (+/- 95% Confidence Interval, CI). (C) Directed line plots for each participant in PCA space from early to later timepoints. Black triangles denote 8 uninfected control samples. (D) Scatterplots of days since index visit at the acute timepoint and the value of PC1 at the acute timepoint or the total distance traveled in PCA space between the acute and late convalescent timepoints (Spearman's correlation with regression line). (E) Heatmap showing the z-score normalized frequency of the log-adjusted feature abundances for the manually gated phenotypic features that change significantly over time (see Table 3.3 for list of features assessed). (F) SCAFFoLD maps showing clusters of cells associated with landmark cell population nodes (black dots). Clusters that significantly change in abundance between the acute and late convalescent timepoints are labeled: increase (red), decrease (blue), or increase and then decrease (green). (G) Heatmap showing the normalized abundance of the clusters (z-score based on % of parent cell type population) that change significantly. Significance in (**E-G**) based on linear mixed effects (LME) model fit with p_adj<0.05. See also **Supplementary Figure 3.1** and **Table 3.2-3.3**.

Figure 2



A
CD14+CD16+ monocytes
***
% of live
Days since index visit
phenotype:
Acute
Early
Late
Ki-67
CD69
CD71
CD4
PD-1
Norm. abundance (%marker+)
B
CD14+ monocytes
Ki-67
HLA-DR
CD38
CD69
CD71
CD86
CD40
PD-1
PD-L1
T-bet
CD4
CD14
CD16
CD11b
CD11c
Norm. expression (median intensity)
Δ cluster freq.
↓
↑then↓
↑
none
Cluster # (panel 1)
C
Cluster 49 abundance
***
%of total CD14+ monocytes
Days since index visit
Cluster 49 (acute visit): Marker expression correlation
MODULE: "proliferative"
"activation"
"maturation state"
Norm. expression (median intensity)
D
Gated on: live/CD45+/ CD3+/γδ TCR-
25.5%
72.3%
CD8 (Nd146Di)
CD4 (Nd145Di)
CCR7 (Er167Di)
CD45RA (Eu153Di)
CD8 non-naive
HLA-DR+CD38+ of non-naive CD8+ T cells
Acute
Early
Late
47.3%
5.0%
0.8%
HLA-DR (Yb174Di)
CD38 (Ce140Di)
E
HLA-DR+CD38+ non-naïve CD8+ T cells
phenotype:
***
% of non-naïve CD8
Days since index visit
ICOS
Ki-67
CTLA4
T-bet
Granzyme B
PD-1
Helios
Norm. abundance (%marker+)
F
CD8+ T cells
Ki-67
HLA-DR
CD38
CD69
ICOS
OX40
CTLA4
TIGIT
PD-1
PD-L1
Granzyme B
T-bet
Eomesodermin
HELIOS
CXCR5
CCR4
CCR6
CX3CR1
CD127
CD25
CD45RA
CCR7
CD27
CD4
CD8
CD3
Norm. expression (median intensity)
Δ cluster freq.
↓
↑then↓
↑
none
Cluster # (panel 2)
G
Gated on live/ CD45+/CD3-/ CD19+
Transitional
3.0%
97.1%
CD38 (Ce140Di)
CD24 (Ho165Di)
Naïve
74.6%
21.3%
IgD (Er170Di)
CD27 (Gd158Di)
CD27-IgD-
non-naïve B cells
Plasmablasts
17.8%
82.3%
CD20 (Yb171Di)
26.5%
37.1%
35.6%
IgM (Yb173Di)
IgD (Er170Di)
IgD-IgM+
IgD+IgM+
IgD-IgM-
Memory B cells
Activated (ABC) and Antibody Secreting (ASC) B Cells
Acute
Early
Late
ABC 26.7%
16.2%
ASC
13.0%
0.4%
10.2%
0.8%
CD20 (Yb171Di)
CD71 (Tm169Di)
H
ASC
**
ABC
p=0.15
% of non-naïve B cells
Days since index visit
ASC: Tbet
*
ABC: CD40
*
ABC: FCRL5
**
ABC: Ki-67
*
% of indicated B cell subset
Days since index visit

Figure 3.2 Transient accumulation of activated immune cells during acute ZIKV infection (A) Frequency (as a % of total live cells) and phenotype (z-scored proportion of cells that express each marker) of CD14+CD16+ monocytes across the course of acute and resolving ZIKV infection. (B) Heatmap showing z-score normalized median expression of indicated markers (rows) for each monocyte-associated cell cluster (columns). Column annotation indicates clusters that significantly decrease (blue), increase (red), increase and then decrease (green), or remain unchanged (gray) in abundance (as a % of CD14+ monocytes; p_adj<0.05). (C) Change in abundance of CD14+ monocyte cluster 49 (as a % of CD14+ monocytes; p_adj=0.0002; left) and Spearman's correlation matrix of marker expression on single cells in CD14+ monocyte cluster 49 from acute visit samples (right). (D) Gating scheme for non-naïve CD8+ T cells that co-express HLA-DR and CD38. Percentages shown are % of parent populations in plotted sample. (E) Frequency (as a % of non-naïve CD8+ T cells) and phenotype (z-scored proportion of cells that express each marker) of HLA-DR+CD38+ non-naïve CD8+ T cells across the course of acute and resolving ZIKV infection. (F) Phenotype (z-scored median expression of each marker) of CD8+ T cell clusters that significantly decrease (blue), increase (red), increase and then decrease (green), or remain unchanged (gray) in abundance. (G) Gating scheme for B cell subsets, including activated and antibody secreting B cells (ABC and ASC, respectively). Percentages shown are % of parent populations in the plotted sample. (H) Frequency (as a % of non-naïve B cells) and phenotype of activated B cells (ABCs) and antibody-secreting cells (ASCs) across the course of acute and resolving ZIKV infection. p_adj<0.05). *p_adj<0.05, **p_adj<0.01, ***p_adj<0.001. A, C, E, H: Red line connects median values at each sampling timepoint (+/- 95% CI). UI=uninfected. N=25 participants. See also **Supplementary Figure 3.3-3.5**.

Figure 3

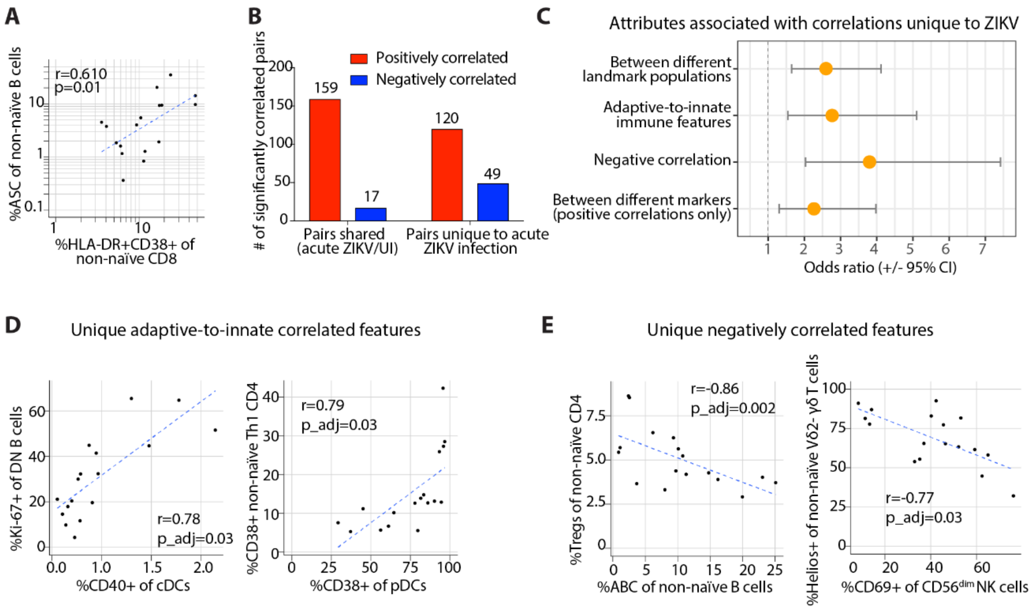
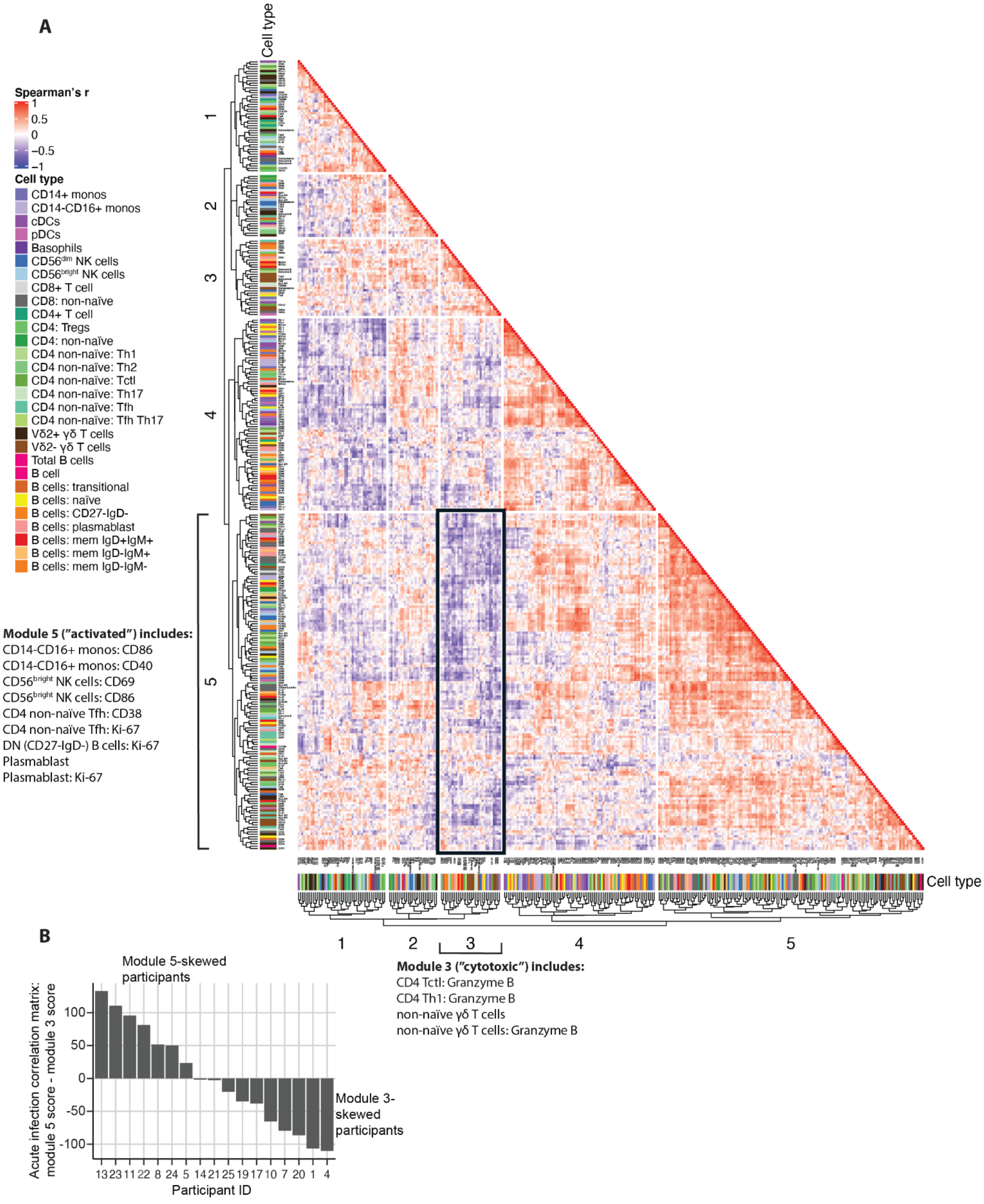


Figure 3.3 Coordinated activation across different cell types in acute ZIKV infection
(A) Scatterplot of the frequency of ASC B cells and CD38+HLA-DR+ CD8+ T cells in acute ZIKV infection with regression line. (B) Number of significant (p_adj<0.05) positive and negative correlations between cellular immune features that are present in acute ZIKV infection, grouped by those that are "unique" to ZIKV versus those "shared" with the uninfected (UI) cohort. (C) Odds ratio (+/- 95% CI) that cellular immune feature correlations unique to ZIKV infection are more likely to be associated with different correlation attributes (compared to the correlations shared with the UI cohort). Correlation plots of select features uniquely correlated in acute ZIKV infection (Spearman's r with correlation line): (D) adaptive-to-innate immune features, (E) negatively correlated features. N=17 participants (anti-ZIKV IgM- at index visit).

Figure 4

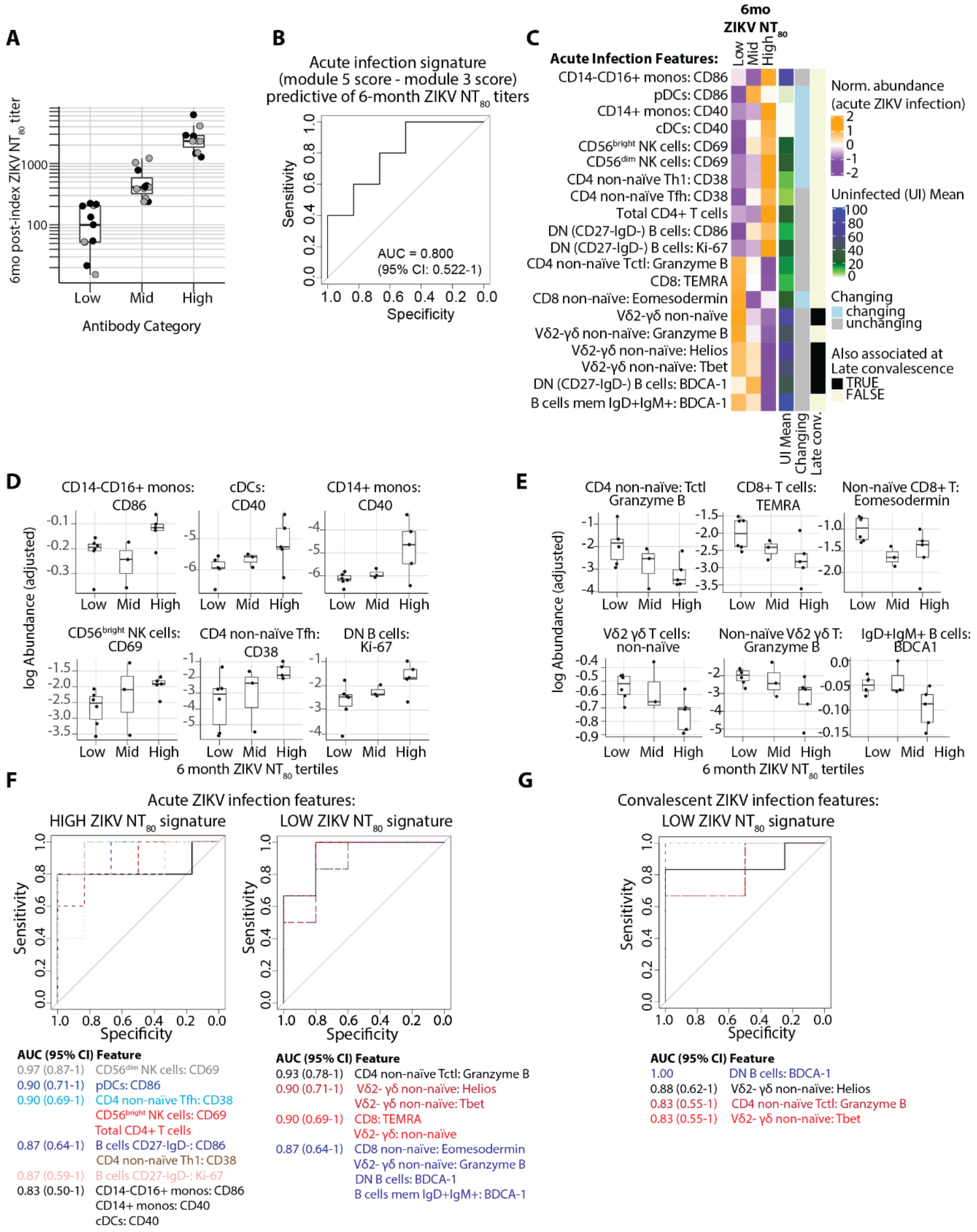


Module 5 ("activated") includes:
CD14-CD16+ monos: CD86
CD14-CD16+ monos: CD40
CD56^{bright} NK cells: CD69
CD56^{bright} NK cells: CD86
CD4 non-naïve Tfh: CD38
CD4 non-naïve Tfh: Ki-67
DN (CD27-IgD-) B cells: Ki-67
Plasmablast
Plasmablast: Ki-67

Module 3 ("cytotoxic") includes:
CD4 Tctl: Granzyme B
CD4 Th1: Granzyme B
non-naïve γδ T cells
non-naïve γδ T cells: Granzyme B

Figure 3.4 Correlated immune cell features during acute ZIKV infection
(A) Correlation heatmap depicting Spearman’s correlation values (no significance cut-off) of all manually gated features from acute ZIKV infection, representing the 17 participants who were ZIKV IgM- ("pre-IgM") at the index visit. Hierarchical clustering was used to group cellular features into five modules. Negatively correlated modules 3 and 5 are indicated with bold outline. (B) Distribution of (module 5 score - module 3 score) values at the acute visit amongst the pre-IgM study participants. N=17 participants (anti-ZIKV IgM- at index visit). See also **Supplementary Figure 3.5**.

Figure 5



A
6mo post-index ZIKV NT80 titer
1000
100
Low
Mid
High
Antibody Category
B
Acute infection signature
(module 5 score - module 3 score)
predictive of 6-month ZIKV NT80 titers
Sensitivity
Specificity
AUC = 0.800
(95% CI: 0.522-1)
C
6mo
ZIKV NT80
Low
Mid
High
Acute Infection Features:
CD14-CD16+ monos: CD86
pDCs: CD86
CD14+ monos: CD40
cDCs: CD40
CD56bright NK cells: CD69
CD56dim NK cells: CD69
CD4 non-naïve Th1: CD38
CD4 non-naïve Tfh: CD38
Total CD4+ T cells
DN (CD27-IgD-) B cells: CD86
DN (CD27-IgD-) B cells: Ki-67
CD4 non-naïve Tctl: Granzyme B
CD8: TEMRA
CD8 non-naïve: Eomesodermin
Vδ2-γδ non-naïve
Vδ2-γδ non-naïve: Granzyme B
Vδ2-γδ non-naïve: Helios
Vδ2-γδ non-naïve: Tbet
DN (CD27-IgD-) B cells: BDCA-1
B cells mem IgD+IgM+: BDCA-1
UI Mean
Changing
Late conv.
Norm. abundance
(acute ZIKV infection)
2
1
0
-1
-2
Uninfected (UI) Mean
100
80
60
40
20
0
Changing
changing
unchanging
Also associated at
Late convalescence
TRUE
FALSE
D
CD14-CD16+ monos: CD86
cDCs: CD40
CD14+ monos: CD40
CD56bright NK cells: CD69
CD4 non-naïve Tfh: CD38
DN B cells: Ki-67
log Abundance (adjusted)
6 month ZIKV NT80 tertiles
E
CD4 non-naïve: Tctl Granzyme B
CD8+ T cells: TEMRA
Non-naïve CD8+ T: Eomesodermin
Vδ2 γδ T cells: non-naïve
Non-naïve Vδ2 γδ T: Granzyme B
IgD+IgM+ B cells: BDCA1
log Abundance (adjusted)
6 month ZIKV NT80 tertiles
F
Acute ZIKV infection features:
HIGH ZIKV NT80 signature
LOW ZIKV NT80 signature
Sensitivity
Specificity
AUC (95% CI) Feature
0.97 (0.87-1) CD56dim NK cells: CD69
0.90 (0.71-1) pDCs: CD86
0.90 (0.69-1) CD4 non-naïve Tfh: CD38
CD56bright NK cells: CD69
Total CD4+ T cells
0.87 (0.64-1) B cells CD27-IgD-: CD86
CD4 non-naïve Th1: CD38
0.87 (0.59-1) B cells CD27-IgD-: Ki-67
0.83 (0.50-1) CD14-CD16+ monos: CD86
CD14+ monos: CD40
cDCs: CD40
AUC (95% CI) Feature
0.93 (0.78-1) CD4 non-naïve Tctl: Granzyme B
0.90 (0.71-1) Vδ2- γδ non-naïve: Helios
Vδ2- γδ non-naïve: Tbet
0.90 (0.69-1) CD8: TEMRA
Vδ2- γδ: non-naïve
0.87 (0.64-1) CD8 non-naïve: Eomesodermin
Vδ2- γδ non-naïve: Granzyme B
DN B cells: BDCA-1
B cells mem IgD+IgM+: BDCA-1
G
Convalescent ZIKV infection features:
LOW ZIKV NT80 signature
Sensitivity
Specificity
AUC (95% CI) Feature
1.00 DN B cells: BDCA-1
0.88 (0.62-1) Vδ2- γδ non-naïve: Helios
0.83 (0.55-1) CD4 non-naïve Tctl: Granzyme B
0.83 (0.55-1) Vδ2- γδ non-naïve: Tbet

Figure 3.5 Distinct cellular immune signatures are associated with the development of high versus low ZIKV neutralizing antibody titers 6 months after infection
(A) ZIKV neutralizing antibody titers (NT_{80}) measured approximately 6 months post-index visit in the overall REDS-III study participants (gray dots) and the sub-cohort studied here (black dots). Participants were divided into tertiles based on these values. (B) Receiver operating characteristic (ROC) curve for predicting high- versus low-titer individuals using the difference between the acute-phase module 5 and module 3 signature scores. (C) Heatmap showing z-score normalized abundance at the acute visit for cellular features that were significantly ($p\_adj<0.05$) increased in high versus low 6-month NT_{80} titer participants at the acute timepoint. Row annotations for each feature indicate: Mean values in a cross-sectional uninfected (UI) control cohort, whether or not the abundance of the feature significantly changed across time between acute to convalescent infection, and whether or not the abundance of the feature was also present at a significantly higher frequency ($p\_adj<0.05$) in the same group (high- versus low-titer participants) at the late convalescent timepoint. Abundance (log-adjusted) of features during acute ZIKV infection associated with high (D) versus low (E) 6-month neutralizing antibody titers. (F) ROC curves for predicting high- versus low-titer individuals using the acute ZIKV cellular features from (**C**) that are associated with high (left) versus low (right) 6-month ZIKV NT_{80} titer. (G) ROC curves for predicting high -versus low- titer individuals using the late convalescent features associated with low 6-month ZIKV neutralizing antibody titers. For (**B**, **F**, **G**): The area under the curve (AUC) value and 95% CI for the features corresponding to each curve are colored by AUC value for each plot. N=14 participants with 6-month ZIKV NT_{80} titer data available (anti-ZIKV IgM- at index visit).

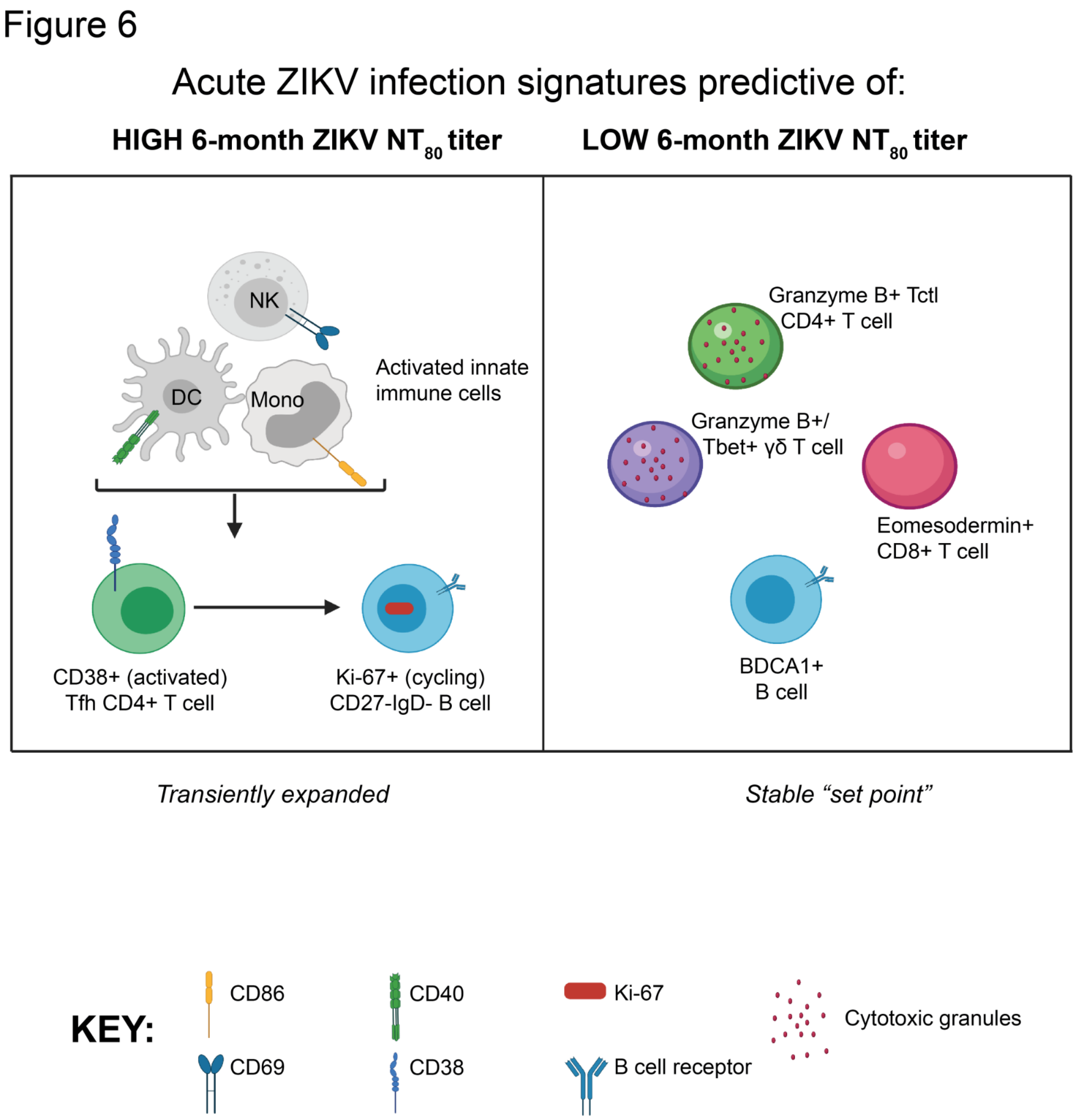
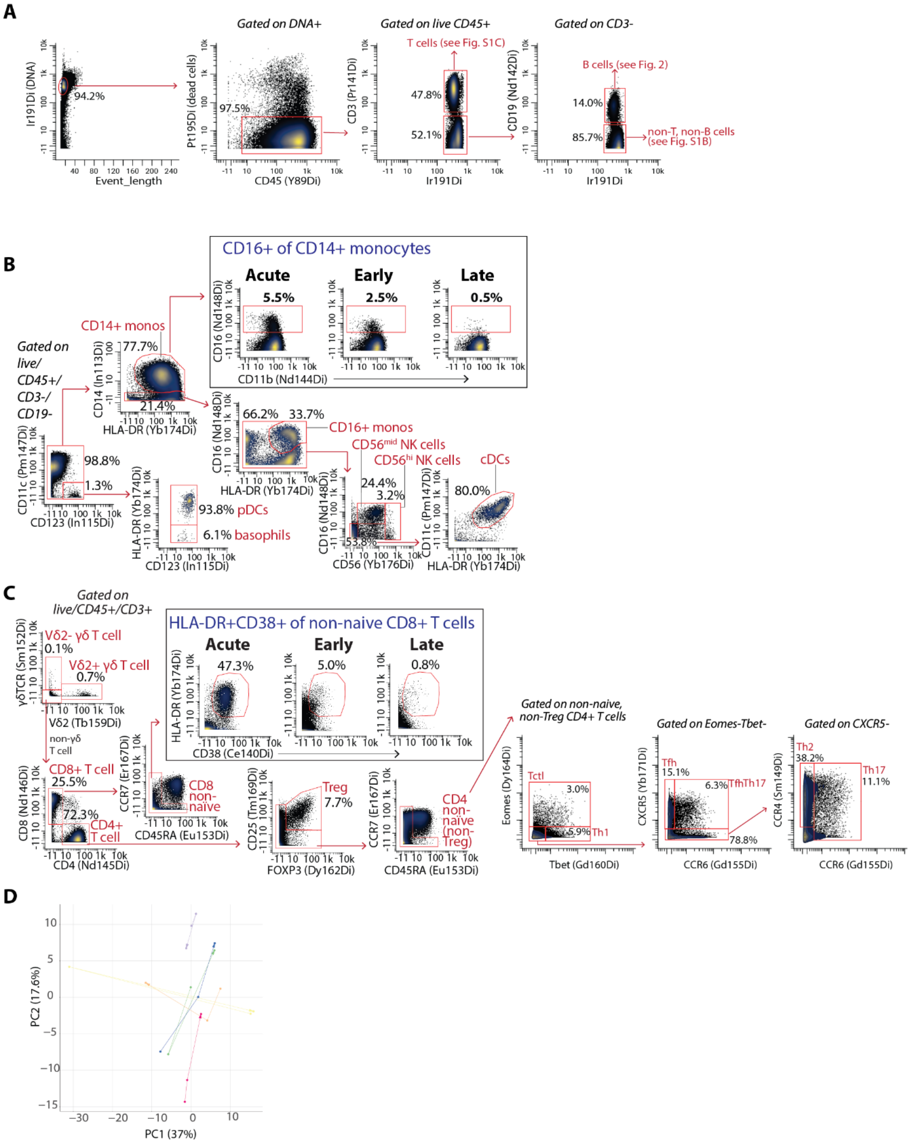


Figure 3.6 Graphical abstract summarizing distinct cellular immune signatures associated with the development of high versus low ZIKV neutralizing antibody titers 6 months after infection

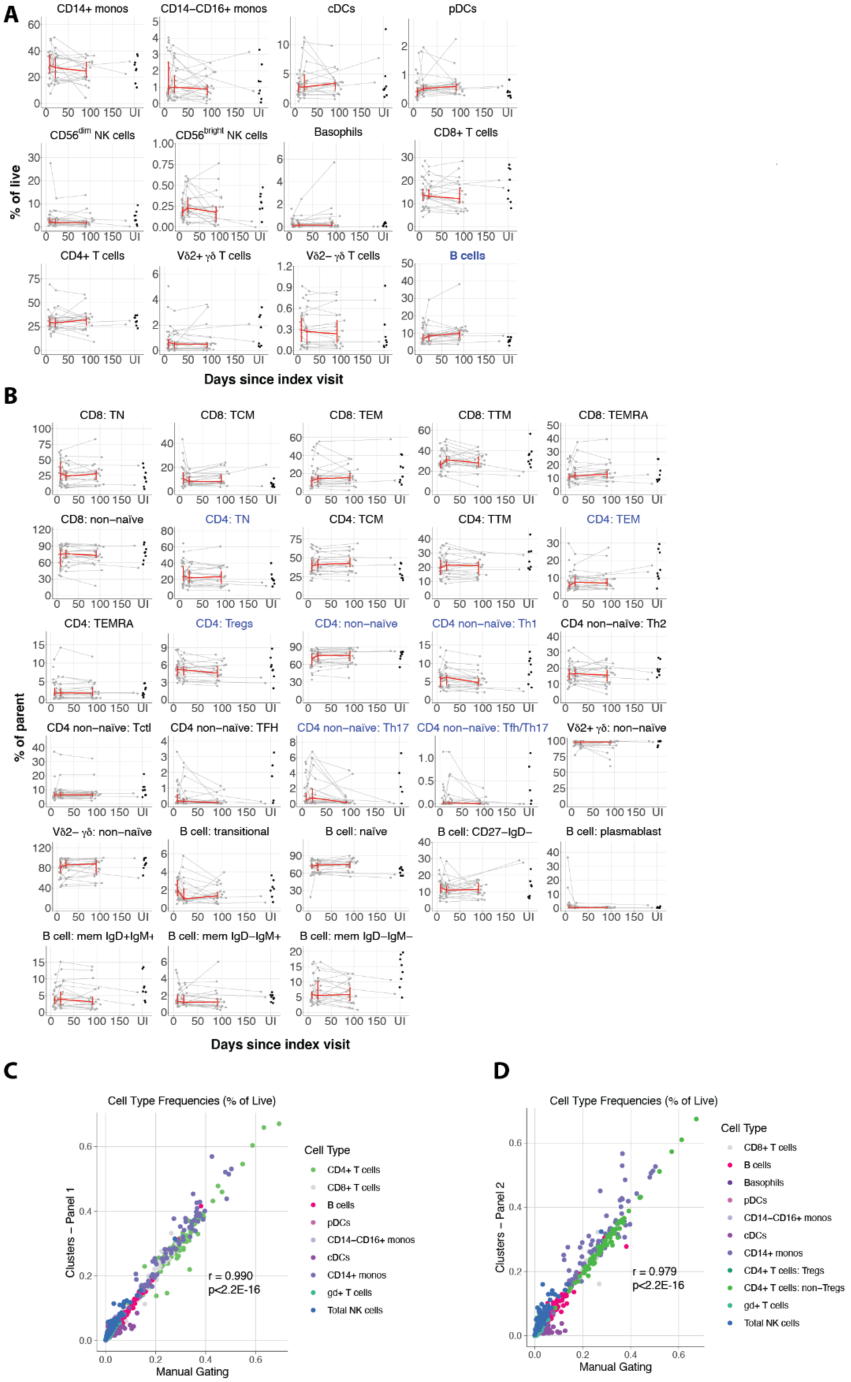
Supplementary Figure 1



Supplementary Figure 3.1 CyTOF gating strategy and Principal Component Analysis (PCA) of uninfected participants

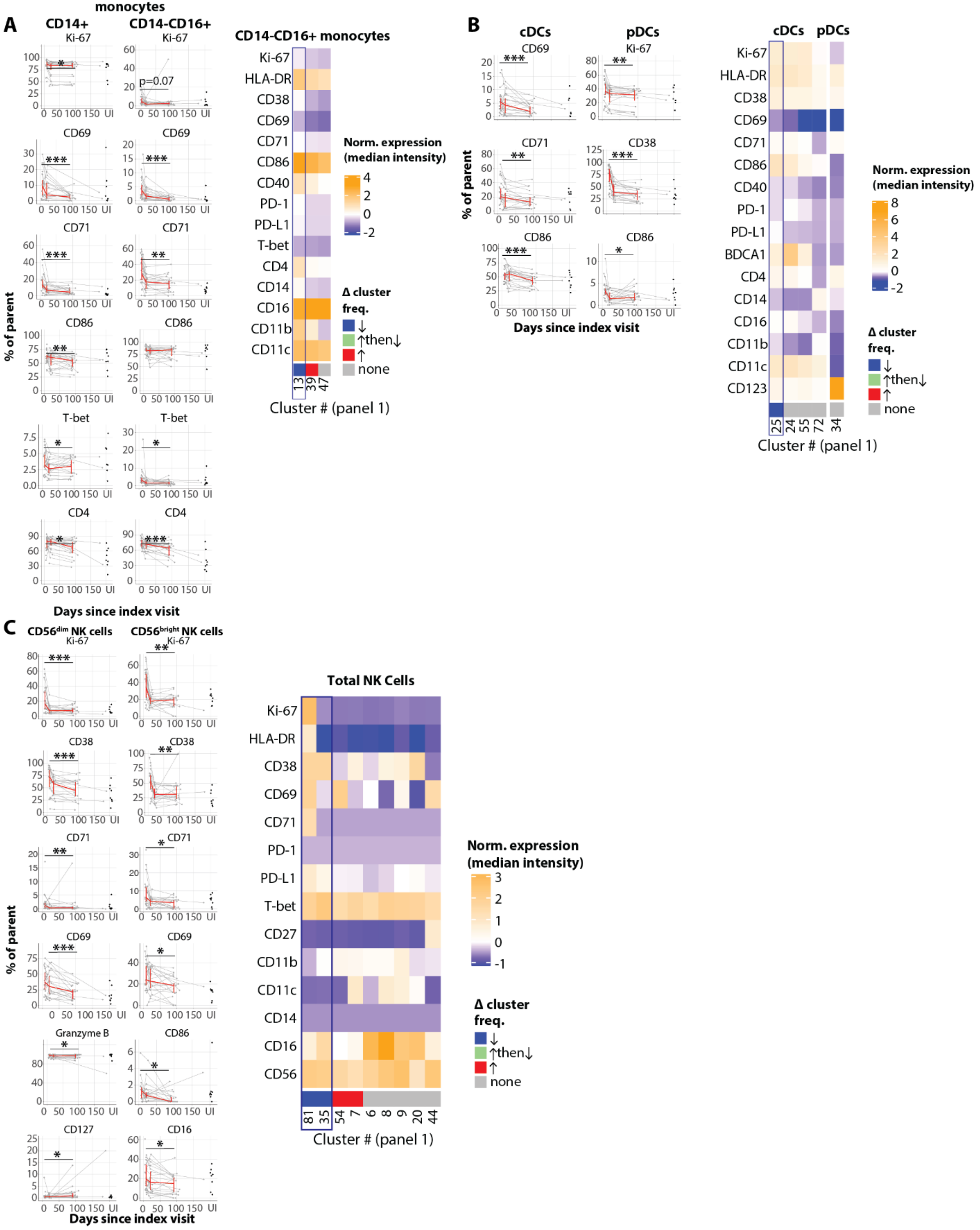
Gating strategy for (A) landmark populations, (B) innate immune cells, and (C) T cells. (D) PCA representation of all manually gated parameters measured on PBMCs from ZIKV-uninfected control participants (N=6) at longitudinal timepoints.

Supplementary Figure 2



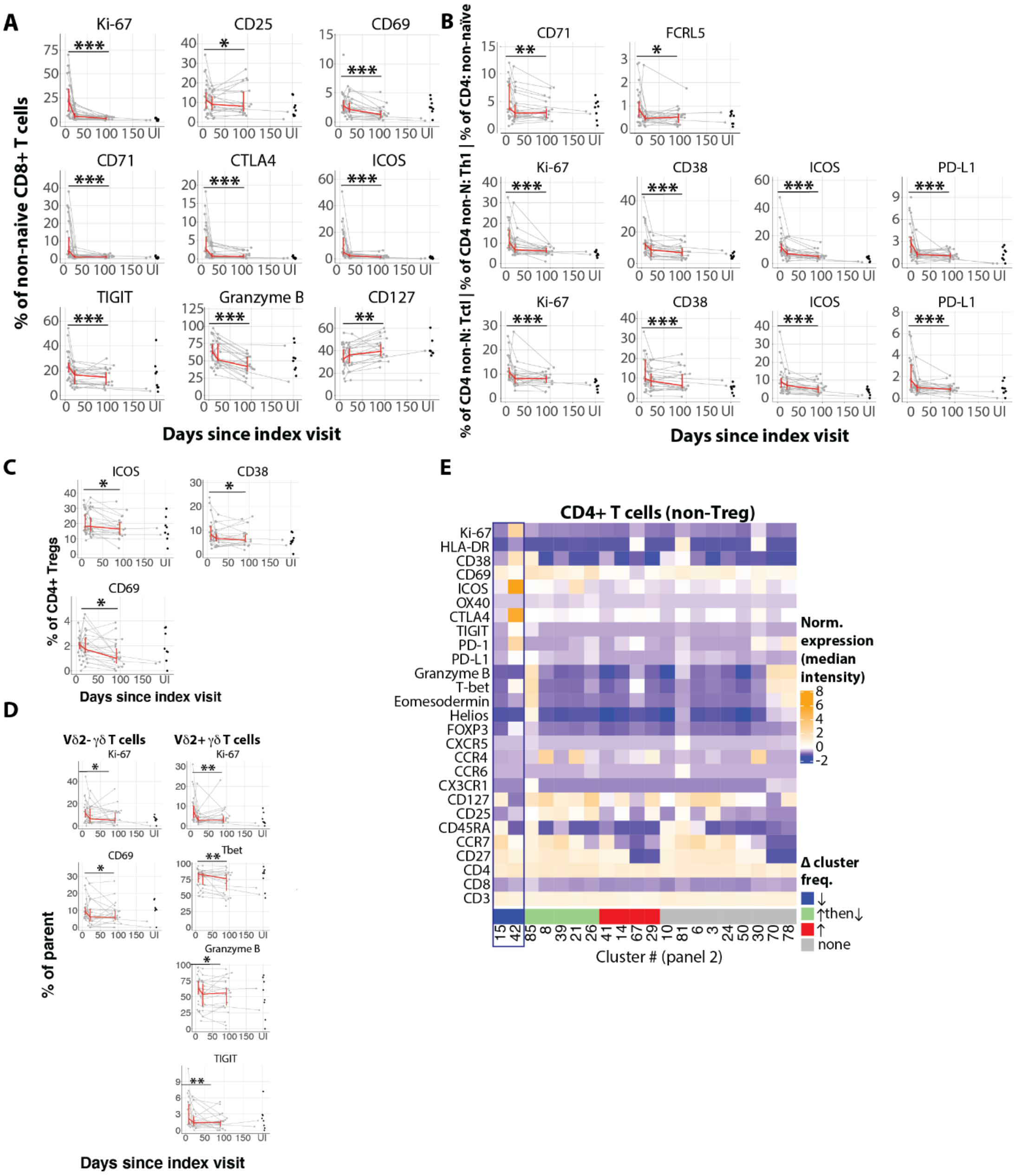
Supplementary Figure 3.2 Landmark and sub-landmark population abundance in acute and convalescent ZIKV infection
Line plots of frequency of indicated (A) landmark cell type and (B) adaptive immune subset for each participant versus time since index visit. Red line connects median values at each sampling timepoint with error bars for +/- 95% CI. Scatter plot for feature abundance from cross sectional uninfected (UI) cohort shown on the far right. Features with $p\_adj < 0.05$ have blue colored titles. p_adj values obtained by LME model fit with Benjamini-Hochberg FDR correction. High concordance in landmark cell population frequencies as measured by manual gating versus SCAFFoLD clustering analysis in Panel 1 (C) and Panel 2 (D). N=25 ZIKV+ and N=8 ZIKV- participants.

Supplementary Figure 3



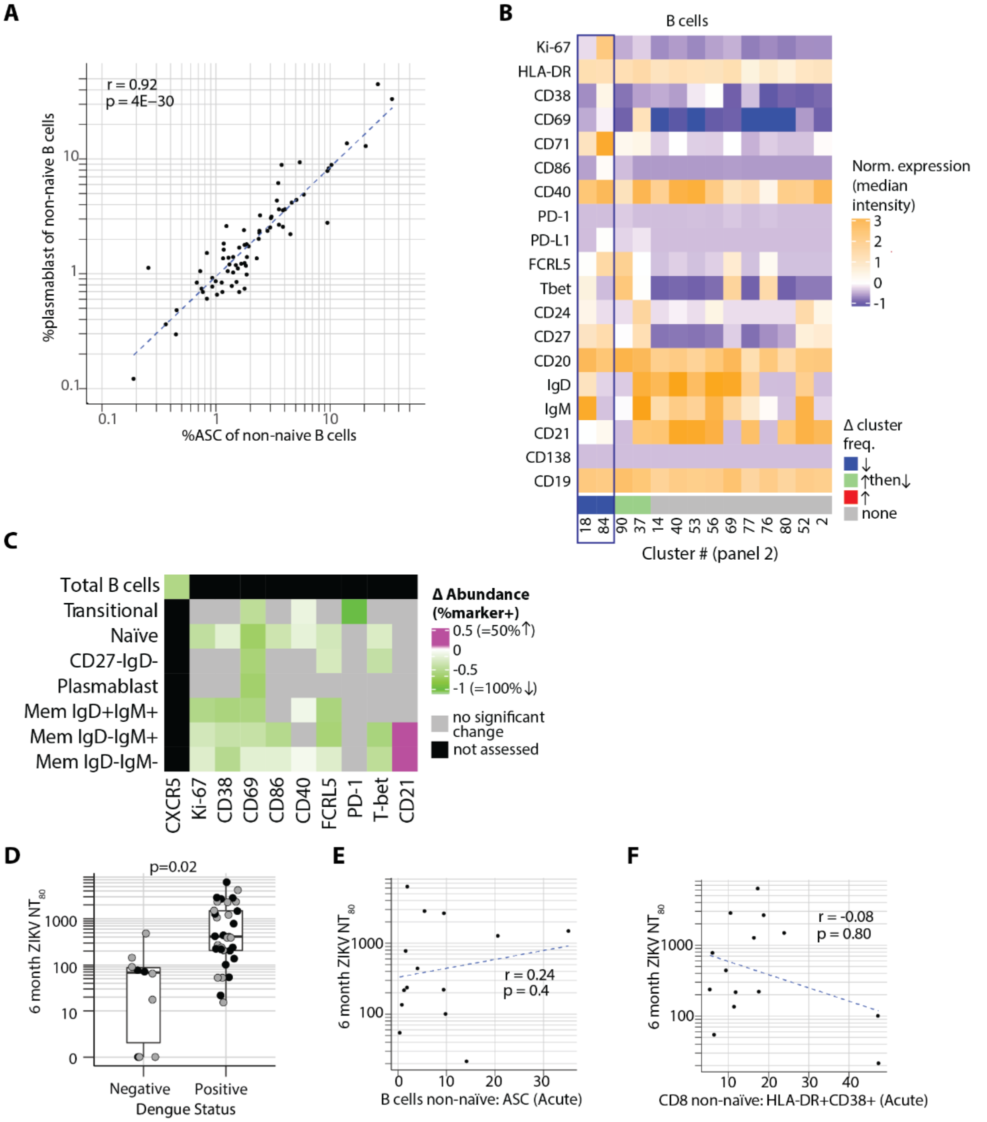
Supplementary Figure 3.3 Innate immune cell features impacted by acute ZIKV infection (A-C, left): Line plots showing frequency of phenotypic features (% of cells that express each marker) versus time since index visit (N=25). Red line connects median values at each sampling timepoint with error bars for +/- 95% CI. Feature abundance from cross-sectional uninfected (UI) cohort (N=8) shown on the far right. *p_adj<0.05, **p_adj<0.01, ***p_adj<0.001 (p_adj values obtained by LME model fit with Benjamini-Hochberg FDR correction). (A-C, right): Heatmaps showing z-score normalized median expression of indicated markers (rows) for each landmark cell population-associated cell cluster (column). Column annotation indicates clusters that significantly decrease (blue), increase (red), increase and then decrease (green), or remain unchanged (gray) in abundance (as a % of the parent population; p_adj<0.05). N=25 ZIKV+ and N=8 ZIKV- participants.

Supplementary Figure 4



Supplementary Figure 3.4 T cell features impacted by acute ZIKV infection
Significantly changing features (shown as % of parent cells that express each marker) for (A) non-naïve CD8+ T cells, (B) non-Treg non-naïve CD4+ T cells, (C) CD4+ Tregs, and (D) γδ T cells for each participant versus time since index visit. Red line connects median values at each sampling timepoint with error bars for +/- 95% CI. Scatterplot for feature abundance from cross-sectional uninfected (UI) cohort shown on the far right. *p_adj<0.05, **p_adj<0.01, ***p_adj<0.001 (p_adj values obtained by LME model fit with Benjamini-Hochberg FDR correction). (E) Phenotype (z-scored median expression of each marker) of non-Treg non-naïve CD4+ T cell clusters that significantly decrease (blue), increase (red), increase and then decrease (green) or remain unchanged in abundance (p_adj<0.05). N=25 ZIKV+ and N=8 ZIKV- participants.

Supplementary Figure 5



Supplementary Figure 3.5 B cell dynamics in ZIKV infection and characteristics associated with ZIKV neutralizing antibody titers 6 months after infection
(A) Correlation plot between the frequency of plasmablast and ASC B cell populations (as a % of non-naïve B cells; Spearman's r with regression line). (B) Phenotype (z-scored median expression of each marker) of B cell clusters that significantly decrease (blue), increase (red), increase and then decrease (green), or remain unchanged (gray) in abundance (as a % of the total B cell population; p_adj<0.05). (C) Relative change in the frequency [(late convalescent - acute)/acute] of the expression of individual activation markers on B cell subsets (median shown). Colors indicate markers with a significant (p_adj<0.05) change in the percent of the parent population that expresses the marker are noted (increase=pink, decrease=green). (D) Difference in 6-month ZIKV NT_{80} between individuals with or without evidence of prior DENV infection at index visit (Wilcoxon Rank Sum test). Individuals from our sub-cohort are colored black and individuals from the larger REDSIII cohort are colored gray. Scatterplots showing 6-month ZIKV NT_{80} titers versus the frequency of (E) ASC B cells or (F) non-naïve CD8+ T cells co-expressing HLA-DR and CD38 at the acute timepoint (Spearman's correlation). (**A-C)**: N=25 participants; E-F: N=14 participants.

Tables

Table 3.1 Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mass cytometry antibodies: metal-antigen (clone)	Self-conjugated unless from Fluidigm	
Y89-CD45 (clone HI30)	Fluidigm	Cat#3089003B; RRID:AB_2661851
In113-CD14 (clone M5E2)	BioLegend	Cat#301802; RRID:AB_314184
In115-CD123 (clone 6H6)	BioLegend	Cat#306002; RRID:AB_2661822
La139-CD33 (clone WM53)	BioLegend	Cat#303402; RRID:AB_314346
Ce140-CD38 (clone HIT2)	BioLegend	Cat#303502; RRID:AB_314354
Pr141-CD3 (clone UCHT1)	BioLegend	Cat#300402; RRID:AB_2661835
Nd142-CD19 (clone H1B19)	BioLegend	Cat#302202; RRID:AB_2661817
Nd143-CXCR3 (clone G025H7)	BioLegend	Cat#353702; RRID:AB_10983073
Nd144-CD11b (clone ICRF44)	BioLegend	Cat#301302; RRID:AB_314154
Nd145-CD4 (clone RPA-T4)	BioLegend	Cat#300502; RRID:AB_314069
Nd146-CD8 (clone RPA-T8)	BioLegend	Cat#301002; RRID:AB_2661818
Sm147-CD11c (clone Bu15)	BioLegend	Cat#337202; RRID:AB_1236381
Nd148-CD16 (clone 3G8)	BioLegend	Cat#302001; RRID:AB_314201
Sm149-CD138 (clone DL-101)	BioLegend	Cat#352302; RRID:AB_10915555
Eu151-CD21 (clone Bu32)	BioLegend	Cat#313502; RRID:AB_416326
Sm152-gdTCR (clone 11F2)	Fluidigm	Cat#3152008B; RRID:AB_2687643
Eu153-CD45RA (clone HI100)	BioLegend	Cat#304102; RRID:AB_314406
Sm154-CD40 (clone 5C3)	BioLegend	Cat#334302; RRID:AB_1236384
Gd156-PDL1 (clone 29E.2A3)	BioLegend	Cat#329702; RRID:AB_940372
Gd157-CD69 (clone FN50)	BioLegend	Cat#310902; RRID:AB_314837
Gd158-CD27 (clone O323)	BioLegend	Cat#302802; RRID:AB_2661825

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Gd160-Tbet (clone 4B10)	BioLegend	Cat#644802; RRID:AB_1595503
Dy161-CTLA4 (clone 14D3)	Fluidigm	Cat#3161004B; RRID:AB_2687649
Dy162-CD80 (clone 2D10.4)	Fluidigm	Cat#3162010B; RRID:AB_2811101
Dy163-CD86 (clone IT2.2)	BioLegend	Cat#305401; RRID:AB_314521
Ho165-CD24 (clone MI5)	BioLegend	Cat#311102; RRID:AB_314851
Er166-NKG2D (clone ON72)	Fluidigm	Cat#3166016B; RRID:AB_2892110
Er167-FCRL5 (clone 509f6)	BioLegend	Cat#340302; RRID:AB_2104586
Er168-Ki67 (clone B56)	Fluidigm	Cat#3168007B; RRID:AB_2800467
Tm169-CD71 (clone CY1G4)	BioLegend	Cat#334102; RRID:AB_1134247
Er170-IgD (clone IA6-2)	BioLegend	Cat#348202; RRID:RRID:AB_10550095
Yb171-CD20 (clone 2H7)	BioLegend	Cat#302302; RRID:AB_314250
Yb172-BDCA1 (clone L161)	BioLegend	Cat#331502; RRID:AB_2661820
Yb173-IgM (clone MHM-88)	BioLegend	Cat#314502; RRID:AB_493003
Yb174-HLA-DR (clone L243)	BioLegend	Cat#307602; RRID:AB_314680
Lu175-PD-1 (clone EH12.2H7)	BioLegend	Cat#329902; RRID:AB_940488
Yb176-CD56 (clone HCD56)	Fluidigm	Cat#3176008B; RRID:AB_2661813
Sm149-CCR4 (clone 205410)	R&D	Cat#MAB1567; RRID:AB_2074395
Nd150-OX40 (clone A019D5)	BioLegend	Cat#351302; RRID:AB_10718513
Eu151-ICOS (clone C398.4A)	BioLegend	Cat#313539; RRID:AB_2810475
Sm154-CX3CR1 (clone 2A9-1)	BioLegend	Cat#341602; RRID:AB_1595422
Gd155-CCR6 (clone G034E3)	BioLegend	Cat#353402; RRID:AB_10918625
Tb159-Vd2 (clone B6)	BioLegend	Cat#331402; RRID:AB_1089226
Dy162-FOXP3 (clone PCH101)	BioLegend	Cat#3162011a; RRID:AB_2687650
Dy164-EOMES (clone WD1928)	ThermoFisher	Cat#14-4877-82; RRID:AB_2572882
Ho165-CD127 (clone A019D5)	BioLegend	Cat#351302; RRID:AB_10718513
Er166-TIGIT (clone A15153G)	BioLegend	Cat#372702; RRID:AB_2632714

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Er167-CCR7 (clone G043H7)	BioLegend	Cat#353202; RRID:AB_10945157
Tm169-CD25 (clone 2A3)	Fluidigm	Cat#3169003B; RRID:AB_2661806
Yb171-CXCR5 (clone RF8B2)	Fluidigm	Cat#3171014B; RRID:AB_2858239
Yb172-Helios (clone 22F6)	BioLegend	Cat#137202; RRID:AB_10900638
Yb173-Granzyme B (clone GB11)	BioRad	Cat#MCA2120; RRID:AB_2114582
Biological samples		
Cryopreserved human PBMCs and plasma	REDS-III study participants	Demographic Data available in **Table 3.2**
Chemicals, peptides, and recombinant proteins		
Cisplatin	Sigma-Aldrich	Cat #P4394
eBioscience FoxP3/Transcription Factor Staining Buffer Set	Thermo Fisher Scientific	Cat #00-5523-00
Maxpar Barcode Perm Buffer	Fulidigm	Cat #201057
Paraformaldehyde	Electron Microscopy Sciences	Cat #15710
Intercalator	Fluidigm	Cat #201103A
Deposited data		
Mass cytometry data	This paper	http://dx.doi.org/10.17632/5cn6cy97b7.1
Software and algorithms		
CellEngine	CellCarta	https://cellcarta.com/cellenginesoftware/
R 3.6.1	The R Foundation	https://www.r-project.org/
premessa 0.1.8	R package	https://github.com/ParkerICI/premessa
flowCore 1.50.0	(80)	RRID:SCR_002205
ggplot2 3.2.1	(82)	RRID:SCR_014601
nlme 3.1-140	(81)	RRID:SCR_015655
factoextra 1.0.5	(88)	RRID:SCR_016692

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FactoMineR 1.42	(83)	RRID:SCR_014602
seriation 1.2.8	(85)	https://cran.r-project.org/package=seriation
ComplexHeatmap 2.1.1	(84)	RRID:SCR_017270
SCAFFoLD	(43)	https://github.com/SpitzerLab/statisticalScaffold
igraph 1.2.4.1	(86)	RRID:SCR_019225
pROC 1.17.0.1	(87)	https://CRAN.R-project.org/package=pROC

Table 3.2 Study participant clinical characteristics

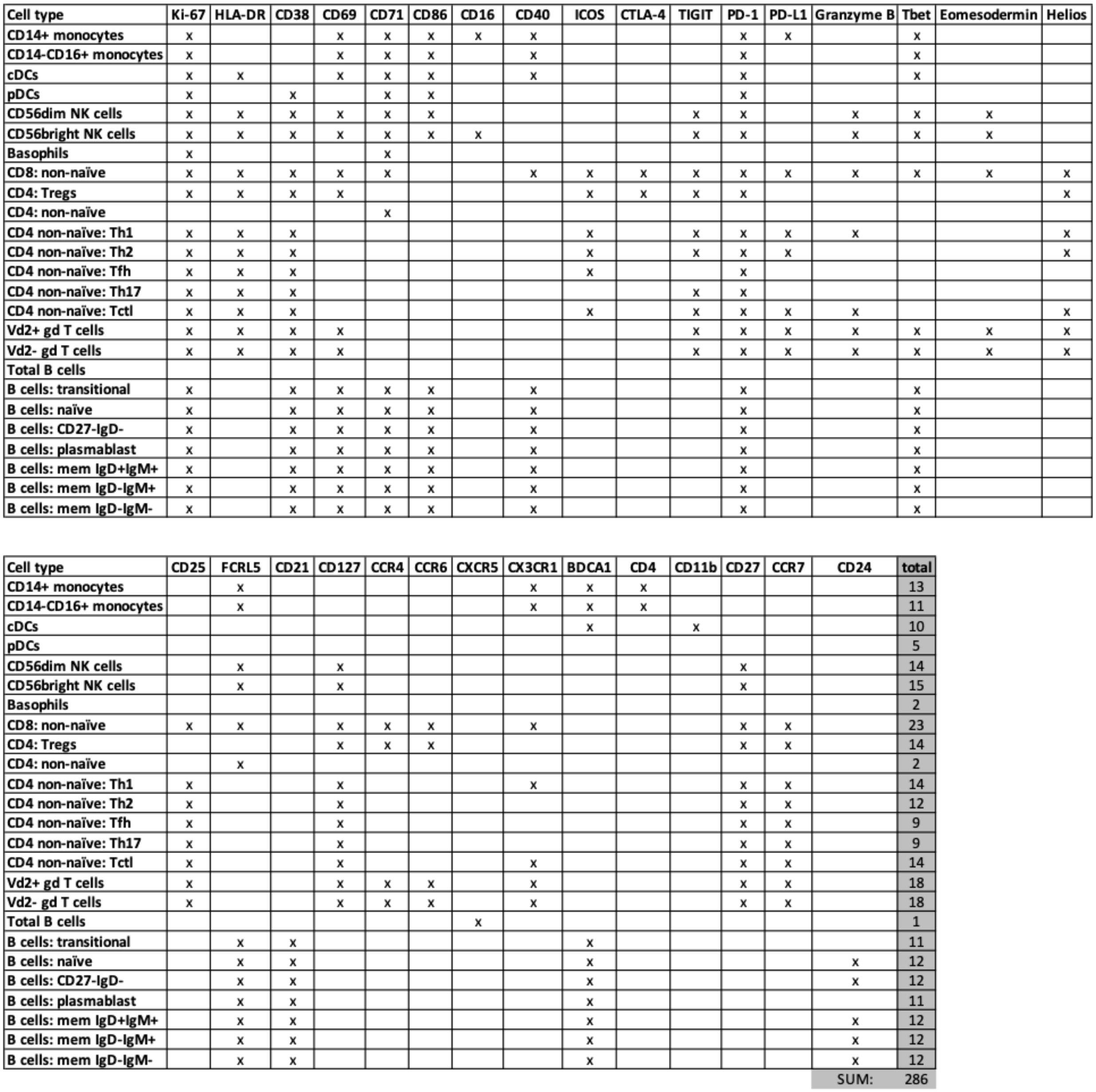
ZIKV-infected

Participant #	Age	Sex	Pre-IgM at index visit	DENV exposed	>=3 symptoms at acute visit	Maximum ZIKV NT80 titer	6mo ZIKV NT80 titer
1	52	Male	Y	Y	TRUE	355.1	54.3
2	24	Male	N	N	FALSE	84.4	0
3	22	Male	N	Y	TRUE	9719.1	2894.4
4	62	Male	Y	Y	FALSE	1069.7	217.1
5	43	Male	Y	N	FALSE	NA	NA
6	54	Male	N	Y	FALSE	13952.6	419.9
7	51	Male	Y	Y	FALSE	2782.7	221.7
8	46	Male	Y	Y	TRUE	688	21.6
9	49	Male	N	N	FALSE	1474.1	71.8
10	37	Male	Y	Y	FALSE	37872.3	6285.7
11	46	Male	Y	Y	TRUE	3086.1	100.2
12	36	Female	N	Y	TRUE	4543	202.9
13	43	Male	Y	Y	TRUE	3405.1	1483.7
14	42	Male	Y	Y	TRUE	2238.7	134.8
15	43	Female	N	N	Unknown	1153	77.1
16	28	Female	N	N	Unknown	NA	NA
17	27	Female	Y	N	TRUE	NA	NA
18	53	Male	N	N	FALSE	1198.7	NA
19	25	Female	Y	Y	FALSE	NA	NA
20	71	Male	Y	Y	FALSE	2091.4	440.9
21	67	Male	Y	Y	FALSE	3752.8	237.6
22	24	Male	Y	Y	FALSE	2206.1	2660.2
23	56	Female	Y	Y	TRUE	9412.2	1278.7
24	44	Male	Y	Y	FALSE	23236.4	2828.5
25	21	Female	Y	Y	TRUE	1490.7	779.1

ZIKV-uninfected

Participant #	Age	Sex
26	49	Male
27	32	Female
28	51	Female
29	60	Male
30	53	Male
31	58	Male
32	42	Female
33	22	Male
34	32	Female
35	20	Male
36	40	Male
37	54	Male
38	44	Male
39	20	Male

Table 3.3 Summary of phenotypic markers assessed on each cell type for manual gating analysis



Cell type	Ki-67	HLA-DR	CD38	CD69	CD71	CD86	CD16	CD40	ICOS	CTLA-4	TIGIT	PD-1	PD-L1	Granzyme B	Tbet	Eomesodermin	Helios
CD14+ monocytes	x			x	x	x	x	x				x	x		x		
CD14-CD16+ monocytes	x			x	x	x		x				x			x		
cDCs	x	x		x	x	x		x				x			x		
pDCs	x		x		x	x						x					
CD56dim NK cells	x	x	x	x	x	x					x	x		x	x	x	
CD56bright NK cells	x	x	x	x	x	x	x				x	x		x	x	x	
Basophils	x				x												
CD8: non-naïve	x	x	x	x	x			x	x	x	x	x	x	x	x	x	x
CD4: Tregs	x	x	x	x					x	x	x	x					x
CD4: non-naïve					x												
CD4 non-naïve: Th1	x	x	x						x		x	x	x	x			x
CD4 non-naïve: Th2	x	x	x						x		x	x	x				x
CD4 non-naïve: Tfh	x	x	x						x			x					
CD4 non-naïve: Th17	x	x	x								x	x					
CD4 non-naïve: Tctl	x	x	x						x		x	x	x	x			x
Vd2+ gd T cells	x	x	x	x							x	x	x	x	x	x	x
Vd2- gd T cells	x	x	x	x							x	x	x	x	x	x	x
Total B cells																	
B cells: transitional	x		x	x	x	x		x				x			x		
B cells: naïve	x		x	x	x	x		x				x			x		
B cells: CD27-IgD-	x		x	x	x	x		x				x			x		
B cells: plasmablast	x		x	x	x	x		x				x			x		
B cells: mem IgD+IgM+	x		x	x	x	x		x				x			x		
B cells: mem IgD-IgM+	x		x	x	x	x		x				x			x		
B cells: mem IgD-IgM-	x		x	x	x	x		x				x			x		

Cell type	CD25	FCRL5	CD21	CD127	CCR4	CCR6	CXCR5	CX3CR1	BDCA1	CD4	CD11b	CD27	CCR7	CD24	total
CD14+ monocytes		x						x	x	x					13
CD14-CD16+ monocytes		x						x	x	x					11
cDCs									x		x				10
pDCs															5
CD56dim NK cells		x		x								x			14
CD56bright NK cells		x		x								x			15
Basophils															2
CD8: non-naïve	x	x		x	x	x		x				x	x		23
CD4: Tregs				x	x	x						x	x		14
CD4: non-naïve		x													2
CD4 non-naïve: Th1	x			x				x				x	x		14
CD4 non-naïve: Th2	x			x								x	x		12
CD4 non-naïve: Tfh	x			x								x	x		9
CD4 non-naïve: Th17	x			x								x	x		9
CD4 non-naïve: Tctl	x			x				x				x	x		14
Vd2+ gd T cells	x			x	x	x		x				x	x		18
Vd2- gd T cells	x			x	x	x		x				x	x		18
Total B cells							x								1
B cells: transitional		x	x						x						11
B cells: naïve		x	x						x					x	12
B cells: CD27-IgD-		x	x						x					x	12
B cells: plasmablast		x	x						x						11
B cells: mem IgD+IgM+		x	x						x					x	12
B cells: mem IgD-IgM+		x	x						x					x	12
B cells: mem IgD-IgM-		x	x						x					x	12
														SUM:	286

Materials and Methods

Viral Load and Antibody Measurements. ZIKV viral load, antibody levels, and ZIKV and DENV neutralizing antibody measurements were performed as described previously (9,41). In brief, ZIKV viral load was measured by quantitative PCR. Anti-Zika virus IgM and IgG were measured by antibody-capture ELISA using recombinant ZIKV antigen kindly provided by the US Centers for Disease Control and Prevention (CDC) and as previously described (72,73). ZIKV neutralizing titers were measured using a ZIKV reporter viral particle neutralization titration assay (Integral Molecular, Philadelphia, PA) (74), and index donations were tested for pre-existing DENV IgG with the Detect IgG ELISA (InBios; Seattle, WA).

PBMC Preparation and Mass Cytometry Staining. Whole peripheral blood was collected at the clinical sites, shipped overnight at ambient temperature to Vitalant, San Francisco, CA, USA, where they were processed and cryopreserved within 24 h of collection and then stored in liquid nitrogen as previously described (42). Mass cytometry experiments were performed over the course of five separate experiments, with normalization between experiments performed as outlined below. PBMCs were thawed, and only samples with >70% viability were used for analysis (most were >90% viable after thawing by the Muse Cell Analyzer [Millipore Sigma, Burlington, MA, USA]) (75,76). We stained 2-4 million cells per panel in two mass cytometry panels, following a previously published protocol (44) with the following modifications. Briefly, we marked dead cells by incubating the samples for one minute with 25mM Cisplatin (Sigma-Aldrich, St. Louis, MO, USA) in phosphate buffered saline (PBS) plus EDTA, performed surface staining with metal-tagged antibodies in PBS with 0.5% bovine serum albumin (BSA) for 30 minutes at room temperature, fixed and permeabilized cells following manufacturer's instructions for the eBioscience Foxp3/Transcription Factor Staining Buffer Set (Thermo Fisher

Scientific, Waltham, MA, USA), barcoded samples using mass-tag cellular barcoding reagents diluted in Maxpar Barcode Perm Buffer (Fluidigm, South San Francisco, CA, USA) as described previously (44), combined up to twenty barcoded samples into a single tube, performed intracellular staining with antibodies diluted in eBioscience Foxp3/Transcription Factor kit perm wash (Thermo Fisher Scientific), fixed cells in freshly prepared 2% paraformaldehyde (Electron Microscopy Sciences, Hatfield, PA, USA) in the presence of a DNA intercalator (77), and then washed and ran cells on the Fluidigm CyTOF 2 mass Cytometer within one week of staining.

Mass Cytometry Data Processing.

Data Quality Control. Following data acquisition, the FCS files were normalized across experiments using bead standards and the data normalization algorithm using the R package 'premessa.' The live cell events were debarcoded using a single-cell debarcoding algorithm (78) and we analyzed >25,000 (mostly >50,000) cells per sample. From the individual sample files, normalization beads were excluded based on Ce140 and Eu153 signals, single cell events were identified based on Ir191 DNA signal measured against event length, and CD45- or Pt195+ dead cells were excluded. Potential batch effects were minimized by including samples from the same individual in the same experiment. Spillover between the Yb173 and Yb174 channels was compensated based on the CyTOF metal purity matrix (79) using flowcore (80). Gating was performed using CellEngine (CellCarta, Montreal, Canada).

Manual Gating. Traditional hierarchical gating was applied to identify 12 "landmark" immune populations: CD14+ "classical" monocytes, CD14-CD16+ "non-classical" monocytes, classical and plasmacytoid dendritic cells [cDC and pDC, respectively], basophils, CD56^{bright} and CD56^{dim} natural killer cells, regulatory CD4+ T cells, non-regulatory CD4+ T cells, CD8+ T cells, gd T cells as stained by either a pan-γδ T cell receptor (TCR) antibody or an antibody that only recognizes gd T cells with the Vδ2 chain (see **Supplementary Fig. 3.1** for gating strategy) as well as well-defined adaptive immune subsets (see **Supplementary Fig. 3.2** for the identity of these populations). Within each of the "parent" cell types, we manually gated positive and negative populations of biologically relevant phenotypic markers from the two mass cytometry panels (see **Table 3.3** for markers assessed on each "parent" population). For each of the parent cell types, we only included phenotypic markers for which we could clearly gate a positive population above background antibody staining levels.

Clustering by Statistical SCAFFoLD. We generated SCAFFoLD maps using the Scaffold R package. As described previously (43,44), using all of the live CD45+ leukocytes collected across participants and timepoints for each staining panel, we applied an unsupervised clustering algorithm based on the CLARA clustering algorithm to partition cells into a user-defined number of clusters (100 clusters per staining panel). We excluded Ki-67 and Granzyme B to avoid having functional markers cluster cells across cell types together. Landmark populations were gated as outlined in **Supplementary Fig. 3.1** (for cluster analysis, NK cells were treated as one population). We next generated force-directed graphs (SCAFFoLD maps) to visualize the association of each cluster with its likely parent landmark population. We excluded from our downstream analysis clusters that contained <20 cells in >80% of samples (12 clusters in Panel 1, 2 clusters in Panel 2) as well as clusters that contained cells that did not have the expected expression of classical landmark population (e.g., we excluded a cluster of cells that clustered with the CD8+ T cells but appeared to co-express the B cell marker, CD19 and may potentially represent doublets [median 0.09% of total CD8+ T cells at the acute timepoint]; all together, these 9 clusters in Panel 1 and 7 clusters in Panel 2 comprised 0.08% and 0.13% of the total live population at the acute timepoint). Cell clusters were thus determined to be "reliably" assigned to landmark cell populations if they were not excluded based on these criteria and if they were identified in Panel 1 for innate immune cells (total 17 classical and 3 non-classical monocyte, 9 NK cell, 4 cDC, 1 pDC clusters) and B cells (total 14 clusters) and Panel 2 for T cell phenotypes (total 6 CD4+ Treg, 20 non-Treg CD4+, 14 CD8+, and 2 γδ T cell clusters). In the SCAFFoLD maps depicted, a representative map from one participant at timepoint 1 is shown.

Quantification and Statistical Analysis

Change in Manually Gated Population and Cell Cluster Frequencies over Time. To measure the change in abundance of manually gated cell features (e.g., landmark and sub-landmark populations and populations expressing individual phenotypic markers) and cell clusters, the frequency of each feature (expressed as a % of the parent population) was log transformed with a constant factor of 1/10E6 or 1/10E3, respectively. Log-transformed values were adjusted for participant age and sex using a linear regression and the residuals (log-adjusted abundance) were used in downstream analyses. For age and sex, the median (± standard deviation) contribution of each of these factors to the variance for individual features was 1.63 (±5.83)% and 1.17 (±3.16)%, respectively. The change over time for the log-adjusted feature abundance between the "acute," "early convalescent" and "late convalescent" visits was assessed using a linear mixed effect (LME) model with the nlme R package (81) with log-transformed days since index visit as a fixed effect and participant ID as a random effect. The p values for each group of features were adjusted for multiple testing correction by Benjamini Hochberg with an FDR cutoff of 5% for a significant effect of time since index visit on feature abundance. For 95% confidence interval graphs, line graphs were generated in R using the package ggplot2 (82). The 95% confidence intervals for the median values were calculated by bootstrapping with 1000 iterations.

PCA Analysis. The log-adjusted manually gated features that were present across all ZIKV infected and cross-sectional uninfected samples (281 of 324 total features in the dataset) were used for principal component analysis with the function PCA (parameters: "scale.unit = TRUE", ncp = 5) from the R package FactoMineR (83). The samples were visualized in PCA space with PC1 and PC2 values as the coordinates using factoextra and ggplot2 in R.

Heatmaps. Heatmaps were made in R using the package ComplexHeatmap (84). For the manually gated features and cluster features summary heatmaps, the row/column orders, respectively, were determined using the R package seriation (85) with the traveling salesperson problem (TSP) method.

Network Correlation Analysis. Pairwise Spearman correlations were calculated on the log-adjusted feature abundances from samples at the acute visit for participants (n=17) who were previously exposed to Dengue and in an early stage of infection (pre-IgM at the time of Index visit). The p values were adjusted with the Benjamini-Hochberg method. The correlation matrix was hierarchically clustered using complete linkage based on Euclidean distance to create correlation modules. For the relationship between modules, the average value was calculated across all significant correlations (p_adj<0.05) between features within each module. For the module 5 - module 3 score, each module score is the sum of the z-score scaled log-adjusted cellular features within the module. The 95% confidence intervals for the correlation in the infected samples for each pairwise feature comparison was calculated using bootstrapping with 1000 iterations. For each significant correlation (p adjusted<0.05), the correlation was categorized as “shared” with the uninfected cohort if the correlation value in the uninfected cohort fell within the 95% confidence interval from the infected samples or had the same sign as the infected correlation and a magnitude greater than the 95% confidence interval magnitude maximum. Otherwise, the correlation was categorized as “unique.” Fisher’s exact test was used to determine odds ratio for correlations being unique to ZIKV as the exposure (versus being “shared” with the uninfected) and the indicated correlation attribute as the outcome. The circular network graph was visualized using ggplot2 and the marker network graph was visualized with igraph (86).

Antibody Associations. The NT_{80} titers at the 6-month timepoint of the DENV-exposed, ZIKV+ individuals from the larger REDS III cohort were classified into antibody tertiles. The association between age and sex and the 6-month NT_{80} titer groupings was assessed on the entire REDS-III cohort using one-way ANOVA and a Chi-square test of independence, respectively. To test the association between cellular immune phenotypes and ZIKV neutralizing antibody titers, we used acute or late-convalescent visit samples from participants who had not yet formed IgM at the index visit, were DENV-exposed, and who had 6-month NT_{80} titers available (n=14). Exact permutation tests were used to test for significant differences in the log-adjusted cellular features (age- and sex-adjusted) between samples from participants in the high versus low tertiles (n=5 in high group and n=6 in low group). The association between CMV IgG seropositivity and 6-month NT_{80} titers was assessed using the Wilcoxon Rank Sum test based on a larger subset of REDS-III study participants for whom CMV serostatus were available (n=10 CMV+, n=23 CMV-).

Antibody Associations Predictive Modeling. We used pROC (87) to plot ROC curves with log-adjusted feature abundance at the acute or late convalescent visit as the predictor and 6-month NT_{80} antibody titer category (e.g., “High” or “Low”) as the response for each participant. The 95% CI for the AUC values were computed with the default “DeLong” method.

References

1. Honein MA, Dawson AL, Petersen EE, Jones AM, Lee EH, Yazdy MM, et al. Birth defects among fetuses and infants of US women with evidence of possible zika virus infection during pregnancy. JAMA. 2017;317(1):59–68.
2. Zorrilla CD, García García I, García Fragoso L, De La Vega A. Zika virus infection in pregnancy: maternal, fetal, and neonatal considerations. J Infect Dis. 2017;216(suppl_10):S891–6.
3. Nithiyanantham SF, Badawi A. Maternal infection with Zika virus and prevalence of congenital disorders in infants: systematic review and meta-analysis. Can J Public Health. 2019;110(5):638–48.
4. Bhatnagar J, Rabeneck DB, Martines RB, Reagan-Steiner S, Ermias Y, Estetter LBC, et al. Zika virus RNA replication and persistence in brain and placental tissue. Emerging Infect Dis. 2017;23(3):405–14.
5. Osuna CE, Lim S-Y, Deleage C, Griffin BD, Stein D, Schroeder LT, et al. Zika viral dynamics and shedding in rhesus and cynomolgus macaques. Nat Med. 2016;22(12):1448–55.
6. Coffey LL, Pesavento PA, Keesler RI, Singapuri A, Watanabe J, Watanabe R, et al. Zika virus tissue and blood compartmentalization in acute infection of rhesus macaques. PLoS ONE. 2017;12(1):e0171148.
7. Barzon L, Percivalle E, Pacenti M, Rovida F, Zavattoni M, Del Bravo P, et al. Virus and antibody dynamics in travelers with acute zika virus infection. Clin Infect Dis. 2018;66(8):1173–80.

8. Calvet GA, Kara EO, Giozza SP, Bôtto-Menezes CHA, Gaillard P, de Oliveira Franca RF, et al. Study on the persistence of Zika virus (ZIKV) in body fluids of patients with ZIKV infection in Brazil. BMC Infect Dis. 2018;18(1):49.
9. Stone M, Bakkour S, Lanteri MC, Brambilla D, Simmons G, Bruhn R, et al. Zika virus RNA and IgM persistence in blood compartments and body fluids: a prospective observational study. Lancet Infect Dis. 2020;20(12):1446–56.
10. Lazear HM, Diamond MS. Zika virus: new clinical syndromes and its emergence in the western hemisphere. J Virol. 2016;90(10):4864–75.
11. Rodriguez-Barraquer I, Costa F, Nascimento EJM, Nery N, Castanha PMS, Sacramento GA, et al. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. Science. 2019;363(6427):607–10.
12. Shan C, Xie X, Shi P-Y. Zika virus vaccine: progress and challenges. Cell Host Microbe. 2018;24(1):12–7.
13. Abbink P, Stephenson KE, Barouch DH. Zika virus vaccines. Nat Rev Microbiol. 2018;16(10):594–600.
14. Richner JM, Diamond MS. Zika virus vaccines: immune response, current status, and future challenges. Curr Opin Immunol. 2018;53:130–6.
15. Ngono AE, Shresta S. Immune response to dengue and zika. Annu Rev Immunol. 2018;36:279–308.
16. Abbink P, Larocca RA, Visitsunthorn K, Boyd M, De La Barrera RA, Gromowski GD, et al. Durability and correlates of vaccine protection against Zika virus in rhesus monkeys. Sci Transl Med. 2017;9(420):eaao4163.

17. Larocca RA, Abbink P, Peron JPS, Zanotto PM de A, Iampietro MJ, Badamchi-Zadeh A, et al. Vaccine protection against Zika virus from Brazil. Nature. 2016;536(7617):474–8.

18. Andrade P, Gimblet-Ochieng C, Modirian F, Collins M, Cárdenas M, Katzelnick LC, et al. Impact of pre-existing dengue immunity on human antibody and memory B cell responses to Zika. Nat Commun. 2019;10(1):938.

19. Abbink P, Larocca RA, De La Barrera RA, Bricault CA, Moseley ET, Boyd M, et al. Protective efficacy of multiple vaccine platforms against Zika virus challenge in rhesus monkeys. Science. 2016;353(6304):1129–32.

20. Barouch DH, Thomas SJ, Michael NL. Prospects for a zika virus vaccine. Immunity. 2017;46(2):176–82.

21. Priyamvada L, Quicke KM, Hudson WH, Onlamoon N, Sewatanon J, Edupuganti S, et al. Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus. Proc Natl Acad Sci USA. 2016;113(28):7852–7.

22. Dejnirattisai W, Supasa P, Wongwiwat W, Rouvinski A, Barba-Spaeth G, Duangchinda T, et al. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus. Nat Immunol. 2016;17(9):1102–8.

23. Grifoni A, Pham J, Sidney J, O'Rourke PH, Paul S, Peters B, et al. Prior dengue virus exposure shapes T cell immunity to zika virus in humans. J Virol. 2017;91(24):e01469-17.

24. Wen J, Elong Ngono A, Regla-Nava JA, Kim K, Gorman MJ, Diamond MS, et al. Dengue virus-reactive CD8+ T cells mediate cross-protection against subsequent Zika virus challenge. Nat Commun. 2017;8(1):1459.

25. Lim MQ, Kumaran EAP, Tan HC, Lye DC, Leo YS, Ooi EE, et al. Cross-Reactivity and Anti-viral Function of Dengue Capsid and NS3-Specific Memory T Cells Toward Zika Virus. Front Immunol. 2018;9:2225.

26. Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nat Immunol. 2009;10(1):116–25.

27. Tan Y, Tamayo P, Nakaya H, Pulendran B, Mesirov JP, Haining WN. Gene signatures related to B-cell proliferation predict influenza vaccine-induced antibody response. Eur J Immunol. 2014;44(1):285–95.

28. Nakaya HI, Hagan T, Duraisingham SS, Lee EK, Kwissa M, Rouphael N, et al. Systems Analysis of Immunity to Influenza Vaccination across Multiple Years and in Diverse Populations Reveals Shared Molecular Signatures. Immunity. 2015;43(6):1186–98.

29. Li S, Sullivan NL, Rouphael N, Yu T, Banton S, Maddur MS, et al. Metabolic phenotypes of response to vaccination in humans. Cell. 2017;169(5):862-877.e17.

30. Popper SJ, Strouts FR, Lindow JC, Cheng HK, Montoya M, Balmaseda A, et al. Early transcriptional responses after dengue vaccination mirror the response to natural infection and predict neutralizing antibody titers. J Infect Dis. 2018;218(12):1911–21.

31. Koutsakos M, Rowntree LC, Hensen L, Chua BY, van de Sandt CE, Habel JR, et al. Integrated immune dynamics define correlates of COVID-19 severity and antibody responses. Cell Rep Med. 2021;2(3):100208.

32. Hu Z, Glicksberg BS, Butte AJ. Robust prediction of clinical outcomes using cytometry data. Bioinformatics. 2019;35(7):1197–203.

33. Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z, Germain RN, et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. Cell. 2014;157(2):499–513.

34. Lum F-M, Lye DCB, Tan JJL, Lee B, Chia P-Y, Chua T-K, et al. Longitudinal Study of Cellular and Systemic Cytokine Signatures to Define the Dynamics of a Balanced Immune Environment During Disease Manifestation in Zika Virus-Infected Patients. J Infect Dis. 2018;218(5):814–24.

35. Barros JB de S, da Silva PAN, Koga R de CR, Gonzalez-Dias P, Carmo Filho JR, Nagib PRA, et al. Acute Zika Virus Infection in an Endemic Area Shows Modest Proinflammatory Systemic Immunoactivation and Cytokine-Symptom Associations. Front Immunol. 2018;9:821.

36. da Silva MHM, Moises RNC, Alves BEB, Pereira HWB, de Paiva AAP, Morais IC, et al. Innate immune response in patients with acute Zika virus infection. Med Microbiol Immunol. 2019;208(6):703–14.

37. Michlmayr D, Andrade P, Gonzalez K, Balmaseda A, Harris E. CD14+CD16+ monocytes are the main target of Zika virus infection in peripheral blood mononuclear cells in a paediatric study in Nicaragua. Nat Microbiol. 2017;2(11):1462–70.

38. Lai L, Rouphael N, Xu Y, Natrajan MS, Beck A, Hart M, et al. Innate, T-, and B-Cell Responses in Acute Human Zika Patients. Clin Infect Dis. 2018;66(1):1–10.

39. Cimini E, Castilletti C, Sacchi A, Casetti R, Bordoni V, Romanelli A, et al. Human Zika infection induces a reduction of IFN-γ producing CD4 T-cells and a parallel expansion of effector Vδ2 T-cells. Sci Rep. 2017;7(1):6313.

40. Grifoni A, Costa-Ramos P, Pham J, Tian Y, Rosales SL, Seumois G, et al. Cutting Edge: Transcriptional Profiling Reveals Multifunctional and Cytotoxic Antiviral Responses of Zika Virus-Specific CD8+ T Cells. J Immunol. 2018;201(12):3487–91.
41. Williamson PC, Biggerstaff BJ, Simmons G, Stone M, Winkelman V, Latoni G, et al. Evolving viral and serological stages of Zika virus RNA-positive blood donors and estimation of incidence of infection during the 2016 Puerto Rican Zika epidemic: an observational cohort study. Lancet Infect Dis. 2020;20(12):1437–45.
42. Musso D, Richard V, Teissier A, Stone M, Lanteri MC, Latoni G, et al. Detection of Zika virus RNA in semen of asymptomatic blood donors. Clin Microbiol Infect. 2017;23(12):1001.e1-1001.e3.
43. Spitzer MH, Gherardini PF, Fragiadakis GK, Bhattacharya N, Yuan RT, Hotson AN, et al. An interactive reference framework for modeling a dynamic immune system. Science. 2015;349(6244):1259425.
44. Spitzer MH, Carmi Y, Reticker-Flynn NE, Kwek SS, Madhireddy D, Martins MM, et al. Systemic immunity is required for effective cancer immunotherapy. Cell. 2017;168(3):487-502.e15.
45. Michlmayr D, Kim E-Y, Rahman AH, Raghunathan R, Kim-Schulze S, Che Y, et al. Comprehensive immunoprofiling of pediatric zika reveals key role for monocytes in the acute phase and no effect of prior dengue virus infection. Cell Rep. 2020;31(4):107569.
46. Chandele A, Sewatanon J, Gunisetty S, Singla M, Onlamoon N, Akondy RS, et al. Characterization of Human CD8 T Cell Responses in Dengue Virus-Infected Patients from India. J Virol. 2016;90(24):11259–78.

47. Wang Z, Zhu L, Nguyen THO, Wan Y, Sant S, Quiñones-Parra SM, et al. Clonally diverse CD38+HLA-DR+CD8+ T cells persist during fatal H7N9 disease. Nat Commun. 2018;9(1):824.
48. Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science. 2020;369(6508).
49. Ellebedy AH, Jackson KJL, Kissick HT, Nakaya HI, Davis CW, Roskin KM, et al. Defining antigen-specific plasmablast and memory B cell subsets in human blood after viral infection or vaccination. Nat Immunol. 2016;17(10):1226–34.
50. Johnson JL, Rosenthal RL, Knox JJ, Myles A, Naradikian MS, Madej J, et al. The Transcription Factor T-bet Resolves Memory B Cell Subsets with Distinct Tissue Distributions and Antibody Specificities in Mice and Humans. Immunity. 2020;52(5):842-855.e6.
51. Sutton HJ, Aye R, Idris AH, Vistein R, Nduati E, Kai O, et al. Atypical B cells are part of an alternative lineage of B cells that participates in responses to vaccination and infection in humans. Cell Rep. 2021;34(6):108684.
52. Lau D, Lan LY-L, Andrews SF, Henry C, Rojas KT, Neu KE, et al. Low CD21 expression defines a population of recent germinal center graduates primed for plasma cell differentiation. Sci Immunol. 2017;2(7):eaai8153.
53. McElhaney JE, Zhou X, Talbot HK, Soethout E, Bleackley RC, Granville DJ, et al. The unmet need in the elderly: how immunosenescence, CMV infection, co-morbidities and frailty are a challenge for the development of more effective influenza vaccines. Vaccine. 2012;30(12):2060–7.

54. Merani S, Pawelec G, Kuchel GA, McElhaney JE. Impact of aging and cytomegalovirus on immunological response to influenza vaccination and infection. Front Immunol. 2017;8:784.
55. Bowyer G, Sharpe H, Venkatraman N, Ndiaye PB, Wade D, Brenner N, et al. Reduced Ebola vaccine responses in CMV+ young adults is associated with expansion of CD57+KLRG1+ T cells. J Exp Med. 2020;217(7):e20200004.
56. Chng MHY, Lim MQ, Rouers A, Becht E, Lee B, MacAry PA, et al. Large-Scale HLA Tetramer Tracking of T Cells during Dengue Infection Reveals Broad Acute Activation and Differentiation into Two Memory Cell Fates. Immunity. 2019;51(6):1119-1135.e5.
57. Rouers A, Chng MHY, Lee B, Rajapakse MP, Kaur K, Toh YX, et al. Immune cell phenotypes associated with disease severity and long-term neutralizing antibody titers after natural dengue virus infection. Cell Rep Med. 2021;2(5):100278.
58. Rahil Z, Leylek R, Schürch CM, Chen H, Bjornson-Hooper Z, Christensen SR, et al. Landscape of coordinated immune responses to H1N1 challenge in humans. J Clin Invest. 2020;130(11):5800–16.
59. Michlmayr D, Pak TR, Rahman AH, Amir E-AD, Kim E-Y, Kim-Schulze S, et al. Comprehensive innate immune profiling of chikungunya virus infection in pediatric cases. Mol Syst Biol. 2018;14(8):e7862.
60. Kazer SW, Aicher TP, Muema DM, Carroll SL, Ordovas-Montanes J, Miao VN, et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. Nat Med. 2020;26(4):511–8.
61. Ndhlovu ZM, Kamya P, Mewalal N, Kløverpris HN, Nkosi T, Pretorius K, et al. Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point. Immunity. 2015;43(3):591–604.

62. Takata H, Buranapraditkun S, Kessing C, Fletcher JLK, Muir R, Tardif V, et al. Delayed differentiation of potent effector CD8+ T cells reducing viremia and reservoir seeding in acute HIV infection. Sci Transl Med. 2017;9(377).

63. Townsley SM, Donofrio GC, Jian N, Leggat DJ, Dussupt V, Mendez-Rivera L, et al. B cell engagement with HIV-1 founder virus envelope predicts development of broadly neutralizing antibodies. Cell Host Microbe. 2021;29(4):564-578.e9.

64. Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scott M, Hagan T, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science. 2020;369(6508):1210–20.

65. Liu C, Martins AJ, Lau WW, Rachmaninoff N, Chen J, Imberti L, et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. Cell. 2021;184(7):1836-1857.e22.

66. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin J-B, Olsson A, et al. Robust T Cell Immunity in Convalescent Individuals with Asymptomatic or Mild COVID-19. Cell. 2020;183(1):158-168.e14.

67. Andrews SF, Chambers MJ, Schramm CA, Plyler J, Raab JE, Kanekiyo M, et al. Activation Dynamics and Immunoglobulin Evolution of Pre-existing and Newly Generated Human Memory B cell Responses to Influenza Hemagglutinin. Immunity. 2019;51(2):398-410.e5.

68. Diamond MS, Ledgerwood JE, Pierson TC. Zika virus vaccine development: progress in the face of new challenges. Annu Rev Med. 2019;70:121–35.

69. Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. Nat Med. 2021;27(7):1205–11.

70. Maciejewski S, Ruckwardt TJ, Morabito KM, Foreman BM, Burgomaster KE, Gordon DN, et al. Distinct neutralizing antibody correlates of protection among related Zika virus vaccines identify a role for antibody quality. Sci Transl Med. 2020;12(547):eaaw9066.

71. Hassert M, Harris MG, Brien JD, Pinto AK. Identification of protective CD8 T cell responses in a mouse model of zika virus infection. Front Immunol. 2019;10:1678.

72. Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, et al. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. Emerging Infect Dis. 2008;14(8):1232–9.

73. Williamson PC, Linnen JM, Kessler DA, Shaz BH, Kamel H, Vassallo RR, et al. First cases of Zika virus-infected US blood donors outside states with areas of active transmission. Transfusion. 2017;57(3pt2):770–8.

74. Whitbeck JC, Thomas A, Kadash-Edmondson K, Grinyo-Escuer A, Stafford LJ, Cheng C, et al. Antigenicity, stability, and reproducibility of Zika reporter virus particles for long-term applications. PLoS Negl Trop Dis. 2020;14(11):e0008730.

75. Owen RE, Sinclair E, Emu B, Heitman JW, Hirschkorn DF, Epling CL, et al. Loss of T cell responses following long-term cryopreservation. J Immunol Methods. 2007;326(1–2):93–115.

76. Odorizzi PM, Jagannathan P, McIntyre TI, Budker R, Prahl M, Auma A, et al. In utero priming of highly functional effector T cell responses to human malaria. Sci Transl Med. 2018;10(463):eaat6176.

77. Ornatsky OI, Lou X, Nitz M, Schäfer S, Sheldrick WS, Baranov VI, et al. Study of cell antigens and intracellular DNA by identification of element-containing labels and

metallointercalators using inductively coupled plasma mass spectrometry. Anal Chem. 2008;80(7):2539–47.

78. Zunder ER, Finck R, Behbehani GK, Amir E-AD, Krishnaswamy S, Gonzalez VD, et al. Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. Nat Protoc. 2015;10(2):316–33.
79. Han G, Spitzer MH, Bendall SC, Fantl WJ, Nolan GP. Metal-isotope-tagged monoclonal antibodies for high-dimensional mass cytometry. Nat Protoc. 2018;13(10):2121–48.
80. Ellis B, Haaland P, Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J, et al. flowCore: Basic structures for flowcytometry data. R package. 2019. Available from: https://doi.org/doi:10.18129/B9.bioc.flowCore
81. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear andNonlinear Mixed Effects Models. R package. 2019. Available from: https://rdrr.io/cran/nlme/
82. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, editor. Springer-Verlag New York; 2016. 213 p.
83. Lê S, Josse J, Husson F. FactoMineR: A Package for Multivariate Analysis. Journal of Statistical Software. 2008;25(1), 1–18.
84. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847-9.
85. Hahsler M, Hornik K, Buchta C. Getting things in order: An introduction to the R package seriation. Journal of Statistical Software. 2008;25(3),1–34.
86. Csardi G. The igraph software package for complex network research. InterJournal, ComplexSystems. 2006;1695(5):1-9.

87. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

88. Kassambara, A. and Mundt, F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package. 2017. Available from: https://cran.r-project.org/package=factoextra

Chapter 4: Immunosuppressive Signature in Peripheral CD14+ Myeloid Cells Predicts Resistance to Immunotherapy in Prostate Cancer

Elizabeth E. McCarthy[1,2,3,4,5,6,13], Bridget P. Keenan[8,11,12], David S. Lee[15], Yang Sun[4], Arielle Illano[11,12], Frances Fan[11,12], Whitney Tamaki[11,12], Kathryn Allaire[11,12], Alexander Cheung[8,11,12], Lawrence Fong[2,8,9,11,12,13,14], Matthew Spitzer[1,2,7,8,9,10], Chun Jimmie Ye[2,3,4,7,9,10,13,14]

[1]Department of Otolaryngology-Head and Neck Surgery, University of California, San Francisco; San Francisco, CA 94143, USA

[2]Department of Microbiology and Immunology, University of California, San Francisco; San Francisco, CA 94143, USA

[3]Institute for Human Genetics, University of California, San Francisco; San Francisco, CA 94143, USA

[4]Division of Rheumatology, Department of Medicine, University of California, San Francisco; San Francisco, CA 94143, USA

[5]Medical Scientist Training Program, University of California, San Francisco; San Francisco, CA 94143, USA

[6]Biological and Medical Informatics Graduate Program, University of California, San Francisco; San Francisco, CA 94143, USA

[7]J. David Gladstone-UCSF Institute for Genomic Immunology, San Francisco, CA 94158, USA

[8]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA

[9]Parker Institute for Cancer Immunotherapy, San Francisco, CA 94129, USA

[10]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

[11]Division of Hematology/Oncology, University of California, San Francisco, San Francisco, CA 94143, USA

[12]Cancer Immunotherapy Program, University of California, San Francisco, San Francisco, CA 94143, USA

[13]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94143, USA

[14]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA 94143, USA

[15]Department of Genome Sciences, University of Washington, Seattle, WA 98105, USA

Introduction

Immune checkpoint receptor inhibitors (ICIs), including those targeting cytotoxic T lymphocyte antigen 4 (CTLA-4), programmed cell death 1 (PD-1), or PD-1 ligand 1 (PD-L1), have been approved based on improved overall survival in multiple malignancies, particularly those with high mutational burden due to microsatellite instability (MSI)/mismatch repair deficiency (MMRD) (1–5). However, in most solid tumors, ICIs as monotherapies are efficacious in only ~20% of patients (6). In the tumor microenvironment, several biomarkers including the expression of checkpoint receptors, amount of T cell infiltration, and mutational landscape (7–9) are predictive of response to ICIs. Recently, many features of the circulating immune landscape have also been shown to be biomarkers predictive of response to immunotherapy (10) and even to reveal mechanisms of immune response to the tumor through recapitulating features of tumor infiltrating immune cells (11) and secondary lymphoid organs (12). Defining the molecular features of circulating immune cells that are predictive of response to immunotherapy has enormous potential for clinical biomarker profiling due to the easily accessible repeated sampling to study the changes in response to treatment (13).

In men, prostate cancer is the most common cancer and the second most common cause of cancer deaths in the United States with significant disparities in clinical outcomes between ethnicities (14). While the survival rate for patients with localized disease is nearly 100%, the survival rate for those with distant metastatic prostate cancer is only 30%. Currently, no ICIs are approved for prostate cancer patients except for tumors with high mutational burden due to MSI/MMRD. Two large trials of ipilimumab, an anti-CTLA-4 checkpoint receptor inhibitor, in metastatic castration-resistant prostate cancer (mCRPC) failed to reach the primary endpoint of an increase in overall survival (15,16). The low level of immune cell infiltration into prostate tumors (17) and the immunosuppressive features of the prostate tumor microenvironment (18) contributed to the failure of ICIs as monotherapies in prostate cancer (19). However, the improved progression-free survival and prostate-specific antigen (PSA) response in the treatment arm in the ipilimumab monotherapy trials suggest that ICIs may have unrealized potential as treatments for advanced prostate cancer.

Currently, sipuleucel-T, an antigen presenting cell vaccine, is the only immunotherapy approved for the treatment of mCRPC (20). Standard sipuleucel-T treatment consists of three cycles spaced two weeks apart. In each cycle, the patient undergoes leukapheresis and the resulting cell product is co-cultured with a fusion protein made of prostatic acid phosphatase (PAP), a prostate cancer antigen, and granulocyte macrophage colony-stimulating factor (GM-CSF) for three days before re-infusion into the patient. Paradoxically, while sipuleucel-T showed an improvement in overall survival that led to its approval, it failed to improve time to disease progression. Although sipuleucel-T can invoke tumor shrinkage in some patients, it may even elicit a tolerogenic T cell response and increase tumor burden in others (21). This presents an opportunity for examining the immune signatures and cell-cell interactions that predict either type of response in prostate cancer with potential broader significance in other cancers.

The recent success of using two ICIs together in prostate cancer in early clinical trials (22) highlights the utility of combination immunotherapy. Combination immunotherapy that targets the myeloid and lymphoid compartment is a rational and compelling strategy. The *ex vivo* activation of antigen presenting cells (APCs), which are primarily CD14+, that occurs with sipuleucel-T leads to the *in vivo* activation of B and T cell response that predicts increased overall survival (23).

Immunosuppressive myeloid cells play an important role in the pathogenesis of prostate cancer both as suppressive tumor associated macrophages (24) and peripherally as circulating myeloid derived suppressor cells (MDSCs). MDSCs are a heterogeneous group of cells that are classified as monocytic or granulocytic origin and are characterized by their ability to inhibit T cell proliferation and cytokine release (25). Circulating MDSC levels increase with increased prostate cancer stage and grade (26–28).

Interferon (IFN) signaling has a complex role within the MDSC compartment. Acute IFN signaling can elicit potent anti-tumor activity in part through suppression of MDSCs, but chronic IFN signaling in tumor cells and many immune cell types, including MDSCs, has been shown to be pro-tumor (29–32). Chronic IFN signaling genes are a subset of interferon signaling genes (ISGs) which maintain high expression for several days after an acute high dose of type I IFN in response to viral infection (33). Chronic IFN signaling genes are also upregulated during chronic low level type I IFN signaling both in both a normal setting, i.e., tonic IFN signaling (34), or a malignant setting, i.e., pro-tumor interferon-related DNA damage response (IRDS) genes (35,36). The exact set of ISGs which are turned on during chronic IFN signaling vary both with cell type and the underlying cellular environment. While chronic IFN signaling in tumor cells has been shown to contribute to immunotherapy resistance (32) the effect of chronic IFN signaling in myeloid cells, in particular the heterogeneous MDSC subset, on resistance to immunotherapy has not been well characterized.

Genetic multiplexing (37) can increase the throughput and decrease the batch effects of single cell experiments. Cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) (38) is used to simultaneously obtain gene and protein expression from the same cell. By using the novel combination of these methods (multiplexed CITE-seq), we simultaneously profiled protein and transcriptome expression from the same cell in ~400,000 peripheral blood mononuclear cells (PBMCs) from longitudinal peripheral blood samples from the prostate cancer immunotherapy (PCI) cohort, which contains participants with mCRPC receiving immunotherapy treatment with sipuleucel-T and ipilimumab.

We used this single cell profiling dataset for unbiased discovery of immunosuppressive transcriptional signatures within the myeloid compartment. We determined that the composition across these myeloid states associates with clinical response as measured by percent change in prostate-specific antigen (PSA), a serum marker used as a proxy for prostate tumor response to therapy. In particular, we identified a co-expressed chronic IFN and complement signature within the CD14+ myeloid compartment as a stable immune set point that predicts PSA response. Within the non-naïve CD8+ T cells, we found a T_{pex}-like CD8+ T cell cluster that was enriched in responders to immunotherapy. Our results present a CD14+ myeloid signature as a potential biomarker to identify participants with resistance to immunotherapy and suggest that T_{pex}-like cells are important mediators of response to ipilimumab in mCRPC.

Results

Inflammatory-Related Pathways are Upregulated in the Myeloid Compartment in Pre-Treatment mCRPC Samples Compared to Healthy Controls

We profiled PBMCs from 31 male clinical trial participants (21 white, 3 Hispanic, 3 black, and 5 unknown ethnicity; mean age is 65.8 years) with a median of 4 time points per participant. These PBMCs were collected during a clinical trial investigating the efficacy of serially combining sipuleucel-T with ipilimumab either immediately (13 participants) or with a three week cycle delay (18 participants). Each cohort received 4 doses of ipilimumab (10 mg/kg) spaced three weeks apart (**Fig. 4.1A** (left)).

The samples were processed using a genetic pooling strategy to enable multiplexed CITE-seq, simultaneous single-cell profiling of 99 cell-surface protein markers and transcriptomes (**Fig. 4.1A** (right)). We used a combination of the protein expression for canonical immune cell markers (e.g., CD3, CD19, CD14, and CD16) and differentially expressed genes between clusters resulted in identifying 14 cell types from 20 leiden clusters (**Fig. 4.1B**).

When we compared the healthy controls and pre-treatment mCRPC samples, we identified a monocyte/cDC cluster, composed of several canonical myeloid cell types (CD14+ classical monocytes, cDCs, etc.), that was almost entirely composed of healthy control samples compared to a CD14+ MHC Class II^{lo} cluster that was enriched in the mCRPC samples (**Fig. 4.1C**). The differentially expressed genes (e.g., *SLC7A11*, *NFE2L2*, *NQO1*, *ANXA5*) in the CD14+MHC Class II^{lo} cluster compared to the monocyte/cDC cluster reflect several genes previously shown to be upregulated in MDSCs (39–41). We used gene set enrichment analysis (GSEA) to find inflammation related pathways, namely response to interleukin-1 (IL-1) and *in vitro* generation of macrophages from monocytes in culture, that were enriched in the genes upregulated in the CD14+MHC Class II^{lo} cluster (**Fig. 4.1D**). The Monocyte/cDC cluster had higher expression many cDC associated (*CLEC10A*, *FCER1A*, *CD1D*, and *CD74*) and monocyte associated (*FCGR3A*, *LYZ*, *CD4*) genes.

CD14+ Myeloid Cells from mCRPC have Lower MHC Class II Expression and Occupy Distinct Cell States Compared Compared to Healthy Controls

In order to improve the separation of the myeloid subsets, we subsetted and re-clustered the myeloid cells (**Fig. 4.1E**). We used marker genes to identify the canonical myeloid cell types (**Fig. 4.1I**). Importantly, canonical myeloid subtypes (e.g., cDC, pDC, and CD16+ monocyte) from mCRPC and healthy controls were clustered together. This allowed us to refine the myeloid cell states that were enriched in the mCRPC versus healthy control samples. The majority of clusters with significant differential abundance from an exact permutation test were CD14+ clusters that were enriched in the pre-treatment mCRPC samples (cluster 0, 2, 3, 9, and 13) or healthy control samples (cluster 4 and 5) (**Fig. 4.1F-G**). The cells in the CD14+ clusters from mCRPC samples had significantly ($p = 3.4E-21$) lower expression of MHC Class II versus those from healthy controls (**Fig. 4.1H**).

mCRPC Samples Cellular Distribution across Myeloid Clusters Associated with Response to Immunotherapy

We next explored if the myeloid states we identified in our refined myeloid clustering also captured heterogeneity within our longitudinal mCRPC samples. Thus we quantified the counts of each mCRPC and healthy control sample across each of the myeloid clusters. This categorical data was visualized with correspondence analysis, a dimensionality reduction technique similar to principal component analysis (PCA) but which appropriately accounts for the compositional nature of the underlying data (42). The sample level visualization captured the separation of the healthy control samples (triangles) and pre-treatment (crosses) and post-treatment (dots) mCRPC samples we had already shown and also heterogeneity within the mCRPC samples (**Fig. 4.2A** (top)). We used k-means clustering to cluster the samples based on their coordinates in this reduced space and found three groupings (Group A, C, and D) of mCRPC samples.

We used the maximal percent change observed from longitudinal serum PSA values to classify the mCRPC samples into participants with any or no response in accordance with the most recent Prostate Cancer Clinical Trials Working Group (PCWG3) guidelines (43). Participants with a negative percent change in PSA at any point were classified as “Any response” versus participants who only had positive percent change in PSA who were classified as “No response”. We found a significant association (Pearson's chi-squared test, X^2 (2, N = 77) = 30.0, p = 0.000028.) between mCRPC group label and PSA response category with samples from Group D more likely than those from Group C to be from participants with “Any response” (**Fig. 4.2B**). The samples within group D also had a lower sample level percent change in PSA compared to those in Group C (**Fig. 4.2C**).

The visualization of the myeloid cluster coordinates within the correspondence analysis space suggested that the Group C samples had increased abundance in cluster 3 while Group D samples had increased abundance in clusters 0, 2, 9, and 13 (designated the “group D clusters”) (**Fig. 4.2A** (bottom)).This qualitative observation was confirmed by a quantitative analysis of the cluster abundances which showed the Group C samples had higher abundance for cluster 3 and lower abundance for the summed group D cluster abundances and vice versa for the group D samples (**Fig. 4.2D**).

mCRPC Participants with Resistance to Immunotherapy Upregulate Chronic Interferon Signature in CD14+ Myeloid Compartment

We used differential expression analysis to identify the significantly differentially expressed genes between cells from group C samples within the group C cluster (cluster 3) and cells from Group D samples from the group D clusters. The cells from the Group C cluster upregulated many ISGs. (**Fig. 4.2E**). Many of these ISGs overlapped with those in chronic IFN signatures described in viral (U-ISGF3 signature (44)), normal (Tonic IFN (34)), and malignant (IRDS/interferon-driven inhibitory ligands (IDILs) (32,35)) settings. Additionally, the single-cell score for the gene set significantly upregulated (log fold change > 1.5 and adjusted p value < 0.05) genes in the group C cluster (i.e., Cluster 3 sig.) were more highly correlated with the chronic IFN signatures compared to an acute IFN gene set (ISGF3 signature (44)) that contains

ISGs that are upregulated by acute IFN signaling and not by chronic interferon signaling (**Fig. 4.2F**). Thus, the ISGs upregulated in cluster 3 (Group C cluster) seem to represent genes upregulated by chronic rather than acute IFN signaling.

The cluster 3 abundance significantly positively correlates with higher percent change in PSA, representing increased tumor burden, across the mCRPC samples (**Fig. 4.2G**). We wanted to investigate if the chronic IFN signature in cluster 3 was driving this association. Thus, we gated out the cells in cluster 3 which did not express the chronic IFN signature captured by the U-ISGF3 gene set (which had the highest correlation with the cluster 3 signature). Surprisingly, the cluster 3 abundance of cells which did not overexpress the chronic IFN signature was also significantly positively correlated with higher percent change in PSA (**Fig. 4.2H**). Thus, we hypothesized that other signatures in cluster 3 were contributing to the association with resistance to immunotherapy.

Tensor Decomposition Reveals Chronic Interferon Signature and Complement Signature within the CD14+ Myeloid Cell Type

We used single-cell interpretable tensor decomposition (scITD) (45) as an orthogonal unbiased method to discover gene signatures within the mCRPC samples. We used pseudobulked counts across the myeloid cell types (**Fig. 4.1E**), B, T, and NK cell types from samples from each mCRPC participant as the input tensor (**Fig. 4.3A**). We used only one sample from each participant in order to power the discovery of gene programs that captured inter-individual rather than intra-individual variance. The tensor decomposition uncovered four multi-cellular gene programs (i.e., factors) with contributions from genes across each input cell type whose expression (i.e., factor sample scores) captured variance across the input samples. We projected those factors onto the tensor from all the mCRPC samples to get scores for each factor across all the samples.

We show the scores from Factor 2 and the genes with significant loadings in Factor 2 as an illustrative example. Samples with higher expression of the genes with positive Factor 2 loadings have a positive Factor 2 score while samples with lower expression of the genes with positive loadings (and higher expression of the genes with negative loadings, if applicable) have a negative Factor 2 score (**Fig. 4.3B** (left)). Many of the annotated genes with significant positive loadings across cell types are ISGs which overlap with the chronic IFN genes we previously identified as upregulated in cluster 3 (**Fig. 4.3B** (right)). The Factor 2 sample scores are also significantly positively correlated with the cluster 3 signature (**Fig. 4.3C** (left)). Thus Factor 2 seems to capture a chronic IFN signature across cell types that is correlated with the chronic IFN signature within the CD14+ myeloid compartment that is enriched in cluster 3.

The Factor 3 scores were also significantly positively correlated with the cluster 3 signature while the other two factors were not (**Fig. 4.3C** (right)). Factor 3 had the highest number of significant positive gene loadings in the CD14+ cell type. Many of the significant positive loading genes for Factor 3 in the CD14+ cells are part of the complement system (*CA2*, *FDX1*, *LGALS3*, *FCN1*, *PRCP*, *C3AR1*).

Co-expression of Chronic Interferon Signature and Complement Signature within CD14+ Myeloid Cells Predicts Resistance to Immunotherapy

We wanted to investigate whether the CD14+ cell gene signatures captured by Factor 2 (chronic IFN signature) and Factor 3 (complement signature) could independently predict immunotherapy response. Thus, we scored the expression of each factor signature, including genes with significant positive loadings for the CD14+ cell type for each factor, across the CD14+ cells and gated cells as expressing Factor 2 only (Factor 2 sig. > 0 and Factor 3 sig. < 0) or Factor 3 only (Factor 2 sig. < 0 and Factor 3 sig. > 0). We used six-fold cross validation to train a support vector model with a linear kernel as a classifier with the abundance of CD14+ cells expressing either Factor 2 or Factor 3 as the input variable and immunotherapy response category as the binary output variable. We used the Receiver Operating Characteristic (ROC) metric to evaluate the model performance and found that neither factor was predictive of immunotherapy response status compared to a random classifier (Factor 2 Area under the curve (AUC) = 0.53 and Factor 3 AUC = 0.41) (**Fig. 4.3D**).

Within cluster 3, in addition to cells that expressed only the Factor 2 or Factor 3 signature, there was a fraction of cells (31%) which co-expressed both signatures (**Fig. 4.3E**). We also found an enrichment in the percent of CD14+ cells co-expressing Factor 2 and Factor 3 signatures in samples from participants with resistance to immunotherapy (**Fig. 4.3F**). Thus, we hypothesized that the co-expression of Factor 2 and Factor 3 could be a predictor for immunotherapy response. The abundance of CD14+ cells co-expressing Factor 2 and Factor 3 signatures reliably predicted immunotherapy resistance (AUC = 0.79) (**Fig. 4.3G**). The ability to predict immunotherapy response from pre-treatment and on-treatment samples suggests this CD14+ state which co-expresses the Factor 2 and Factor 3 signatures is a stable immune set point present at baseline in mCRPC participants who are resistant to therapy which persists during immunotherapy treatment.

T_{pex}-like Cells Enriched in mCRPC Immunotherapy Responders

While the tensor decomposition captured two factors whose co-expression in CD14+ cells was increased in immunotherapy resistant mCRPC, it did not identify signatures that were enriched in the responders. The gene signature in the Group D clusters enriched in responders (**Fig. 4.2E**) contained inflammatory chemokines (*CXCL3*) and cytokines (*IL1B* and *IL1A*) which have been previously shown to be pro-tumor (46,47). Thus, we hypothesized that since the CD14+ cells in the responders did not have an immunostimulatory signature, they represented a permissive myeloid compartment where the lack of presumably immunosuppressive chronic IFN/complement CD14+ signature could allow for anti-tumor activity by other cell types. Thus, we focused on other immune compartments that could have immunostimulatory/anti-tumor signatures.

Cytotoxic CD8+ T cells are crucial targets of the response to checkpoint receptor inhibitors, including ipilimumab. Thus, we focused on the non-naïve CD8+ T cells by subsetting and clustering them (**Fig. 4.4A**). Of the 13 clusters, only cluster 6 was significantly enriched (adjusted p value = 0.03, linear mixed effect model) in participants with any response compared to participants with no response to immunotherapy (**Fig. 4.4B**). Interestingly, cluster 6 had the highest expression of *TCF7* across all the non-naïve CD8+ T cell clusters (**Fig. 4.4C**). Recently, TCF1+ (encoded by the *TCF7* gene) CD8+ T cells which also express some exhaustion markers (e.g., PD1) have emerged as "stem-like" or progenitor exhausted cells (termed T_{pex}) which are enriched in the tumor draining lymph node and can travel to the tumor to mediate the response to immunotherapy (48,49). To explore if cluster 6 represented a T_{pex}-like cell state, we scored gene signatures from a meta-analysis of tumor-infiltrating T cells (50) for terminally exhausted CD8+ T cells (Terminal T_{ex}) and *TCF7*+ T cells also expressing exhaustion markers (*TCF7*+ T_{ex}) (**Fig. 4.4D-E**). The majority of cells (70%) in cluster 6 only expressed the *TCF7*+ T_{ex} signature and did not express the Terminal T_{ex} signature (*TCF7*+ $T_{ex} > 0$ and Terminal $T_{ex} < 0$) (**Fig. 4.4F**).

Overall, our data suggests that resistance to immunotherapy in mCRPC is mediated by an immunosuppressive CD14+ myeloid cell state marked by the co-expression of a chronic interferon signature and complement signature. In contrast, in responders, an inflammatory CD14+ myeloid state represents a permissive myeloid signature which allows for anti-tumor T_{pex}-like CD8+ T cells to mediate a response to immunotherapy (**Fig. 4.4G**).

Discussion

Here we present single-cell profiling of PBMCs from an mCRPC cohort receiving a combined immunotherapy regime of sipuleucel-T and ipilimumab. We found an inflammatory CD14+ myeloid signature that was enriched in the pre-treatment mCRPC samples compared to the healthy controls. Human MDSCs are generally classified into three subtypes: early-stage MDSC (Lin-HLA-DR-CD33+), polymorphonuclear-MDSC (CD14-CD11b+CD15+), and monocytic-MDSC (M-MDSC) (CD11b+CD14+HLA-DR$^{low/-}$CD15-) (25). The CD14+ myeloid cells from the mCRPC samples had low expression of MHC Class II suggesting they overlap with the M-MDSC cell type. Thus, our dataset adds to the small set of studies providing single-cell profiling of MDSC/MDSC-like cells (51,52).

We performed a myeloid focused sub clustering to define subtypes within the heterogeneous CD14+ myeloid compartment. We found a co-expressed chronic interferon and complement signature in the CD14+ myeloid cells which predicted resistance to immunotherapy. Previously, Keenan et al. described a circulating CD14+ myeloid population with an inflammatory signature which was associated with resistance to anti-PD-1 therapy in biliary cancer (53). The CD14+ myeloid compartment in our mCRPC cells had a similar expression of inflammatory markers but only a subset expressed the predictive chronic interferon and complement signature. Thus our results may present a refinement of this previously described peripheral immunosuppressive CD14+ myeloid population. The complement signature is an intriguing therapeutic target since *C3AR1* is associated with macrophage infiltration in prostate adenocarcinoma (54) and has been proposed as a potential immune checkpoint receptor target (55). Recently, Boukhaled et al. showed that epigenetically programmed high interferon response capacity in CD4 effector T cells predicted resistance to anti-PD1 therapy (56). The chronic interferon signature we profiled in the CD14+ myeloid compartment was part of a multicellular pattern that was expressed across all the cell types (including T cells) in our tensor decomposition analysis. Future studies should focus on the contributions of chronic interferon signaling in different cell types to immunotherapy resistance.

In the responders, we found an increased level of T_{pex}-like CD8+ T cells compared to participants with no PSA response. T_{pex} cells in the periphery have emerged as important drivers of response to ICIs. Interestingly, in the context of murine chronic viral infection, TCF1 has been shown to promote T cell stemness through opposing type I IFN signaling (57). Thus chronic interferon signaling could be a pre-established immune set-point driven by past infections, tumor factors, or commensal microbes (58) which favors the development of T cell exhaustion over the maintenance of T cell stemness.

Our study adds to a growing collection of blood biomarker signatures that predict or associate with response to immunotherapy (10,12,59–62). The predictive chronic interferon/complement signature we described could be used as a biomarker to select mCRPC cohorts who do not have a pre-programmed resistance to immunotherapy. Cohort selection will be a powerful tool for realizing the potential of current ICIs within prostate cancer and other solid tumors (63). However even within the mCRPC participants with any response to immunotherapy, there were only a small number of complete responders (N=3 with a percent change in PSA less than -50%). Thus, for most responders, the decline in PSA represented a short-lived response to immunotherapy. Our findings suggest that distinct new targeted therapeutic approaches may be needed for individuals who are predicted to have resistance versus those who are partial responders to current immunotherapy therapies to ultimately improve response to immunotherapy in prostate cancer and other solid tumors.

Figures

Figure 1

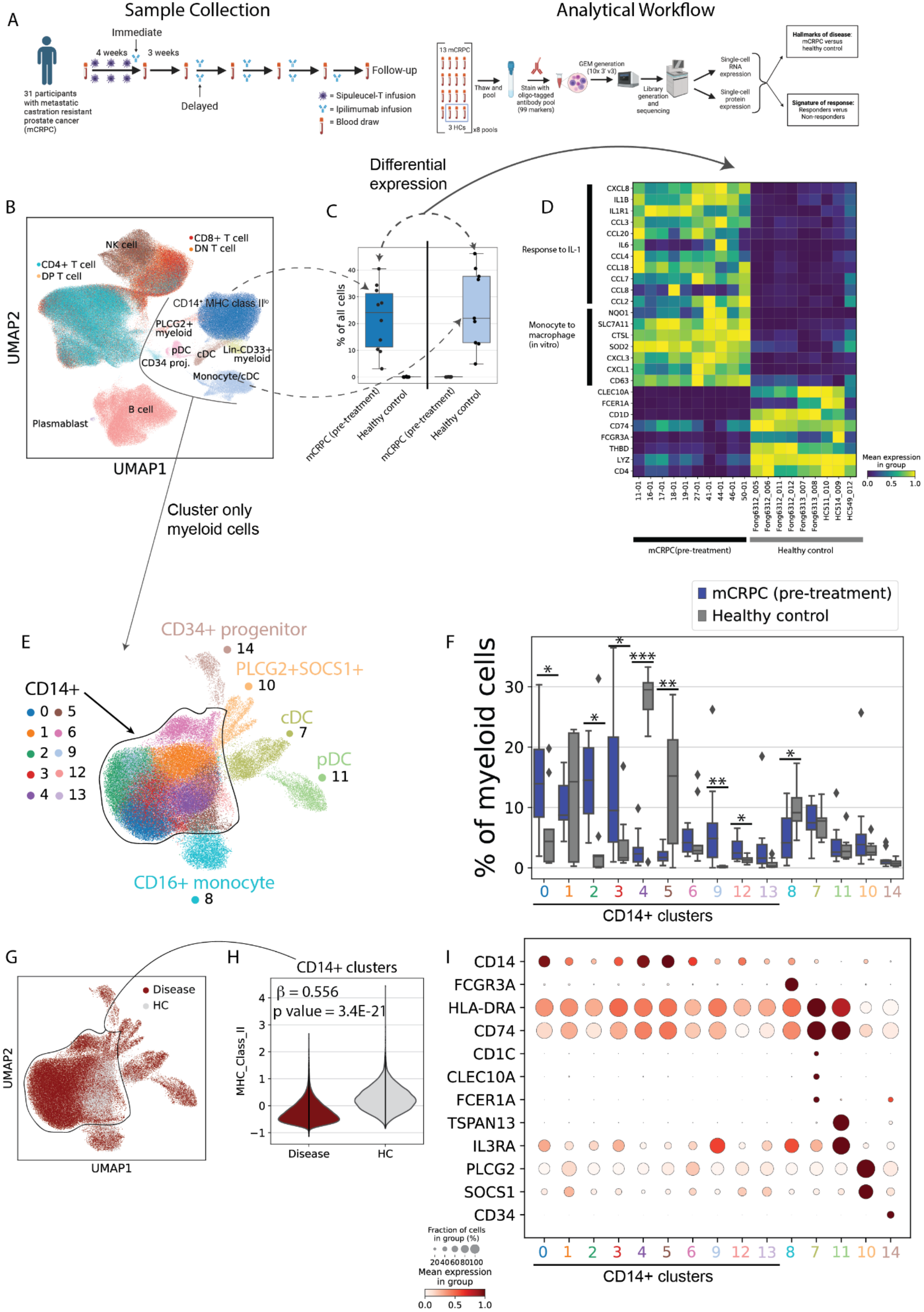


Figure 4.1 Inflammatory myeloid signature is a hallmark of metastatic prostate cancer in the peripheral immune compartment

(A) Combined immunotherapy regime with sipuleucel-T and ipilimumab and blood sampling timeline for participants with metastatic castration-resistant prostate cancer (mCRPC) in the PCI cohort (left). Pooling strategy for analysis of mCRPC and healthy control (HC) samples with multiplexed CITE-seq (right). (B) UMAP of all cells colored and labeled with cell type. (C) Box plot of abundance of each indicated cell type in the 10 pre-treatment mCRPC samples versus the 9 HC samples. (D) Heatmap showing standard scale normalized expression for the indicated genes (row labels) for pre-treatment mCRPC and HC samples. Row annotations identify enriched pathways that the genes are included in. (E) UMAP of myeloid cells colored by leiden cluster and labeled with annotated cell type. (F) Box plot of abundance of each indicated myeloid leiden cluster in the 10 pre-treatment mCRPC samples (blue) versus the 9 HC samples (gray). Significance indicated by asterisk for p value (exact permutation test) < 0.05 (*), < 0.1 (**), < 0.001 (***). (G) UMAP of myeloid cells colored by disease status (mCRPC – Disease [red] or Healthy Control – Healthy [gray]). (H) Violin plot of single cell scores for MHC Class II gene set for mCRPC (Disease) or Healthy Control (HC) cells from CD14+ myeloid clusters from (**E**/**G**). Effect size (β) and p value are from a linear mixed-effects model. (I) Dot plot showing expression of labeled genes for each cluster from (**E**) is shown by percentage of cells with expression greater than zero (dot size) and mean expression for cells with nonzero expression (color).

Figure 2

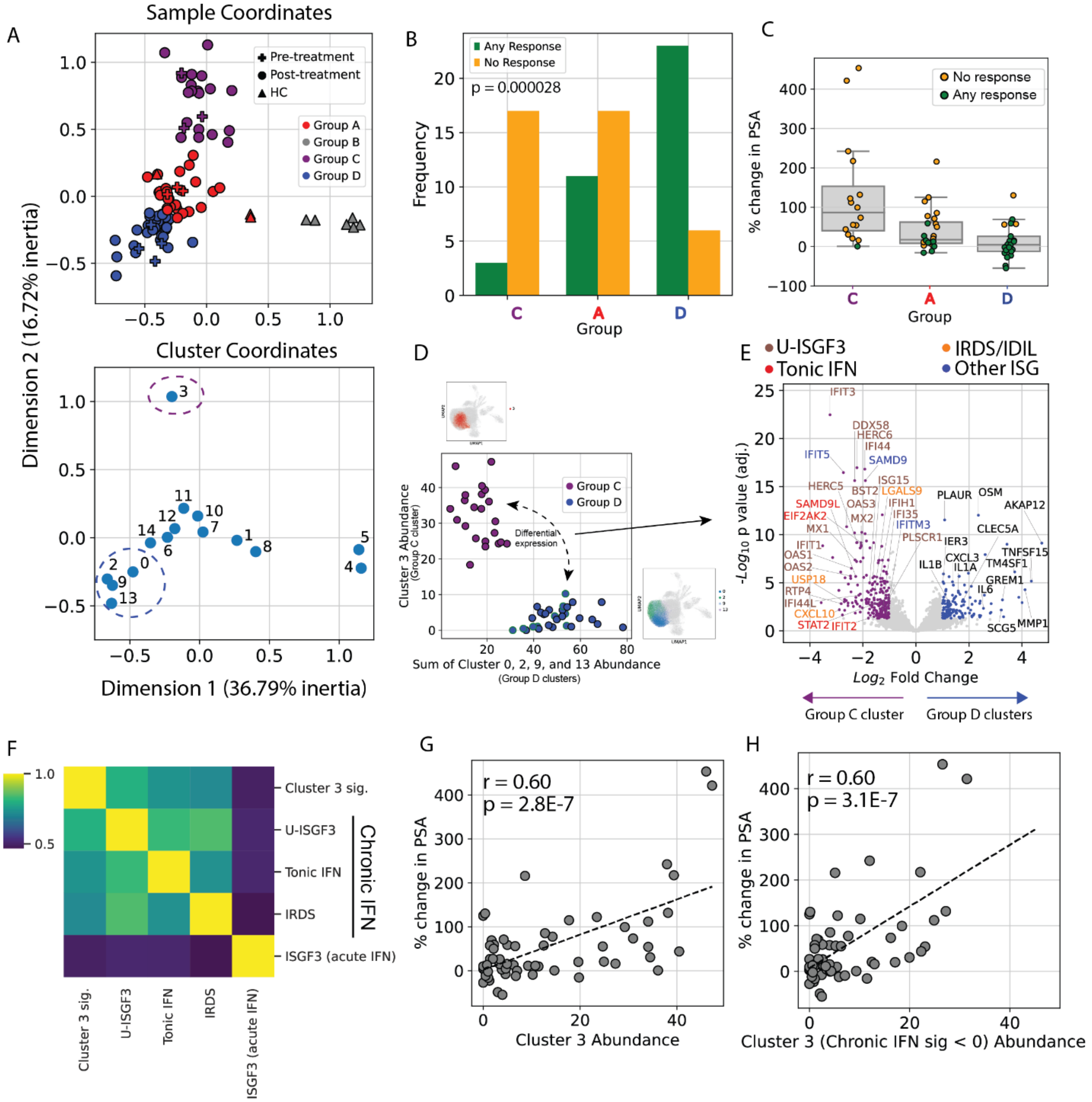

A
Sample Coordinates
Pre-treatment
Post-treatment
HC
Group A
Group B
Group C
Group D
Dimension 2 (16.72% inertia)
Cluster Coordinates
3
11
10
12
7
14
1
6
8
2
0
9
13
5
4
Dimension 1 (36.79% inertia)
B
Any Response
No Response
p = 0.000028
Frequency
C
A
D
Group
C
No response
Any response
% change in PSA
C
A
D
Group
D
Group C
Group D
Cluster 3 Abundance (Group C cluster)
Differential expression
Sum of Cluster 0, 2, 9, and 13 Abundance (Group D clusters)
E
U-ISGF3
Tonic IFN
IRDS/IDIL
Other ISG
IFIT3
DDX58
HERC6
IFI44
IFIT5
SAMD9
HERC5
ISG15
BST2
LGALS9
SAMD9L
OAS3
IFIH1
EIF2AK2
MX2
IFI35
MX1
IFITM3
PLSCR1
IFIT1
OAS1
OAS2
USP18
RTP4
IFI44L
CXCL10
STAT2
IFIT2
PLAUR
OSM
AKAP12
CLEC5A
IER3
TNFSF15
CXCL3
IL1B
IL1A
TM4SF1
GREM1
IL6
SCG5
MMP1
$-Log_{10}$ p value (adj.)
Log_2 Fold Change
Group C cluster
Group D clusters
F
Cluster 3 sig.
U-ISGF3
Tonic IFN
IRDS
ISGF3 (acute IFN)
Chronic IFN
G
r = 0.60
p = 2.8E-7
% change in PSA
Cluster 3 Abundance
H
r = 0.60
p = 3.1E-7
% change in PSA
Cluster 3 (Chronic IFN sig < 0) Abundance

Figure 4.2 Chronic interferon signature in CD14+ myeloid compartment captures mCRPC sample heterogeneity

(A) Correspondence analysis plot based on compositional data for each sample across the myeloid clusters in **Fig. 1E** showing coordinates for each sample (top) and for each cluster (bottom). Dots for each sample are colored according to group (Group A - D) assigned by k-means clustering on plot coordinates, and the dot shape indicates sample identity as pre-treatment (+) and during treatment (●) mCRPC samples or HC (▲). (B) Barplot of the number of samples from participants with no response (orange) or any response (green) within each myeloid group from **A**. The p value is from Pearson's chi-squared test. (C) Boxplot of percent change in PSA for samples in each group from **(A)**. Dots for each sample are colored according to the response for the participant of origin [no response (orange) or any response (green)]. (D) Scatter plot showing cluster 3 abundance (Group C enriched cluster) versus the sum of abundances in cluster 0, 2, 9, and 13 (Group D enriched clusters) for samples in Group C (dots colored purple) and Group D (dots colored blue). UMAPs highlighting each set of clusters are shown along each axis. (E) Volcano plot of differentially expressed genes from pseudobulked counts for cluster 3 from group C samples versus counts for group D clusters (cluster 0, 2, 9, and 13) from group D samples. Dots are colored as significantly overexpressed (adjusted p value < 0.05 and $\log_2$ fold change > 1) in Group C cluster (purple), Group D clusters (blue), or not significantly different between the groups (gray). Labeled genes which are significantly overexpressed in Group C are colored according to the interferon gene sets they belong to as indicated in the legend. (F) Hierarchical clustering of Pearson's correlation values between single-cell scores in all myeloid cells for the three gene sets used to define chronic interferon (IFN) signaling, the acute IFN only gene set, and the cluster 3/group C signature from **E**. Scatter plot of percent change in PSA versus cluster 3 abundance (G) and cluster 3 abundance for cells not expressing the U-ISGF3 chronic interferon signature (H) for each mCRPC sample. Dashed line is the best fit line and Pearson's correlation coefficient r and p value are given in overlaid text.

Figure 3

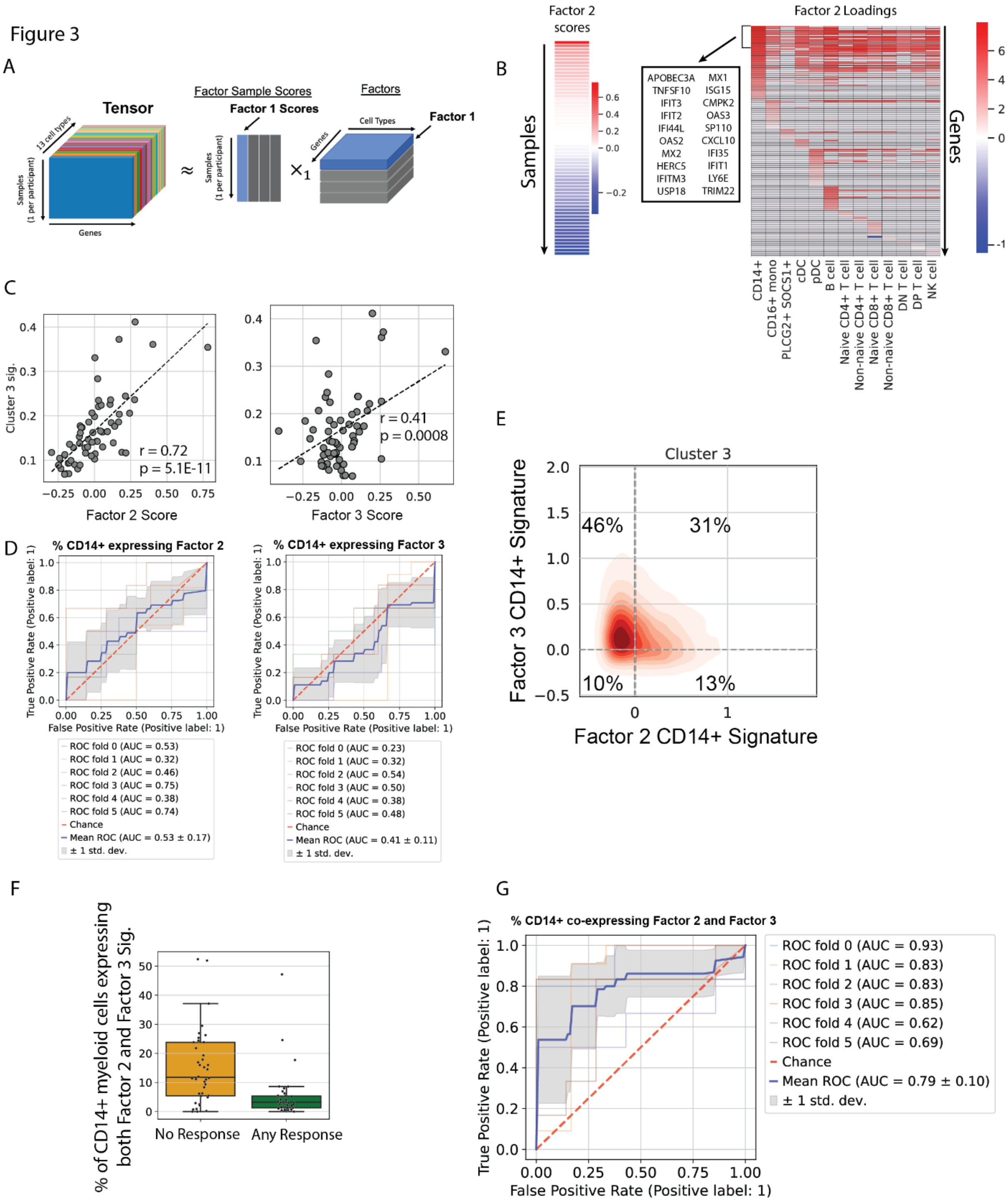

A
Tensor
13 cell types
Samples (1 per participant)
Genes
≈
Factor Sample Scores
Factor 1 Scores
Samples (1 per participant)
×1
Factors
Cell Types
Genes
Factor 1
B
Factor 2 scores
Samples
0.6
0.4
0.2
0.0
−0.2
Factor 2 Loadings
APOBEC3A MX1
TNFSF10 ISG15
IFIT3 CMPK2
IFIT2 OAS3
IFI44L SP110
OAS2 CXCL10
MX2 IFI35
HERC5 IFIT1
IFITM3 LY6E
USP18 TRIM22
Genes
6
4
2
0
−1
CD14+
CD16+ mono
PLCG2+ SOCS1+
cDC
pDC
B cell
Naive CD4+ T cell
Non-naive CD4+ T cell
Naive CD8+ T cell
Non-naive CD8+ T cell
DN T cell
DP T cell
NK cell
C
Cluster 3 sig.
0.4
0.3
0.2
0.1
r = 0.72
p = 5.1E-11
−0.25 0.00 0.25 0.50 0.75
Factor 2 Score
0.4
0.3
0.2
0.1
r = 0.41
p = 0.0008
−0.25 0.00 0.25 0.50
Factor 3 Score
D
% CD14+ expressing Factor 2
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
ROC fold 0 (AUC = 0.53)
ROC fold 1 (AUC = 0.32)
ROC fold 2 (AUC = 0.46)
ROC fold 3 (AUC = 0.75)
ROC fold 4 (AUC = 0.38)
ROC fold 5 (AUC = 0.74)
Chance
Mean ROC (AUC = 0.53 ± 0.17)
± 1 std. dev.
% CD14+ expressing Factor 3
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
ROC fold 0 (AUC = 0.23)
ROC fold 1 (AUC = 0.32)
ROC fold 2 (AUC = 0.54)
ROC fold 3 (AUC = 0.50)
ROC fold 4 (AUC = 0.38)
ROC fold 5 (AUC = 0.48)
Chance
Mean ROC (AUC = 0.41 ± 0.11)
± 1 std. dev.
E
Cluster 3
Factor 3 CD14+ Signature
2.0
1.5
1.0
0.5
0.0
−0.5
46%
31%
10%
13%
0
1
Factor 2 CD14+ Signature
F
% of CD14+ myeloid cells expressing both Factor 2 and Factor 3 Sig.
50
40
30
20
10
0
No Response
Any Response
G
% CD14+ co-expressing Factor 2 and Factor 3
True Positive Rate (Positive label: 1)
1.0
0.8
0.6
0.4
0.2
0.0
0.00 0.25 0.50 0.75 1.00
False Positive Rate (Positive label: 1)
ROC fold 0 (AUC = 0.93)
ROC fold 1 (AUC = 0.83)
ROC fold 2 (AUC = 0.83)
ROC fold 3 (AUC = 0.85)
ROC fold 4 (AUC = 0.62)
ROC fold 5 (AUC = 0.69)
Chance
Mean ROC (AUC = 0.79 ± 0.10)
± 1 std. dev.

Figure 4.3 CD14+ myeloid cells co-expressing chronic interferon signature and complement signature are predictive of immunotherapy response resistance
(A) Schematic of tensor decomposition for the pseudobulked counts across 13 cell types across all cells. (B) Heatmaps showing Factor 2 sample scores (left) and Factor 2 gene loading values (right). (C) Scatter plot of cluster 3/group C signature from **E** versus Factor 2 (left) or Factor 3 (right) samples scores. Dashed line is the best fit line and Pearson's correlation coefficient r and p value are given in overlaid text. (D) ROC curves for predicting no versus any response individuals using the percentage of cells in CD14+ clusters expressing only Factor 2 CD14+ cell type signature (left) or only Factor 3 CD14+ cell type signature (right). Legend shows AUC values for each six-fold cross validation. (E) Density plot for Factor 2 CD14+ cell type signature versus Factor 3 CD14+ cell type signature in cells from cluster 3. (F) Boxplot of percentage of cells in CD14+ clusters expressing Factor 2 CD14+ cell type signature and Factor 3 CD14+ cell type signature in samples from no (orange) or any (green) response individuals. (G) ROC curves for predicting no versus any response individuals using the percentage of cells in CD14+ clusters expressing Factor 2 CD14+ cell type signature and Factor 3 CD14+ cell type signature. Legend shows AUC values for each six-fold cross validation.

Figure 4

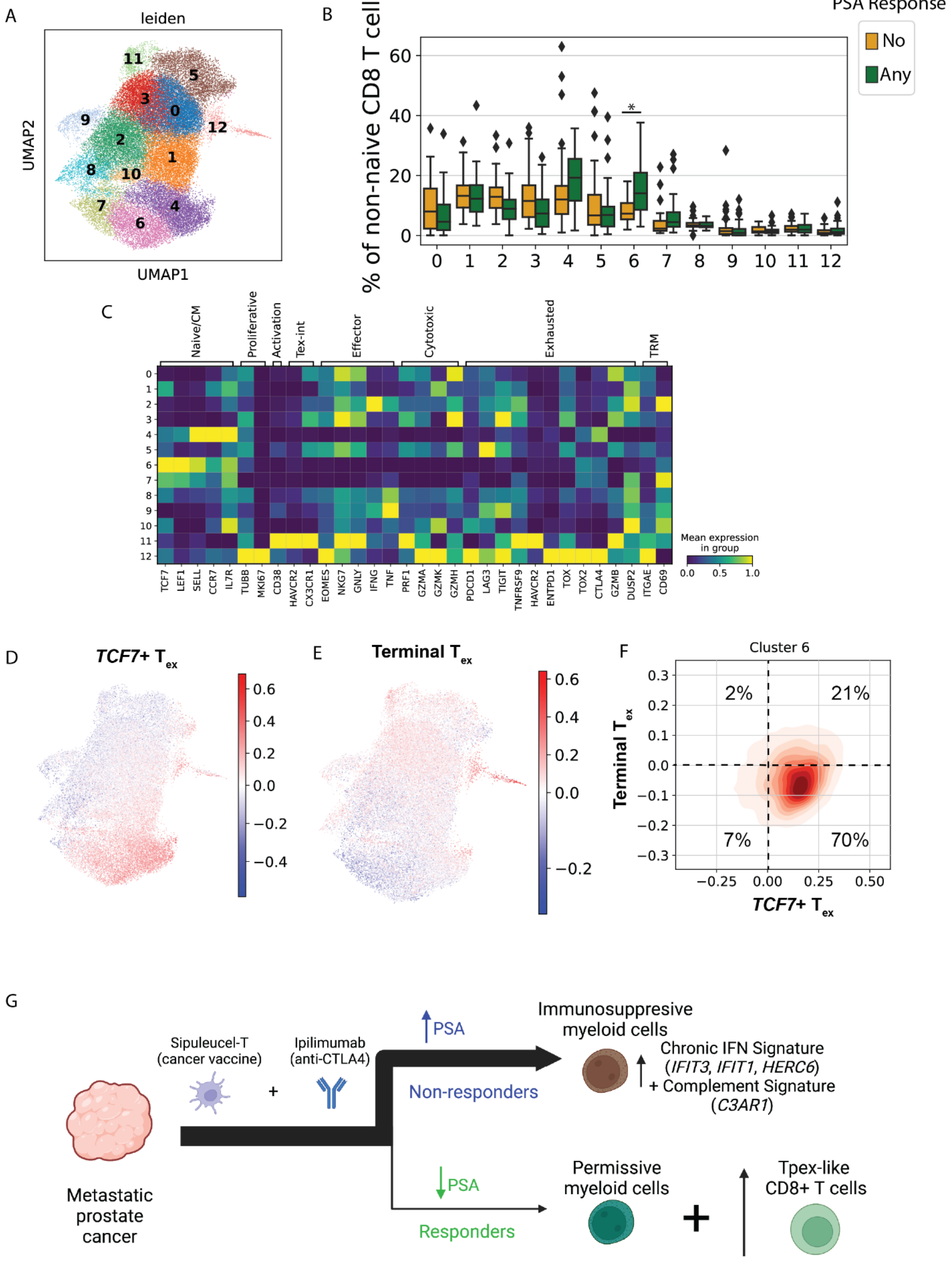

A
leiden
11
5
3
0
9
12
2
1
8
10
7
4
6
UMAP2
UMAP1
B
% of non-naive CD8 T cells
60
40
20
0
*
0 1 2 3 4 5 6 7 8 9 10 11 12
PSA Response
No
Any
C
Naive/CM
Proliferative
Activation
Tex-int
Effector
Cytotoxic
Exhausted
TRM
0 1 2 3 4 5 6 7 8 9 10 11 12
TCF7 LEF1 SELL CCR7 IL7R TUBB MKI67 CD38 HAVCR2 CX3CR1 EOMES NKG7 GNLY IFNG TNF PRF1 GZMA GZMK GZMH PDCD1 LAG3 TIGIT TNFRSF9 HAVCR2 ENTPD1 TOX TOX2 CTLA4 GZMB DUSP2 ITGAE CD69
Mean expression in group
0.0 0.5 1.0
D
TCF7+ T_ex
0.6
0.4
0.2
0.0
−0.2
−0.4
E
Terminal T_ex
0.6
0.4
0.2
0.0
−0.2
F
Cluster 6
Terminal T_ex
0.3 0.2 0.1 0.0 −0.1 −0.2 −0.3
2%
21%
7%
70%
−0.25 0.00 0.25 0.50
TCF7+ T_ex
G
Sipuleucel-T (cancer vaccine)
+
Ipilimumab (anti-CTLA4)
Metastatic prostate cancer
PSA
Non-responders
Immunosuppresive myeloid cells
Chronic IFN Signature (IFIT3, IFIT1, HERC6)
+ Complement Signature (C3AR1)
PSA
Responders
Permissive myeloid cells
+
Tpex-like CD8+ T cells

Figure 4.4 T_{pex}-like non-naïve CD8+ T cell cluster associated with response to immunotherapy

(A) UMAP of non-naïve CD8+ T cells colored by leiden cluster. (B) Boxplot of percentage of non-naïve CD8+ T cells in each cluster in **A** for responders versus non-responders. Significance indicated by asterisk for p value (linear mixed effect model) < 0.05 (*). (C) Heatmap showing standard scale expression of marker genes for clusters in **A**. UMAP of non-naïve CD8+ T cells colored by single-cell score of *TC7*+ T_{ex} (D) and Terminal T_{ex} (E) CD8+ T cell signatures (F) Density plot for *TC7*+ T_{ex} and Terminal T_{ex} single-cell scores in cells from cluster 6. (G) Graphical abstract summarizing distinct cellular immune signatures associated with the resistance or response to immunotherapy in mCRPC.

Materials and Methods

Single-cell RNA and ADT Library Preparation and Sequencing. The 99x antibody pool was prepared by combining 2 mL of each antibody from the 99x Abseq panel (BD Cat. no. 564220) and dialyzing into 100 mL of staining buffer using an Amicon Ultra-0.5 device. Antibody pool was kept at 4 °C during cell thawing.

Frozen PBMCs from participants in each of the 8 multiplexed pools (~16 participants/pool; 13 genetically distinct samples from the PCI cohort and 3 healthy age and gender-matched healthy controls) were each thawed into 10 mL of CHM media (500 ml RPMI plus 25 mL filtered human serum, 2.5 mL L-Glutamine, 5 mL Pen-Strep, 5 mL Sodium Pyruvate, and 5 mL Non-essential Amino Acids) then centrifuged and resuspended in 10 mL of CHM media with DNAseI (15 U / mL; Roche Cat. no. 04536282001). The cells were incubated in the DNAseI media for 30 minutes at 37 °C before being resuspended in CHM media for cell counting using a Cellometer Auto T4. Equal number of cells from each sample were combined to create a pool of 1,000,000 cells. The pooled cells were stained at room temperature for 10 minutes with the Human TruStain FcX blocking reagent (5 mL reagent; BD / 95 mL staining buffer (2% BSA/0.02% Tween in 1X PBS)) before staining for 45 minutes with the antibody pool at 4 °C. Stained cells were washed three times with 2 mL of staining buffer for each wash. The resulting cell pool was resuspended in CHM media and filtered through a Flowmi strainer (Sigma Cat. no. BAH136800040) to remove cell clumps. The single cell suspension was diluted to yield a 3,906 cell / mL solution of which 20 mL was used to load 78,125 cells/well over 4 wells per pool onto the 10x Genomics controller for a target capture rate of 30-40% of loaded cells/well for a yield of ~4,000 cells/sample.

Droplet-based paired single-cell gene expression (GEX) library prep was performed using the 10x Genomics Chromium Single Cell 3' v.3 kit per manufacturer's instructions from 10x Genomics with the addition of a 0.5 mL of a 4 mM of an additive primer (5'-CAGACGTGTGCTCTTCCGATCT) at the cDNA amplification step for the generation of the antibody derived tag (ADT) cDNA from the BD Abseq panel. The supernatant from the post cDNA amplification reaction cleanup was used to prep the ADT cDNA libraries per manufacturer's instructions from BD as detailed in Neely et al. (64). The resulting GEX and ADT libraries were sequenced on an Illumina Novaseq 6000 sequencer with paired-end (PE) reads with a target of ~25,000 PE reads per cell for both the GEX and ADT libraries from each well.

Bulk RNA Sequencing for Genotyping. RNA extraction from at least 40,000 cells was performed with the Qiagen Rneasy Mini Kit per manufacturer's instructions. cDNA was prepared and amplified using the previously published SMARTSeq2 protocol using the Illumina Nextera XT DNA Library Preparation Kit. The resulting libraries were sequenced on an Illumina HiSeq 4000 sequencer with paired-end (PE) reads with a target of 25,000,000 reads per sample. Genotypes for each participant were extracted from the resulting fastqs as described in (65). Briefly, quality control for the raw fastq reads was completed with FastQC v.0.11.8 (66) and low quality reads were trimmed with Trim Galore v.0.4.4_dev as a wrapper for Cutadapt v.1.18 (67). The filtered reads were aligned using the STAR v.2.4.2a_modified (68) with the default settings to the GRCh38 transcriptome. GATK v.4.0.6.0 (69) was used to call SNPs to yield genotypes for each individual.

Alignment of Single-cell Sequencing Data. The GEX and ADT fastq files were aligned using CellRanger v.3.0.1 with the default settings to the GRCh38 transcriptome or a customized reference file from BD for the Abseq antibodies, respectively, resulting in 873,755 cells (based on the CellRanger cell calling algorithm from the GEX libraries) with counts for 33,538 genes and 99 proteins.

Demultiplexing. The aligned GEX sequencing reads were demultiplexed with freemuxlet (70), as previously described (71), to identify cells from each participant based on host genetics and to exclude doublets from the data. The genotypes estimated from bulk RNA-seq were compared to those estimated by freemuxlet to match the sample ids with the freemuxlet ids. By repeating this protocol over 8 pools, 98% of cells from a total of 117 samples were successfully demultiplexed.

Cell and Gene Filtering. The GEX and ADT count matrices from all pools were combined and filtered (72). Low quality cells (17% of total cells) were filtered out based on having number of genes detected or total counts of unique molecular identifiers (UMIs) less than 3 median absolute deviations (MAD) below the median and with percent of total UMIs mapping to mitochondrial genes greater than 3 MAD above the median for each 10x well. Genetic doublets (24% of high quality cells) were removed by only keeping cells which were called as singlets for the same participant from the consensus of freemuxlet and souporcell v.2.0 (73). To detect non-genetic doublets, we ran the scrublet well on cells from each individual well and used a Gaussian mixture model to cluster the simulated doublet scores into two distributions. The minimum doublet score of the distribution with the higher peak doublet score was used as the threshold to mark "neotypic" doublets based on the doublet score for each cell assigned by scrublet v.0.2.3. These neotypic doublets (1.1% of high quality cells) were also removed. We filtered out 7,031 genes that were detected in less than 3 cells.

To avoid overly stringent filtering based on low counts of *HBB*, we filtered out red blood cells (RBCs) (1.0% of high quality singlets) based on clustering cells in each well into RBC and non-RBC distributions based on raw *HBB* counts using a Gaussian mixture model implemented with sklearn v.0.24.1 (74). Platelets (raw *PF4* count > 0) (2.6% of high quality singlets) were also filtered out. Finally, cells from healthy control samples that were genetically female (5.2% of high quality singlets) and cells from samples (N=7) with less than 1000 cells (0.5% of filtered singlets) were also removed to leave a matrix of 408,783 cells and 26,867 genes.

Processing for ADT Counts. The ADT counts were CLR normalized by cell.

Normalization, Batch Correction, and Visualization. The raw counts were normalized to 10,000 counts and log1p transformed. For downstream analyses, we excluded all RBC related genes (*HBA1*, *HBA2*, *HBB*, *HBG1*, *HBG2*, *HBQ1*, *HBD*, *HBM*, *HBZ*) and *XIST*, to avoid clustering cells based on detection of these genes. We identified 1,850 highly variable genes which were used with the default settings in scanpy v.1.7.1 for principal component (PC) analysis. These PC coordinates were used as the input to Harmony v.0.0.5 (75) for batch correction with each pool as a batch. The top 20 batch corrected PCs were used for nearest neighbor detection with scanpy.pp.neighbors. The neighborhood graph was used for clustering with the leiden algorithm (76) and dimensionality reduction with uniform manifold approximation and projection (UMAP) (77). This resulted in 20 clusters which were collapsed into 14 cell types. We used the expression of marker genes and proteins for the B cell, myeloid, and NK cell type annotations. For the T cells, we automatically gated positive and negative cells using a Gaussian mixture model implemented with sklearn v.0.24.1 to gate CD4+ T cells (CD3+CD4+CD8-), CD8+ T cells (CD3+CD4-CD8+), DP T cells (CD3+CD4+CD8+), and DN T cells (CD3+CD4-CD8-).

Sub-clustering of Myeloid Cells. We first subsetted 80,280 myeloid cells. We assigned cell cycle phase to each cell using the scanpy function sc.tl.score_genes_cell_cycle with the gene list from (78). We regressed out percent of counts mapped to mitochondrial genes, percent of counts mapped to ribosomal genes. We identified 1,551 highly variable genes which were used with the default settings in scanpy v.1.7.1 for principal component (PC) analysis. These PC coordinates were used as the input to Harmony v.0.0.5 for batch correction with cell cycle phase, disease status, and pool as a batch, iteratively. The top 20 batch corrected PCs were used for nearest neighbor detection with scanpy.pp.neighbors, clustering with the leiden algorithm, and visualization with UMAP. This resulted in 15 clusters which were collapsed into 6 cell types.

Sub-clustering of Non-naïve CD8+ T Cells. We first subsetted 45,780 non-naïve CD8+ T cells. Naïve CD8 T cells were removed by gating CD45RA+*CCR7*+ cells from the CD8+ T cells with a Gaussian mixture model implemented with sklearn v.0.24.1. We assigned cell cycle phase to each cell using the scanpy function sc.tl.score_genes_cell_cycle with the gene list from (78). We identified 1,603 highly variable genes which were used with the default settings in scanpy v.1.7.1 for principal component (PC) analysis. These PC coordinates were used as the input to Harmony v.0.0.5 for batch correction with cell cycle phase, disease status, and pool as a batch, iteratively. The top 20 batch corrected PCs were used for nearest neighbor detection with scanpy.pp.neighbors, clustering with the leiden algorithm, and visualization with UMAP to identify and graph 13 clusters.

Differential Gene Expression. Pseudobulked samples were normalized with the variance stabilizing transformation (VST) function from DESeq2 v.1.22.2 (79). Differential gene expression was done with a negative binomial model with multiple testing correction with Benjamini-Hochberg implemented via DESeq2.

Gene Set Enrichment Analysis. The differential gene list was filtered to remove genes with NA for the adjusted p value or log fold change. The filtered gene list was used to create ranked gene lists with the sign(log fold change) times the -log10(raw p value) as the ranking metric. The ranked list was used as input to look for gene set enrichment in pathways from the BP: subset of GO gene sets of the C5: ontology gene sets and the ImmuneSigDB subset of the C7: immunologic signature gene sets from the Human Molecular Signatures Database (MSigDB) (80) in the 'classic' mode with the GSEAPreranked tool from GSEA v.4.1.0 (http://www.broad.mit.edu/gsea/) with the default settings.

Differential MHC Class II Expression. MHC Class II expression single cell scores were assigned using the log-normalized scaled gene counts and the gene list (*HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DQA2*, *HLA-DQB2*, *HLA-DOB*, *HLA-DMB*, *HLA-DMA*, *HLA-DOA*, *HLA-DPA1*, *HLA-DPB1*, *CD74*) as inputs to the score_genes function from scanpy. Significant difference in MHC Class II expression for cells from *CD14*+ clusters from mCRPC or healthy control (HC) samples was determined with a linear mixed effect model MHC Class II score ~ disease category (mCRPC or HC) + leiden cluster with a random intercept for each sample implemented with statsmodels v.0.13.2 (81).

Correspondence Analysis. We implemented the correspondence analysis (CA) from prince v.0.7.1 (82). We performed k-means clustering with sklearn v.0.23.1 on the standard scaled CA coordinates. We determined the number of clusters using the cluster number from 1 to 10 clusters with the minimum silhouette score and minimum sum of square error.

PSA Analysis. We calculated percent change in PSA from the baseline PSA value closest to after the completion of sipuleucel-T (Day 0) [range of baseline PSA: -16 days to 19 days; median days = -3.5. Participants with any negative percent change in PSA were assigned as "Any response" and participants with only positive percent change in PSA were assigned as "No response”. For the association between PSA response status and CA plot myeloid group, we used a chi-square test from scipy v.1.8.1 (83).

Gene Set Scoring. For single-cell scoring of gene signatures, we used the indicated gene list to create single-cell scores from the log-normalized scaled gene counts with the score_genes function from scanpy v.1.9.1.

Tensor Decomposition. We used pool distributed samples (one from each participant) for the initial tensor decomposition with scITD v.1.0.2. Only samples with at least 10 cells in every cell type were included in the initial tensor decomposition and for the projection of the factors.

Predictive Modeling. We created a support vector classifier model with sklearn v. 1.1.1. For the training/test splits, we partitioned all the samples from a particular participant into either the training or test set to avoid inflated performance due to overfitting in the training set on samples from a participant and then testing on samples from the same participant in the test set.

Other. We used scipy v.1.8.1 for Pearson correlation coefficient calculation, NumPy v.1.22.4 (84) for the best fit line analysis, pandas v.1.4.3 (85) for data frame manipulation, and matplotlib v.3.5.2 (86) and seaborn v.0.11.2 (87) for the visualizations. Figure 4.1A and 4.4G were created with BioRender.com.

References

1. Hodi FS, O’Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, et al. Improved survival with ipilimumab in patients with metastatic melanoma. The New England journal of medicine. 2010;363(8):711–23.

2. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature. 2014;515(7528):563–7.

3. Powles T, Eder JP, Fine GD, Braiteh FS, Loriot Y, Cruz C, et al. MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. Nature. 2014;515(7528):558–62.

4. Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, et al. Nivolumab in previously untreated melanoma without BRAF mutation. The New England journal of medicine. 2015;372(4):320–30.

5. Sahin IH, Akce M, Alese O, Shaib W, Lesinski GB, El-Rayes B, et al. Immune checkpoint inhibitors for the treatment of MSI-H/MMR-D colorectal cancer and a perspective on resistance mechanisms. British journal of cancer. 2019;121(10):809–18.

6. Chiou VL, Burotto M. Pseudoprogression and Immune-Related Response in Solid Tumors. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2015;33(31):3541–3.

7. Davis AA, Patel VG. The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors.

Journal for immunotherapy of cancer. 2019;7(1):278.

8. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. The New England journal of medicine. 2014;371(23):2189–99.

9. Bonaventura P, Shekarian T, Alcazer V, Valladeau-Guilemond J, Valsesia-Wittmann S, Amigorena S, et al. Cold Tumors: A Therapeutic Challenge for Immunotherapy. Frontiers in immunology. 2019;10:168.

10. Kagamu H, Kitano S, Yamaguchi O, Yoshimura K, Horimoto K, Kitazawa M, et al. CD4+ T-cell Immunity in the Peripheral Blood Correlates with Response to Anti-PD-1 Therapy. Cancer immunology research. 2020;8(3):334–44.

11. Tobin RP, Jordan KR, Kapoor P, Spongberg E, Davis D, Vorwald VM, et al. IL-6 and IL-8 Are Linked With Myeloid-Derived Suppressor Cell Accumulation and Correlate With Poor Clinical Outcomes in Melanoma Patients. Frontiers in oncology. 2019;9:1223.

12. Spitzer MH, Carmi Y, Reticker-Flynn NE, Kwek SS, Madhireddy D, Martins MM, et al. Systemic Immunity Is Required for Effective Cancer Immunotherapy. Cell. 2017;168(3):487–502.e15.

13. Nixon AB, Schalper KA, Jacobs I, Potluri S, Wang IM, Fleener C. Peripheral immune-based biomarkers in cancer immunotherapy: can we realize their predictive potential? Journal for immunotherapy of cancer. 2019;7(1):325.

14. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians. 2020;70(1):7–30.

15. Kwon ED, Drake CG, Scher HI, Fizazi K, Bossi A, van den Eertwegh AJM, et al. Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. The lancet oncology. 2014;15(7):700–12.

16. Beer TM, Kwon ED, Drake CG, Fizazi K, Logothetis C, Gravis G, et al. Randomized, Double-Blind, Phase III Trial of Ipilimumab Versus Placebo in Asymptomatic or Minimally Symptomatic Patients With Metastatic Chemotherapy-Naive Castration-Resistant Prostate Cancer. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2017;35(1):40–7.

17. Zhao SG, Lehrer J, Chang SL, Das R, Erho N, Liu Y, et al. The Immune Landscape of Prostate Cancer and Nomination of PD-L2 as a Potential Therapeutic Target. Journal of the National Cancer Institute. 2019;111(3):301–10.

18. Vitkin N, Nersesian S, Siemens DR, Koti M. The Tumor Immune Contexture of Prostate Cancer. Frontiers in immunology. 2019;10:603.

19. Laccetti AL, Subudhi SK. Immunotherapy for metastatic prostate cancer: immuno-cold or the tip of the iceberg? Current opinion in urology. 2017;27(6):566–71.

20. Kantoff PW, Higano CS, Shore ND, Berger ER, Small EJ, Penson DF, et al. Sipuleucel-T immunotherapy for castration-resistant prostate cancer. The New England journal of medicine. 2010;363(5):411–22.

21. Hagihara K, Chan S, Zhang L, Oh DY, Wei XX, Simko J, et al. Neoadjuvant sipuleucel-T

induces both Th1 activation and immune regulation in localized prostate cancer. Oncoimmunology. 2019;8(1):e1486953.

22. Patel A, Fong L. Immunotherapy for Prostate Cancer: Where Do We Go From Here?-PART 2: Checkpoint Inhibitors, Immunotherapy Combinations, Tumor Microenvironment Modulation, and Cellular Therapies. Oncology. 2018;32(6):e65–e73.

23. Sheikh NA, Petrylak D, Kantoff PW, Dela Rosa C, Stewart FP, Kuan LY, et al. Sipuleucel-T immune parameters correlate with survival: an analysis of the randomized phase 3 clinical trials in men with castration-resistant prostate cancer. Cancer immunology, immunotherapy: CII. 2013;62(1):137–47.

24. Lissbrant IF, Stattin P, Wikstrom P, Damber JE, Egevad L, Bergh A. Tumor associated macrophages in human prostate cancer: relation to clinicopathological variables and survival. International journal of oncology. 2000;17(3):445–51.

25. Bronte V, Brandau S, Chen SH, Colombo MP, Frey AB, Greten TF, et al. Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards. Nature communications. 2016;7:12150.

26. Mehra N, Seed G, Lambros M, Sharp A, Fontes MS, Crespo M, et al. Myeloid-derived suppressor cells (MDSCs) in metastatic castration-resistant prostate cancer (CRPC) patients (PTS). Annals of oncology: official journal of the European Society for Medical Oncology / ESMO. 2016;27(suppl_6):VI257.

27. Chi N, Tan Z, Ma K, Bao L, Yun Z. Increased circulating myeloid-derived suppressor cells correlate with cancer stages, interleukin-8 and -6 in prostate cancer. International journal of

clinical and experimental medicine. 2014;7(10):3181–92.

28. Idorn M, Køllgaard T, Kongsted P, Sengeløv L, Thor Straten P. Correlation between frequencies of blood monocytic myeloid-derived suppressor cells, regulatory T cells and negative prognostic markers in patients with castration-resistant metastatic prostate cancer. Cancer immunology, immunotherapy: CII. 2014;63(11):1177–87.

29. Fenton SE, Saleiro D, Platanias LC. Type I and II Interferons in the Anti-Tumor Immune Response. Cancers. 2021;13(5):1037.

30. Taleb K, Auffray C, Villefroy P, Pereira A, Hosmalin A, Gaudry M, et al. Chronic Type I IFN Is Sufficient To Promote Immunosuppression through Accumulation of Myeloid-Derived Suppressor Cells. Journal of immunology . 2017;198(3):1156–63.

31. Jorgovanovic D, Song M, Wang L, Zhang Y. Roles of IFN-γ in tumor progression and regression: a review. Biomarker research. 2020;8:49.

32. Benci JL, Xu B, Qiu Y, Wu TJ, Dada H, Twyman-Saint Victor C, et al. Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. Cell. 2016;167(6):1540–54.e12.

33. Cheon H, Wang Y, Wightman SM, Jackson MW, Stark GR. How cancer cells make and respond to interferon-I. Trends in cancer research. 2022;S2405-8033(22):194-7.

34. Mostafavi S, Yoshida H, Moodley D, LeBoité H, Rothamel K, Raj T, et al. Parsing the Interferon Transcriptional Network and Its Disease Associations. Cell. 2016;164(3):564–78.

35. Khodarev NN, Beckett M, Labay E, Darga T, Roizman B, Weichselbaum RR. STAT1 is

overexpressed in tumors selected for radioresistance and confers protection from radiation in transduced sensitive cells. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(6):1714–9.

36. Weichselbaum RR, Ishwaran H, Yoon T, Nuyten DSA, Baker SW, Khodarev N, et al. An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(47):18490–5.

37. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nature biotechnology. 2018;36(1):89–94.

38. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nature methods. 2017;14(9):865–8.

39. Schröder M, Loos S, Naumann SK, Bachran C, Krötschel M, Umansky V, et al. Identification of inhibitors of myeloid-derived suppressor cells activity through phenotypic chemical screening. Oncoimmunology. 2017;6(1):e1258503.

40. Beury DW, Carter KA, Nelson C, Sinha P, Hanson E, Nyandjo M, et al. Myeloid-Derived Suppressor Cell Survival and Function Are Regulated by the Transcription Factor Nrf2. Journal of immunology . 2016;196(8):3470–8.

41. Liechtenstein T, Perez-Janices N, Gato M, Caliendo F, Kochan G, Blanco-Luquin I, et al. A highly efficient tumor-infiltrating MDSC differentiation system for discovery of

anti-neoplastic targets, which circumvents the need for tumor establishment in mice. Oncotarget. 2014;5(17):7843–57.

42. Vermeesch P. Exploratory analysis of provenance data using R and the provenance package. Minerals (Basel, Switzerland). 2019;9(3):193.

43. Scher HI, Morris MJ, Stadler WM, Higano C, Basch E, Fizazi K, et al. Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2016;34(12):1402–18.

44. Cheon H, Holvey-Bates EG, Schoggins JW, Forster S, Hertzog P, Imanaka N, et al. IFNβ-dependent increases in STAT1, STAT2, and IRF9 mediate resistance to viruses and DNA damage. The EMBO journal. 2013;32(20):2751–63.

45. Mitchel J, Grace Gordon M, Perez RK, Biederstedt E, Bueno R, Ye CJ, et al. Tensor decomposition reveals coordinated multicellular patterns of transcriptional variation that distinguish and stratify disease individuals [Internet]. bioRxiv [Preprint]. 2022 [cited 2022 Dec 06]: 47 p. Available from: https://doi.org/10.1101/2022.02.16.480703

46. Zhang L, Zhang L, Li H, Ge C, Zhao F, Tian H, et al. CXCL3 contributes to CD133(+) CSCs maintenance and forms a positive feedback regulation loop with CD133 in HCC via Erk1/2 phosphorylation. Scientific reports. 2016;6:27426.

47. Gelfo V, Romaniello D, Mazzeschi M, Sgarzi M, Grilli G, Morselli A, et al. Roles of IL-1 in Cancer: From Tumor Progression to Resistance to Targeted Therapies. International journal of molecular sciences. 2020;21(17):6009.

48. Connolly KA, Kuchroo M, Venkat A, Khatun A, Wang J, William I, et al. A reservoir of stem-like CD8$^+$ T cells in the tumor-draining lymph node preserves the ongoing antitumor immune response. Science immunology. 2021;6(64):eabg7836.

49. Zehn D, Thimme R, Lugli E, de Almeida GP, Oxenius A. 'Stem-like' precursors are the fount to sustain persistent CD8+ T cell responses. Nature immunology. 2022;23(6):836–47.

50. Zheng L, Qin S, Si W, Wang A, Xing B, Gao R, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science. 2021;374(6574):abe6474.

51. Alshetaiwi H, Pervolarakis N, McIntyre LL, Ma D, Nguyen Q, Rath JA, et al. Defining the emergence of myeloid-derived suppressor cells in breast cancer using single-cell transcriptomics. Science immunology. 2020;5(44):eaay6017.

52. Veglia F, Hashimoto A, Dweep H, Sanseviero E, De Leo A, Tcyganov E, et al. Analysis of classical neutrophils and polymorphonuclear myeloid-derived suppressor cells in cancer patients and tumor-bearing mice. The Journal of experimental medicine. 2021;218(4):e20201803.

53. Keenan BP, McCarthy EE, Ilano A, Yang H, Zhang L, Allaire K, et al. Circulating monocytes associated with anti-PD-1 resistance in human biliary cancer induce T cell paralysis. Cell reports. 2022;40(12):111384.

54. Lawal B, Tseng SH, Olugbodi JO, Iamsaard S, Ilesanmi OB, Mahmoud MH, et al. Pan-Cancer Analysis of Immune Complement Signature C3/C5/C3AR1/C5AR1 in Association with Tumor Immune Evasion and Therapy Resistance. Cancers. 2021;13(16):4124.

55. Wang Y, Zhang H, He YW. The Complement Receptors C3aR and C5aR Are a New Class of Immune Checkpoint Receptor in Cancer Immunotherapy. Frontiers in immunology. 2019;10:1574.

56. Boukhaled GM, Gadalla R, Elsaesser HJ, Abd-Rabbo D, Quevedo R, Yang SYC, et al. Pre-encoded responsiveness to type I interferon in the peripheral immune system defines outcome of PD1 blockade therapy. Nature immunology. 2022;23(8):1273–83.

57. Wu T, Ji Y, Moseman EA, Xu HC, Manglani M, Kirby M, et al. The TCF1-Bcl6 axis counteracts type I interferon to repress exhaustion and maintain T cell stemness. Science immunology. 2016;1(6):eaai8593.

58. Ayala AV, Matsuo K, Hsu CY, Terrazas MC, Chu H. Commensal bacteria promote type I interferon signaling to maintain immune tolerance [Internet]. bioRxiv [Preprint]. 2021 [cited 2022 Dec 06]: 20 p. Available from: https://doi.org/10.1101/2021.10.21.464743

59. Krieg C, Nowicka M, Guglietta S, Schindler S, Hartmann FJ, Weber LM, et al. Author Correction: High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. Nature medicine. 2018;24(11):1773–5.

60. Sinha M, Zhang L, Subudhi S, Chen B, Marquez J, Liu EV, et al. Pre-existing immune status associated with response to combination of sipuleucel-T and ipilimumab in patients with metastatic castration-resistant prostate cancer. Journal for immunotherapy of cancer. 2021;9(5):e002254.

61. Gebhardt C, Sevko A, Jiang H, Lichtenberger R, Reith M, Tarnanidis K, et al. Myeloid Cells and Related Chronic Inflammatory Factors as Novel Predictive Markers in Melanoma

Treatment with Ipilimumab. Clinical cancer research: an official journal of the American Association for Cancer Research. 2015;21(24):5453–9.

62. Santegoets SJ, Stam AG, Lougheed SM, Gall H, Jooss K, Sacks N, et al. Myeloid derived suppressor and dendritic cell subsets are related to clinical outcome in prostate cancer patients treated with prostate GVAX and ipilimumab. Journal for immunotherapy of cancer. 2014;2:31.

63. Signorelli D, Giannatempo P, Grazia G, Aiello MM, Bertolini F, Mirabile A, et al. Patients Selection for Immunotherapy in Solid Tumors: Overcome the Naïve Vision of a Single Biomarker. BioMed research international. 2019;2019:9056417.

64. Neely J, Hartoularos G, Bunis D, Sun Y, Lee D, Kim S, et al. Multi-Modal Single-Cell Sequencing Identifies Cellular Immunophenotypes Associated With Juvenile Dermatomyositis Disease Activity. Frontiers in immunology. 2022;13:902232.

65. Bunis DG, Bronevetsky Y, Krow-Lucal E, Bhakta NR, Kim CC, Nerella S, et al. Single-Cell Mapping of Progressive Fetal-to-Adult Transition in Human Naive T Cells. Cell reports. 2021;34(1):108573.

66. Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data. 2010. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

67. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17(1):10–12.

68. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.

69. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491–8.

70. popscle. Github. 2020. Available from: https://github.com/statgen/popscle

71. Zarinsefat A, Hartoularos G, Rychkov D, Rashmi P, Chandran S, Vincenti F, et al. Single-Cell RNA Sequencing of Tocilizumab-Treated Peripheral Blood Mononuclear Cells as an in vitro Model of Inflammation. Frontiers in genetics. 2020;11:610682.

72. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome biology. 2018;19(1):15.

73. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. Nature methods. 2020;17(6):615–20.

74. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python [Internet]. arXiv [Preprint]. 2012 [cited 2022 Dec 06]: 6 p. Available from: https://arxiv.org/abs/1201.0490

75. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nature methods. 2019;16(12):1289–96.

76. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Scientific reports. 2019;9(1):5233.

77. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nature biotechnology. 2018;37:38–44.

78. Tirosh I, Izar B, Prakadan SM, Wadsworth MH 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189–96.

79. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

80. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545–50.

81. Seabold S, Perktold J. Statsmodels: Econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference. SciPy; 2010.

82. Halford M. prince: Python factor analysis library (PCA, CA, MCA, MFA, FAMD). Github. 2021. Available from: https://github.com/MaxHalford/prince

83. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods. 2020;17(3):261–72.

84. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature. 2020;585(7825):357–62.

85. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. SciPy; 2010.

86. Hunter. Matplotlib: A 2D Graphics Environment. 2007;9:90–5.

87. Waskom M. seaborn: statistical data visualization. Journal of open source software. 2021;6(60):3021.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Elizabeth McCarthy
715D5FB155DD4E9...
Author Signature

12/2/2022
Date