Development and application of methods for the analysis of high-dimensional genetic and immunologic data

By

Wyatt James McDonnell

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology and Immunology

30 June 2019

Nashville, Tennessee

Approved:

Christopher R. Aiken, PhD, chair

David W. Haas, MD

Spyros A. Kalams, MD

John J. Karijolich, PhD

John R. Koethe, MD

Dedication

To Frank Steiner, who taught me that chance favors the prepared mind;
To Frank Herbert, who knew the value of being a generalist;
To my peers and colleagues, who made this work possible;
To my family and loved ones, who are so generous of their time and love;
And to those who came before me.

Acknowledgments

Science is made up of so many things that appear obvious after they are explained.
~ Pardot Kynes, from Frank Herbert's novel Dune (1965)

My first failure in graduate school was underestimating exactly how difficult it is to undertake and complete a PhD. The PhD is an endeavor that 1) is impossible to start, progress through, or finish alone, 2) is astonishingly unpredictable, and 3) is duplicitous in that it can be as invigorating as it is harrowing and as intriguing as it is tedious. While institutional and federal funding from VUMC/VU and the NIH financed this work, all the funding in the world could never replace the bright, collegial, and generous people who made my completion of this degree possible. This dissertation is the product of those individuals, who I have been so fortunate to work with, and who I briefly thank below (recognizing in advance that I will fail to thank these people as well as I would like):

- Simon Mallal, MBBS: Thank you for years of stimulating conversation, for championing your research team, and for providing me with opportunities that most graduate students never experience. Thank you for reminding me that you do have to be a little crazy to stay in this business. Thank you for always being willing to listen, for being receptive to constructive criticism, and not allowing the power dynamic of the advisor-advisee relationship to get in the way of honest conversation. I have learned so much from you, and hope to continue doing so for a very long time.

- Elizabeth Phillips, MD: Your keen intellect, working memory, and work ethic were and are an inspiration to me. Your enthusiasm and passion for translating findings into research are second-to-none, and your insightful questions along the way have prompted me to improve both my work and the quality of the questions I ask of others.

- John Koethe, MD: I cannot imagine a scenario where this thesis would be complete without you. Thank you for truly going above and beyond in mentoring and supporting me, reminding me to think big and to ask myself "Does it matter?", and for bringing me under your wing.

- Spyros Kalams, MD: You helped instill in me a healthy level of skepticism, without the cynicism that typically accompanies. Thank you for teaching me

the importance of showing and inspecting all of the data, the value of finely matched controls and rigorous experimental design, and for stimulating conversation spanning from science fiction, AI, politics, and beyond.

- Ivelin Georgiev, PhD: I will never be able to thank you enough for taking a risk and allowing me to be "half a graduate student" in your lab. I am better for knowing and being mentored by you. Thank you for teaching me how to have a hard conversation, how to get things out of the way up front, and how to be gracious when a graduate student makes a mistake. Thank you for teaching me the importance of making every second of your work day count, and that doing karaoke with your graduate students is one of the best ways to get everyone to loosen up. Thank you for openly sharing with us and letting us learn what it's like to start a lab, how to enjoy the small things, and how to pick things up when they fall apart a little.

- Abha Chopra, PhD: I look up to you in so many ways and am incredibly grateful to have learned from you in so many areas over the past few years. I would be adrift in a sea of sequencing data and clueless as to how any of it worked without you, and I would have zero appreciation for automation, assay QA and QC, and how to identify, eliminate, and overcome bottlenecks in NGS assays. You have been an exceedingly fair and helpful critic of my work, a staunch supporter of building up a strong dry lab, and you have personally kept me going time after time. Your ability to whip an assay up out of thin air astonishes me day after day, and I can't imagine reciprocating late-night teleconferences with anyone else. I know I speak for everyone in the group when I say that none of the machine would work without you.

- Ramesh Ram, PhD: You have, in a very short period of time, become one of my closest peers and mentors and are an inspiration to me. I hope to one day have a tenth of your abilities to write coding solutions as elegant as yours (if I do, I know they won't be as short or as fast). Your tireless commitment to working on these questions made all the analysis, staying up late, and troubleshooting software installation and machine set-up infinitely less painful. Thank you for making me feel that I wasn't working in isolation, thank you for being my sounding board, and thank you for all of your help in troubleshooting and analysis (and especially for catching stupid typos in my code). Thank you for being overwhelmingly generous with your time, for being patient and kind, and for setting the standard when it comes to understanding both the wet and dry lab aspects of an experiment.

- Shay Leary: Your ability to translate code between languages boggles my mind, and I can't tell you how awe-inspiring it's been to watch you maintain, improve, and iterate on VGAS and its design for the past few years. Thank you for always putting your best foot forward when the rest of us are tired and out of ideas, for innovating and inventing new ways for us to visualize the piles of data from all of the various assays, and for always being up to the challenge of taming R and bringing it into VGAS.

- Silvana Gaudieri, PhD: You are the one of the reasons why I joined the group, and I am so privileged to have walked away from every single conversation with you having learned something new. The fact that even the dastardly Jurkat cells behave for you is proof that you have hands of gold, and I will forever save every protocol you've ever sent me and walked me through knowing that every important detail is on the page. I hope that we can share a porter again sometime soon, and in the meantime, I will strive to attain the level of balance that you have achieved in flowing through the academic year unperturbed.

- Mark Pilkinton, MD, PhD: From torque teno virus to Stan Lee to Tolkien, I have so enjoyed every conversation I have had with you. Thank you for teaching me how to think like a BioMek, for working with me to help end the Great Metadata Crisis of 2016-2018, for teaching me the value of an interpretable experiment, and for being a truly stellar mentor and friend in every way.

- Celly Wanjalla, MD, PhD: I am so glad to know you and have learned so very much from you since you came to Vanderbilt. Partnering with you has made me a better scientist in so many ways, and your work ethic and ability to fly through analyses without getting anything wrong puts the rest of us to shame! Thank you for motivating me, for helping me think about my data in all kinds of different ways, and for being my V-SCHoLARS writing buddy!

- Katie White, MD, PhD: Thank you for working with me on getting things managed and organized, for being a champion and advocate of Slack, for brainstorming a million different ways to visualize TCRs and how they go together, and for helping me dream up new ways of interacting with and analyzing all of these data while trying to stay sane.

- Rama Gangula, MSc: The entire research enterprise of everybody I've worked with closely over the past few years would grind to a complete halt without you. I look up to you for so many reasons and will never be able to thank you enough for all that you do. Your notebooks are impeccable and something all scientists should see and learn from, your troubleshooting skills are matched only by your kindness and patience to everyone around you, and I will forever be grateful for the delicious culinary items you've brought to lab meeting.

- Louise Barnett, Cindy Hager, and Rita Smith: Thanks to you, I will forever appreciate the intricacies of flow cytometry, maintaining a ready-for-inspection lab space, and I will always make sure to find the senior technicians like yourselves. You are fonts of knowledge and I deeply appreciate your patience, assistance, and hard work. Thank you for making all of these papers and all of the papers from the group possible.

- Ian Setliff, PhD: I don't know where to start—in a short period of time you have become one of my closest friends and mentors. I look up to you in so many ways and know that I would never have gotten my feet wet in this area if it wasn't for you. Thank you for all of the time we've spent cursing R and Python together, rolling our eyes when IMGT is being slow, running ideas by each other, for caring about the details in the same way that I do, and for all of the ever-so-elegant Greek curses you've taught me.

- Justin Balko, PharmD, PhD, and Doug Johnson, MD: Thank you for bringing me up to speed on years of incredible immunotoxicity research and for broadening my horizons. You expanded my interests into an area I had never considered, and I am immensely grateful to both of you. I needed to be reminded that 80% done is better than 100% perfect and it's never too late to bounce back—I am thankful you have been there to set that example for me.

- Chris Aiken, PhD: If the entire faculty within our department turned over except for you, I would do my very best to choose you as my committee chair again. Thank you for providing a calming presence, listening ears, and a guiding voice during both stressful and happy times throughout my PhD. Thank you also for challenging me to ask myself, "Is this really what's right for me? Is this really in my best interest?" every step of the way.

- David Haas, MD: You have provided me with unyielding support through so many channels at Vanderbilt for the past 3 years that I wouldn't know where to start or finish when it comes to saying thank you. From ensuring that we had the computing equipment to get the job, to enabling me to share my research via TN-CFAR, and including me in the grant writing process, you have been an incredible mentor and colleague.

- John Karijolich, PhD: I will never forget your patience and help while we were struggling with our lentiviral expression system. Thank you for asking keen questions, your enthusiasm, and for being a fantastic role model as a scientist and as an early-career tenure-track faculty member.

- The VANTAGE core laboratory team: Without Angela, Karen, Liv, Matt, Neha, and others, the sequencing data described in this dissertation would never exist. Much of the single cell work would never have happened, and I would never have forayed into bioinformatics, much less developed or maintained an interest in NGS.

- Suman Das, PhD: I am so grateful that you spotted the signs of burnout before myself and everyone else. Letting me share the light through your office window was the kindest gift I have ever received from a colleague, and I am infinitely grateful and in your debt for being able to have a modicum of privacy and quiet in a structurally confined workplace.

- Nora Gilgallon-Keele: You keep the Division alive and moving along. Thank you for always providing a listening ear, keeping a well-stocked candy pile, making the best witch's brew ever, and sharing my desiccated sense of humor. You will forever be my favorite Michigander at Vanderbilt.

- Kay Harding: Nobody knows the ins-and-outs of the Division and the Department like you. Thank you for somehow managing to always find a conference room at the absolute latest hour, miraculously preventing the S-2100 projector remote from vanishing accidentally, making sure that we had enough Ethernet ports, and always providing a desperately-needed smile before our 8 am lab meetings on Mondays.

- Lorie Franklin: You were and are the guardian angel of our Department and graduate program. The whole ship would sink without you, and I doubt you are told that enough. Thank you for making travel reimbursement less

hellacious, for coming through in a pinch every time a conference registration emergency came up, for telling persnickety hotel managers to get off my case, for keeping PMI seminar running smoothly week after week, and for everything else you do.

- Wendy Hess and Lisa Au: No process in academia could be better described as "herding cats" than grant submission—I am so lucky to have known and learned from you both. Thank you for tirelessly responding to emails over your Christmas and countless other holidays, for wading through Dropbox folders hundreds of files deep with tens of versions per file, for converting those files to PDFs, and for making sure that everybody stayed on top of their deadlines.

- Mary Kirby and Adrienne Jenkins: The Division owes you both more than can be described, and I do too. Between navigating 100-message email threads with HR, helping me every time I accidentally locked my keys in, going to bat for me and others with IT, and clarifying hiring paperwork for myself and rotating students, you both saved me precious time week after week of the calendar year. Mary—you will probably never read this, but I hope retirement treats you as kindly as you treated me and that you get to enjoy every second with your family and dogs. Adrienne—I didn't think anybody could've filled Mary's shoes, and I am so grateful to have been so wonderfully wrong about that.

- Gary Henriksen: I'm so thankful that you were my Virgil to the circles of VUMC/VU IT and all of its infrastructure and people. Without you, none of our infrastructure would be properly configured, and your patience, kindness, humor, and diligence in helping myself and others resolve a myriad of IT issues will always be unparalleled in my mind.

Robert Southey is supposed to have said that words are like sunbeams, in that "the more they are condensed, the deeper they burn." In that spirit, I think it best to say, above all, that I am grateful.

Table of Contents

List of Tables

List of Figures

Abbreviations

- APC. Allophycocyanin.
    - Also, antigen-presenting cell.
- ART. Antiretroviral therapy.
- AT. Adipose tissue.
- ATM or ATMs. Adipose tissue macrophage(s).
- BCR. B cell receptor.
- BMDM or BMDMs. Bone marrow-derived macrophage(s).
- CDR3. Complementarity-determining region 3.
- CMV. Cytomegalovirus.
- DAPI. 4',6-diamidino-2-phenylindole.
- DIO. Diet-induced obesity.
- DNA. Deoxyribonucleic acid.
- EBV. Epstein-Barr virus.
- FACS. Fluorescence-activated cell sorting.
- GPR56. G protein-coupled receptor 56.
- HFD. High fat diet.
- IR. Insulin resistance.
- irAE or irAEs. Immune-related adverse event(s).
- isoLG or isoLGs. Isolevuglandin(s).
- kg. Kilogram.
- LFD. Low fat diet.
- MHC-I. Major histocompatibility complex, class I.
- MHC-II. Major histocompatibility complex, class II.
- MMe or MMes. Metabolically activated macrophage(s).
- NADPH. Nicotinamide adenine dinucleotide phosphate reduced/H.
- NGS. Next-generation sequencing.
- Nrf2. Nuclear factor erythroid 2-related factor 2.
- PBS. Phosphate buffered saline.
- PC or PCs. Principal component(s).
- PCA. Principal component analysis.
- PE. Phycoerythrin.
- PerCP. Peridinin-chlorophyll-protein complex.
- PCR. Polymerase chain reaction.
- RNA. Ribonucleic acid.
- scRNA-seq. Single cell RNA sequencing.
- -seq. Suffix for sequencing (*e.g.* RNA sequencing = RNA-seq).

- SVF. Stromal vascular fraction.
- TBHP. *tert*-butyl hydroperoxide.
- TCR. T cell receptor.
- Treg or Tregs. T regulatory cell(s).
- t-SNE. t-distributed stochastic neighbor embedding.
- UMAP. Uniform manifold approximation and projection.

Chapter I

Introduction

An intersection of poverty, HIV, metabolic disease, and immune inflammation: building knowledge of the immune system through the study of known perturbations

Where the 19th century marked the advent of infection-centered immunology and the germ theory of disease[1], the 20th and 21st centuries have ushered in an epidemiological transition coinciding with economic development. Chronic disease, lifestyle diseases, and diseases of aging such as cancer, diabetes, hypertension, and heart disease now account for approximately 70% of human disease.[2] Changes in standards of living, availability of medical care, and poverty can directly explain this transition, as first posited by Abdel Omran in 1971.[3,4] This transition has been so jarring that some diseases, such as beriberi (severe and chronic thiamine deficiency), were incorrectly assumed to be immune-mediated in response to a human pathogen for decades, hindering proper clinical and scientific investigation into treatments.[5] Increases in lifespan and aging as a process of immune senescence (and hence, dysfunction) have certainly increased the prevalence of chronic disease. However, any model attributing this increase predominantly to aging is incomplete. More complete models have correctly incorporated genetic

heritability and “germinating” , chronic, systemic inflammation as central components to a unifying theory of the etiology of these diseases.[6–12] Seminal publications at the turn of the 21st century directly implicated the immune system as an unexpectedly central player in many of these diseases.[13–16]

From a historical and evolutionary perspective, the correlation between immune dysfunction and Omran’s epidemiologic transition makes perfect sense. Energy (nutrition) management and metabolic homeostasis are two of the most important functions that any living organism must maintain. Any organism that cannot endure starvation or the threat of severe infectious disease would be poorly adapted to the immense immunogenic burden that one encounters in daily life.

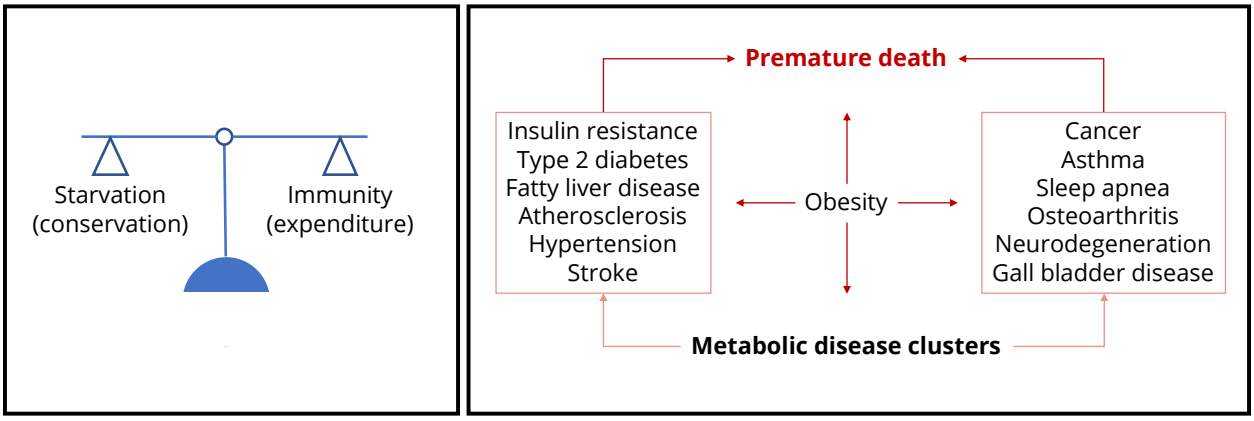


Figure 1. Immune responses and energy conservation are inherently opposed functions whose relationship is contextualized by obesity and inflammation. *Left panel:* Mounting an immune response against a pathogen or immune stimulus requires expenditure of valuable cellular and systemic energy; building these reserves during nutritional surplus is important not only for surviving starvation but for maintaining immune function. *Right panel:* Obesity provides a man-made experiment where type 2 diabetes, hypertension, and other chronic “lifestyle” and diseases of civilization are induced by nutritional excess and subsequent immune activation that cannot occur in the setting of starvation.

This balance explains the stunning conservation and coordination of the immune system, immune organs, immune genes, and immune regulation from flies to man and how an adaptive mechanism of human evolution has become maladaptive in the present day and age (**Figure 1**).[17–19] Religion provides an excellent experiment of man that supports this model; longitudinal prospective studies of individuals undergoing ascetic fasting and their non-fasting peers have revealed that caloric restriction leads to reduced circulatory inflammatory marker expression, improved cholesterol profiles, and caloric restriction has proven itself as a robust way of prolonging life in every major animal model studied.[20–22] Similarly, cross-sectional cohort studies of Cretian farmers have documented a cultural shift away from a Mediterranean diet with concomitant increases in mean BMI, blood pressure, cholesterol, and cardiovascular disease (CVD) risk factors.[23,24]

Given the impossibly long list of variables that could stimulate or repress the immune system—pathogens, symbionts, chemical and environmental agents, diet, stress, and more—it is clear that understanding the contributions of the immune system to these human diseases is only possible by studying 1) experiments of nature, and 2) experiments of man. Examples of the former include the study of individuals with severe combined immunodeficiency and variable immunodeficiency (in that specific immune lineages are dysfunctional or non-functional) manifesting in susceptibility to select bacterial and viral infections.[25–31] In

contrast, notable experiments of man include the adoptive transfer of lymphocytes and their progenitors from one human to another after myeloablative therapy,[32,33] bone marrow transplantation,[34] and most uniquely, antiretroviral therapy (ART) for the treatment of human immunodeficiency virus (HIV) and reversal of the immunodepletion that occurs during natural HIV infection.[35,36] Leveraging the contrast of these two types of experiments provides a framework for understanding the contributions of the immune system to human disease (**Figure 2, next page**). This framework provides a way to design human immunological studies in a way that provides filters and internal validation of the data generated so that the assignment of a cellular population as causal of disease becomes more robust.

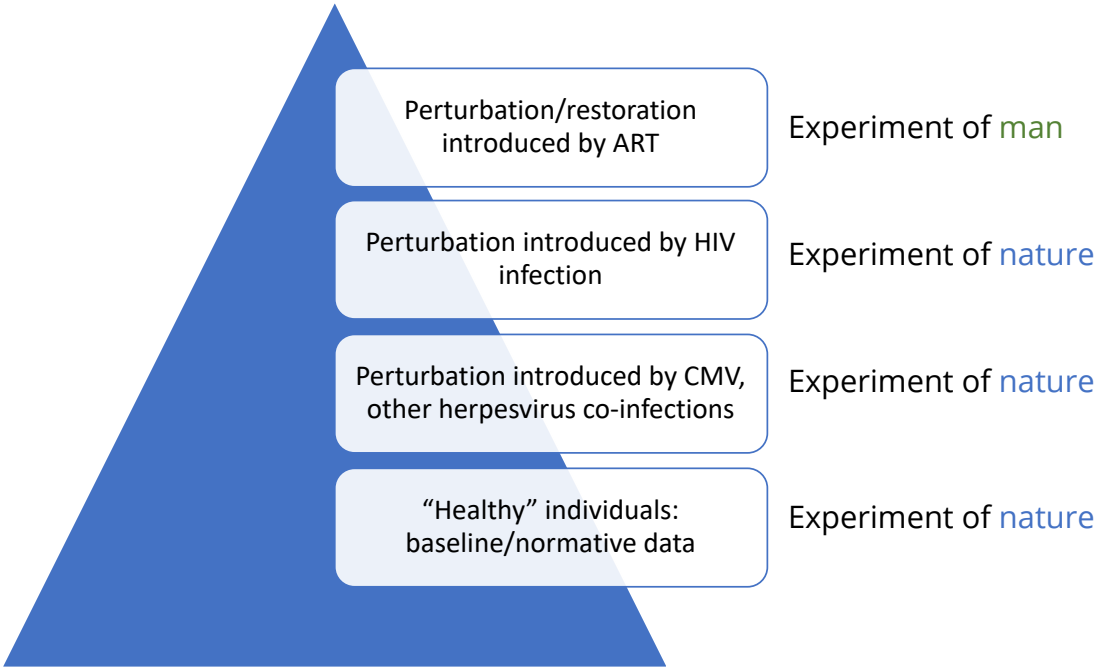


Figure 2. A hierarchy to understand immune-mediated disease. Understanding immune insult and injury is best achieved by building and collecting hierarchical knowledge of the immune system. This process begins by collecting well-curated, multi-modal/multi-omic data from genetically diverse and generally "healthy" individuals (I place "healthy" in quotation marks here to emphasize that the immune system is an exquisitely sensitive sensor that can respond quickly and forcefully to even minor insults in individuals with a perfectly clean bill of health and clinical history). In chapter 2, I apply this hierarchy in a large cohort of veterans living with and without HIV to learn immune features that are associated with mortality. As one moves from the bottom of the pyramid to the top, the number of individuals who can be studied shrinks. Similarly, collecting multi-modal data on a large number of individuals, and collecting a large number of measurements from multiple individuals (much less over time), becomes less and less practical. Accordingly, the generation of cohort and population-level data and accounting for known immune perturbations such as cytomegalovirus (CMV) co-infection becomes essential to 1) reduce type I/discovery errors in small human studies and 2) reveal the true effects and causal diagram of the intervention, disease, or immune component being studied.

Human immunodeficiency virus (HIV) establishes a permanent infection that depletes a specific subset of immune cells—primarily CD4+ T cells, with efficiency varying in different subsets of CD4+ T cells—over time if left untreated.[37–39] While modern ART has made HIV a manageable chronic illness, chronic HIV infection is a condition marked by persistent, low-level inflammation originating from multiple

sources (**Figure 3**).[40–45]

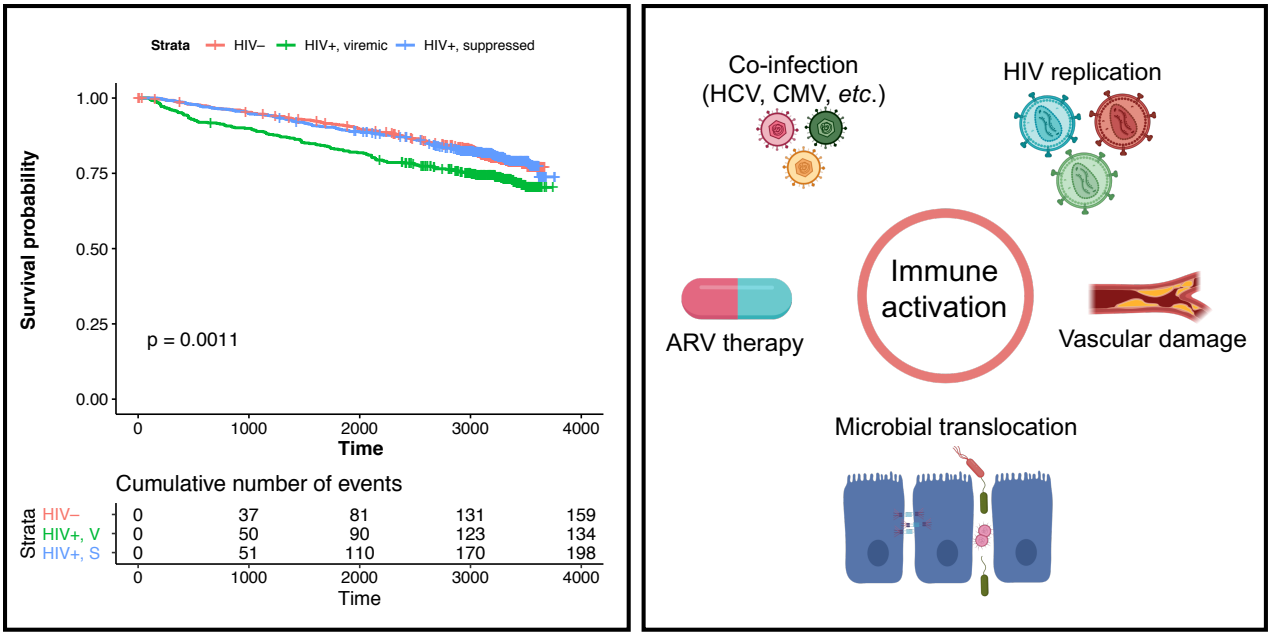


Figure 3. Overview of HIV mortality and mechanisms of immune activation in chronic HIV infection. *Left panel*: Kaplan-Meier hazard curves visualizing survival over time of three groups of veterans from the Veterans Aging Cohort Study Biomarker Cohort: HIV− (red), HIV+ with unsuppressed HIV (viremic, green), and HIV+ with suppressed HIV (suppressed, blue). The survival of PLWH with suppressed HIV is approximately equivalent to that of HIV− individuals, which has been shown elsewhere, as has the fact that PLWH with uncontrolled HIV infection are much more likely to die over time. *Right panel:* Common sources of and contributors to chronic inflammation in PLWH; notably, these responses can be exacerbated in specific ways depending on the composition of the CD4+ T cell compartment, which I provide evidence for in Chapter 2 below.

The portmanteau “inflammaging” (inflammation + aging) has been used to describe this inflammatory state. Indeed, this accelerated “aging” of the immune system correlates with elevations in inflammatory circulating biomarkers such as IL-6, soluble CD163, and soluble CD14 (among others) that predict morbidity and mortality in people living with HIV (PLWH). Given these findings, it is also unsurprising that PLWH experience excess risk of cardiovascular disease, liver, kidney, bone, neurologic, and metabolic diseases even after adjusting for common

clinical variables such as BMI, antiretroviral usage, age, sex, race, and drug use (reviewed in [41]).

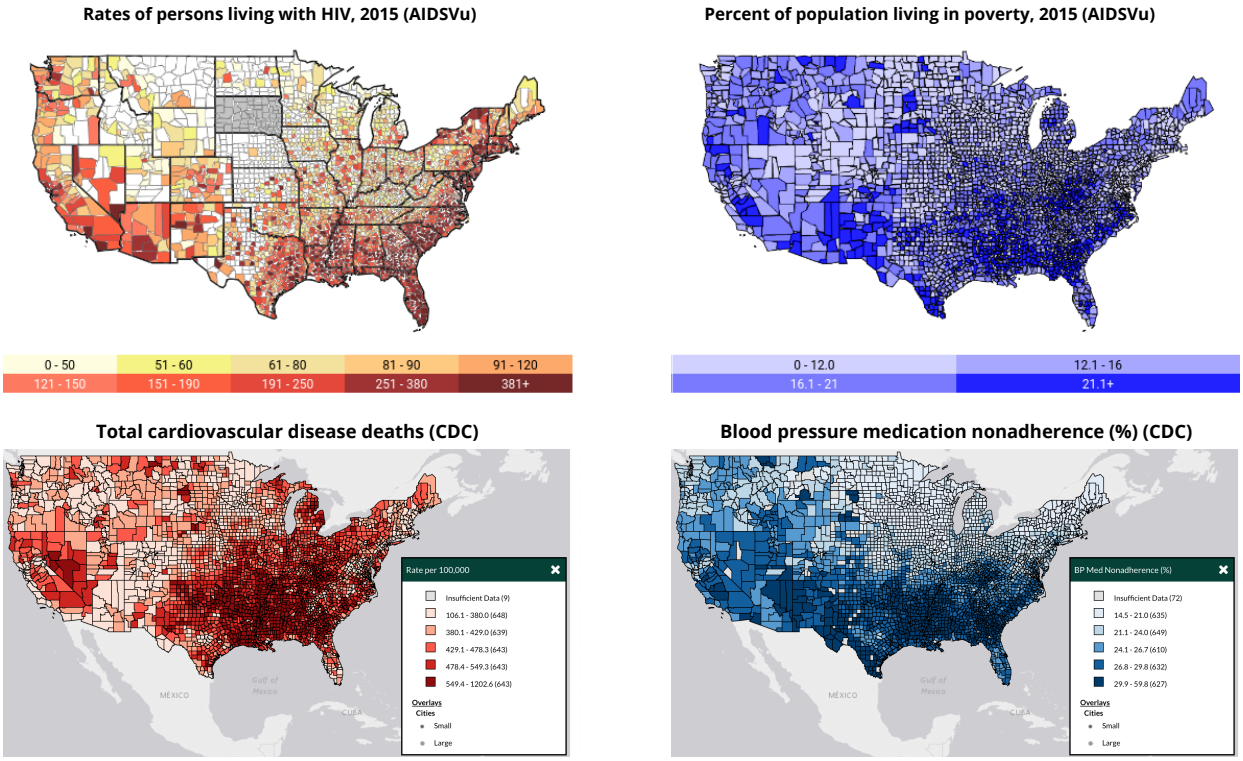


Figure 4. Concurrent epidemics of obesity, cardiovascular disease, and poverty coincide with the HIV epidemic within the United States. Data are sourced from and visualized by AIDSVu (top panels) and the CDC's Interactive Atlas of Heart Disease and Stroke (bottom panels) and reproduced under the Creative Commons CC-BY-NC-SA license.

In addition to the immune alterations that natural HIV infection causes, HIV infection rates are highest where poverty, socioeconomic disparity, cardiovascular disease and obesity are found in the United States (**Figure 4**). This pattern holds true around the world, especially in poor and population-dense regions with limited

access to medical care (**Figure 5**).

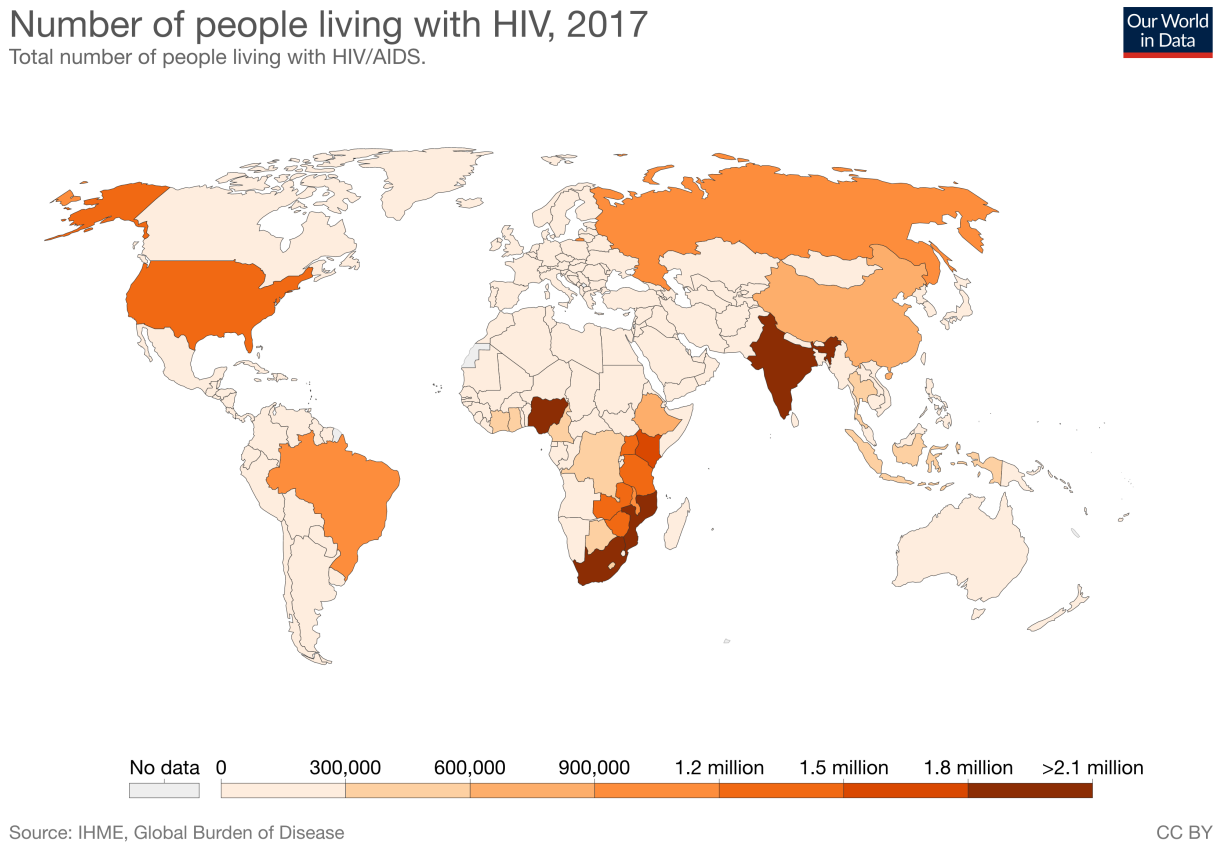


Figure 5. HIV infection is most common in countries where poverty and income inequality, restricted access to medical care, poverty, and high population density are also common. This chart is reproduced from Max Roser and Hannah Ritchie's HIV/AIDS article in Our World in Data (https://ourworldindata.org/hiv-aids) under the Creative Commons BY license.

Indeed, the stigma associated with HIV infection amplifies these factors, with social isolation and exclusion from solidarity networks, which in turn leads to food and livelihood insecurity, failure to adhere to ART, and poor health in cyclical fashion.[46,47] By performing research with PLWH and HIV-negative individuals and taking into account the compounding nature of these overlapping epidemics, it

becomes possible to understand the contribution of the immune system to specific immune-mediated disorders, such as hypertension, diabetes, and frailty. I describe below the genetic variation of this sensitive and specific adaptive immune system, several of its cellular and molecular constituents (namely, the T lymphocyte, who features prominently in this work), and some of the known immune perturbations that arise in chronic viral infection, inflammation, and related immune processes. Indeed, the T cell and B cell receptor (whose affinities *in vivo* range from nanomolar to sub-picomolar), when acting in conjunction with antigen-presenting cells (APCs), form arguably the most sensitive biological detector of non-normality and perturbation.

Genetic variation and the human adaptive immune system

Given this framework of leveraging known immune perturbations to understand the immune system, it becomes important to understand and account for inherent variation within the adaptive immune system itself as well. As a species, *Homo sapiens* is distributed across and throughout an incredibly diverse array of environments, each varying in its temperature, humidity, flora and fauna, sunlight, and other natural variables. This diversity is further amplified by cultural and socioeconomic factors such as population density, family relationships (e.g. do children and their parents live together for much of their lives), how children are

raised, the availability and quality of medical care and resources, and the social perceptions of human disease.

Genetic variation amongst *H. sapiens* builds yet another layer of diversity. Each single haploid cell of the typical human contains approximately 3 billion base pairs of DNA. Of those 3 billion base pairs, approximately 3 million (0.1%) on average will differ between any two humans on the planet at any given time. This variation has enabled researchers to use Mendelian randomization (as genotypes are randomly assigned at birth) to identify causal genetic variants associated with human disease and variation in phenotypes, to approximate ancestry of individuals using single nucleotide polymorphisms (SNPs), to advance forensic science, and much more.[48] However, of all the loci in the human genome, an astonishing amount of genetic diversity is located in genes encoding: 1) the human leukocyte antigen (HLA),[49] 2) the T cell receptor,[50] and 3) the B cell receptor.[51] The genetic variation of these loci alone and of the functional products they encode is one of the most important reasons cohorts focusing on questions relevant to adaptive immunity and its interacting cellular populations should be genotyped and why as many individuals as possible should be studied. Structural variation and polymorphism in these regions and their associations with disease will be likely captured using full allelic typing and whole-genome sequencing rather than SNP genotyping or GWAS given the inheritance patterns and diversity of these loci

(including KIR, HLA, TCR, and BCR genomic regions). Notably, the major experiments of nature and man I describe above have been predominantly agnostic of the critical allelic variation found in these immune receptors (including the killer-immunoglobulin-like receptor (KIR) family).

Genetic diversity in relation to structure and function of HLA

The major histocompatibility complex (MHC), known as human leukocyte antigen (HLA) in humans, is a genomic array found on chromosome 6. There are two main classes of HLA, class I and II, which differ in their expression on immune cell subsets and in human tissue despite their structural similarities. Both classes of HLA bear immunoglobulin-like domains near the membrane-inserting portion of the molecule, and their most distal domains from the membrane form a peptide-binding groove (colloquially referred to as a “hot dog bun”). The *HLA* loci are in strong linkage disequilibrium and encode the majority of glycoproteins responsible for presenting foreign and self-peptides to T cells. Most polymorphic genes contain nucleotide sequences encoding differences of 1-3 amino acids between variants; the *HLA* loci encodes variation of a difference up to (and possibly beyond) 20 amino acid substitutions between allelic variants. These differences are predominantly located in the peptide-binding site (which displays antigens to T cells) and on

exposed surface areas (which most mature T cells bind weakly, and which play a key role in T cell development and selection) (**Figure 6**).

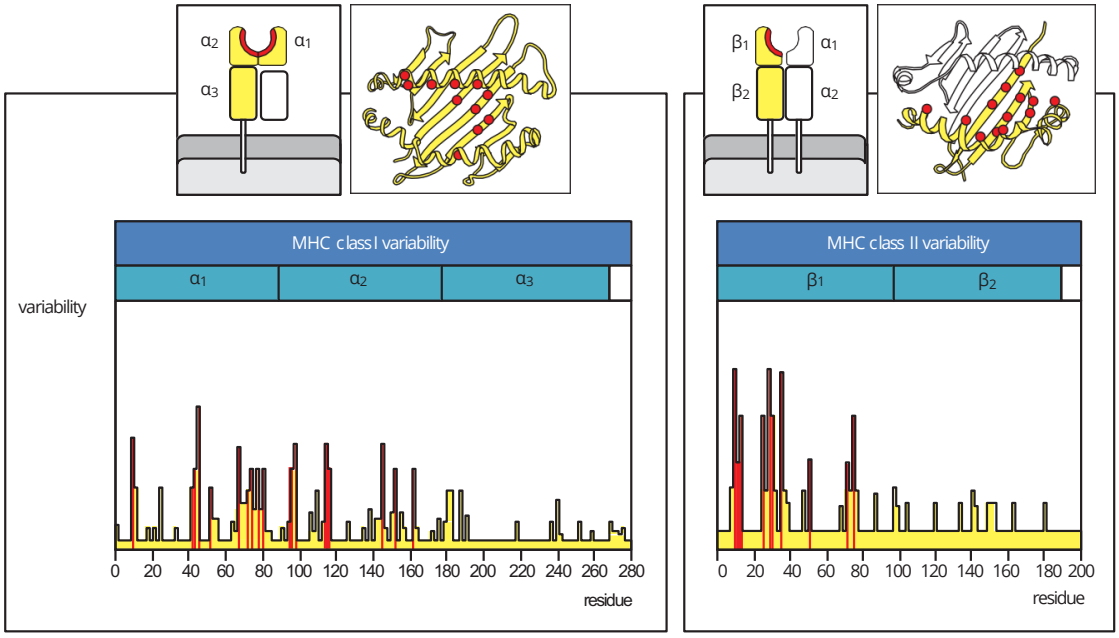


Figure 6. The majority of allelic variation in HLA is found in peptide-binding and surface exposed regions. Most inter-allelic variation in the HLA loci is found in two places: 1) the α1 and α2 domains of HLA class I and 2) the β1 and β2 domains of HLA class II. These domains form the respective peptide grooves of both classes of HLA; the groove of class I HLA alleles accommodates shorter peptides while the groove of class II HLA alleles holds significantly longer peptides. The α chain of HLA class II is particularly invariant. This figure was modified from Janeway's Immunobiology, 9th Edition by Kenneth Murphy and Casey Weaver (© 2017 by Garland Science, Taylor & Francis Group, LLC); it is modified and reproduced with permission of W.W. Norton & Company, Inc.

The HLA genes arose and have maintained their diversity through gene duplication and subsequent divergence. Sequencing of these loci across multiple related and unrelated humans and other species indicate that gene conversion—a process by which segments of multiple genes interchange during meiotic apposition—has enabled HLA genes to accrue multiple mutations rapidly, especially

given that the HLA loci are tandemly arrayed (**Figure 7**).

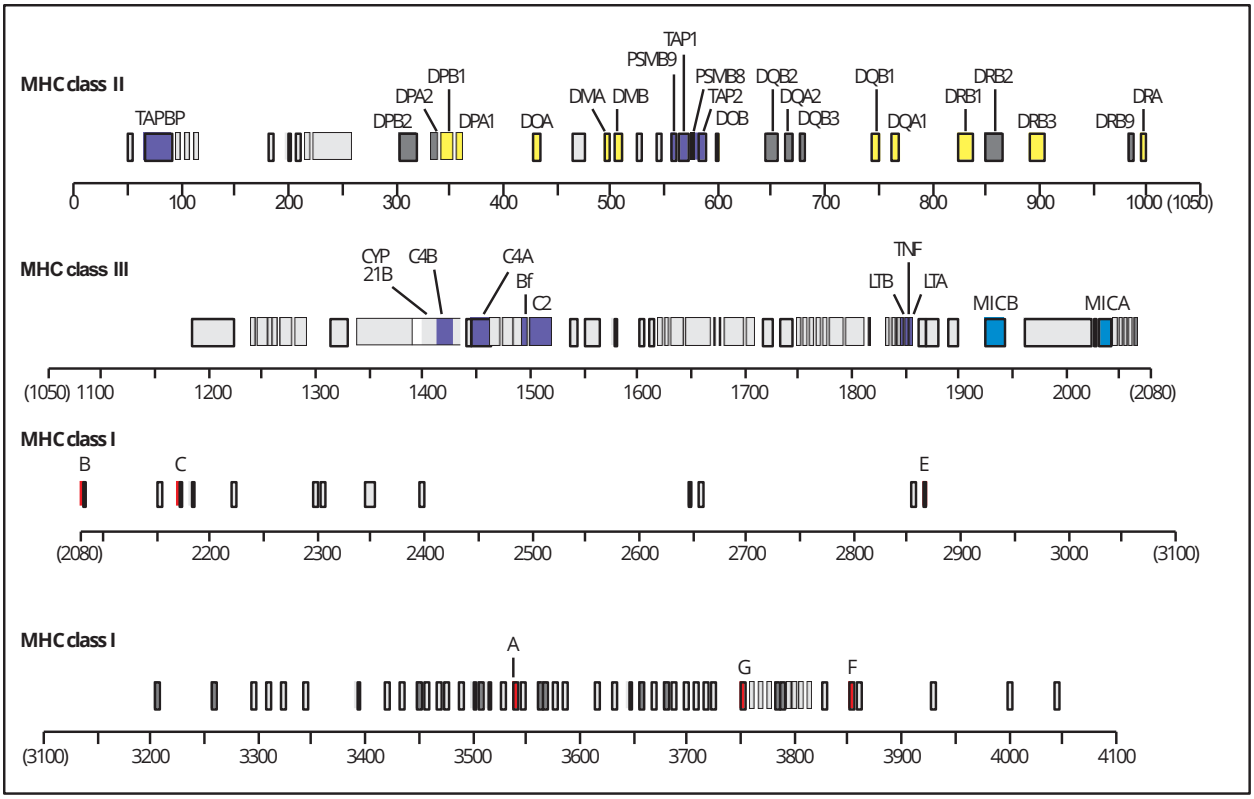


Figure 7. Chromosomal map of the HLA complex. The class I, class II, and class III regions of the HLA complex are shown here with genetic distances in thousands of base pairs. Genes such as *HLA-E*, *HLA-F*, and *HLA-G* shown in the class I region are class I-like genes encoding regulatory class Ib molecules. Genes shown in the class III segment encode complement proteins C4, C2, and factor B (shown as Bf), in addition to genes encoding tumor necrosis factor-α (TNF) and lymphotoxin (LTA, LTB). HLA class I genes are colored in red, except for the MIC genes, which are colored in blue; these are distinct from the other class I-like genes and are under different transcriptional control. HLA class II genes are shown in yellow. Genes unrelated to HLA class I and II in the HLA region with immune functions are colored in violet. Genes in dark gray are pseudogenes related to immune function genes. This figure was modified from Janeway's Immunobiology, 9th Edition by Kenneth Murphy and Casey Weaver (© 2017 by Garland Science, Taylor & Francis Group, LLC); it is modified and reproduced with permission of W.W. Norton & Company, Inc.

This polymorphism is accented by the polygenic nature of the HLA, in that each human on the planet carries multiple allelic variants, each of which are expressed in different combinations. Accordingly, most humans are heterozygous for class I and class II HLA and carry specific haplotypes. This heterozygosity and

diversity pose one of the greatest biological barriers to transplantation, as even siblings frequently carry and express different HLA haplotypes. However, this allelic and haplotypic diversity is, in the final analysis, of utmost importance to a properly coordinated immune response. The unique peptide-binding groove sequences of these allelic variants confers preferential binding specificity to select peptide motifs; carrying and expressing these diverse allelic variants permits the human immune system to present and respond to very broad peptide repertoires; recent analyses confirm that these loci are still under positive selection and diversification, and that class II alleles in particular develop “generalist” peptide-binding motifs in areas with diverse pathogens and high pathogen burden.[52,53] Similar variation also exists in the KIR loci,[54] who were co-discovered along with natural killer (NK) cells via the “missing-self hypothesis”[55]—where absence of the incredibly polymorphic HLA molecule leads to NK cell killing,[56,57] which is turn partially controlled by KIR-HLA interactions that inhibit or activate cytolytic pathways.[58,59] While the urgency of organ transplantation research naturally led to the development of full allelic typing for HLA, the KIR gene family is much newer—even though KIR family members have NK-activatory, NK-inhibitory, and HLA-binding activities.

Genetic diversity of the T cell receptor loci

The T cell receptor (TCR) is a heterodimer comprised most commonly of an α:β or a γ:δ pair of receptor chains. These chains are assembled at the molecular level from up to three parts: a V gene, a D gene, and a J gene. All TCR chains are comprised of both a variable and a constant region. The variable region of the β and δ chains is encoded in three parts—a V gene joined to a D gene, with random nucleotides inserted and deleted on either side, joined again to a J gene, forming a V(D)J junction—while the α and γ chains are forming by the joining of a V gene and a J gene. The V(D)J and VJ junction are commonly (albeit incorrectly) referred to as the third complementarity-determining region (CDR3)—while incorrect from a taxonomic standpoint, this does accurately reflect that the majority of the variability inherent in the somatic combinatorial process that creates a TCR is found in the V(D)J and V-J junctions. While human TCRs do not further diversify through somatic hypermutation (SHM) like B cells, the TCRα chain of the nurse shark has recently been demonstrated to undergo extensive SHM,[60] demonstrating another mechanism through which even more diversification of the TCR can be achieved.

Like many genes, TCR genes are inherited in haplotypes—meaning that, for instance, while there are at least 52 commonly known *TRBV* genes, not all of these genes will be present in each individual or inherited. In addition to this heritable variability, at least two-thirds of these genes are known to contain silent,[61,62] point,[62]

and premature termination/stop codon mutations or mutations that disrupt secondary and tertiary structure necessary for proper TCR pairing and function.[63] Consequently, polymorphism within the TCR loci means that some TCR repertoires contain “holes” in them where certain TCR genes are functionally useless in V(D)J recombination, in additional to “holes” where entire non-human antigens might not be recognized due to thymic exclusion of TCRs responding to them.[64]

One particularly outstanding example of the functional impact of this polymorphism is found in the study of an HLA-B*35:01-restricted response targeting an epitope of Epstein-Barr virus (EBV). Gras and colleagues first identified a “public” TCR—a T cell receptor found in multiple HLA-B*35:01-carrying individuals with identical nucleotide and amino acid sequences—recognizing an epitope of EBV. In all of these individuals, the *TRBV* gene used in these TCRs was *TRBV9*01*; the *TRBV9*02* allele was uniformly absent. By studying the same EBV-specific TCR with the *TRBV9*02* allele—which contains a single amino acid point mutation in the first complementarity-determining region (CDR1)—and replacing its naturally selected *TRBV9*01* allele, they showed conclusively that CDR1 and CDR2 polymorphism increase the affinity of the TCR for HLA. These results were later supported by findings from Robbins and colleagues, who demonstrated the same phenomenon in the setting of melanoma and HLA-A*02:01 natural and *in vitro* polymorphism.[65,66] Polymorphism of the TCR may also play a role in responding to antigenic

polymorphism; Geldmacher and colleagues showed that the TL9 epitope of HIV is recognized by a *TRBV12-3* public TCR when present in HIV subtype C viruses, but that TL9 variant cross-reactive TCRs are only found in HLA-B81-carrying individuals and not HLA-B42-carrying individuals (a striking finding given that HLA-B81 and HLA-B42 differ by just six amino acids).[67]

Recent findings of 37 novel *TRBV* genes in a sample of just 81 Caucasian individuals have renewed interest in the polymorphism of the TCR loci; in that study, 6 allelic profiles representing distinct haplotypes of the *TRBV* loci were clearly identifiable and revealed stark differences in both the number of uncommon *TRBV* alleles and the number of distinct *TRBV* alleles found in individuals of each haplotypic grouping.[68] While the authors of that study speculated that these haplotypes could potentially identify individuals with different risks of immune-related adverse events (irAEs, which I present work on in Chapter), it is worth noting that *TRBV* polymorphism is associated with multiple autoimmune diseases, most notably narcolepsy,[69–71] type I diabetes,[72,73] and multiple sclerosis,[74,75] among others.[76,77] Indeed, multiple groups have recently reported extensive single-nucleotide polymorphism, evidence of increased mutational rate, and population-specific differences in the TCR loci.[78,79] Extension of approaches developed by De Witt *et al*.[80]—who used TCRβ CDR3 sequencing (Adaptive ImmunoSEQ) of 666 unrelated individuals with metadata including age, sex, ethnicity, CMV serostatus,

and full class I and II HLA typing to identify TCRs strongly associated with specific HLA alleles and responses against common pathogens—to these diseases represents a promising avenue of research as well.

The cellular origins of the immune system

In many multicellular, eukaryotic organisms, the presence of diverse and adaptable pathogens has guided and selected for the evolution of a cellular defense response coordinated by incredibly specialized cells. These cells and the organs in which they live and circulate—collectively, the immune system—cooperate to recognize, prevent, respond to, eliminate, and remember disease. All immune cells arise from pluripotent hematopoietic stem cells, which differentiate into common lymphoid progenitor cells and common myeloid progenitor cells.[81–85] These immune stem populations produce the antigen-specific lineages (B cells, T cells, natural killer (NK) cells, and T/B/NK-like innate lymphoid cells) and granulocytic/polymorphonuclear cell lineages, respectively. The bone marrow is home to these developing populations, though certain tissue-resident immune cells originate from the yolk sac or fetal liver (examples include microglial cells in the central nervous system, fetal gamma-delta T cells, and others)[86,87]. Below, I will briefly discuss the developmental origins of one of these adaptive immune lineages—the T lymphocyte, or T cell.

Development of the T lymphocyte and rearrangement/recombination of the T cell receptor

T lymphocytes (so named for their thymic origin) are antigen-specific, HLA-restricted lymphocytes. This means that they recognize and respond to their targets by binding HLA-peptide complexes containing truncated and processed peptide fragments that are loaded into the central groove of a class I or class II HLA molecule. Before a T cell can recognize an HLA-restricted peptide, its hematopoietic progenitor cells and lymphoid precursors must first migrate to the thymus, where they attempt to rearrange a TCR, test the rearranged TCR against a variety of self-antigens (positive and negative selection), and then leave the thymus to further differentiate and respond to antigens in the periphery (**Figure 8, next page**).

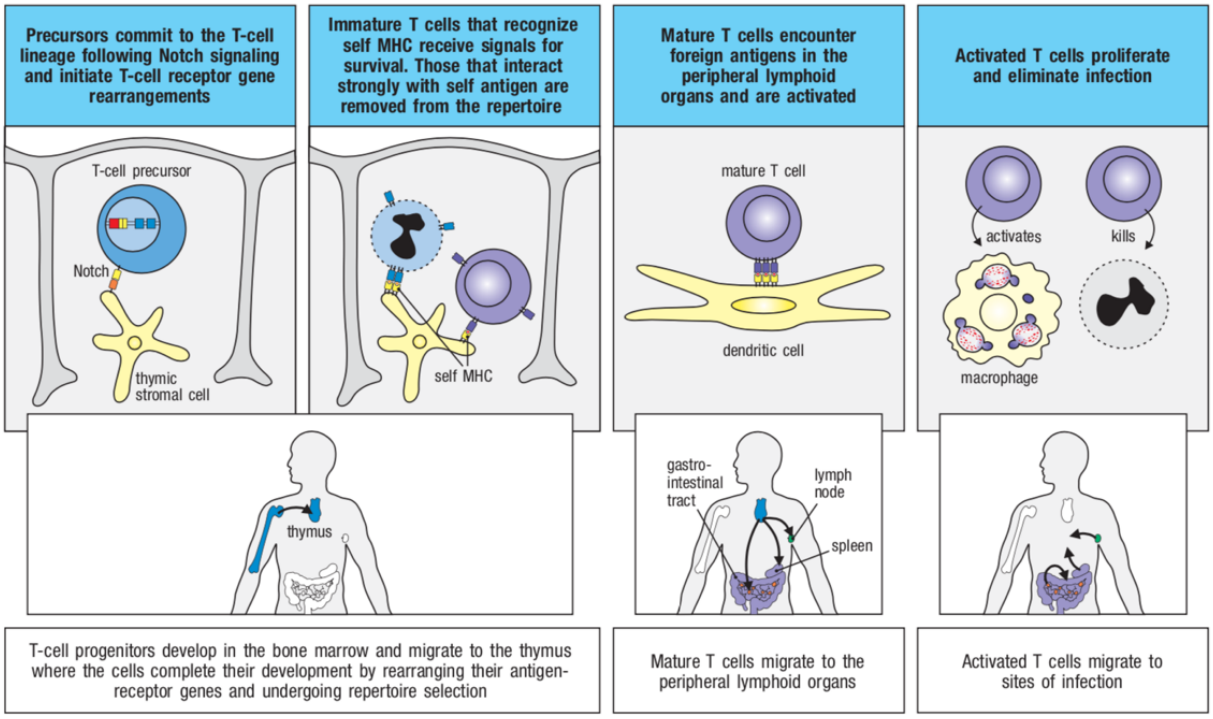


Figure 8. Graphical summary of the process of T lymphocyte development and the timing of TCR rearrangement. This figure is reproduced from Janeway's Immunobiology, 9th Edition by Kenneth Murphy and Casey Weaver (© 2017 by Garland Science, Taylor & Francis Group, LLC) with permission of W.W. Norton & Company, Inc.

Hematopoietic and common lymphoid progenitors that traffic to the thymus successfully commit to the T cell lineage after 1) receiving co-stimulatory Notch signaling from the thymic epithelium, 2) activation of the master T cell lineage GATA3 and TCF1 transcription factors, 3) TCF1-GATA3 co-induced expression of CD3 and RAG1, and 4) inhibition of non-T-cell developmental trajectories by activation of the BCL11B transcription factor.[88–91]

While many T cell progenitors successfully migrate to the thymus, murine studies indicate that only 2-4% of these cells exit the thymus as mature, selected T

cells.[92] T progenitor cells undergo 4 distinct 'double-negative' stages—so named because these T progenitor cells are negative in their surface expression of both CD4 and CD8 (CD4−CD8−). Rearrangement of a *TCRD* segment to a *TRBJ*, followed by rearrangement of a *TRBV* segment to the putative *TCRD-TRBJ* fusion results in a "testable" TCRβ that can pair with a pre-TCRα that can be expressed on the surface with CD3 isoforms and stimulated.[93,94] If a γδ TCR is successfully re-arranged and co-stimulated with CD3, the T cell is directly exported from the thymus, and does not undergo further thymic selection.[93,95–99]

Double-negative T cells bearing the precursor TCRβ-pre-TCRα next undergo rearrangement and recombination of the *TRAJ* and *TRAV* loci and acquire the 'double-positive' (CD4+CD8+) phenotype. CD4+CD8+ T cells can then move forward in development if 1) a putative TCRα is successfully rearranged and 2) can pair with the putative TCRβ and be positively stimulated by binding HLA and HLA-self-peptide complexes with sufficient strength.[100–104]

If the above conditions are met, the almost-mature T cell will transition from a CD3+CD4+CD8+ T cell to a 'single-positive' CD3+CD4+ T cell or a CD3+CD8+ T cell. While positive selection ensures that such cells can bind HLA with sufficient strength to surveil for non-self-peptides and leads to the single-positive developmental stage, negative selection removes those single-positive CD4 and CD8 T cells which react too strongly to self-antigens.[105] This is achieved by thymic

expression of critical self-antigens from multiple tissues by the *AIRE* gene,[106] which ensures these self-targets are shown to maturing T cells. Current evidence suggests that too much negative selection leads to automatic deletion of a given T cell both thymically and in the periphery, though a certain amount of negative selection is tolerable and leads to the lineage selection of CD4 and CD8 T regulatory cells.[107,108]

The adaptive immune system as a sensor of immune perturbation

The immune machinery I describe above comprise a sophisticated device that is capable of recognizing, responding to, and remembering foreign pathogens and immune insults. Chronic viral infections and select diseases of civilization are known to influence this machinery in specific ways (e.g. changing the frequency and phenotype of immune cell subsets, introducing low-grade systemic inflammation) that should be accounted for when designing immunologic studies in the framework I describe above. Below, I list severable notable examples of these immune perturbations, which are of utility to researchers studying the immune system “at scale” as inherent and common immune contrasts, and should be accounted for and used as matching criteria where possible prior to embarking on large-scale clinical immunological studies and single-cell RNA sequencing studies.

Cytomegalovirus and other herpesvirus infections

Cytomegalovirus (CMV) is a ubiquitous human betaherpesvirus that infects approximately half of the population by age 40, and effectively 100% of the population by age 90. Infection with cytomegalovirus shifts the composition of the CD8 and CD4 T cell compartment to a dominantly effector memory (Tem) and effector memory revertant/re-expressing CD45RA (TemRA) phenotype—a finding that has been conclusively demonstrated in 1000 healthy individuals spanning 20-70 years of age, matched by decade and by sex.[109] In one study, CMV serostatus alone explained greater than 50% of variation in the immune responses of monozygotic twins.[110] Responses to just one or two class I- and class II-restricted epitopes of CMV can occupy up to a quarter or more of the entire CD8 and CD4 T cell compartment as indicated by sequencing of the TCR repertoire and the antigen-specific TCR repertoire in members of the general population and in people living with HIV (PLWH).[111–116] This is partially due to an immunologic process known as memory inflation—thought to be unique to CMV—in which frequencies of CMV-specific cells grow to high proportions of the T cell compartment over time (reviewed in [117,118] and recently studied in innovative fashion in murine CMV models in [119,120]). CMV is also one of the first viruses for which antigen-specific natural killer cells—a newly identified phenomena and cellular population—have been described (with varicella zoster virus being the second).[121,122] CMV is a curious virus in that

infection with CMV appears to significantly improve vaccination outcomes in children,[123] while CMV co-infection in the setting of HIV and advanced aging is predictive of significant clinical co-morbidity,[124] cardiovascular disease and mortality,[125,126] frailty,[127,128] and vaccine responsiveness.[129]

HIV infection

Chronic infection with HIV leads to profound and marked depletion of and selection for select T cell phenotypes, including checkpoint "high" cells expressing the co-inhibitory molecules PD-1, LAG-3, TIGIT, and others[130–134]; depletion and infection of the CD4+ T follicular helper population[135]; selective and early depletion of T central memory (Tcm) cells[136]; expansion of CD4+ and CD8+ "activated" CD38+ HLA-DR+ T cells[137–139]. Evidence suggests that these alterations to the CD4+ T cell compartment also lead to preferential depletion of subsets of CD8+ T cells,[140,141] NK cells,[142] and B cells as well.[143]

Hypertension

Hypertension is an extremely common disease, with greater than 40% of adults over 25 years of age qualifying as hypertensive, another 30% as prehypertensive who will develop hypertension within 5 years, and more than 50% of adults given recent American Heart Association guidelines.[144–146] Murine models of hypertension, including use of multiple immunodeficient mouse lineages and

adoptive transfer, clearly demonstrate that hypertension is an immune-mediated disease driven by T lymphocytes and NK cells in a TH17 and TH1-dependent fashion.[147,148] Corresponding genetic and cellular studies in humans have confirmed these results; multiple polymorphisms in CD3 and the *HLA* gene families are associated with elevations in blood pressure,[149–151] and hypertensive humans have markedly higher circulating levels of CD4 and CD8 TH1, T effector memory (Tem), and T effector memory revertant/re-expressing CD45RA (TemRA) T cells.[152–155]

Dissecting the immune system using single-cell RNA sequencing, immune receptor sequencing, and computational analysis

The advent of single-cell RNA sequencing (scRNA-seq) has opened new avenues of investigation of the immune system at high-resolution. scRNA-seq methods differ in their methods and performance, and an explosion of novel scRNA-seq computational methods for analysis has emerged in the past decade. This rapid-moving area is characterized by wet-lab and dry-lab innovation and rapid turnover of conceptualization of even basic ideas as to how scRNA-seq data is best characterized and described. For example, Silverman *et al*. recently demonstrated that the abundant zeroes in scRNA-seq count data originate from distinct technical and biological processes and that description of scRNA-seq data as “zero-inflated”

and analyzing these data with a zero-inflated negative binomial statistical model is inappropriate.[156] The significant wet-lab innovation in this area is best embodied by the innovation of multi-modal approaches (which I describe below) and best illustrated by the $\log_{10}$-scale increases over the past decade in the number of single cells that can be assayed at once (**Figure 9**).

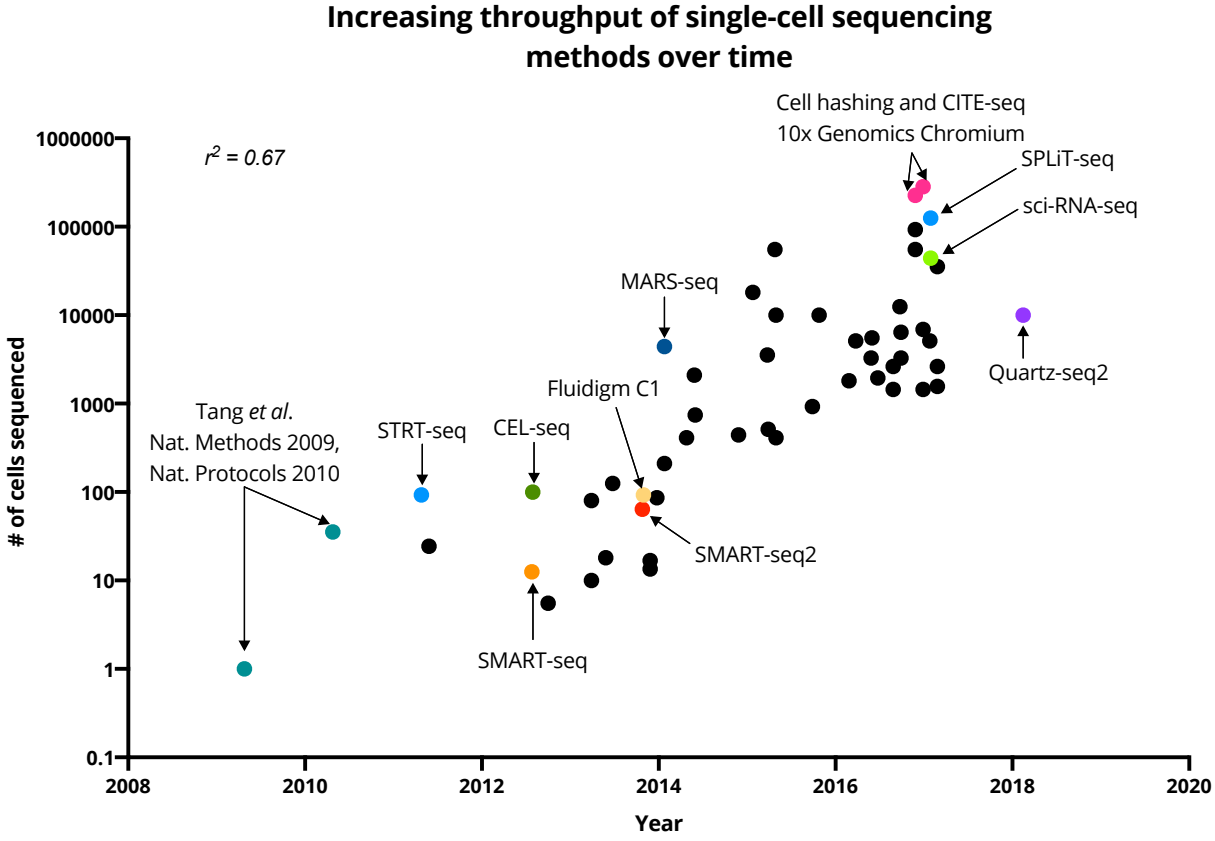


Figure 9. Increases in the throughput of scRNA-seq approaches over time. Since the invention of scRNA-seq in 2009, significant advances in technology and molecular biology have made it possible to measure hundreds of thousands to millions of cells at once. This figure was inspired by Figure 1 of Svensson, Vento-Tormo, and Teichmann[157] but has been updated to reflect the addition of other recent scRNA-seq methods, such as Quartz-seq2, which did not exist at the time of publication of Svensson and colleagues' work.

Existing single-cell RNA sequencing approaches: methods for cell capture and isolation

Currently, there are four common approaches to isolating single cells for single-cell RNA sequencing (scRNAseq): flow cytometry, microfluidic encapsulation within droplets, combinatorial indexing (included here though technically it does not isolate single cells), and physical separation of cells using microfluidic chip technology. The 10X Chromium[158] is the most popular microfluidic encapsulation platform, though the Fluidigm C1 microchip system[159] remains popular for microfluidic capture of rare cell populations. Notable non-commercial approaches using a similar approach, but with distinct barcoding methods, include Drop-seq[160] and inDrop.[161] However, the sensitivity of these methods is inferior to current commercial scRNAseq platforms.[162] Well-based assays such as SMART-seq,[163] SMART-seq2,[164,165] STRT-seq,[166] and STRT-seq-2i[167] rely on the sorting of individual cells, as do pooled amplification methods such as CEL-seq,[168] CEL-seq2,[169] MARS-seq,[170] SCRB-seq,[171] and mcSCRB-seq.[172] Finally, combinatorial methods such as SPLiT-seq[173] and sci-RNA-seq[174] rely on fixation, limiting dilution, and migration of cells between wells in a fashion that ensures that each cell will have received a unique combination of barcodes for reverse transcription—while the throughput of these methods is comparable to highly multiplexed microfluidic encapsulation, their sensitivity is poorest of all existing scRNAseq methods (approximately 200 genes captured per cell).[173,174]

Existing single-cell RNA sequencing approaches: amplification methods

Once individual cells are lysed, reverse transcription converts cellular RNA into cDNA that is amplified through pooled PCR (all microfluidic encapsulation methods, and non-SMART-seq and non-STRT-seq approaches) or through individual cell amplification (SMART-seq1/2, STRT-seq, and Fluidigm C1). The increase in cost is significant for individual cell amplification due to the linear increase in the number of PCR reactions required with the number of input cells. However, this increase in cost can be well worth the significantly higher sensitivity of these methods ($5x10^3$ – $10x10^3$ genes per cell).[162,175,176] Both pooled-PCR and plate-based approaches benefit significantly from automation on liquid handling platforms, which reduce the labor needed to complete an experiment and reduce the technical variation introduced prior to sequencing, which can introduce a sizable batch effect. The throughput of the pooled-PCR approaches is due to the incorporation of a cell barcode (a nucleotide sequence corresponding to each cell and delivered by well or via a gel bead) attached to a longer sequence comprised of a sequencing adapter, a promoter sequence or template-switching oligonucleotide sequence, and a unique molecular identifier (UMIs, used to remove PCR duplicates and used to identify exactly unique transcripts of mRNA, first described in application to scRNA-seq by Islam *et al*.[177]). For phenotyping specific and more rare cellular populations, plate-based methods are typically most appropriate, though the advent of initially

"proteogenomic" CITE-seq,[178] REAP-seq,[179] and ECCITE-seq[180] has enabled elegant multi-modal phenotyping of individual cells in a fashion comparable to flow and mass cytometry. In these methods, oligonucleotide-conjugated CRISPR guides and antibodies are introduced to single cells, enabling the 'indexing' of surface protein and CRISPR interruption and knockdown data to single cell transcriptomes and to individual T and B cell receptors. The number of markers that can be measured in this fashion is theoretically limited only by 1) the edit distance/dissimilarity between the oligonucleotide sequences on the analytes and 2) sequencing depth, in contrast to flow and mass cytometry which are limited by spectral overlap of conjugated fluorophores and mass density of conjugated metals.

Existing single-cell RNA sequencing approaches: sequencing methods

There are two main approaches to sequencing scRNA-seq libraries: "prime" approaches and "full-length" sequencing, so named for their sequencing of the 5' or 3' end of the cDNA or of the full-length cDNA. While Illumina short-read sequencing is the dominant and preferred approach in the literature, it is possible to sequence scRNA-seq libraries using the PacBio and Oxford Nanopore Technologies (ONT) platforms as well via the RAGE-seq and ScISOr-seq approaches,[181,182] which allow for the direct analysis and identification of full-length transcripts by sequencing 10x Chromium-generated full-length cDNA on both long-read and short-read platforms.

Bioinformatic approaches can also enable analysis of isoform-level data with short-read sequencing data,[183–186] though the sequencing cost and throughput of the ONT platform only approaches that of Illumina instrumentation with the PromethION platform (personal communications, Drs. Mick Watson and Nick Loman).

Existing single-cell RNA sequencing approaches: key limitations

Current scRNA-seq approaches are unable to capture non-polyadenylated RNAs (*i.e.* many microRNAs, long non-coding RNAs, and circular RNAs). The capabilities to analyze DNA, protein, CRISPR perturbation, and chromatin accessibility in tandem with scRNA-seq data from the same cell are all relatively new innovations. Innovations in bioinformatic analysis using transfer learning,[187,188] canonical correlation,[189,190] and other machine learning approaches[186,191–193] has made the integration of these data types possible even from distinct datasets. However, all scRNA-seq methods suffer from sensitivity that is lower than bulk RNA-seq, which is unlikely to improve given recent evidence suggesting that addressing this challenge by increased sequencing depth rapidly becomes cost-ineffective (and that detection of certain genes, such as CD4, requires saturating depth),[194] although innovative 'targeting' approaches using cell-barcode-specific locked nucleic acid arrays have demonstrated promise by showing that libraries of individual cells can be amplified

from library pools with greatly improved sensitivity.[195] The cost of scRNA-seq experiments using droplet-based approaches can be significantly decreased by pooling multiple samples stained with unique oligo-tagged antibodies that are captured and reverse-transcribed,[178,180] or by combining samples from donors of different genotypes. In the latter methods, genotype-aware approaches have demonstrated that it is possible to identify cell-type specific *cis* and *trans*-expression quantitative trait loci, and even to demultiplex cells pooled from different donors without any specific genotype information *a priori* due to differences in transcribed RNAs between the donors.[196–198]

Existing single-cell RNA sequencing approaches: select methods to identify and characterize immune cell subsets

Kacser and Waddington's popular metaphor of cellular plasticity[199] noted that cells fall along a continuous developmental landscape. This continuous landscape implies that there are few "specific" markers for a given cell type; this poses a philosophically and informatically difficult problem, especially given that clustering—a notoriously "slippery slope" approach for group identification[200]—is a dominant computational method used to identify distinct cellular groups in scRNA-seq data. Given the size and shape of scRNA-seq data—which typically contains many more features detected than cells measured, a formulation of the $n * p$

problem—dimensional reduction using approaches such as principal component analysis (PCA) or independent component analysis (ICA) prior to identifying groups of single cells has become the recognized standard biostatistical approach in this field. Other popular methods for achieving this include probabilistic methods such as t-SNE and FIt-SNE[201] and non-linear projection methods such as locally linear embedding (LLE),[202] ISOMAP[203], and diffusion mapping.[204,205] All of these approaches are inherently reliant on distance metrics (which quantify how similar or dissimilar each cell in a dataset is to every other cell); only two studies to-date have performed comprehensive benchmarking of these metrics and the effect of metric choice on downstream analysis,[206,207] indicating that treating scRNA-seq count data as compositional data offers strong analytical advantages. Few papers have compared the clustering performance of individual tools as well.[193,208,209]

These reduced components, which explain most of the variation between individual cells and groups of cells, are used in multiple analytical approaches to identify clusters of cells, each of which represents a unique cell type or subtype. There are three main types of clustering analyses, namely:

1. Hierarchical clustering: Hierarchical clustering is a simple and agglomerative (or "greedy") clustering method that assumes clusters of single cells are robustly separated in multidimensional space in approximately equally size groups. While it has been applied in its most

generic form in several scRNA-seq papers,[210–212] more robust approaches with more advanced feature selection and cross-validation methods that "specialize" hierarchical clustering for scRNA-seq data have been established.[212–217]

2. Density-based clustering: Density-based clustering operates by partitioning cells into clusters based on the density of regions of multidimensional space; this is analogous to drawing "gates" in flow cytometry or identifying a breakpoint between two peaks in a one-dimensional histogram. While DBSCAN is the most popular method,[218] there are other variations of density-based clustering that have been applied to scRNA-seq which are more innovative in their downstream feature selection steps.[219,220]
3. Graph-based clustering: A key limitation of hierarchical and density-based clustering approaches is the choice of hyperparameters informing the estimated number of groups or downstream analytical choices where human decision (commonly referred to as "expert opinion") is used to decide how many groups are represented in the data. Graph-based clustering approaches, which have become common place in social network and network analysis (with datasets of similar or larger size than scRNA-seq data), instead automatically detect the number of clusters

within a scRNA-seq dataset by analyzing a graph of the data—where each cell is a node connected to a number of its neighbors (other cells), *k*, in multidimensional space. Another advantage of graph-based approaches is their compatibility with statistical null hypotheses, such as a fully random graph of the same dataset, the overlap in nearest-neighbors between adjacent cells (i.e. how distinct is a cluster), and how likely a cell is to be randomly connected to its nearest neighbors and their nearest neighbors. Graphs easily represent non-linear structure, which also enables the identification of groups of cells with different density, 'shapes' of cells in multidimensional space (*e.g.* circular groups represented by the cell cycle stages, or differentiation trajectories) and scale easily across millions of cells.[221,222] Seurat,[223] PhenoGraph,[224] SCANPY,[225] PAGODA,[226] SC3,[227] and other single-cell software suites all use graph-based clustering approaches given these significant advantages.

Existing single-cell RNA sequencing approaches: case studies of the decomposition of heretofore homogeneous immune populations

Application of the clustering approaches described above to populations thought to be relatively homogeneous in the peripheral blood has led to a Kuhnian revolution

in the re-definition of multiple immune cell types. Examples include but are not limited to:

1. Naïve definition of major murine and human splenic and peripheral immune cell populations: Work by Jaitin *et al*. (the developers of MARS-seq)[170] used simple hierarchical clustering as described above with "purification" of sub-clusters via probabilistic mixture simulation to identify all of the major leukocyte (T, B, NK, dendritic cell, macrophage, and monocyte) populations within the spleens of mice. By stimulating dendritic cells of mice with the bacterial toxin lipopolysaccharide, they also identified type I interferon-dependent and independent subsets of plasmacytoid dendritic cells in the spleen. This work was later validated in stunning detail in the peripheral blood of humans by Villani *et al*, who successfully revealed 6 human dendritic cell subsets and 4 monocyte subsets. Importantly, these works revealed that the major immune populations could be defined without any *a priori* knowledge of the surface markers of these immune cells.

2. Redefinition of innate lymphoid cells (ILCs): Bjorklund and colleagues[228] and Gury-BenAri *et al*.[229] profiled human tonsil-derived ILCs and murine ILCs of the gastrointestinal tract and independently identified, in addition to the

known ILC-1/2/3 subsets, 2 new ILC subtypes. The first new ILC subtype expresses interferon gamma in a ROR-gamma-t-dependent fashion akin to TH17 cells, while the second expresses IL-2 and CCL22. The results of Gury-BenAri *et al*.[229] in particular are extremely valuable to the scientific community, as they show that the microbe-free gastrointestinal tract in drug-treated and germ-free mice leads ILC-1, ILC-2, and ILC-3 cells to converge on a single phenotype—implying that these mice cannot ever be an appropriate choice of model organism for the study of ILC biology.

3. Identification of fate bifurcation of malaria-specific T cells *ex vivo*: The work of Stubbington and colleagues[183] represents some of the most innovative computational analysis in scRNA-seq to-date. By analyzing splenic CD4+ T cells longitudinally over the course of *Salmonella typhimurium* infection, they showed 1) that it is possible to reconstruct and quantify expression of the full-length T cell receptor of individual TCRs from SMART-seq2 data, 2) that *S. typhimurium*-reactive CD4+ T cells occupy a functional spectrum ranging from effector memory to TH1 and activated and proliferating cells, and 3) that individual clones of CD4+ T cells branch into and seed different functional compartments of the antigen-specific T cell response. Stubbington *et al*. later extended this approach to the analysis of single B cells;[184] their method was

subsequently used by Croote *et al*. to study the somatic hypermutation lineage and transcriptomic trajectory of the human peanut-specific immunoglobulin E response.[230]

Existing single-cell RNA sequencing approaches: select methods and case studies to identify topologies and trajectories of differentiation and their branch points

As all cells in multicellular organisms arise from a limited number of progenitor and stem cell populations, all cells can be thought of as being positioned along a continuous, probabilistic, multi-branched developmental tree. This type of analysis is commonly referred to as a pseudotime analysis or a trajectory inference. The term 'pseudotime' emerged with the publication of Monocle by Trapnell *et al.*,[231] who demonstrated that latent temporal features can be extracted from cells sampled and sequenced at the same time (*i.e.* it is possible to identify differentiation processes at the single-cell level using cross-sectional data alone).

Trajectory inference can be a misnomer in that some trajectory inference methods rely on user-specified "shape" of the scRNA-seq data, including how many branches the cells should divide into and whether the process is linear, branching, or shaped otherwise. In this sense, the "shape" of the data itself is not inferred in some methods. For this reason, I suggest that pseudotime analysis is the more appropriate term. This is partially because pseudotime tools can be coarsely

divided into sensible computational groupings based on whether or not they infer the topology of the data and which topologies they can successfully identify:

1. Methods optimized for pre-specified trajectories: These methods were some of the first pseudotemporal algorithms developed for and applied to scRNA-seq, and typically assume linear/straight topologies or two-branched/bifurcating topologies. Examples include NBOR (which was applied to identify lineage markers of pre-dendritic cell populations, the branching points of conventional mouse dendritic cell populations, and showed that dendritic cell lineage commitment occurs in the bone marrow),[232] Waterfall,[233] Monocle,[231] diffusion pseudotime (which successfully gene signatures of multiple immune hematopoietic progenitor populations and their regulatory checkpoints, which was thought to be an intractable problem using scRNA-seq alone)[234], and Wishbone (which successfully identified the checkpoints of murine T cell development within the thymus).[235] Monocle, Waterfall, and NBOR are graph-based approaches, while diffusion pseudotime performs direct dimension reduction via diffusion mapping, and Wishbone uses supervised random walks along graphs and networks; this demonstrates that multiple mathematical models are appropriate for pseudotemporal

analysis of these shapes and types of data. More advanced forms of these algorithms such as FateID and RaceID3[236] can use supervised learning to statistically identify branch points, cells that biased progenitor populations, and more.

2. Methods that automatically detect both the shape and ordering of the trajectory: Recent innovations in graph theory have led to the development of scRNA-seq tools that significantly outperform the methods listed above. These graph-based methods do so for many reasons, but the chief reasons include 1) better modeling of the undersampling or absence of intermediary states in scRNA-seq data, 2) better representation of the disconnect between populations (i.e. not all trajectories should be connected), and 3) incorporation of probabilistic data that allows better modeling of the uncertainty of cells along a given trajectory. The most notable of these methods include partition-based graph abstraction (PAGA, which more accurately models topologies than any current method, a result independently validated by Saelens *et al*.[237]),[238] and Monocle 2 (which uses reverse graph-embedding to identify the shape and disconnectedness of a scRNA-seq data and successfully

identified rare progenitor populations that differentiate into novel subtypes due to mutations in major transcription factors).[239]

3. Next-generation trajectory inference methods (beyond pseudotime): Two highly innovative methods are worth describing in more detail on their own, as they utilize the unique features of two biological processes to reconstruct temporal lineages from cross-sectional data. The first of these, RNA velocity,[240] utilizes the ratio of un-spliced to spliced RNA, which is easily detected in scRNA-seq data, to directly estimate the first derivative of gene expression (gene expression over time). As a result, RNA velocity can not only identify the temporal relationship between cells, but it can quantify how far apart in time cells are from each other and even forecast where a cell will be along a differentiation trajectory and in relation to other cells up to 8 hours ahead of time. The second of these methods uses the rapid accrual of heritable mutations in mitochondrial DNA (mtDNA) to show that highly accurate lineages are present in both scRNA-seq and scATAC-seq data. Indeed, these naturally-occurring endogenous barcodes can be used to reconstruct single-cell clonal histories in mixture experiments, previously published datasets of the immune system, human tumors, and more.[241,242] These approaches

represent the future of lineage tracing in humans at single-cell resolution without the need for lentiviral barcoding or use of genetic data with lower signal to noise ratio such as single-nucleotide variation, copy number variation, or other methods.

Conclusion

The immune system and the methods I describe in this Chapter represent a highly configurable, customizable system with innumerable methods to study its composition, function, and states. The size and depth of immune datasets generated using these and other methods have begun to rapidly establish these approaches as gold standards. In the chapters below, I demonstrate application of several of these methods in the context of the immune perturbation framework I describe above to show the utility and value of studying immune-mediated human disease in this fashion.

Chapter II.

Learning the immune parameters predictive of mortality in HIV+ and HIV− veterans

N.B. This work comprises a manuscript in preparation. Here, we adapt and apply techniques originally intended for single-cell analysis to 80 immune cell measurements from 2,332 veterans living with and without HIV. Using these techniques, we build an integrated immune phenotype map that accounts for changes in these immune populations explained by HIV status, medication usage, cytomegalovirus serostatus, and other factors. Applying novel clustering approaches and statistical models to this map allowed us to identify distinct immune risk groups that differ significantly in their all-cause mortality over time and in common clinical parameters. These methods and preliminary results were submitted as an abstract to the 2019 Next-Generation Genomics conference.

Abstract

The number, distribution, and functional properties of immune cells change with aging. Absolute counts of adaptive immune cell subsets predict mortality and co-morbidity in the settings of HIV infection and transplantation, where they are used in clinical care. However, the extent to which the overall profile and distribution of adaptive immune cells describe, predict, and co-vary with the risk and extent of chronic disease and all-cause mortality in the general population is unclear. Here, we apply dimensional reduction, trajectory, and consensus clustering methods—originally designed for single-cell RNA-seq analysis—to flow cytometry measurements of 80 circulating immune cell subsets collected from 2005-2007 and relate these to subsequent mortality in 2,332 US Veterans with and without HIV in the Veterans Aging Cohort Study (VACS).

We identify 7 distinct groupings of immune cell subsets which differ significantly in their accompanying risk of all-cause mortality over 20,000 years of follow-up (550 events) after adjusting for HIV status, hypertensive and statin medication, and CMV serostatus using Harmony, an algorithm designed to adjust data in multidimensional space for known effects.[243] These groups also differed in their estimated glomerular filtration rates, CD4 nadir (for HIV+ veterans), and their anti-cytomegalovirus IgG titers. Higher mortality risk correlated with increased proportions of CD4 T effector memory and terminally differentiated cells and decreased CD4 T central memory and naïve cells. Using unsupervised trajectory inference through the `dynverse` workflow,[237] we subsequently identified a CD4 T cell differentiation trajectory that also correlated with risk of mortality. Finally, further position along this trajectory and elevations in CD4 T effector memory populations significantly correlated with the VACS index, a previously validated independent predictor of all-cause mortality, cause-specific mortality, and other outcomes. These results comprise an age-independent immune mortality signature that generalizes from the HIV+ population to the general population. Collectively, they suggest that routine monitoring of adaptive immune cells such as subsets of CD4 T cells may have broad clinical utility.

Introduction

Aging is associated with an apparent deterioration of the immune system and its ability to respond to threats and dampen chronic inflammation; it is unknown whether or not this state is reversible, and if so, to what extent.[244] Cross-sectional cohort- and population-level studies have revealed that the composition and functional capacity of the innate and adaptive branches of the immune system change with age and that these changes correlate with the development of "inflammaging" (premature aging of the immune system due to low-grade circulating inflammation).[245–247] Notably, the Newcastle octogenarian study—a prospective study of 751 individuals—identified cytomegalovirus (CMV) seropositivity and concomitant elevations in CD4 T effector memory RA+ (TemRA) cells as predictors of coronary heart disease, myocardial infarction, and stroke.[125] Some of these changes appear to be partially heritable,[109,248,249] though an equal or larger number of these changes are associated with ecological stimuli such as smoking, CMV infection, and variation in exposure to circulating influenza viruses and vaccination.[109,110] It is worth noting that a significant percent of variation in the serological response to common human pathogens can be explained by genetic associations and alterations in the human leukocyte antigen (HLA) and killer immunoglobulin-like receptor (KIR) genes.[250–253]

The current standard of clinical care in chronic HIV infection includes routine monitoring of CD4 cells. The frequency and phenotype of these CD4 cells is predictive of all-cause mortality, cause-specific mortality, frailty, and more. CD4 count also predicts the extent of immunosuppression and durable loss of T cell functionality after transplantation,[254–256] risk of post-transplantation complication by infection and cancer,[254,257,258] loss of allograft function,[254,259] and development of atherosclerosis after transplantation.[260] Even in HIV-negative individuals, low levels of CD4 T cells lead to opportunistic infections and reduced levels of NK cells and B cells.[261] Clinical administration of recombinant IL-7 can reverse this trajectory in individuals suffering from idiopathic CD4 lymphopenia,[262] and also increases the frequency of CD4 T cells and memory in HIV+ individuals.[263] However, the extent to which changes in the phenotype of the CD4 T cell population and other adaptive and innate immune subsets remains unclear. Recent work prospectively monitoring 135 healthy individuals over 5 years also identified an age-associated signature of immune aging associated with mortality and cardiovascular mortality in the Framingham cohort, though it is not known if this signature will generalize to co-morbid individuals and people living with HIV (PLWH).

Here, we study 79 immune cell subsets, circulating inflammatory markers, and clinical variables of 2,332 American veterans, of whom 1,525 are PLWH. Using pseudotemporal embedding, we identify an age-independent signature of CD4 T

cell immune aging associated with all-cause mortality. We also identify 7 groups of veterans who are similar in their immune composition. All 7 groups differ from each other in their mortality, clinical parameters such as their contemporaneous estimated glomerular filtration rate, their liver function (FIB-4 score), and in their anti-CMV IgG titer. We also provide evidence supporting both beneficial and pathologic roles of CMV co-infection which reconciles disparate evidence the literature in relation to CMV.

Methods

Study participants

The Veterans Aging Cohort Study Biomarker Cohort (VACS-BC) is a prospective cohort and subset of the larger VACS study comprised of 1525 HIV+ and 853 HIV− American veterans who provided blood samples and consented to DNA analysis between 2005 and 2007. More than 80% of the cohort are middle-aged (41-64 years of age), and the cohort is predominantly (95%) male and of African-American ancestry (68%). As part of this study, specimens from all individuals within VACS-BC have been assayed for genotype (Illumina HumanOmniExpress BeadChip 980K), circulating biomarkers of inflammation (IL-6, d-dimer, and soluble CD14), the alcohol consumption biomarker phosphatidylethanol/PEth, and immune cell subsets. These variables are linked to clinical phenotypes of cardiovascular disease,

cancer, liver and renal disease, frailty, metabolic disease, smoking, alcohol consumption, and opioid use that have been adjudicated by multiple physicians.

Flow cytometry analyses

6 flow cytometry panels were used to quantify and phenotype the presence of 79 immune cell subsets spanning **1)** CD4 and CD8 T cells and their memory subsets (naïve, Tcm, Tem, and TemRA), **2)** γδ T cells and NK cells, **3)** Th1 and Th2 T cells, **4)** T regulatory cells, and **5)** monocyte and granulocyte subsets. These panels were comprised of the following markers:

1. CD4 and CD8 memory T cell panel: PerCP-Vio700 CD4, PerCP-Vio700 CD8, APC-Vio CD45RA, VioBlue CD45RO, PE CD28, APC CD57, FITC CD27, and VioGreen live/dead.
2. Secondary CD4 and CD8 memory T cell panel: APC-Vio CD4, PerCP-Vio CD8, PE CD279, FITC CD45RO, APC CD127, VioBlue CD197, PE-Vio CD27, and VioGreen live/dead.
3. γδ T cells and NK cell panel: VioBlue CD3, PE γδ T cell receptor, PEVio770 CD56, PerCPVio CD16, FITC CD27, APCVio770 CD19, APC CD5, and VioGreen live/dead.
4. Th1 and Th2 cell panel: APC CD4, FITC CD8, PEVio770 IFN, PE IL-4, APC-Vio770 IL-17, VioGreen CD45RA, PerCP-Vio700 CD45RO, VioBlue live/dead.

5. Regulatory T cell panel: PerCP-Vio700 CD4, FITC CD25, VioBlue CD6, PE FoxP3, PE-Vio770 CD127, APC CD73, APC-Vio770 CD39, VioGreen live/dead.
6. Monocyte panel: PE-Vio CD14, PerCP-Vio CD16, FITC CD163, APC-Vio TIE2, PE CX_3CR1, APC HLA-DR, VioBlue live/dead.

Missing variable imputation

We used multivariate imputation with chained equations (MICE) with fully conditional specifications and predictive mean matching to impute any missing flow cytometry measurements (absent due to insufficient sample remaining), and CD4 count, CD8 count, viral load/HIV RNA, and CD4 nadir for HIV+ veterans. We used 5 iterations to allow the algorithm sufficient time to converge and estimate an appropriate imputed value. This approach is ideal as it will not impute a value below or above the observed range, and it does not introduce further collinearity or correlation between variables. Imputations were performed separately for HIV+ and HIV− veterans to account for immune alterations introduced by chronic HIV infection.

CMV serostatus assessment

We used the Diamedix CMV IgG immunoassay to quantify the serological response against CMV in all veterans in the VACS-BC. We classified individuals with a

response greater than 8 EU/mL as CMV-seropositive based on the included internal controls and standards of this assay.

Permutation and jackstraw analysis

To reduce the resulting data to an appropriate number of dimensions, we used the jackstraw test implemented in R to identify which principal components (PCs) explain a significant amount of variance and which PCs represent noise. This test identified the first 6 PCs as meaningful.

Dimensional reduction and UMAP embedding

We used truncated singular value decomposition to calculate the first 6 PCs of the flow cytometry measurements using the `irlba` package in R.[264] We then used these 6 PCs as input to the uniform manifold approximation and projection (UMAP) algorithm to visualize in 2 dimensions where each veteran resides relative to the others in multidimensional space. We used a minimum distance of 10^{-3} and allowed the selection of 15 nearest-neighbors for each data point. UMAP demonstrates extremely accurate performance in representing similarly high-dimensional data such as mass cytometry and single-cell RNA sequencing data appropriately.[265]

Correction of immune PCs using Harmony

We applied the Harmony algorithm[243] to correct the 6 calculated PCs for variation explained by 1) HIV serostatus and viral suppression, 2) CMV serostatus, and 3) statin and antihypertensive medication usage (as our primary outcome was all-cause mortality, we sought to compensate for cardiovascular disease co-morbidity).

Immune trajectory analysis

We used `dynverse`[237] to automatically select the most appropriate pseudotemporal algorithm to embed the immune subset measurements. `dynverse` selected the `SCORPIUS`[266] algorithm; we compared this trajectory to a trajectory we independently assembled using the `destiny` package[205] with parameters described previously.[267]

Consensus clustering

We used an ensemble of clustering methods to perform consensus clustering on the Harmony-corrected and uncorrected principal components, thus identifying groups of veterans who are similar in their immune composition. We used the diceR package to do so,[268] using the stochastic block model (which automatically identifies the appropriate number of communities), affinity propagation, and non-negative matrix factorization. We used 10-fold resampling, holding out 20% of the

cohort at a time, to boost the stability of the identified clusters. We tested the statistical validity of these clusters using the `sigclust` package,[269] which simulates data similarly distributed to the observed data and then calculates the probability that a given clustering solution exists given the same proportion of clusters existing in the randomly-generated data, where each cluster is a subset of a multivariate Gaussian distribution.

Assignment of common clinical parameters

We classified HIV+ veterans as viremic if their HIV RNA was higher than 500 copies/mL as previously described in VACS analyses.[270,271] We also stratified the HIV+ branch of the cohort by CD4 count lower than 50, between 50 and 200, between 201 and 400, and higher than 400 cells/mL. We stratified the entire cohort by HIV and CMV serostatus, and by quartiles of the anti-CMV titer (in addition to using CMV titer as a continuous measure and predictor of hazard). We calculated the VACS index as previously described. We also calculated a modified VACS using an approximation of CD4 count in HIV− veterans using the proportion of CD4 T cells as live lymphocytes (data shown below in **Results**) and excluding HIV RNA.

Additional statistical testing

We used the log-rank test to test for significant between-cluster differences in survival over time. We also used the Cox method to estimate the effect sizes of cluster membership in relation to important clinical covariates including age, gender, BMI, smoking, hepatitis C co-infection, and self-reported alcohol consumption. We used Schoenfeld's test to test proportionality of all predictors in the model in relation to transformed-time interactions. We used the `smoothHR` package to fit and visualize non-parametric splines of key variables in relation to hazard of mortality risk while adjusting for other immune subsets and the clinical variables listed above, to calculate the optimal degrees of freedom to include in our Cox models, and to calculate the confidence intervals of these hazard curves.[272] We also use exact permutation as implemented in the `lmPerm` package to assess the fit of regression models using the immune subset in relationship to each other (linear regression), to HIV and CMV serostatus (logistic regression), and CD4 count (Poisson regression).

Results

HIV+ veterans achieving viral suppression experience similar risk of mortality compared to HIV− veterans

As previously reported, we observed similar survival between virally suppressed PLWH and HIV− veterans (**Figure 10**). Both viremic PLWH and PLWH with lower CD4 counts experienced significantly higher risks of mortality over time (**Figure 10**).[273–276]

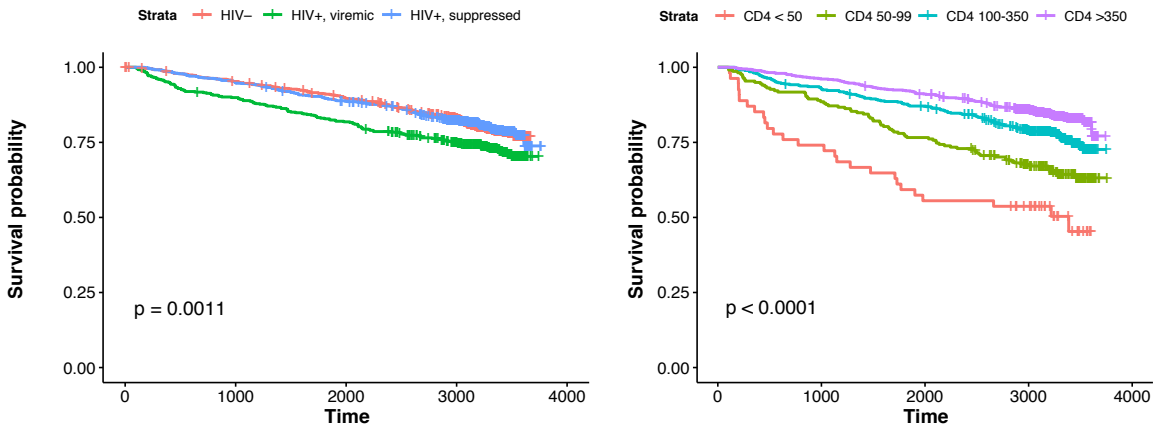


Figure 10. Survival of PLWH by suppression status in comparison to HIV-negative veterans. Left: Viremic PLWH experience a higher risk of mortality over time compared to suppressed PLWH and HIV-negative individuals. Right: CD4 count at baseline predicts all-cause mortality in VACS-BC, in agreement with previously published studies.

Signatures of cytomegalovirus infection differ between HIV+ and HIV− individuals

We observe significant increases ($p < 0.0001$, one-way ANOVA) in the CD8 TemRA population in CMV-seropositive individuals, as previously reported (**Figure 11**).

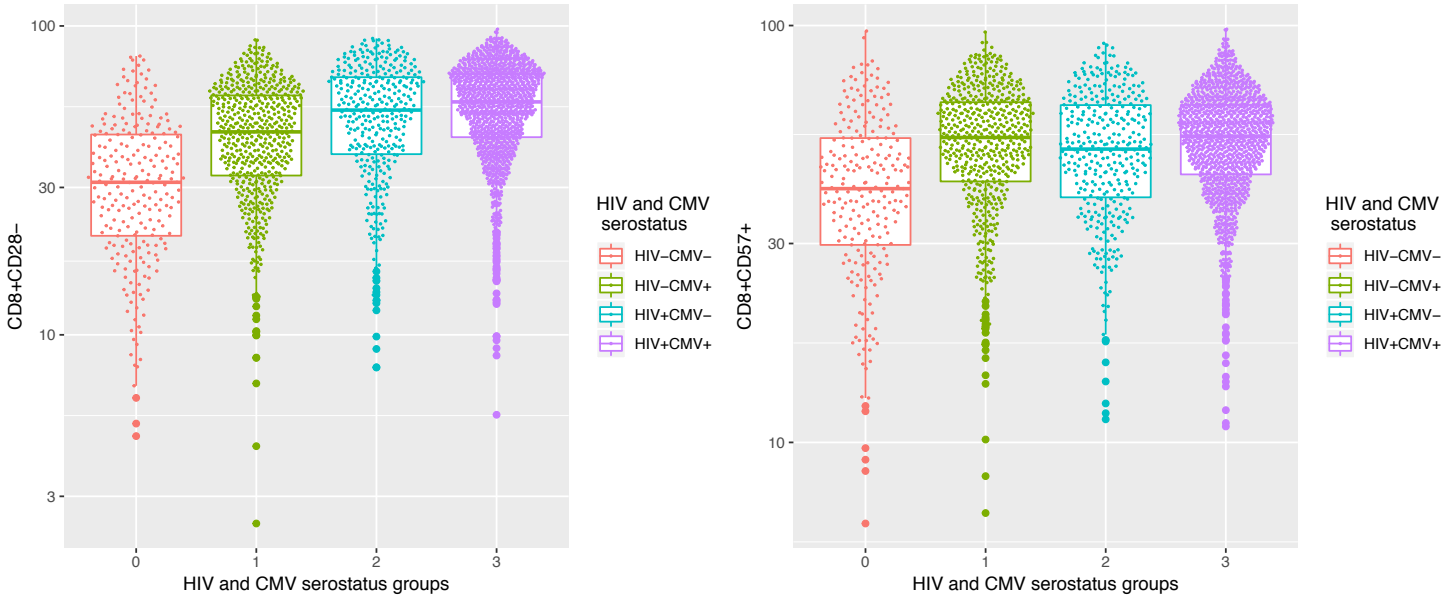


Figure 11. CMV+ individuals have higher levels of CD8 TemRA in peripheral blood. Using two definitions of terminally differentiated CD8+ T cells, we confirm that both HIV+CMV+ and HIV−CMV+ individuals carry higher proportions of CD8 TemRA cells.

We also confirm, as in previous reports, that HIV−CMV+ individuals have higher proportions of CD4 TemRA populations; however, we report for the first time that this population does not appear to distinguish between CMV+ and CMV− PLWH even when adjusting for CD4 count at baseline (**Figure 12**, see next page). This may be due to the presence of HIV-specific T cells with a TemRA phenotype, selective depletion of non-TemRA CD4 subsets in chronic HIV infection (as previously suggested), or other reasons; however, this cannot be established without longitudinal studies of both PLWH and HIV− individuals, ideally capturing seroconversion of newly CMV-seropositive PLWH.

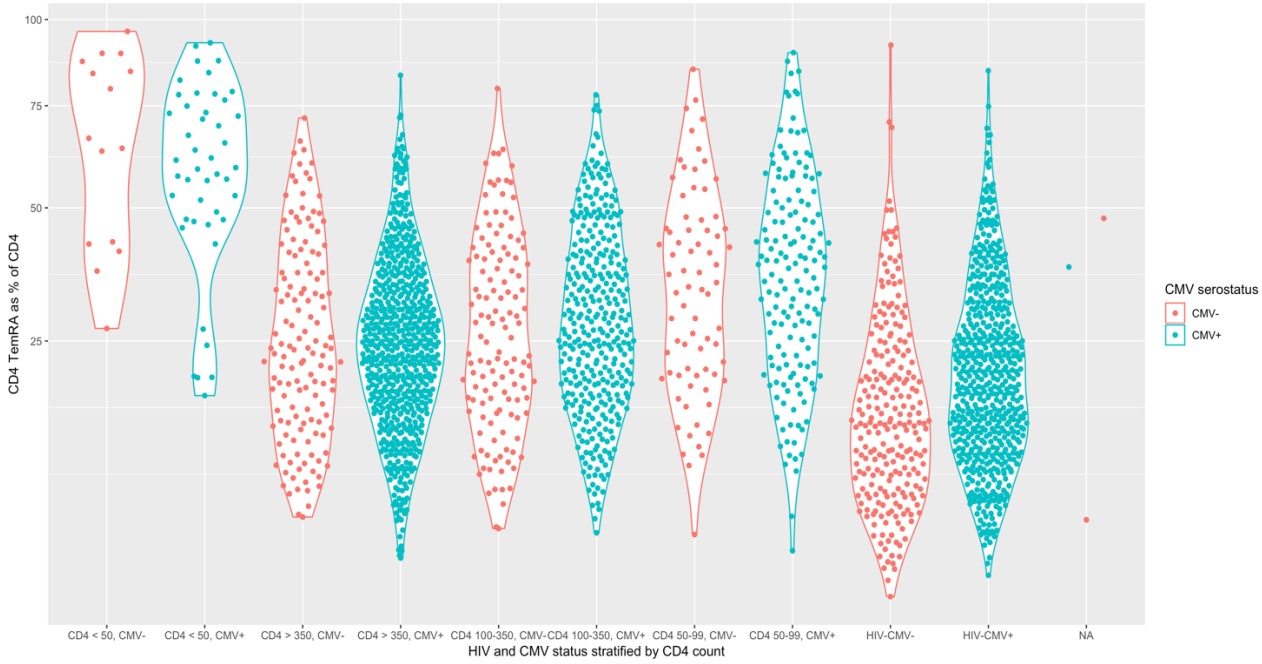


Figure 12. Proportion of CD4 TemRA as % CD4 does not distinguish between CMV+ and CMV− PLWH.

We also document evidence supporting a survival benefit of CMV seropositivity (OR 0.82, CI 0.68-1.0, p = 0.054), in agreement with early-life literature suggesting that CMV+ individuals benefit from enhanced serologic immunity and murine literature that MCMV+ mice respond more robustly to heterologous infection (**Figures 13 and 14, see next page**).

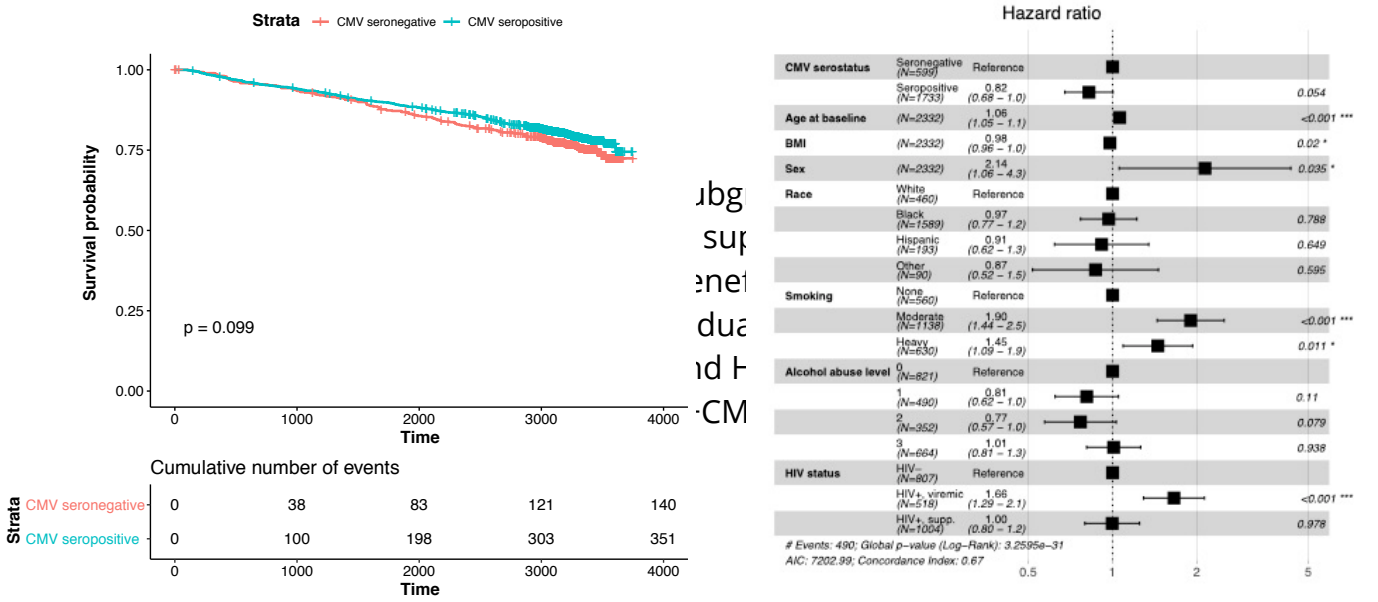


Figure 13. CMV-seropositive individuals in VACS-BC survive longer than CMV-seronegative individuals. Left: Kaplan-Meier survival curve of CMV+ vs. CMV− individuals. Right: Cox model including CMV serostatus and common clinical parameters.

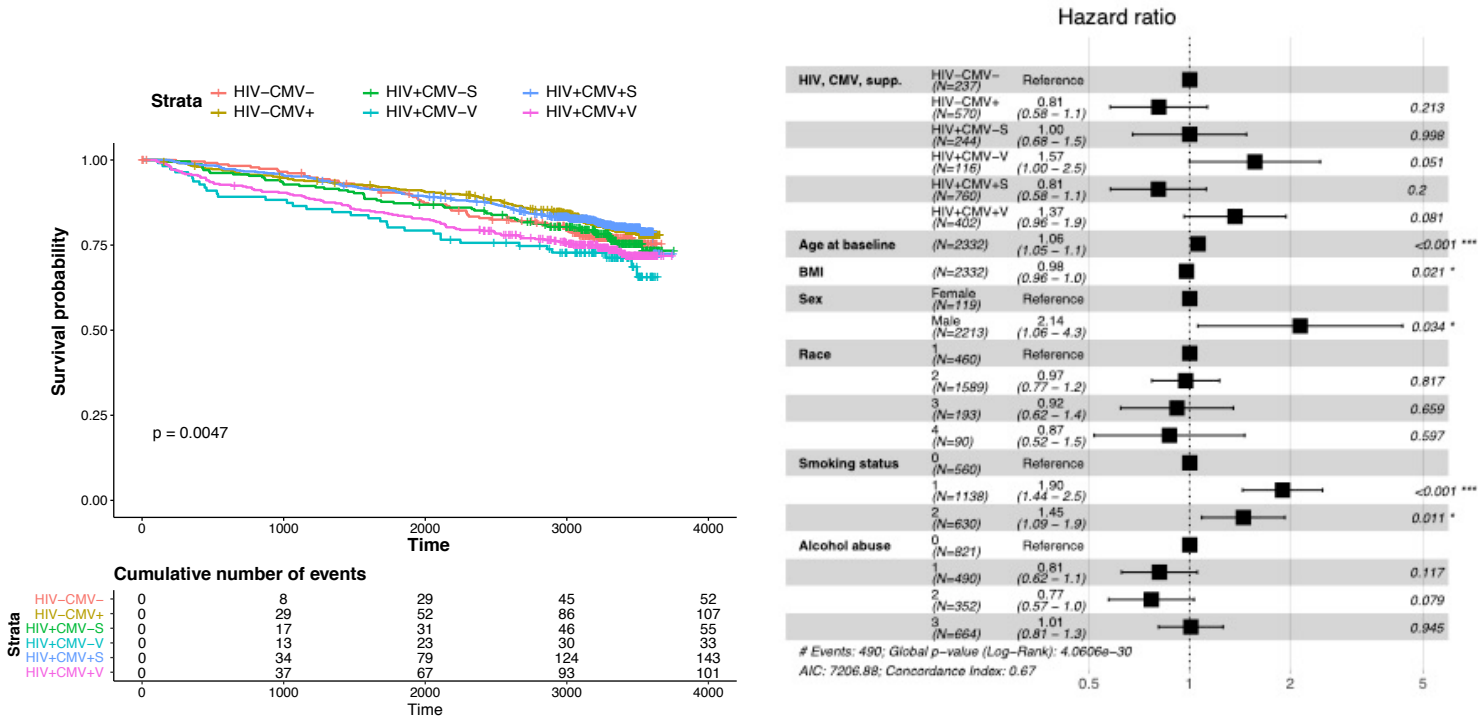


Figure 14. CMV-seropositivity and viral suppression enhance survival in PLWH and in HIV− persons. Left: Kaplan-Meier curve stratified by HIV status, CMV status, and viral suppression status for PLWH, with log-rank test for ordered stratification of survival curves (p = 0.0047). Right: Cox model (global $p = 4x10^{-30}$) incorporating stratification groups by HIV, CMV, and viral suppression with common clinical parameters and other predictors of mortality.

To better understand the potential benefit of CMV seropositivity in chronic HIV infection—which contradicts some reports in the literature of expanded HIV reservoirs, decreased survival, and enhanced cardiovascular disease in CMV+ individuals—we also examined the lowest CD4 count (CD4 nadir) in the clinical record of all PLWH in VACS BC, stratified by CMV serostatus. Indeed, HIV+CMV+ had higher CD4 nadirs than their HIV+CMV− counterparts, even when stratifying by their viral suppression status (**Figure 15**).

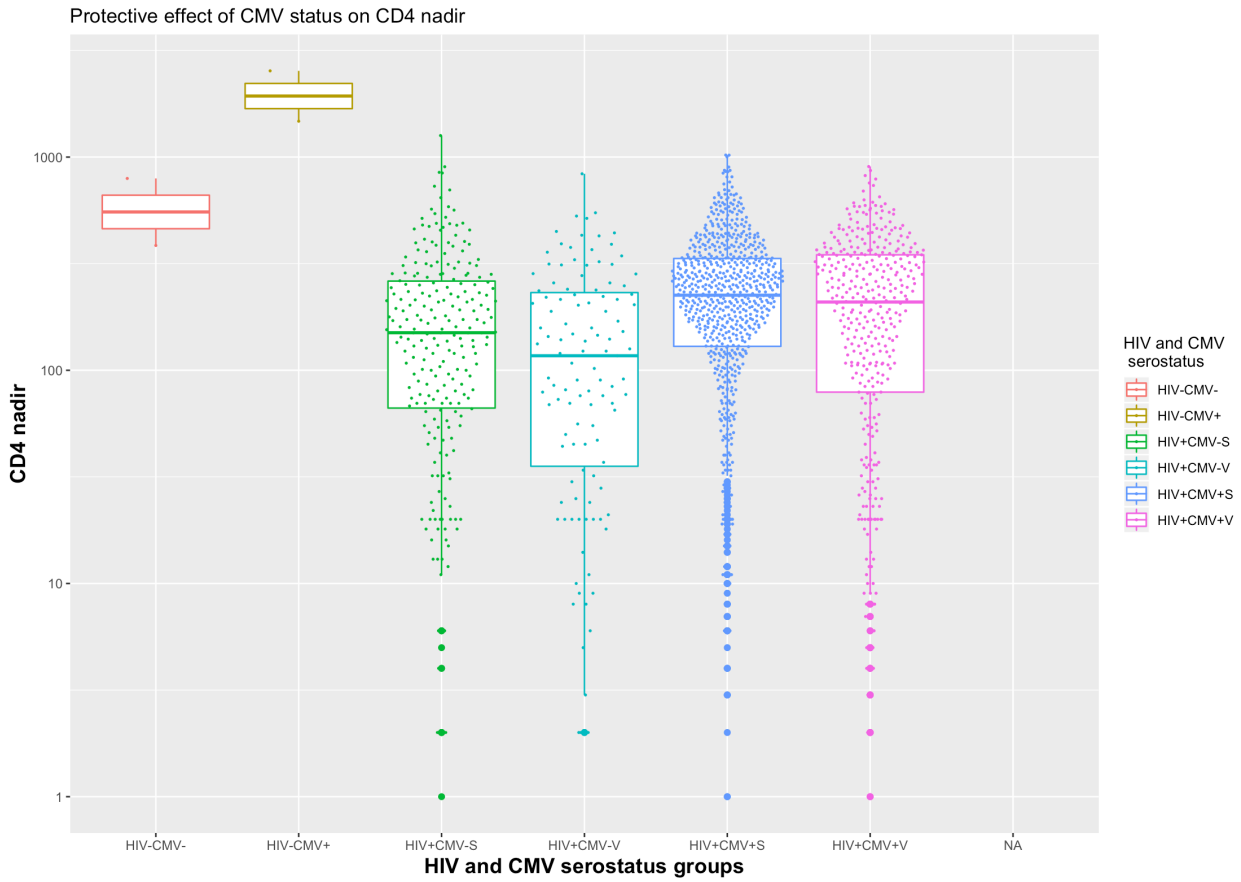


Figure 15. CMV-seropositive PLWH in VACS BC had higher historic CD4 nadirs than their CMV-seronegative counterparts.

However, given that previous reports have established CMV seropositivity as a predictor of both mortality and co-morbidity in both PLWH and the general population, we inspected the anti-CMV IgG response and the proportion of CD4 TemRA cells as continuous predictors of mortality while adjusting for age, BMI, race, smoking, alcohol abuse, HIV status and viral suppression, and hepatitis C co-infection. We found that increases in anti-CMV IgG titer and increases in the proportion of CD4 cells with the TemRA phenotype independently increased risk of hazard of death after adjusting for the variables listed above (**Figure 14** and **Figure 16**), which we also analyzed as quartiles (**Figure 17**).

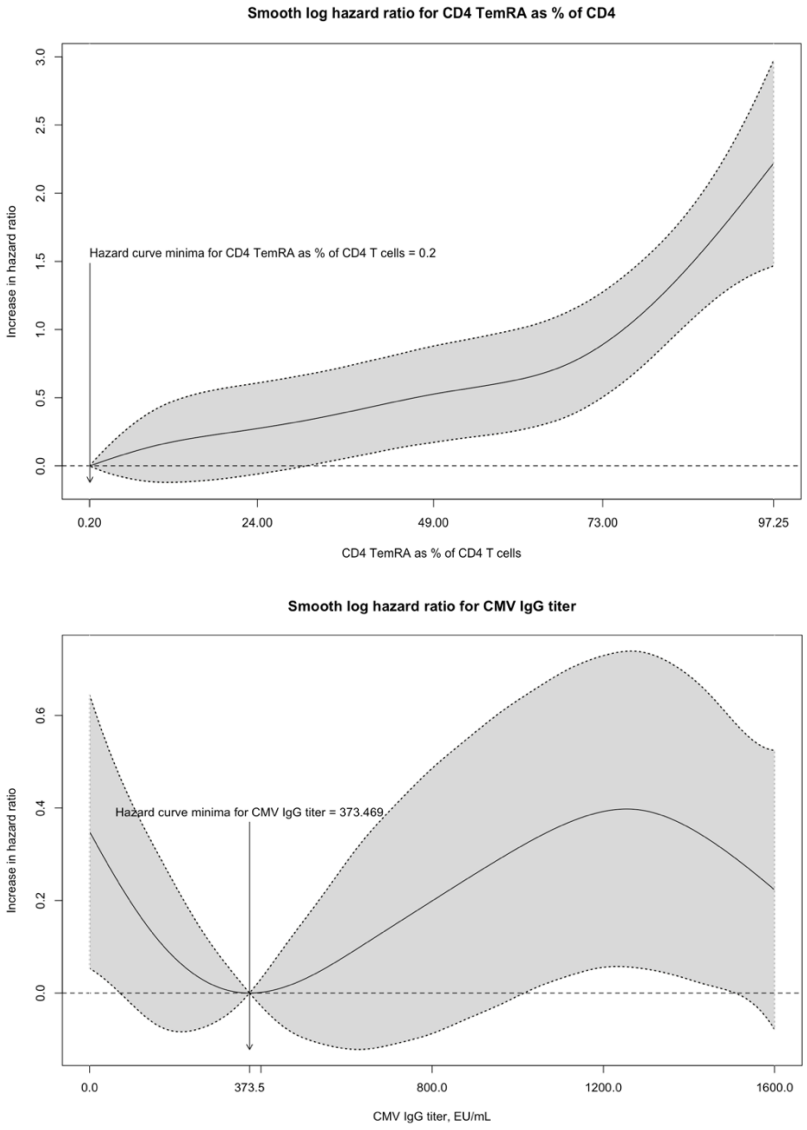


Figure 16. Smoothed hazard ratio of death for CD4 TemRA proportion (top) and anti-CMV IgG titer (bottom) after adjusting for common clinical parameters.

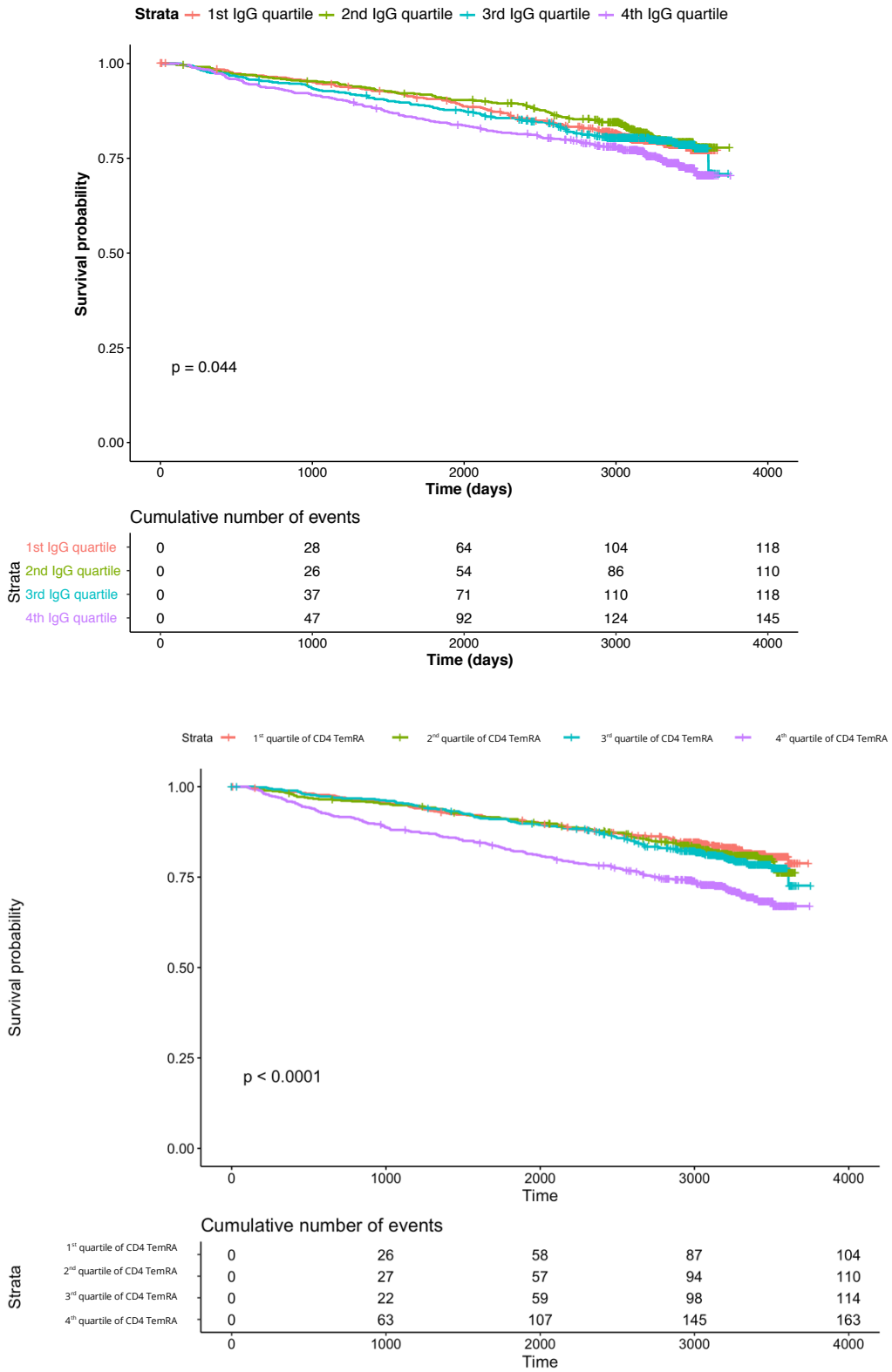


Figure 17. Median survival decreases with concomitant increases in the anti-CMV IgG titer and CD4 TemRA proportion.

We also assessed the relationship between these two variables (shown in **Figure 18**), demonstrating, for the first time for our knowledge, the relationship between the anti-CMV IgG titer and CD4 TemRA cells in both PLWH and in HIV− persons.

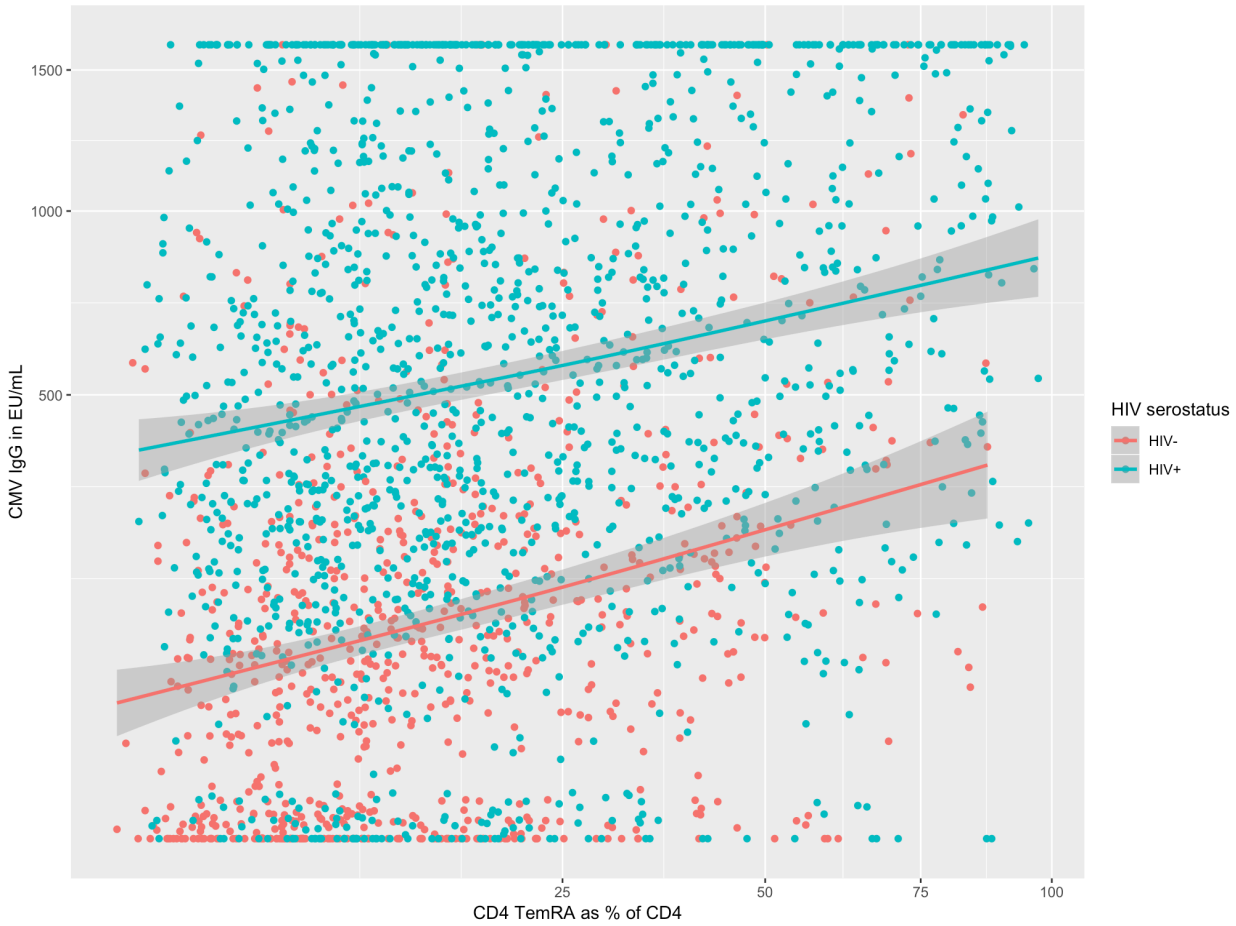


Figure 18. CD4 TemRA cells and anti-CMV IgG titer correlate in both HIV+ and HIV− veterans.

Identifying an age-independent CD4-based trajectory of immune aging

Given the association between elevations in CMV titer and CD4 TemRA cells and all-cause mortality, we next performed unsupervised analyses using clustering and pseudotime trajectory inference approaches designed for single-cell analysis and visualized these clusters using the UMAP algorithm (see **Methods**). Briefly, we used statistical testing to identify an appropriate number of principal components (PCs)

to reduce the 2,332 by 79-dimensional data to via principal component analysis (PCA) and then embedded these PCs with UMAP for visualization purposes (**Figure 19**). We then adjusted for HIV status, hypertensive and statin medication usage, and CMV serostatus using the Harmony algorithm, which uses local linear regression to adjust for categorical factors after dividing observations into biologically similar neighborhoods. Using the corrected PCs, we performed consensus clustering using automatic community detection to identify groups of both HIV+ and HIV− veterans who are similar in their immune profiles while accounting for distance in multidimensional space explained by the variables listed above.

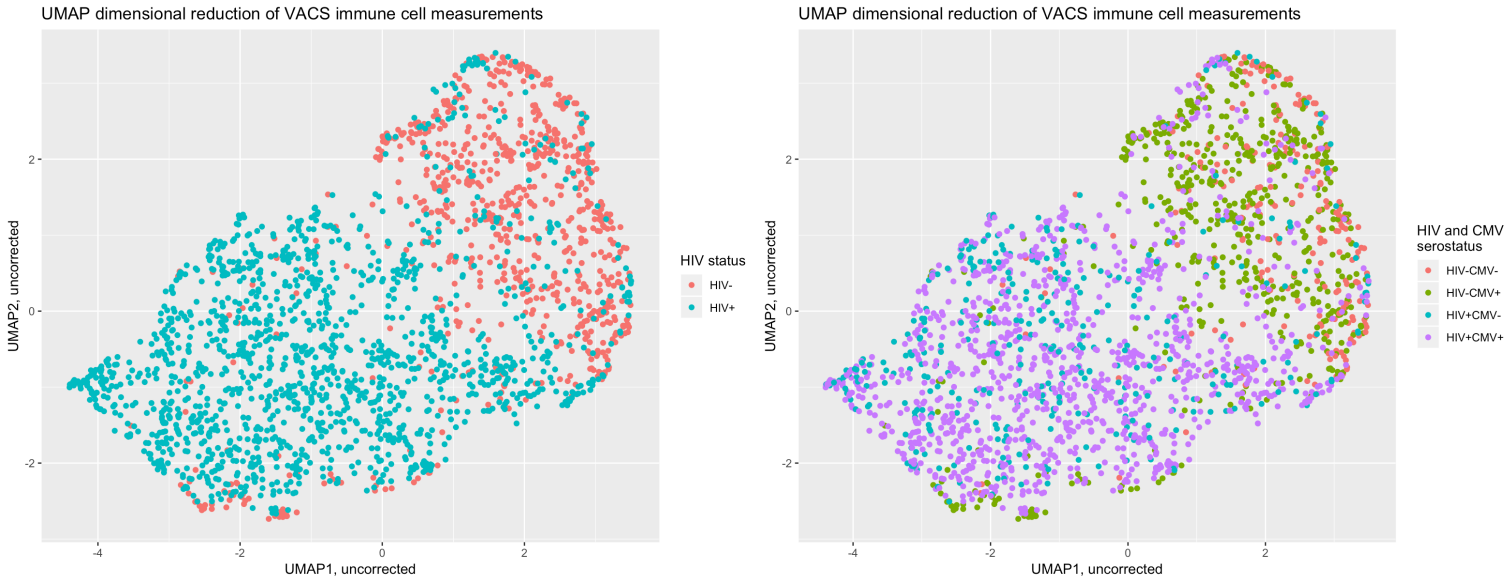


Figure 19. The uncorrected UMAP of the VACS-BC based on immune cell measurements. Left: UMAP visualization of the 2,332 veterans in VACS BC; distance between two points represents the similarity between two individuals based on their 79 immune cell subsets. Notably, HIV status is the dominant present effect in the uncorrected map, which is to be expected. Right: Though we do observe differences in select immune cell subsets by CMV serostatus, these differences do not significantly skew the topology of the data in multidimensional space.

We observed that location on the uncorrected UMAP not only separated veterans by HIV status, but that increases in cell populations historically associated with chronic HIV infection were strongly represented and also highest in individuals with high VACS index scores, which are independent predictors of mortality (**Figure 20**).

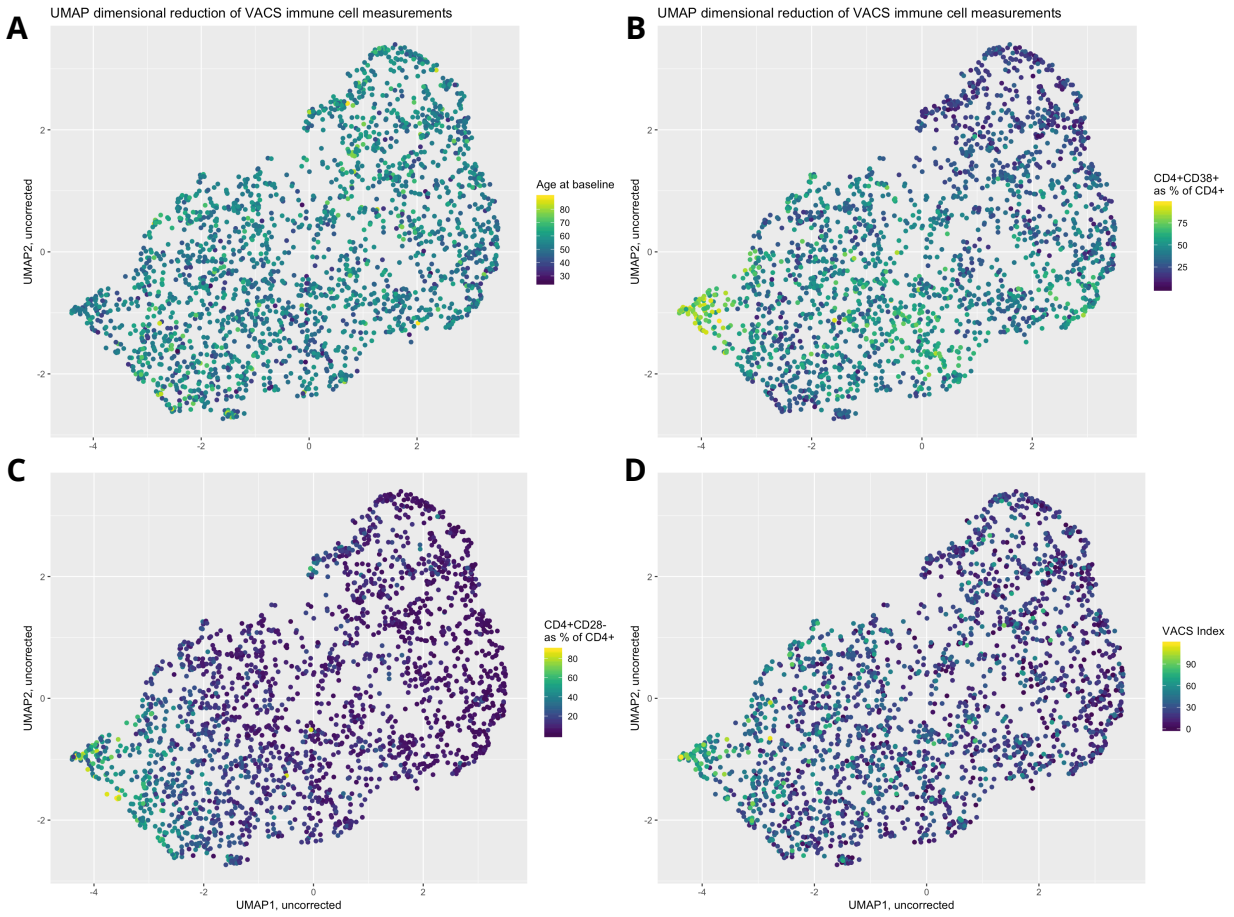


Figure 20. UMAP analysis reveals correlations between VACS index and select CD4 T cell populations associated with HIV status and mortality in PLWH and an age-independent.The topology of the immune cell measurements is not predominantly explained by changes in age. B) HIV+ individuals have higher proportions of activated CD4+CD38+ T cells as previously reported, as well as C) higher proportions of terminally differentiated CD4 T cells. D) Elevations in activated and differentiated CD4 T cells correlate with higher VACS index scores, an independent predictor of mortality.

Our consensus clustering analysis revealed 7 groups of veterans with similar immune cell compositions, and that use of the Harmony algorithm successfully adjusted for HIV status (**Figure 21**). These groups of individuals different significantly in their survival over a decade later (over 20,000 years) and also in their median VACS index scores (**Figure 22**).

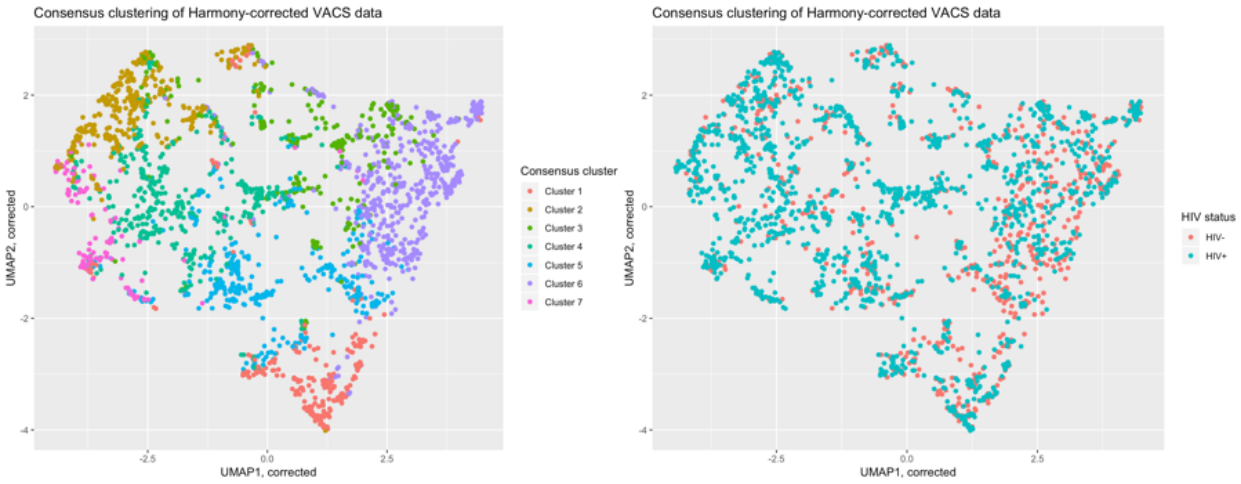


Figure 21. Corrected consensus clusters of veterans based on their immune measurements. Left: The 7 consensus clusters identified by community detection on the first 6 corrected principal components of VACS-BC and its 79 immune cell measurements. Right: The corrected principal components, as visualized with UMAP, are effectively adjusted for HIV serostatus and other clinical parameters using Harmony.

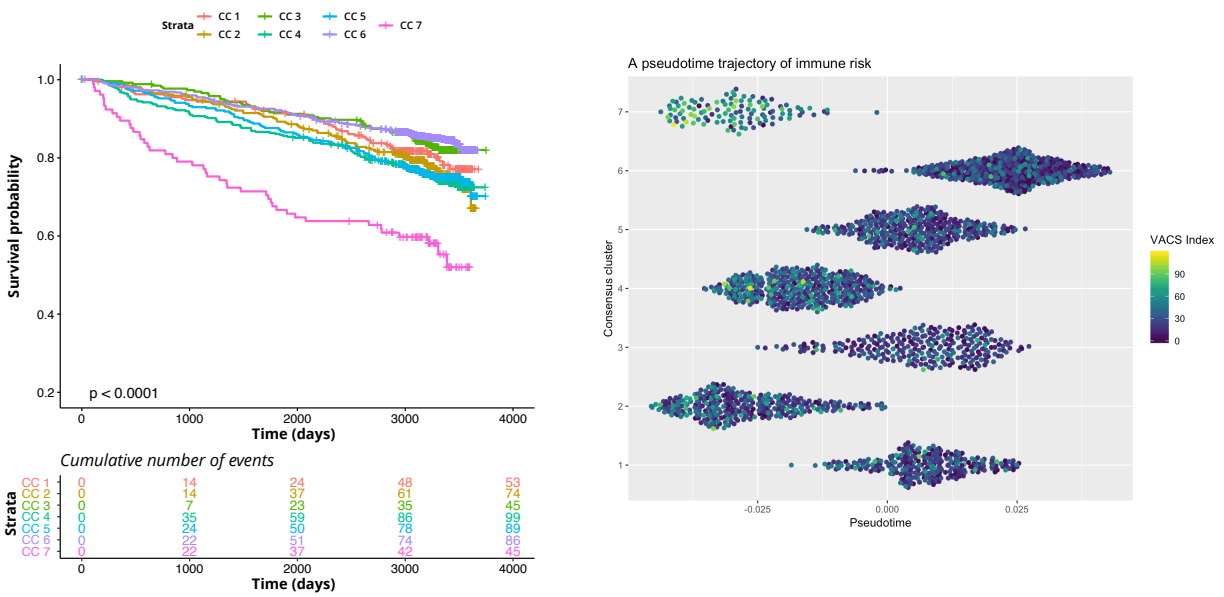


Figure 22. Variation in immune composition is associated with differential risk of all-cause mortality and VACS index. Left: Differences in survival over time between the 7 consensus clusters identified in **Figure**. Right: The 7 consensus clusters differ in their position in pseudotime as calculated by SCORPIUS and destiny, and also in their VACS index scores, an all-cause and cause-specific predictor of mortality in both PLWH and in HIV-negative persons.

We next embedded the corrected PCs in pseudotime using the destiny and SCORPIUS algorithms (see **Methods**) to test whether or not this immune signature correlated significantly with age. While pseudotime embedding clearly revealed an immune differentiation trajectory (**Figure 23A**), this signature weakly correlated with age (**Figure 23B**) and more strongly correlated with the VACS index (**Figure 23C**). Deviation from the best-fit line in the age-pseudotime correlation was best explained by the second dimension of pseudotime (**Figure 23B**) and leftward position (further along pseudotime) independently captured the CMV titer-CD4 TemRA mortality signature identified above (**Figure 23D**), in that the lowest-

pseudotime consensus clusters are markedly higher in their anti-CMV titers than their high-pseudotime peers. This is remarkable given that these clusters are approximately balanced in their ratios of CMV+:CMV− individuals (**Figure 24**). When we decomposed this signature using jackstraw testing to identify immune cell subsets associated with the main PCs, and using PERMANOVA to identify immune cell subsets best associated with each cluster, we found that this signature was best described by progressive CD4 differentiation (**Figure 25**). Finally, in a Cox model comprised of position in pseudotime, baseline age, sex, race, body mass index, hepatitis C co-infection, smoking and alcohol consumption, kidney and liver function, CMV serostatus and CMV IgG titer, HIV status and viral suppression status, and cocaine and intravenous drug use, position in pseudotime was the dominant predictor of mortality (odds ratio 4705, 95% C.I. 4.32−5,129,000). This Cox model did not violate the proportional hazards assumption (global fit $p = 0.5023$; potentially violating variables HCV co-infection [$p = 0.0334$] and baseline age [$p = 0.0537$]).

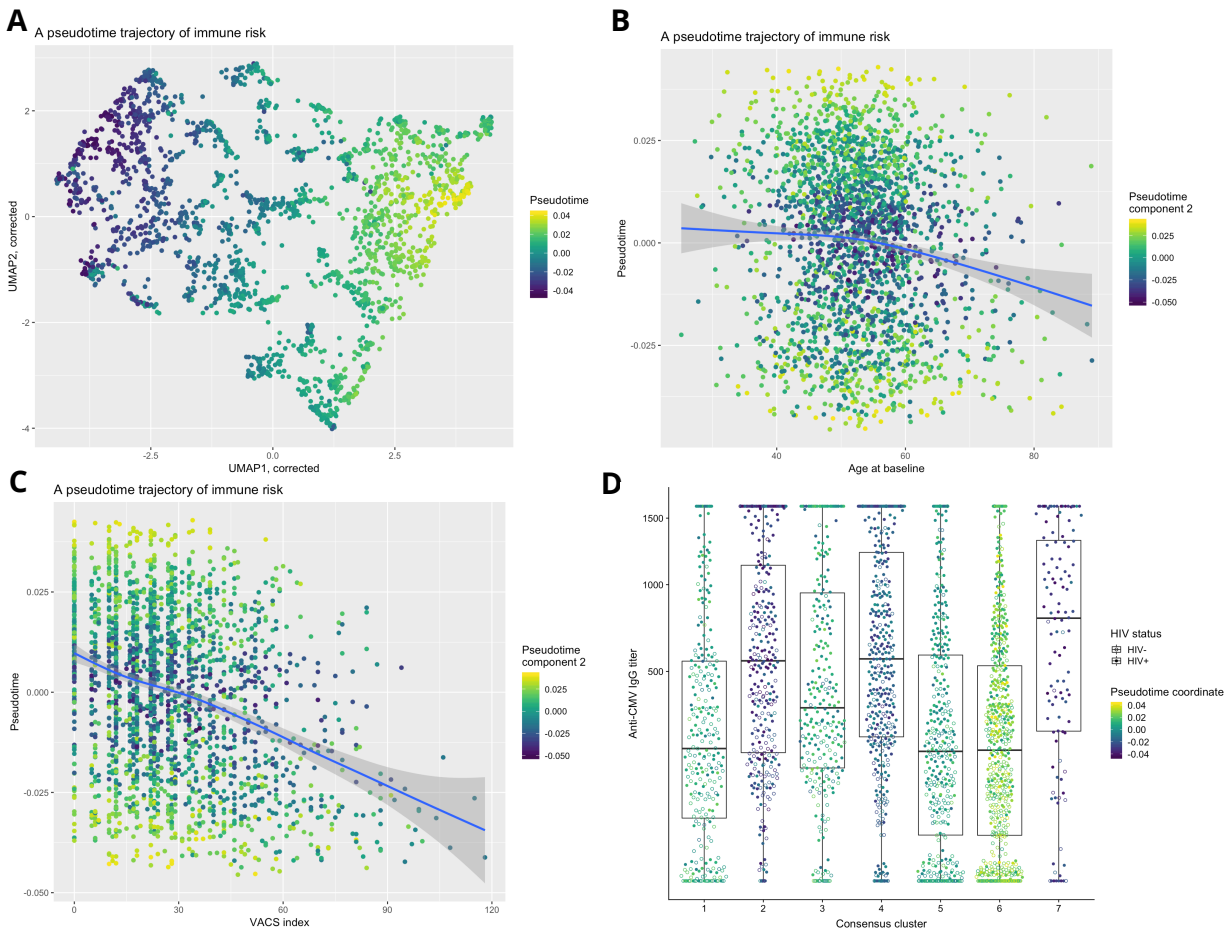


Figure 23. Trajectory analysis reveals an age-independent immune trajectory associated with mortality. Extraction of latent temporal data from this cross-sectional cohort via the pseudotime algorithms destiny and SCORPIUS reveals a strong immune aging trajectory that is only weakly B) associated with age. Residual variance between age at enrollment in VACS-BC and pseudotime is best explained by the second dimension in pseudotime. C) This pseudotemporal immune aging signature correlates strongly with the VACS index, an all-cause and cause-specific predictor of mortality in both HIV+ and HIV− individuals. D) Pseudotime also captures the increase in the anti-CMV IgG titer, which is independently associated with all-cause mortality as shown above. Clusters with the highest mortality (2, 4, and 7) are positioned farthest back in pseudotime and have significantly ($p < 0.0001$) higher anti-CMV titers than the lower-mortality clusters.

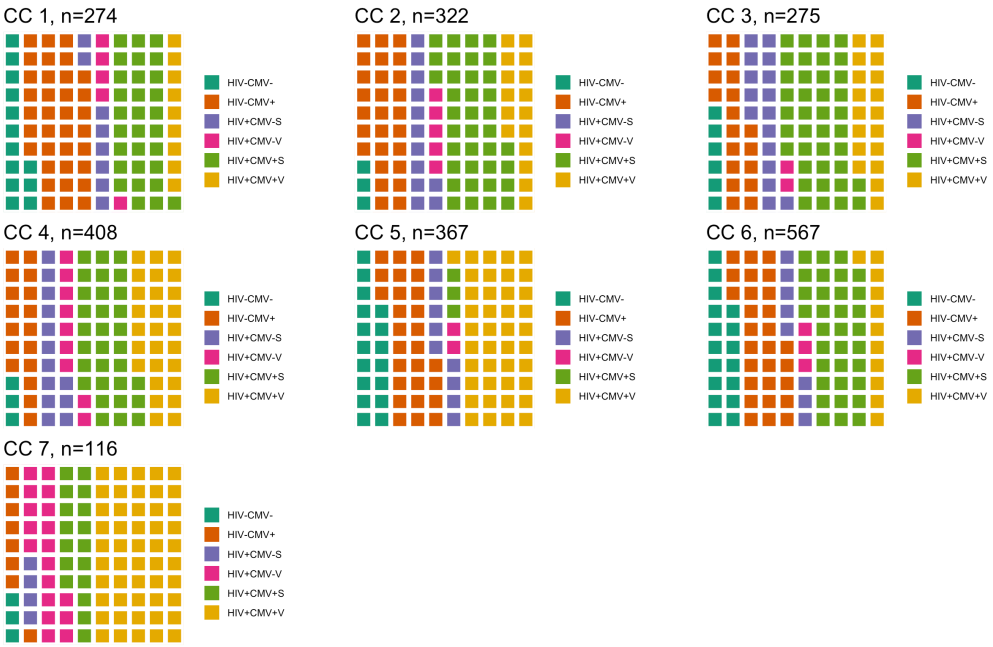


Figure 24. Compositional analysis of the consensus clusters by HIV status, CMV status, and suppression status. The consensus clusters we identified are approximately balanced by these three variables with the exception of CC 7, which is mostly comprised of viremic CMV+ PLWH.

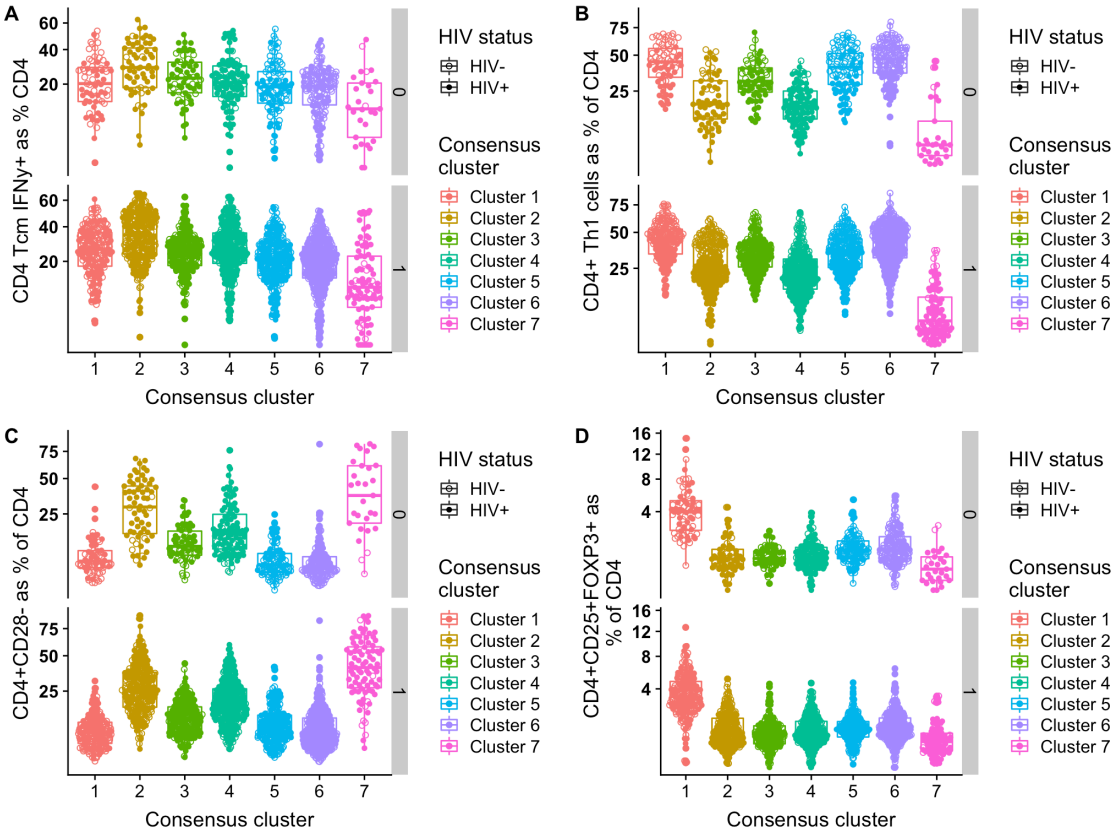


Figure 25. The age-independent immune aging trajectory is defined by CD4 effector and regulatory populations. Using PERMANOVA and random-forest variable importance trees, we identified branching points (inter-cluster transition points) that corresponded significantly with multiple immune cell subsets, including A) CD4 Tcm cells expressing IFN-y, B) CD4+ Th1-cytokine producing cells, C) terminally differentiated CD4+ T cells, and D) T regulatory cells. All plots are faceted by CMV serostatus (CMV− = 0, CMV+ = 1).

Discussion

Most studies to-date studying aging and the immune system in relation to mortality are restricted by 1) their cross-sectional nature, 2) study of a small number of individuals, and 3) lack of relevant clinical data and sufficient follow-up time. Our study is limited with respect to its cross-sectional nature of the baseline immune measurements, though the size and complexity of the VACS-BC and the adjudicated longitudinal follow-up make this study one of the first to analyze adaptive cellular

immune cell subsets at this scale, with the added contrast of HIV and CMV infection, and carefully adjudicated clinical outcomes. Despite these limitations, we successfully extracted an age-independent immune aging signature characterized by alterations in the ratio of CD4 effector memory T cells in relation to the rest of the CD4 compartment. This signature was robust across both HIV+ and HIV− individuals, and correlated with the VACS index, a simple and robust measure that can be calculated from widely available clinical data.

Our results demonstrate that simple measurements of the immune system, using widely-available flow cytometry (in comparison to more expensive and less widely available mass cytometry, single-cell RNA sequencing, and similar methods), can predict mortality years in advance. These measurements serve as guideposts along an immune trajectory of CD4 differentiation that is only weakly correlated with age—suggesting that there is clinical utility in studying the immune system directly to identify individuals at high risk of mortality. Our results are concordant with those recently reported by Alpert *et al*., in that the immune aging signature we identified is present and clearly separates groups of adults between the ages of 40 and 60, despite the fact that the VACS-BC cohort is mostly male (95%) in contrast to the approximate balance by sex in the study design of Alpert *et al*. Our pseudotime analyses provide support for a continuous immune landscape that diverges at certain points, and that divergence at these points leads to “hills and valleys” with

increased and decreased risks of mortality. More simply put, these results show that healthy immune aging is a significant, if not the dominant, component of healthy aging even in a highly co-morbid population.

The relationship between CD4 nadir and the composition of the CD4 compartment is interesting, and while the flow cytometry data we present here are cross-sectional we speculate that a higher proportion of T effector memory cells prior to HIV infection could potentially insulate newly-infected PLWH from more rapid CD4 depletion (as suggested by our CMV analyses). It also seems reasonable that as CD4 T cells are progressively depleted during chronic HIV infection that CD4 Tem and similar cells would be dominant in the CD4 compartment as our data seem to suggest (**Figure 26**), or that CD4 Tem and TemRA cells are resistant to HIV infection as described previously in a limited study of elite controllers.[277] This represents one possible explanation as to why individuals who reached low CD4 nadirs historically suffered from immune reconstitution diseases, as restoration of a cytotoxic and cytolytic CD4 T cell population without a compensating CD4 Treg population could lead theoretically lead to immune dysregulation, morbidity, and mortality.

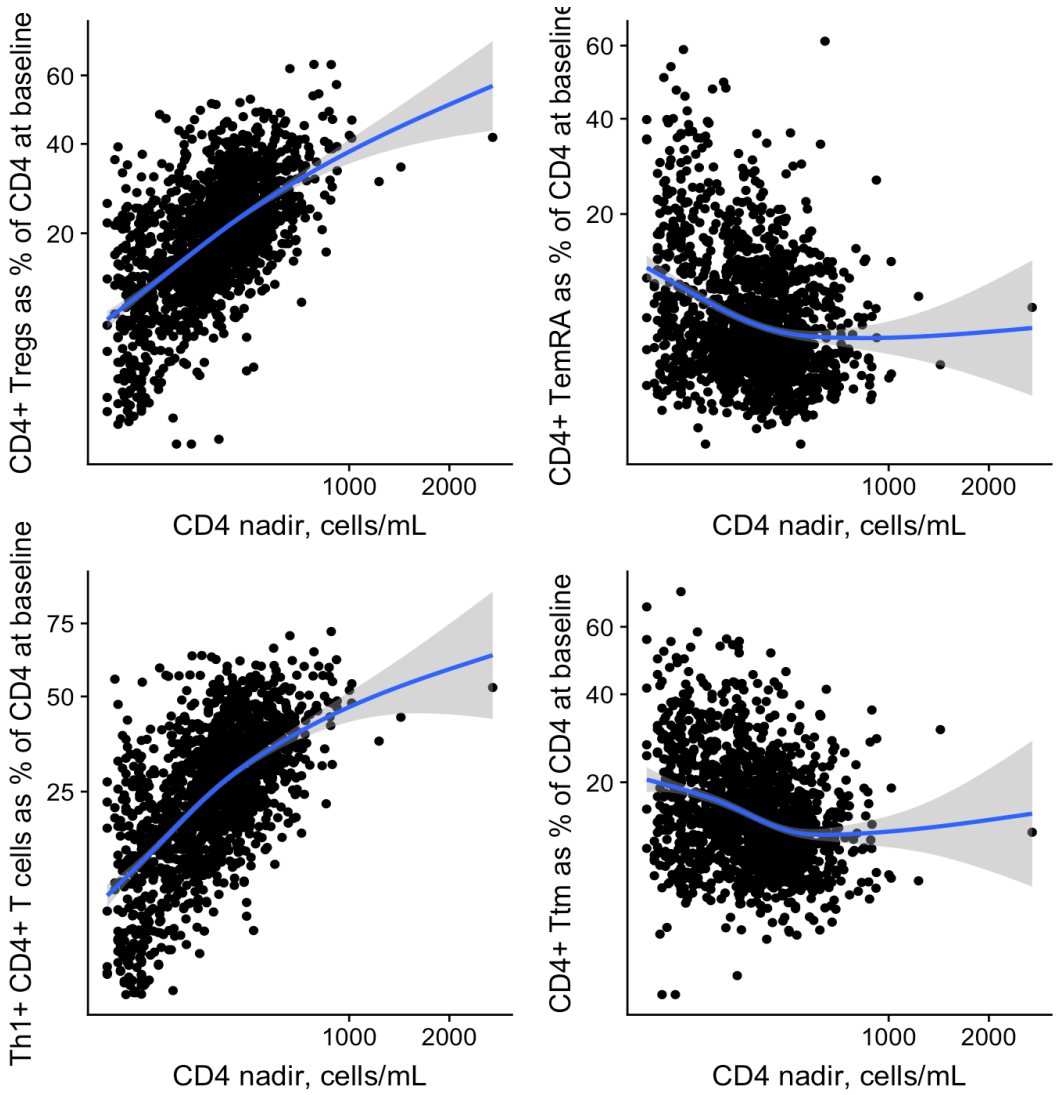


Figure 26. Data in support of a model of concomitant CD4 Tem and TemRA enrichment with progressive depletion of the CD4+ T cell compartment. These data show that key immunomodulatory T cells such as Th1 and Treg subsets are positively correlated with CD4 nadir, while CD4 TemRA and pre-TemRA populations such as T transitional memory (Ttm) are negatively correlated with CD4 nadir. These data suggest a model where individuals suffering from lower CD4 counts and nadirs, and who experience elevated risk of mortality, acquire a CD4 compartment with functional gaps that can only be partially fulfilled by effector CD4+ T cells that are more resistant to HIV infection and therefore depletion by HIV.

While we studied the relationship between all-cause mortality and differentiation of the CD4 T cell compartment as a primary outcome in this study, cause-specific mortality, which are included as part of the VACS-BC, are likely to reveal additional stratifications. Specifically, analysis of cardiovascular mortality and

cancer mortality may reveal differential risk of mortality in the 7 immune consensus groups that we identified—if so, decomposing the risk in these groups to univariate predictors (*e.g*. as with CD4 TemRA proportion as we demonstrate in this work) that are clinically translatable could provide easily deployable biomarkers of subclinical cardiovascular disease and could potentially identify individuals who should be monitored more carefully for development of cancer. There is limited evidence from clinical trials that immunotherapy and reduction in specific CD4 immune populations can improve cardiovascular outcomes, and application of these types of analyses to such trials could reveal mechanisms by which the immune system contributes or does not contribute to specific disease processes.

Chapter III

High CD8 T-cell receptor clonality and altered CDR3 properties are associated with elevated isolevuglandins in adipose tissue during diet-induced obesity

N.B. This work was published in its entirety as McDonnell *et al.* **Diabetes** 2018; 67:2361-2376. doi: 10.2337/db18-0040. In this study, we successfully documented concomitant changes in the T cell receptor repertoire of adipose tissue T cells that accompanied a higher burden of isolevuglandin-bearing macrophages. This work identifies potential therapeutic targets of glucose intolerance.

Abstract

Adipose tissue (AT) CD4+ and CD8+ T cells contribute to obesity-associated insulin resistance. Prior studies identified conserved T-cell receptor (TCR) chain families in obese AT, but the presence and clonal expansion of specific TCR sequences in obesity has not been assessed. We characterized AT and liver CD8+ and CD4+ TCR repertoires of mice fed a low-fat diet (LFD) and high-fat diet (HFD) using deep sequencing of the TCRβ chain to quantify clonal expansion, gene usage, and CDR3 sequence. In AT CD8+ T cells, HFD reduced TCR diversity, increased the prevalence of public TCR clonotypes, and selected for TCR CDR3 regions enriched in positively charged and less polarized amino acids. Although TCR repertoire alone could distinguish between LFD- and HFD-fed mice, these properties of the CDR3 region of AT CD8+ T cells from HFD-fed mice led us to examine the role of negatively charged and nonpolar isolevuglandin (isoLG) adduct-containing antigen-presenting cells within AT. IsoLG-adducted protein species were significantly higher in AT

macrophages of HFD-fed mice; isoLGs were elevated in M2-polarized macrophages, promoting CD8+ T-cell activation. Our findings demonstrate that clonal TCR expansion that favors positively charged CDR3s accompanies HFD-induced obesity, which may be an antigen-driven response to isoLG accumulation in macrophages.

Introduction

The stromal vascular fraction (SVF) of adipose tissue (AT) contains immune cells that contribute to the paracrine signaling milieu that modulates local inflammation and adipocyte function[278]. This has been posited to be the primary mechanism by which inflamed AT induces insulin resistance (IR) because inflammatory cytokines are known to interfere with insulin signaling. AT macrophages (ATMs) are believed to serve as the primary cell type responsible for inflammation-induced IR, with AT T cells serving as modulators of ATM activation. However, additional evidence suggests that adaptive immune responses, mediated by both B cells[279,280] and T cells[281–284], also contribute directly to adverse changes in adipocyte metabolic fitness in obesity. Regulatory T cells (Tregs) are reduced in obese AT compared with lean AT,[283] are protective against AT inflammation and IR, and possess a distinctive gene repertoire.[283] Experimental models using adoptive transfer of CD4+ T cells into lymphocyte-free $Rag^{-/-}$ mice reversed weight gain and IR, and depletion of CD8+ T cells in AT reduced ATM density and improved insulin sensitivity. These

findings suggest that immunotherapy could represent a novel approach to treatment of metabolic disease.[281,285] Although the etiology of T-cell expansion is unclear at present, in vitro studies have found that obese fat independently activates CD8+ T cells and induces proliferation, whereas lean fat has little effect.[281] Spectratyping analyses of AT CD8+ T cells in the setting of obesity also have described a narrowed repertoire of T-cell receptor (TCR) Vα and Vβ chain families compared with cells in lean tissue, suggesting that these CD8+ T cells may undergo oligoclonal expansion in AT.[281,286] An important limitation of spectratyping is that it cannot detect individual TCRs, does not provide complementarity-determining region 3 (CDR3) sequences, and only describes deviations from the normally Gaussian distribution of amino acid or nucleotide length. Spectratyping also was used to demonstrate that proinflammatory CD4+ T helper 1 (TH1) cells with a biased TCR Vα repertoire expand in AT, but CD4+ TH17 cells do not.[284] Furthermore, progressive reductions in CD4+ FoxP3+ anti-inflammatory T cells with increasing obesity were detected.[284] Although spectratyping in these studies allowed the authors to suspect clonal expansion, it did not allow them to quantify TCR repertoire diversity, detect specific clones, examine CDR3 sequences and properties, or examine clonal overlap. Similarly, although analysis of CDR3α sequences from 98 single AT Tregs showed that AT Tregs have a unique V gene repertoire,[283] the diversity observed and its relation to the native repertoire were

less clear because the transgenic mice in this study only carried a single Vα gene, were designed to produce a distinctly narrowed TCR repertoire, and had a lower-than-normal thymic output.[287] Because prior studies demonstrated that obese mice have increased AT T-cell density and a shift toward a proinflammatory phenotype, we hypothesized that high-fat feeding would induce an increase in AT CD4+ and CD8+ T cells in conjunction with the clonal expansion of specific TCR clonotypes. We performed high-throughput TCR DNA sequencing of the AT and livers of mice fed either a high-fat diet (HFD) or a low-fat diet (LFD). The TCR repertoires of mice fed an HFD are markedly enriched for public clonotypes in the CD8 and CD4 TCR repertoires of the AT. This overlap in TCRs between HFD samples distinguished between lean and obese mice, suggesting selection both for more public clonotypes and for autoreactive T cells. We find that the CD8+ TCR repertoire of mice fed an HFD is more clonal and characterized by more charged and less polar CDR3s. These HFD-induced features of the T-cell response led us to examine possible sources of neoantigens within AT, specifically the immunogenic and well-characterized isolevuglandin (isoLG) protein adducts, which are the sole lipid neoantigen species that induce a T-cell response in hypertension.[288] We detected an increase in immunogenic isoLGs, a family of negatively charged protein adducts generated from reactive gamma-ketoaldehydes, in CD206-expressing macrophages isolated from AT of HFD-fed mice. Moreover, isoLGs were elevated in M2-polarized

macrophages in vitro, and cocultures of these macrophages with T cells promoted CD8+ T-cell activation. These observations demonstrate that dietary modifications affect the adaptive immune response in AT in a systemic fashion and provide novel insights into potential mechanisms by which HFD and obesity may lead to pathogenic T-cell responses directed toward modified self-peptides.

Methods – Wet lab

Mice and diets

Male C57BL/6J mice were purchased from The Jackson Laboratory. Studies were divided into three cohorts. The first cohort was used for AT T-cell studies and TCR repertoire analysis, the second was used for isoLGs, and the third was used for measuring isoLGs in ATM subsets and coculture studies. At 8 weeks of age, mice were placed on a 10% LFD for 9 weeks. Subsequently, mice were randomized into either the LFD or the HFD group. Both diets were purchased from Research Diets (HFD: D12492, LFD: D12450B; New Brunswick, NJ). Mice were fed ad libitum and given free access to water. All animal procedures were performed with approval from the institutional animal care and use committee of Vanderbilt University.

Glucose tolerance testing

Mice were fasted for 5 h, and basal blood glucose levels were measured (0 min) before intraperitoneal administration of 1.5 g dextrose/kg lean body mass. Blood glucose was assessed at 15, 30, 45, 60, 90, and 150 min after injection.

AT SVF isolation

The SVF was isolated from epididymal fat pads through collagenase digestion and differential centrifugation as previously described.[289]

Liver nonparenchymal cell isolation

Liver was excised and minced in 1 mg/mL collagenase in PBS. Minced liver was incubated on a shaker at 37°C for 30 min. The cell suspension was filtered and spun at 800g for 10 min at 4°C. The cell pellet was suspended in 33% Percoll. The Percoll gradient was centrifuged at 800g for 30 min at room temperature. Red blood cells were lysed using ammonium-chloride-potassium lysing buffer. Cells were centrifuged at 800g for 5 min at 4°C. The cells were suspended in FACS buffer and stained for flow cytometry. The following primary fluorophore-conjugated antibodies, along with isotype controls, were used to characterize AT and liver T-cell populations: PerCP-Cy5.5–conjugated anti-mouse CD45, antigen-presenting cell (APC)–conjugated anti-mouse TCRβ, Alexa Fluor 700–conjugated anti-mouse CD4,

and V500-conjugated anti-mouse CD8a (all from eBioscience). DAPI was added immediately before analysis to enable discrimination between live and dead cells.

FACS analysis

FACS was performed on a BD FACSAria III flow cytometer (BD Biosciences) at the Vanderbilt Medical Center Flow Cytometry Core shared resource, and data were analyzed using FlowJo software (Tree Star). The gating strategy is shown in **Figure 27**.

TCR sequencing

Genomic DNA from sorted CD4+ and CD8+ T cells was isolated using the QIAamp DNA Blood Mini Kit (QIAGEN). DNA from AT CD4+ and CD8+ T cells and liver CD8+ T cells was used for bulk TCRb CDR3 region amplification and sequencing using the ImmunoSEQ assay (Adaptive Biotechnologies, Seattle, WA). In this method, bias-controlled V and J gene primers are used to amplify rearranged V(D)J segments for sequencing.[290]

IsoLG adduct quantification

ATMs were analyzed by flow cytometry using the following antibodies: PerCP-Cy5.5–conjugated anti-mouse CD64, PE-Cy7–conjugated anti-mouse CD86, Alexa Fluor

700– conjugated anti-mouse CD11c, PE-conjugated anti-mouse CD206, APC-conjugated anti-mouse F4/80, and PECF594- conjugated anti-mouse CD45 (Becton Dickinson). We used intracellular staining with the single-chain antibody D-11 to detect isoLG protein adducts. The D-11 single-chain variable fragment antibody was labeled with Alexa Fluor 488 using the APEX Alexa Fluor 488 Antibody Labeling Kit (Invitrogen). Cells labeled with surface antibodies were fixed and permeabilized for intracellular detection of isoLGs using the FIX & PERM Cell Fixation and Permeabilization Kit (Invitrogen). Dead cells were excluded from analysis using a LIVE/DEAD Fixable Dead Cell Stain Kit (Invitrogen). For each experiment, we gated on single live cells and used fluorescence-minus-one controls for each fluorophore to establish the gates. Data were analyzed using FlowJo software. The gating strategy is shown in **Figure 27**.

Bone Marrow–Derived Macrophage Polarization and T-Cell Coculture

Bone marrow-derived macrophages (BMDMs) isolated from C57BL/6J mice were obtained as described by Trouplin *et al*.[291] On day 6, fully differentiated BMDMs were split and loaded into 24-well plates at a density of 700,000 cells per well in L929-conditioned media. Cells were allowed to adhere overnight and then polarized as follows: For M1 polarization, BMDMs were stimulated for 24 h with interferon-g (100 ng/mL; R&D Systems) and lipopolysaccharide (10 ng/mL; Sigma). For M2

polarization, BMDMs were treated for 96 h with interleukin 4 (10 ng/mL; R&D Systems) and interleukin 13 (10 ng/mL; R&D Systems). For metabolic polarization (metabolically activated macrophages [MMes]), BMDMs were treated for 24 h with 30 mmol/L glucose, 10 nmol/L insulin, and 0.4 mmol/L palmitic acid as described by Hill *et al*.[292] For *tert*-butyl hydroperoxide (TBHP) treatment, BMDMs were treated with 1 mmol/L TBHP for 30 min, after which the TBHP media were replaced with fresh media. T cells were isolated from spleens of LFD- and HFD-fed mice using APC magnetic beads (Miltenyi) to isolate TCRb-APC–labeled cells. Polarized BMDMs were cocultured with isolated pan-T cells from LFD- and HFD-fed mice at a ratio of 1:2. BMDMs and T cells were collected. Immunostaining for flow cytometry was performed using the following antibodies: for macrophage panel, PE-conjugated anti-mouse CD45, APC-conjugated anti-mouse F4/80, and intracellular staining with the single-chain antibody D-11 to detect isoLG protein adducts, and for activated T-cell panel, PerCP-Cy5.5–conjugated anti-mouse CD45, fluorescein isothiocyanate–conjugated anti-mouse CD4, PECF594-conjugated anti-mouse CD69, and PE-Cy7–conjugated anti-mouse CD8 (BD Biosciences) and DAPI. Gating strategies for BMDMs and activated T cells are shown in **Figures 27, 38, 39, and 40**.

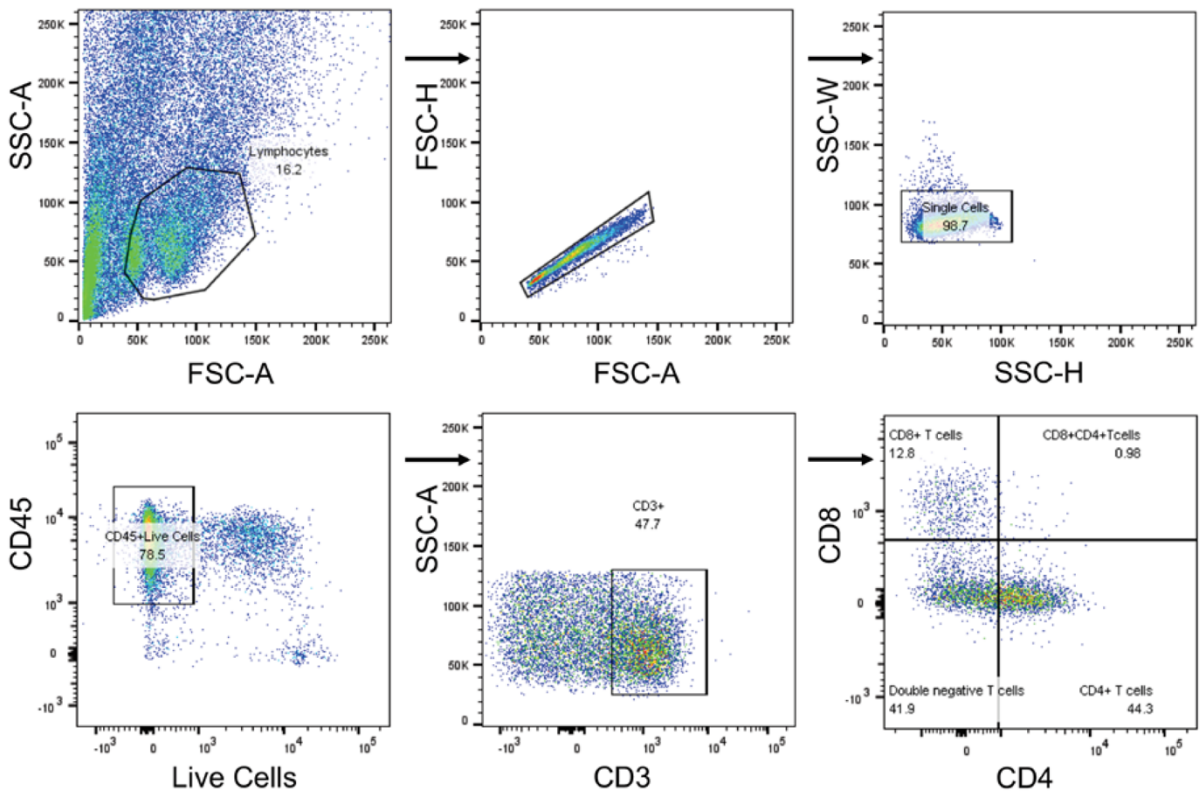


Figure 27. Gating strategy for AT T cells.

Methods – Dry lab

Data import and preprocessing

All TCR sequences, including nonproductive sequences, from immunoSEQ were processed and imported using VDJTools as previously described.[293] Amino acid sequences were analyzed because the amino acid sequence of a TCR determines its structural properties.

Repertoire similarity

VDJTools was used to calculate a pairwise distance matrix between samples using the geometric mean overlap $F_{ij} = \sqrt{f_{ij} f_{ji}}$, such that $f_{ij} = \sum_{k=1}^{N} \emptyset_{ik}$ sums to the frequencies of TCRs shared between two samples present in the first as the distance metric. This matrix was used to perform complete linkage hierarchical clustering through the hclust function in base R version 3.4.0 software. Permutation testing confirmed significant factors in clustering. Each sample was downsampled in a density-dependent fashion to 10,000 clonal sequences to confirm significant factors in clustering and significant structure. Mice with < 1,000 clonotypes detected in any sample were excluded from analysis.

Repertoire overlap

The VDJTools JoinSamples routine was used to detect clonotypes overlapping between samples with shared amino acid sequences.

Clonal homeostasis

The frequency of each clonotype within each sample was calculated using VDJTools. The clonal space, defined by the sum of frequencies, occupied by the top 1–3, 4–10, 11–20, 21–50, 51–100, and bottom 101–n clonotypes was calculated per sample as described previously.[294,295] These clonal bins were log-transformed and tested for

differences in the ratio of their geometric means between LFD and HFD. The CD4 and CD8 clonal bins were then condensed per mouse and used as input for hierarchical clustering. Average linkage was used as the clustering method, and absolute correlation was used as the distance metric. For visualization, the geometric means for each set of samples was calculated per clonal bin and normalized to range from 0 to 1.

t-Distributed Stochastic Neighbor Embedding

Clonal homeostatic proportions were fed into the t-distributed stochastic neighbor embedding (t-SNE) algorithm with principal component analysis (PCA) before projection. Perplexity was set at 2, θ at 10^{-5} , learning rate at 1, and perplexity was exaggerated for the first 50 iterations to restrain the *t*-SNE algorithm in a size-dependent fashion. Five thousand iterations were allowed. Calculations were performed using the Rtsne package.

PCA

Homeostatic proportions were centered and scaled before projection. Calculations were performed using the prcomp function in base R.

P value-validated clustering

Clonal homeostatic proportions were clustered using complete linkage for clustering and absolute correlation for distance. Calculations were performed using the pvclust R package.[296]

Multidimensional scaling

The pairwise distance matrix using geometric mean overlap was used as input for nonmetric multidimensional scaling using VDJTools[293] and the isoMDS function from the MASS R package (20).

Repertoire publicity analysis

Public clonotypes were defined as those shared among at least three mice to account for sampling depth. All public clonotypes were retrieved using VDJTools,[293] and the frequency within each mouse was converted to a binary value representing present (1) or absent (0). For each public clonotype, each dietary group was then assigned a binary value representing whether it was greater than (1), less than (0), or equally prevalent to (0) the other dietary group. These values were then used to construct a 2×2 contingency table representing the proportion of public TCRs for each dietary group, in similar fashion to analysis of concentration and association.[297,298]

Sequence logo plots

Seq2Logo was used to generate three types of sequence logo plots. The median TCRβ length was calculated, and these sequences were used to generate logo plots using Shannon entropy (the probability of an amino acid appearing), the Kullback-Leibler divergence (probability with enrichment and depletion), and the probability-weighted Kullback-Leibler divergence (probability of each amino acid multiplied by its weight). Where indicated, Hobohm's first algorithm (nearest-neighbor selection clustering) was used to bin TCRs with similar sequences, using a threshold of 20% to account for the diversity and divergence of the TCRβ chain. A weighted prior correction of 200 pseudocounts on the basis of the BLOSUM62 matrix was used to correct for amino acids detected at very low levels as previously described.[299] Where appropriate, each mouse's TCR repertoire also was downsampled in density-dependent fashion to 10,000 clonal sequences to draw comparisons between samples and to cross-validate the multidimensional structure and similarity of data.

Statistics

Significance was set *a priori* at $P < 0.05$.

Weight and clonality

Repeated-measures ANOVA with Bonferroni multiple comparisons testing was used to assess changes between diet groups using GraphPad Prism version 6.05 software (GraphPad Software, La Jolla, CA) and R version 3.4.0 (RStudio version 1.0.143).

Glucose tolerance testing

Two-way ANOVA with post-hoc Bonferroni-Šídák multiple comparisons testing was used to assess differences between dietary groups at each time point.

Clonal homeostasis

The Mann-Whitney *U* test was used to assess differences in the geometric mean ratios of clonal homeostasis bins using GraphPad Prism software.

Factor analysis in clustering

Significant factors in hierarchical clustering were assessed using permutation testing in VDJTools for repertoire similarity.

Cluster analysis in clonal homeostasis

Cluster significance for clonal homeostatic signatures was assessed using multistep-multiscale bootstrap resampling in the pvclust R package.[296]

CDR3 amino acid and nucleotide profiling

CDR3 properties were calculated per sample and per dietary group using VDJTools.[293] The Mann-Whitney *U* and Wilcoxon signed rank tests were used to assess differences in CDR3 properties using GraphPad Prism and R software. *P* values were corrected using the Benjamini-Hochberg false discovery rate correction.

Repertoire publicity

We tested the 2×2 contingency table computed as described above using Pearson's χ^2 test. The *P* value for this test was calculating using Monte Carlo simulation with 1,000 bootstraps in R. *P* values were adjusted using the Bonferroni correction to control the family-wise error rate.

Results

High-fat feeding reduces CD8+ T-cell diversity in AT

An HFD increased body weight (**Figure 28A**), AT mass (**Figure 28B**), and liver mass **Figure 28C**). Although fasting blood glucose was similar among diet groups (**Figure 28D**), plasma insulin was significantly higher in the HFD-fed mice (**Figure 28E**), and HFD-fed mice had impaired glucose tolerance (**Figure 28F**).

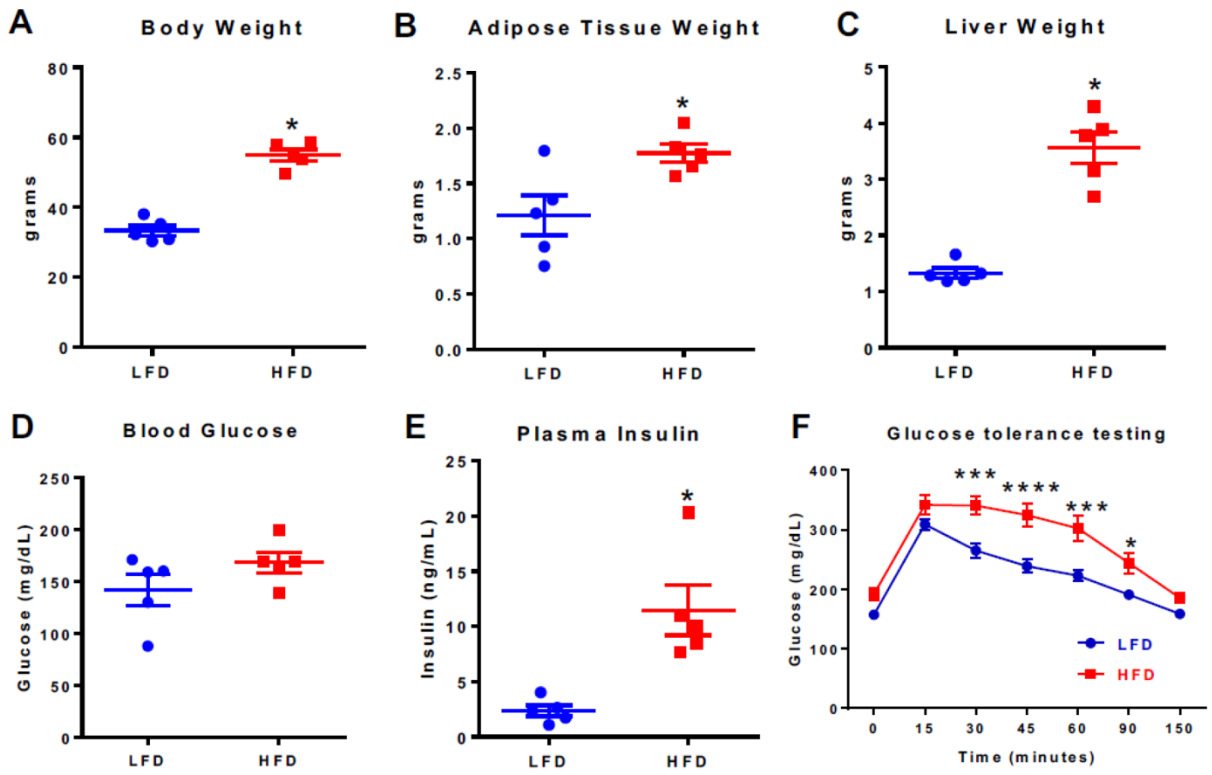


Figure 28. Dietary and metabolic changes associated with LFD and HFD. Male C57BL/6J mice were placed on a 10% low-fat diet (LFD) or a 60% high-fat diet (HFD) as described in the Methods. (A) Mice on HFD demonstrated a significant increase in body weight, (B) epididymal AT weight, and (C) liver weight (one-way ANOVA with Bonferroni correction). (D) Fasting blood glucose did not differ between the groups (one-way ANOVA with Bonferroni correction). (E) Plasma insulin concentrations were elevated in the HFD mice (one-way ANOVA with Bonferroni correction). (F) Glucose tolerance testing revealed that the glucose tolerance of HFD-fed mice was significantly impaired (two-way ANOVA with post-hoc Bonferroni-Šídák multiple comparisons testing). Data are presented as mean ± SEM, n = 4-5 mice/group.

We next compared the T-cell density (cells/gram tissue) in AT and liver. Similar to previous studies,[284] the density of AT CD8+ T cells was elevated by threefold in HFD-fed mice compared with LFD-fed controls (**Figure 29A**). In contrast, AT CD4+ T-cell and liver CD8+ T-cell populations were not significantly different between groups (**Figure 29B, Figure 29C**).

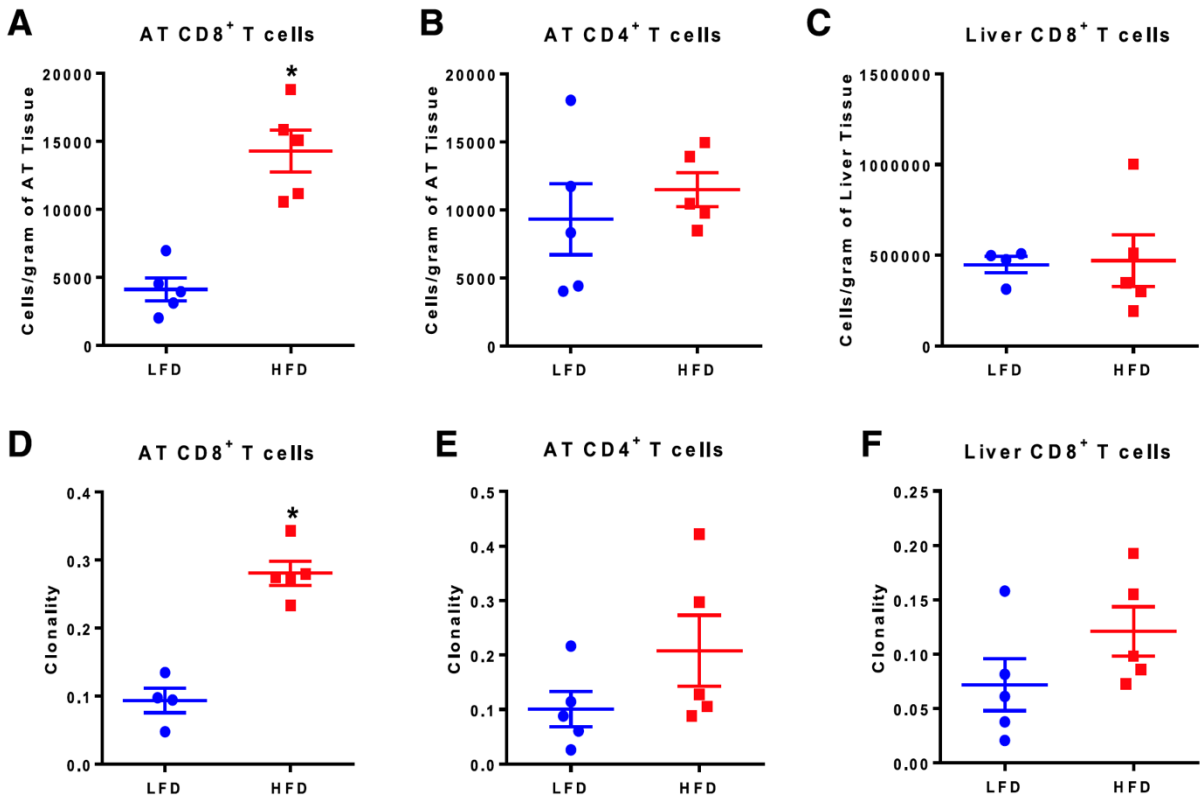


Figure 29. HFD leads to increased T-cell number and clonality. A and B: AT CD8+ number was increased in HFD mice (two-tailed t test) (A), whereas CD4+ T cells were not significantly increased (B). C: There was no difference in the number of liver CD8+ T cells. D: The TCR repertoires of AT CD8+ T cells of HFD-fed mice are significantly more clonal than those of LFD-fed mice (P,0.05, two-tailed Student t test). E: AT CD4+ clonality did not differ between HFD and LFD (P = 0.18, two-tailed Student t test). F: Liver CD8+ T cells of HFD-fed mice are not significantly clonal compared with LFD-fed mice (P = 0.18, two-tailed Student t test). Data points represent the clonality of each mouse as calculated using normalized Shannon entropy. Data are mean 6 SEM (n = 4–5 mice/group). *P ≤ 0.05.

TCRβ chain amplification and deep sequencing of AT CD8+ and CD4+ T cells as well as liver CD8+ T cells yielded three repertoires of V(D)J gene TCR sequences per mouse. These TCR sequences were used to calculate overall repertoire clonality as measured by Shannon entropy using the immunoSEQ Analyzer. AT CD8+ T cells from HFD-fed mice had a higher clonality score, indicating less diversity, compared with LFD-fed mice (**Figure 29D**). Although there was somewhat increased clonality in AT CD4+ T cells (**Figure 29E**) and liver CD8+ T cells (**Figure 29F**), this difference was not significant; because we did not detect differences in the liver CD4+ populations, we did not sequence these T cells.

T-cell clonal distributions differ between mice fed an HFD or LFD

Our observation of the marked clonality in the CD8+ T cells from AT of HFD-fed mice led us to ask whether diet-induced obesity (DIO) affected the relative contributions of the major and minor clonotypes within a given TCR repertoire. We first examined this clonal homeostatic space within each mouse using the Mann-Whitney *U* test after downsampling to account for sampling depth. The top clonotypes of mice fed an HFD occupied significantly higher proportions of clonal space ($P < 0.001$) (**Figure 30A**), whereas the bottom clonotypes occupied significantly less clonal space ($P < 0.001$) in CD8+ T cells from AT and liver. The proportions of these TCRs from the AT and liver also distinguish between lean and

obese mice on the basis of *t*-SNE (**Figure 30B**), PCA (**Figure 30C**), and hierarchical clustering ($P < 0.05$) (**Figure 30D**). These multidimensional analyses demonstrate that HFD distorts the size of the largest and smallest clonotypes of the repertoire (**Figure 30, see next page**).

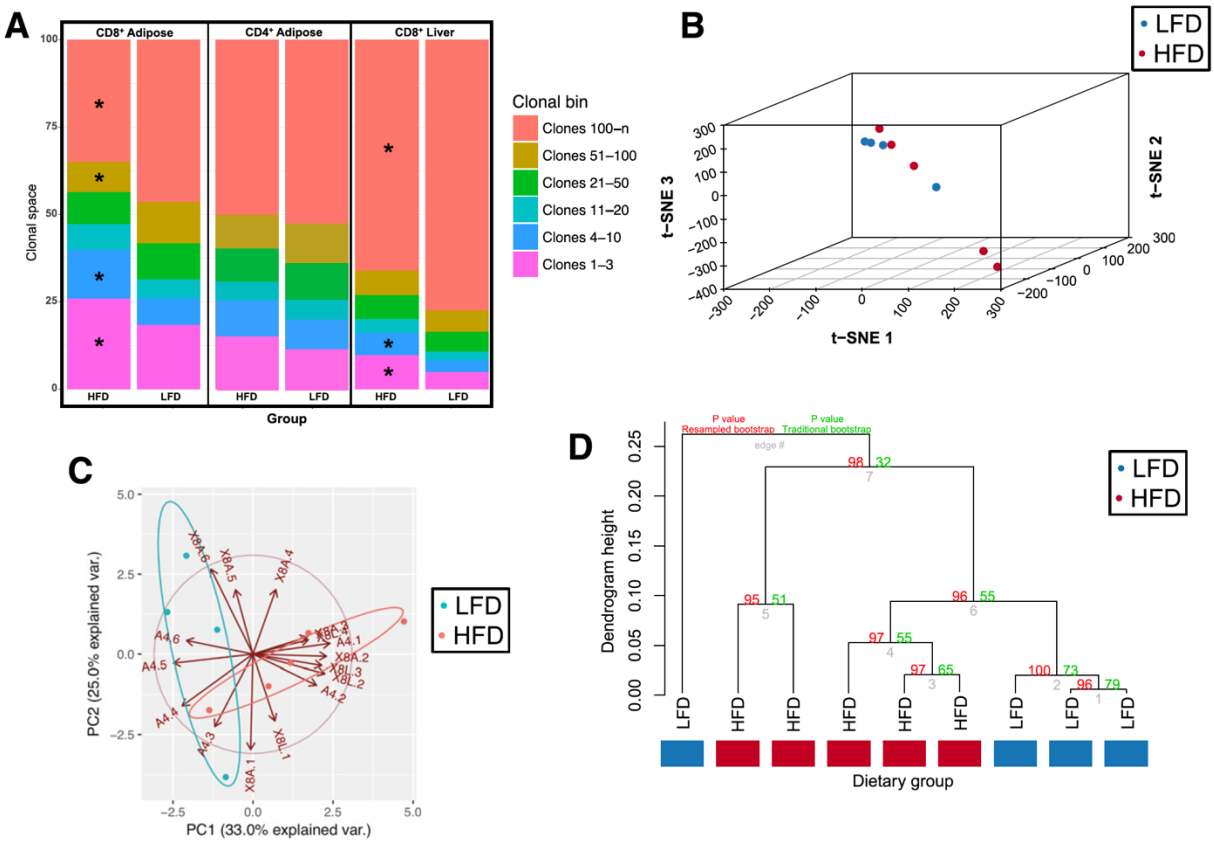


Figure 30. Clonal homeostatic space differs between dietary groups and is a distinctive signature in multidimensional space. A: The space within the repertoire occupied by the clonotypes within each clonal bin was calculated for eachmouse. The geometric mean for each of these proportions was calculated for each diet group, and the clonal homeostatic proportions occupied were tested using the two-tailed Mann-Whitney U test. The clonal proportions distinguish between dietary conditions using modern machine learning methods measuring several multidimensional distances. B: Two-dimensional t-SNE projection of clonal homeostatic proportions from the AT reveals that lean and obese mice differ in clonal homeostasis as measured in a two-dimensional projection. PCA with centering and scaling was performed before t-SNE, and 15 dimensions were retained for embedding with the t-SNE algorithm. C: PCA projection of clonal homeostatic proportions reveals that the repertoire space occupied by both highly and lowly abundant clonotypes distinguishes between lean and obese mice. D: Hierarchical clustering using average linkage and absolute correlation reveals that clonal homeostatic proportions distinguish between lean and obese mice. Values shown at each node demonstrate the confidence that a cluster truly exists as measured using multiscale bootstrap resampling and traditional bootstrapping. Multiscale bootstrap resampling validates that these characteristic differences are not expected to distinguish between lean and obese mice by chance alone. *P ≤ 0.05. PC, principal component; var., variance.

Shared clonotypes and gene usage distinguish between TCR repertoires from mice fed LFD or HFD

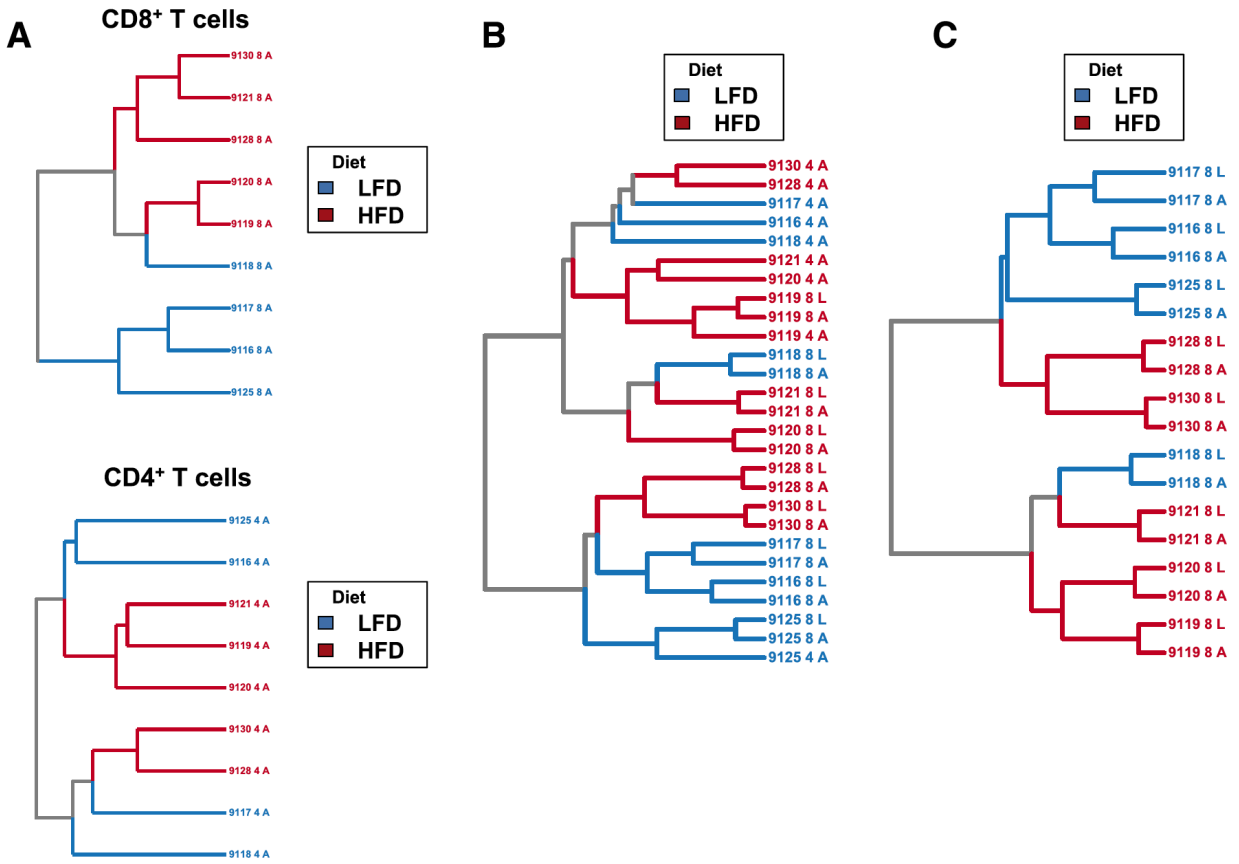


Figure 31. Dietary group is predictive of repertoire similarity and synchronously alters the TCR repertoire in AT. Geometric mean overlap was used to calculate pairwise distance matrices on the basis of shared clonotypes with identical CDR3 amino acid sequences. Labels on each dendrogram correspond to a specific mouse, tissue type, and cell type (4 or 8). A: AT CD8+ (upper) and CD4+ repertoires (lower) share more TCRs within diet groups than between diet groups ($P < 0.0001$ for CD8+ and CD4+, permutation of pairwise distance matrix). This suggests that sets of shared TCRs are enriched and depleted in a diet-dependent fashion. B: Diet significantly affects repertoire similarity across multiple cell types and within multiple tissues ($P = 0.01$, permutation of pairwise distance matrix). TCRs shared between mice within multiple tissue and T-cell types distinguish between lean and obese mice. Diet-dependent repertoire similarity is observed in liver in addition to AT. C: CD8+ T-cell repertoires from liver and AT reflect simultaneous clonal expansion and sharing within each mouse and show that diet significantly affects repertoire similarity within the CD8+ compartment ($P < 0.05$, permutation of pairwise distance matrix). A, AT; L, liver.

Hierarchical clustering on repertoire similarity of TCR clonotypes sharing the same amino acid sequence successfully distinguished between both AT CD8+ TCR ($P < 0.05$) and AT CD4+ TCR repertoires of LFD- and HFD-fed mice ($P < 0.001$) (**Figure 31A, see previous page**). The impact of dietary group on repertoire similarity held true when pooling TCR repertoires from AT and liver and from both CD8+ and CD4+ T cells ($P = 0.01$) (**Figure 31B**). Expansion of both major and minor clonotypes led to synchronous changes in CD8+ T cells between AT and liver repertoires within each mouse (**Figure 31C**) in a diet-dependent fashion ($P < 0.05$). Each mouse's AT and liver repertoires were most similar to each other, demonstrating the reproducibility of repertoire changes that are tissue specific.

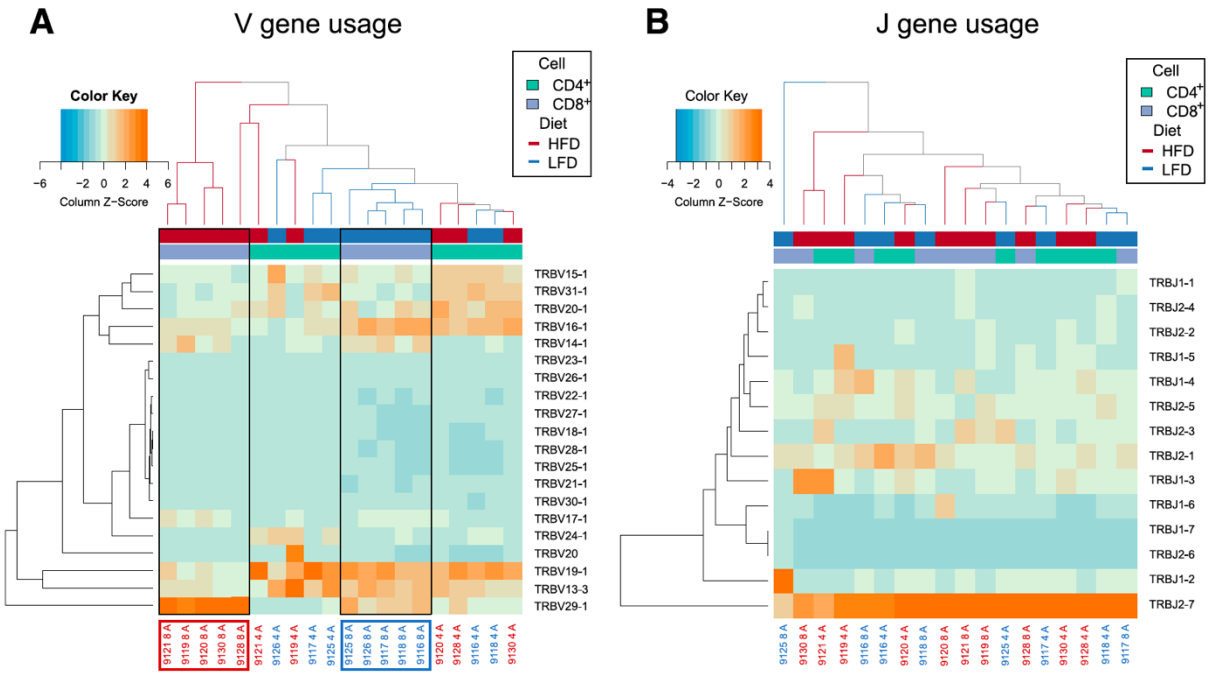


Figure 32. CD8+ TCR repertoires within AT are genetically distinct by dietary group. Z-score–normalized heat maps were built using VDJTools. Labels underneath the heat maps correspond to individual mice, and rows correspond to a given Vb (TRBV) gene. The heat map is colored on the basis of Z-score and how many SDs from the mean a given sample is in its expression of a given Vb gene family. A: Obesity leads AT CD8+ T cells to use distinct sets of V gene segments. These changes do not appear to distinguish between CD4+ T cells of lean and obese mice, although CD4+ T cells appear to use different sets of Vb genes than CD8+ T cells regardless of dietary condition. B: Obesity does not affect the J segments used within the CD8+ TCR repertoires of mice to the same extent that V segments do. This is unsurprising because there are fewer Jb (TRBJ) genes than Vb genes and because the Jb region frequently accounts for less nucleotide diversity than the Vb region.

Our observation of the dissimilarity of the TCR repertoires between dietary groups led us to examine whether Vβ gene usage differed in AT between LFD- and HFD-fed mice. The AT CD8+ T-cell repertoires of each dietary group demonstrated differential usage of several Vβ genes (**Figure 33A**). Neither LFD nor HFD affected Jβ gene usage in the AT (**Figure 33B**). Similar to other groups, we observed differential Vβ gene usage between CD4+ and CD8+ T cells,[300,301] which also distinguished

between CD4+ and CD8+ TCR repertoires (**Figure 32, see previous page**).

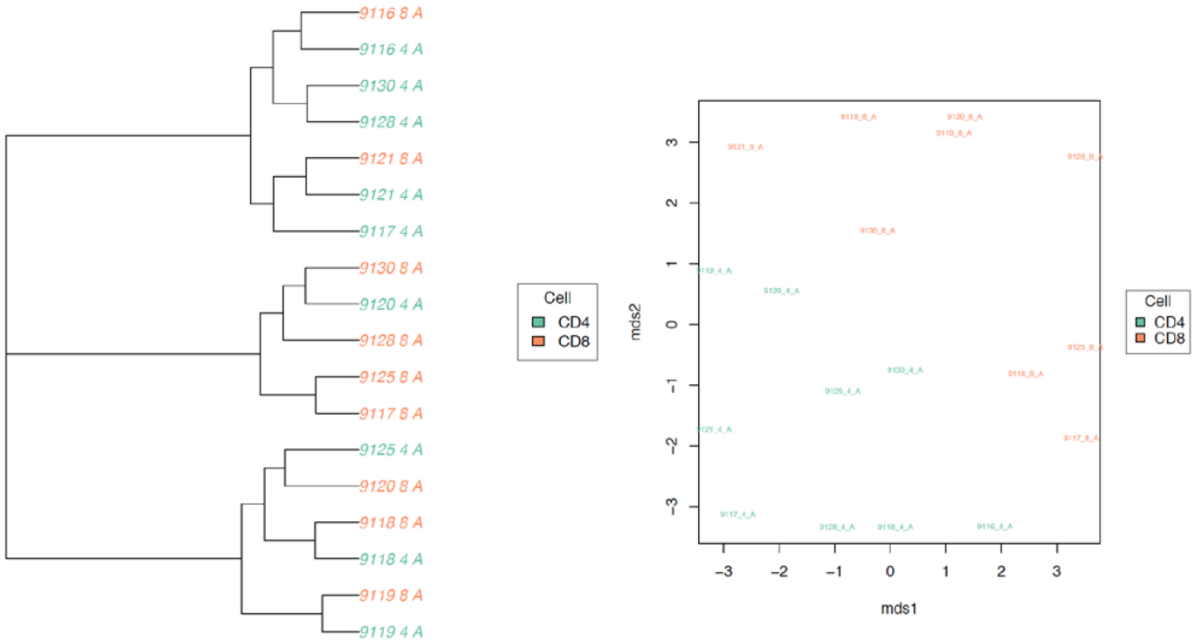


Figure 33. Differential TRBV usage distinguishes between the CD4+ and CD8+ TCR repertoires. Hierarchical clustering using geometric mean overlap reveals that CD4+ and CD8+ repertoires possess characteristically unique sets of TCRs, and that these repertoires are distinct in multidimensional space.

CD4+ TCRs expressed a decrease in charge and increase in polarity compared with CD8+ TCRs ($P < 0.05$) (**Figure 34A, see next page**). They also had characteristically shorter N-D gene-N regions ($P < 0.05$) (**Figure 34B, see next page**).

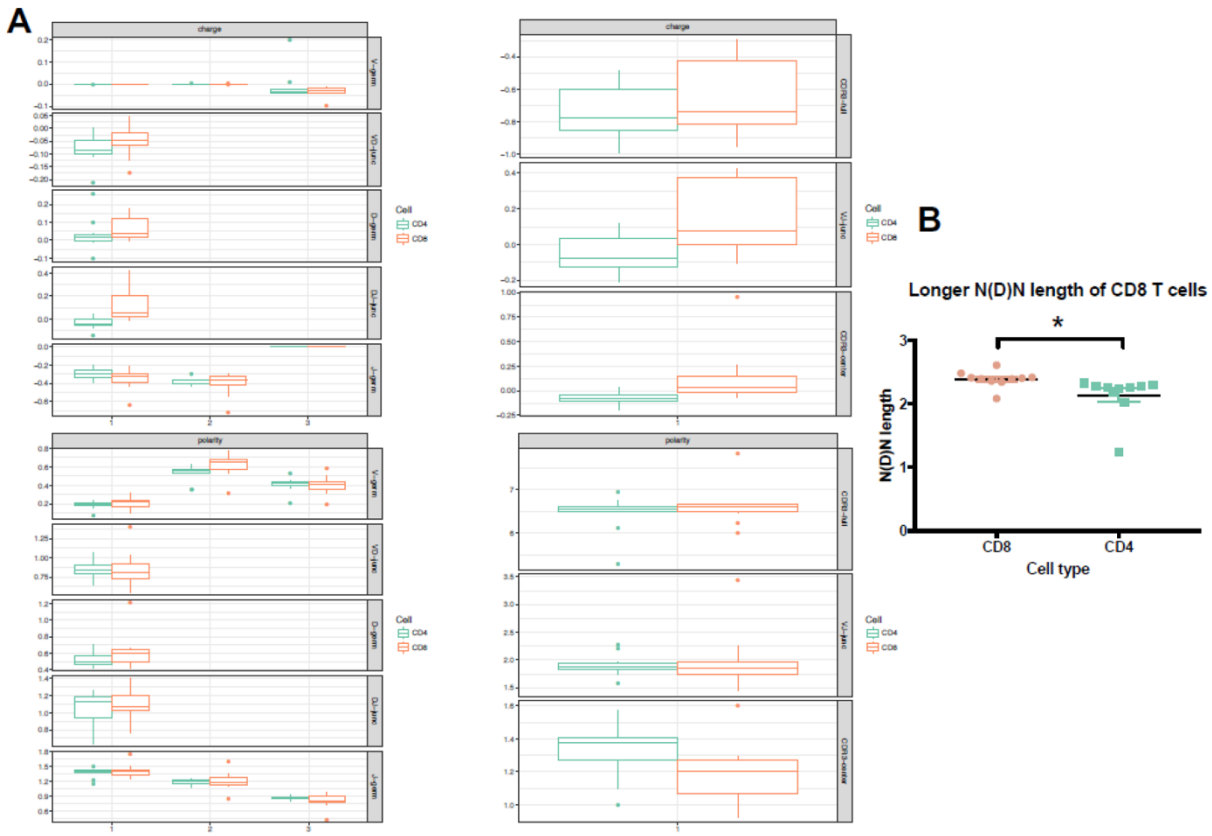


Figure 34. Physicochemical and genetic features distinguishing CD4+ repertoires from CD8+ repertoires in mice. CD4+ TCR repertoires utilize CDR3s with decreased charge and polarity in comparison to CD8+ TCR repertoires ($P < 0.05$, Wilcoxon signed-rank test). N(D)N length of CD8 and CD4 T cells ($P < 0.05$, Wilcoxon signed-rank test).

HFD increases the prevalence of public T-cell clonotypes

The observation that diet strongly influences TCR repertoire similarity led us to ask whether HFD-fed mice were more likely to share clonotypes than LFD-fed mice. To test this hypothesis, we used a specialized Monte Carlo–validated χ^2 test (**Figure 35**). This test proved to be robust to both *P* value permutation and probability proportionate to size downsampling in addition to extremely conservative *P* value correction. HFD mice were more likely to have public clonotypes shared among at

least three mice in their CD8+ ($P < 0.001$) and CD4+ AT T-cell repertoires ($P = 0.01$). This difference was not observed in the liver CD8+ TCR repertoire.

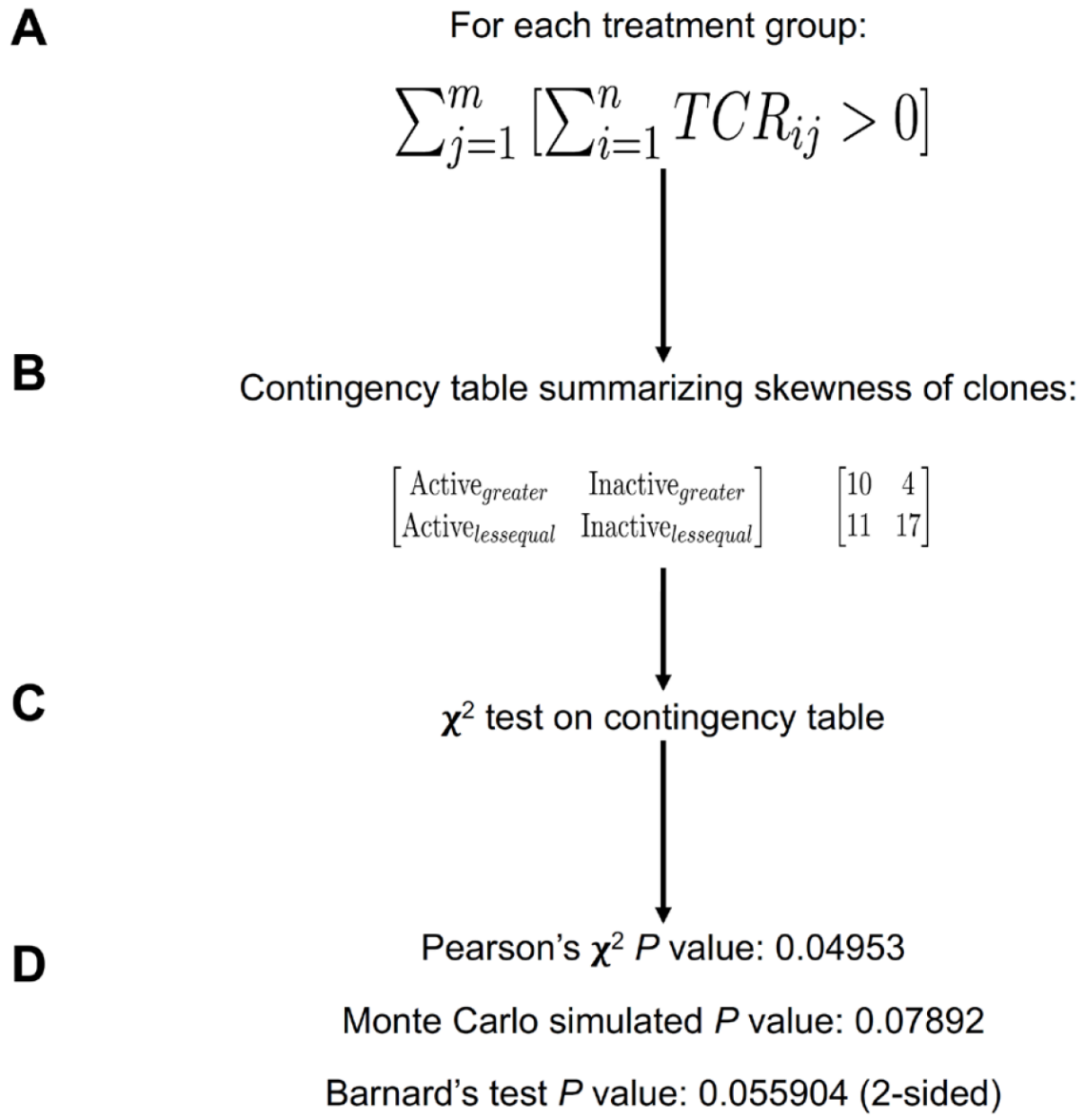


Figure 35. Calculation of repertoire publicity measure. (A) Each TCR is counted as present or absent within a given sample. (B) A contingency table is built by counting whether a given TCR is predominantly present within a treatment group (1), equally present between treatment groups (0), or present within the other treatment group (0). (C) Contingency table testing is performed to evaluate the null hypothesis that public clonotypes are evenly distributed between treatment groups. Either Pearson's χ2 test or the Barnard test are appropriate to implement, though Fisher's exact test is not recommended due to its requirement that marginals of the contingency table remain equal. If Pearson's χ2 test is used, Monte Carlo simulation should be used to ensure that the P-value observed would still be considered significant within the probability null distribution.

Regardless of dietary condition and tissue type, we identified 15,419 TCRs shared between any two samples, 9,034 TCRs shared between any two CD8+ liver samples, 1,126 TCRs shared between any two CD4+ AT samples, and 826 TCRs shared between any two CD8+ AT samples. We also detected previously reported clonotypes that have been associated with obesity and IR in mouse models in both LFD- and HFD-fed mice (**Table 1**).

Table 1—Detection of previously reported T-cell clonotypes associated with pathogenesis in mouse models of obesity and diabetes

CDR3 AA	Reference	HFD Liver CD8	HFD Adipose CD8	HFD Adipose CD4	LFD Liver CD8	LFD Adipose CD8	LFD Adipose CD4	Annotation	*P* value
CASSLGSGNTLYF	26	4	1	1	3	0	0	Expanded in pancreas	NS
CASSDAGQYEQYF	26	1	0	0	2	0	0	Expanded in pancreas	NS
CASSDSAETLYF	26	3	3	0	1	0	0	Expanded in pancreas	*
CASGDEGYEQYF	26	1	0	1	0	0	0	Expanded in pancreas	NS
CASGDAREQYF	26	2	1	0	3	0	0	Expanded in pancreas and B-cell scaffolds	NS
CASGSSYEQYF	26	2	0	1	2	0	0	Expanded in β-cell scaffolds	NS
CASSDAGTTNERLFF	26	1	0	1	0	0	0	Expanded in β-cell scaffolds	NS
CASGSSYEQYF	26	2	0	1	2	0	0	Expanded in β-cell scaffolds	NS
CASSLGTGDEQYF	27	2	1	1	1	0	1	Diabetogenic clone BDC 4.12	NS
CASSGTGGQNTLYF	28	0	0	0	3	0	0	Diabetogenic clone AI15.A3, β-cell autoreactive	*
CASSGTGGQNTLY	28	0	0	0	3	0	0	Diabetogenic clone AI5.F5, β-cell autoreactive	*
CASSLGGYEQYF	29	3	1	3	5	1	2	Proinsulin reactive	NS
CASSRDNTEVF	29	3	0	1	2	0	0	Insulin reactive	NS
CASSLTGNTGQLY	29	2	0	4	1	0	0	Insulin reactive	*
CASSPDNYEQY	29	1	1	1	2	0	0	Insulin reactive	NS
CASSRHQDTQYF	30	1	0	0	2	0	0	Specificity not known	NS
CASSLGGYEQY	31	3	1	2	5	1	2	Reactive against insulin	NS

*$P \leq 0.05$.

Table 1. Detection of previously reported T-cell clonotypes associated with pathogenesis in murine models of obesity and diabetes.

HFD alters physicochemical properties of the TCR repertoire

In addition to being more clonal, the TCR repertoire of mice fed an HFD demonstrated marked physicochemical differences from mice fed an LFD. To assess these differences, we systematically compared the CDR3 of every TCR

detected in LFD- and HFD-fed mice and $CD4^+$ to $CD8^+$ TCRs. Obesity resulted in an elevation in charge ($P = 0.01$) (**Figure 36A**) and a decrease in polarity in $CD8^+$ TCRs from AT ($P < 0.05$) (**Figure 36B**). These repertoire-level changes do not appear to originate from germline selection and instead derive from the junctional regions of the CDR3, where TCR diversity is generated at the nucleotide level (**Figure 36C,D**).

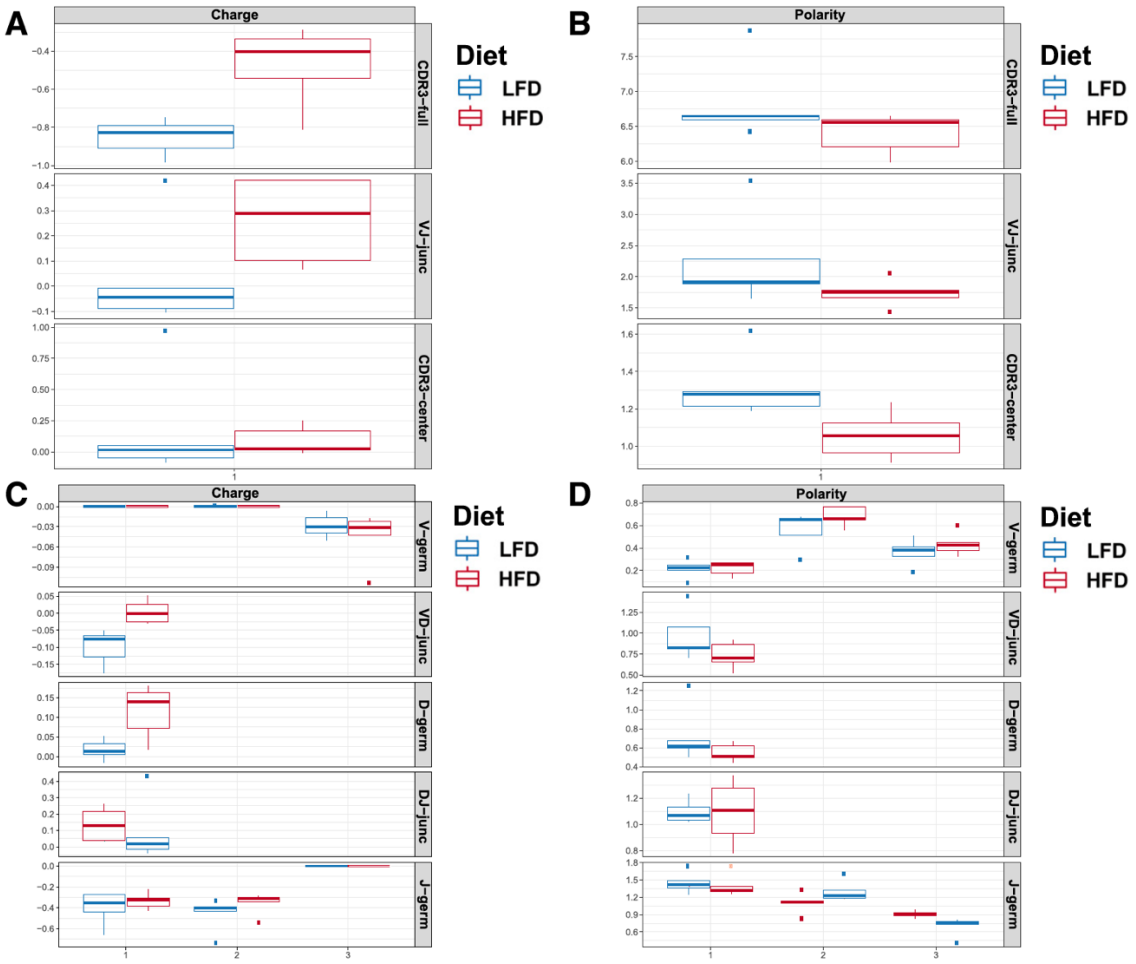


Figure 36. CD8+ TCR repertoires of mice fed an HFD differ in physicochemical properties compared with those of mice fed an LFD. A: AT CD8+T-cell repertoires of obese mice are significantly higher in charge (P = 0.01, Student t test). This elevation in charge is due to amino acids encoded in the junctional regions of the TCR, where diversity is generated. B: The CD8+ T-cell repertoire within AT also is markedly lower in polarity in obese mice (P < 0.01, Student t test). Unlike the changes observed in charge, the difference in polarity appears to arise from germline selection. C: Elevations in charge arise from the junctional regions of the TCR, where TCR diversity is generated at the nucleotide level. Repertoires were assessed using three scanning bins to account for variance along each region and to identify signal arising from each genetic region of the TCR. D: The increased polarity of AT CD8+ TCR repertoires from obese mice appears to be more closely related to changes selected at the germline level, unlike changes in charge. germ, germline; junc, junction.

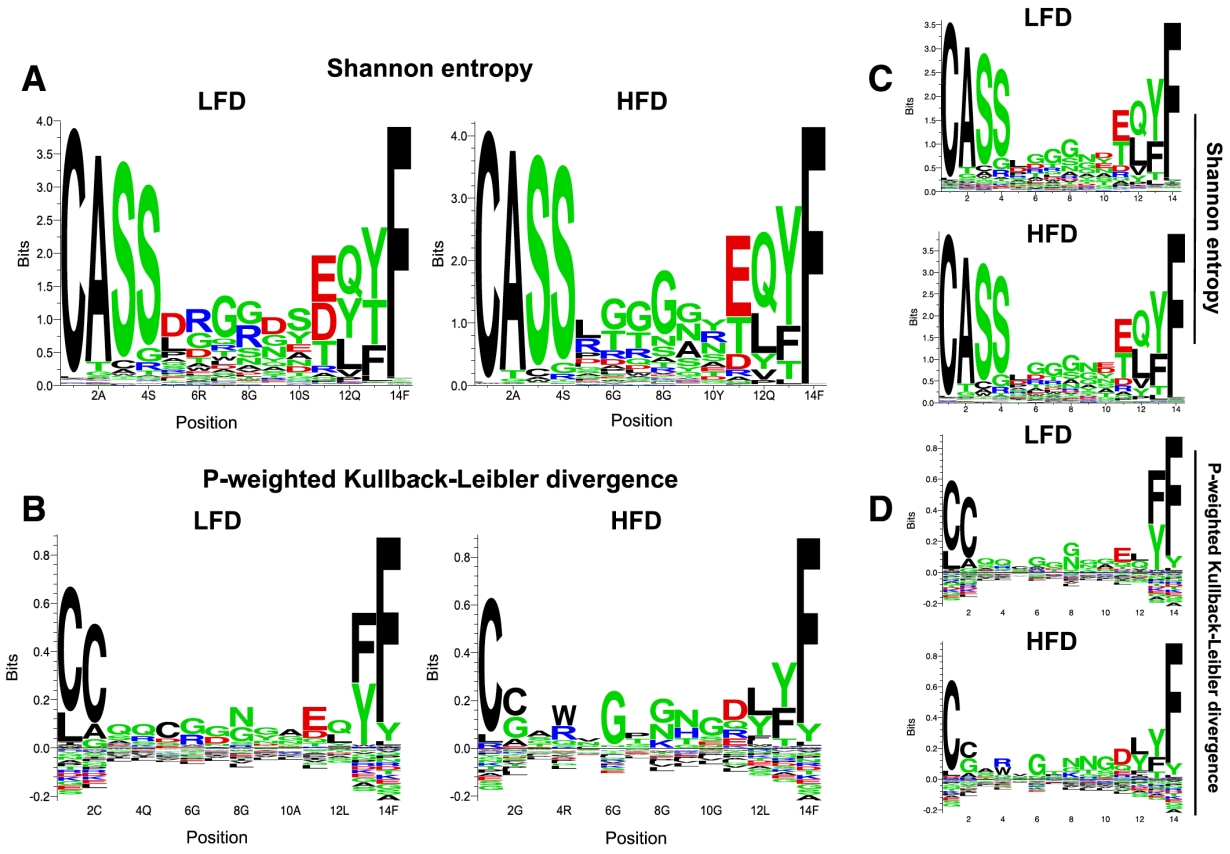


Figure 37. Weighted and unweighted measures of entropy demonstrate selection for positively charged amino acids within the CDR3 of HFD TCR repertoires. The median CDR3 length was 14 amino acids, which we analyzed using several sequence entropy methods. A: Shannon entropy reveals enrichment of arginine and depletion of negatively charged amino acids in the CD8+ repertoires of HFD-fed mice. B: P-weighted Kullback-Leibler divergence with Hobohm1 clustering reveals that other positively charged amino acids are enriched within HFD CD8+ TCR repertoires from AT and that negatively charged amino acids are correspondingly depleted. C and D: Unweighted analyses using Shannon entropy (C) and P-weighted Kullback-Leibler divergence (D) show that these amino acid enrichments are due to clonotypes present at higher frequencies, although there is still enrichment for arginine, lysine, and histidine, all of which carry a positive charge at physiological pH. This indicates that obesity-induced changes in charge and polarity at the repertoire level are driven not only by highly expanded clonotypes but also by clonotypes that are present at lower frequencies.

To examine more closely the differences in charge and polarity, we used sequence entropy measurements to detect enriched and depleted amino acids in the CDR3s of CD8+. Weighted Shannon entropy (**Figure 37A**) and Kullback-Leibler divergence (**Figure 37B**) demonstrated that these changes are due to an enrichment of

arginine and depletion of negatively charged amino acids, with other positively charged amino acids selected for at positions 8, 9, and 11. Unweighted Shannon entropy (**Figure 37C**) and Kullback-Leibler divergence (**Figure 37D**), where each TCR is given equal weight, revealed that both lysine and histidine contribute strongly to the increase in charge observed.

HFD leads to isoLG accumulation in ATMs

The physicochemical alterations in the CD8+ TCR repertoires of HFD-fed mice led us to consider how the CDR3 of the TCRβ chain could interact with the local biochemical environment of AT. Because we observed an increased charge and decreased polarity, we chose to examine the isoLG content of APCs within AT because these highly reactive g-ketoaldehydes carry both negatively charged phospholipid content and nonpolar lipid content (**Figure 38A**). IsoLGs have been shown to rapidly adduct to lysines on self-proteins, forming isoLG protein adducts that are processed and presented as neoantigens, and induce an autoimmune-like state in hypertension.[288] We isolated total leukocytes from AT and performed flow cytometry to identify isoLG protein adducts in ATMs (**Figure 38A**).

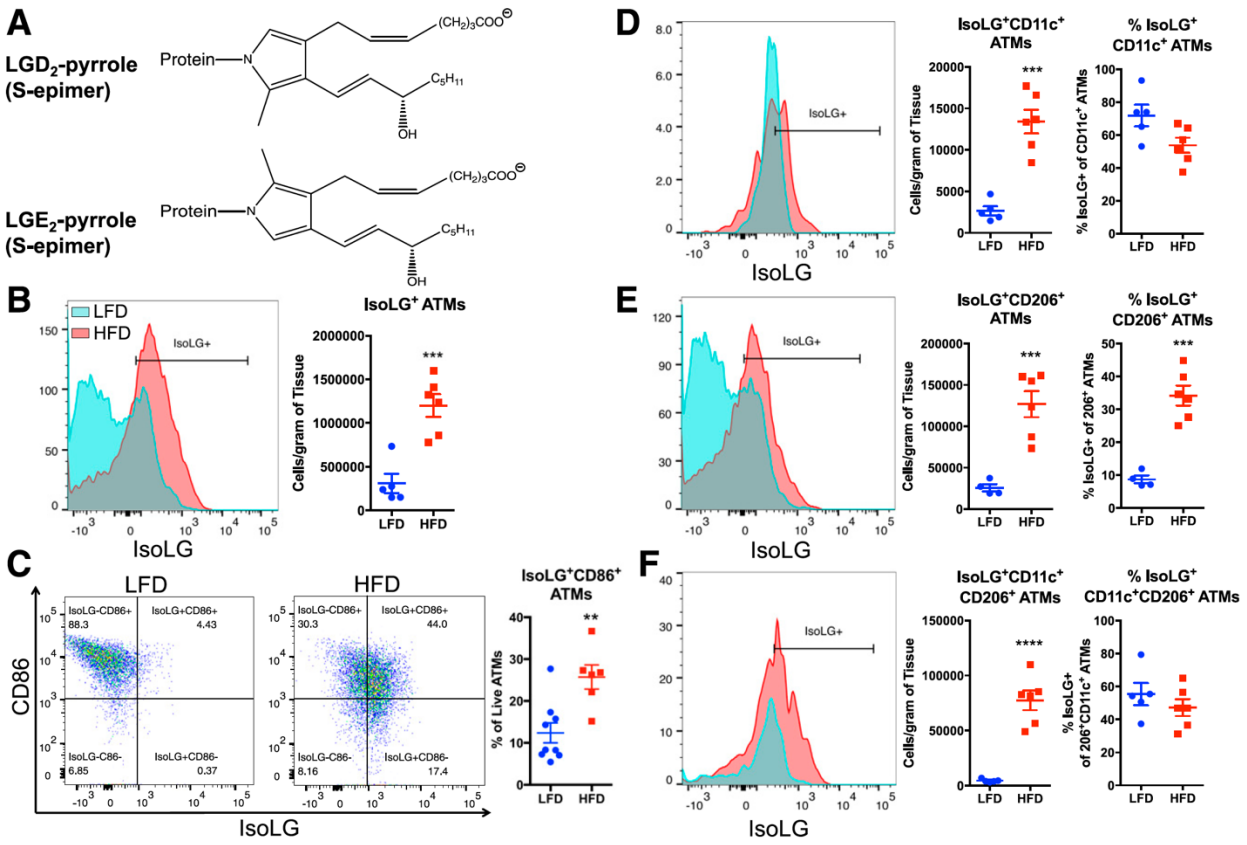


Figure 38. HFD induces formation of immunogenic isoLGs and expression of B7 ligand CD86 in ATMs. Total leukocytes were isolated from AT of mice fed an LFD or HFD, and isoLG content was quantified using previously described gating schema. A: Structure of the major isoLG protein adducts. The major intermediates are a pair of pyrrolic epimers whose covalently bonded nitrogen originates from lysine residues on proteins. Phospholipids attached to isoLGs derive from polyunsaturated acids, arachidonic acids, and phospholipids thereof. B: Representative flow cytometry plot and average data showing intracellular staining for isoLG protein adducts in ATMs using the single-chain antibody D-11 single-chain variable fragment. C: Representative FACS plots and average data showing surface expression of CD86 and isoLGs in ATMs (proportion of ATMs staining CD86+isoLG+). D–F: Representative flow cytometry plots and average data showing the number of isoLG+ cells per gram of tissue and percentage of isoLG+ cells per ATM subsets CD11c+ (D), CD206+ (E), and CD11c+CD206+ (F) of HFD-fed mice. Data are mean ± SEM (n = 5-10 mice/group). **P < 0.01, ***P < 0.001, ****P < 0.0001.

As previously reported by other research groups,[302,303] HFD increased total ATM density and ATM subsets CD11c+ (M1-like ATMs), CD206+ (M2-like ATMs), and double-positive populations (**Figure 39**). After HFD, ATMs showed increased isoLG-adduct formation (**Figure 38B**) and expression of the costimulatory molecule B7 ligand CD86 (**Figure 38C**). Regardless of ATM polarization state, HFD led to

significant elevations of isoLGs in the number of cells per gram of AT of CD11c+, CD206+, and CD11c+CD206+ ATMs compared with LFD (**Figure 38D–F**). However, HFD only increased the percentage of isoLG+ cells in CD206+ ATMs (**Figure 38E**).

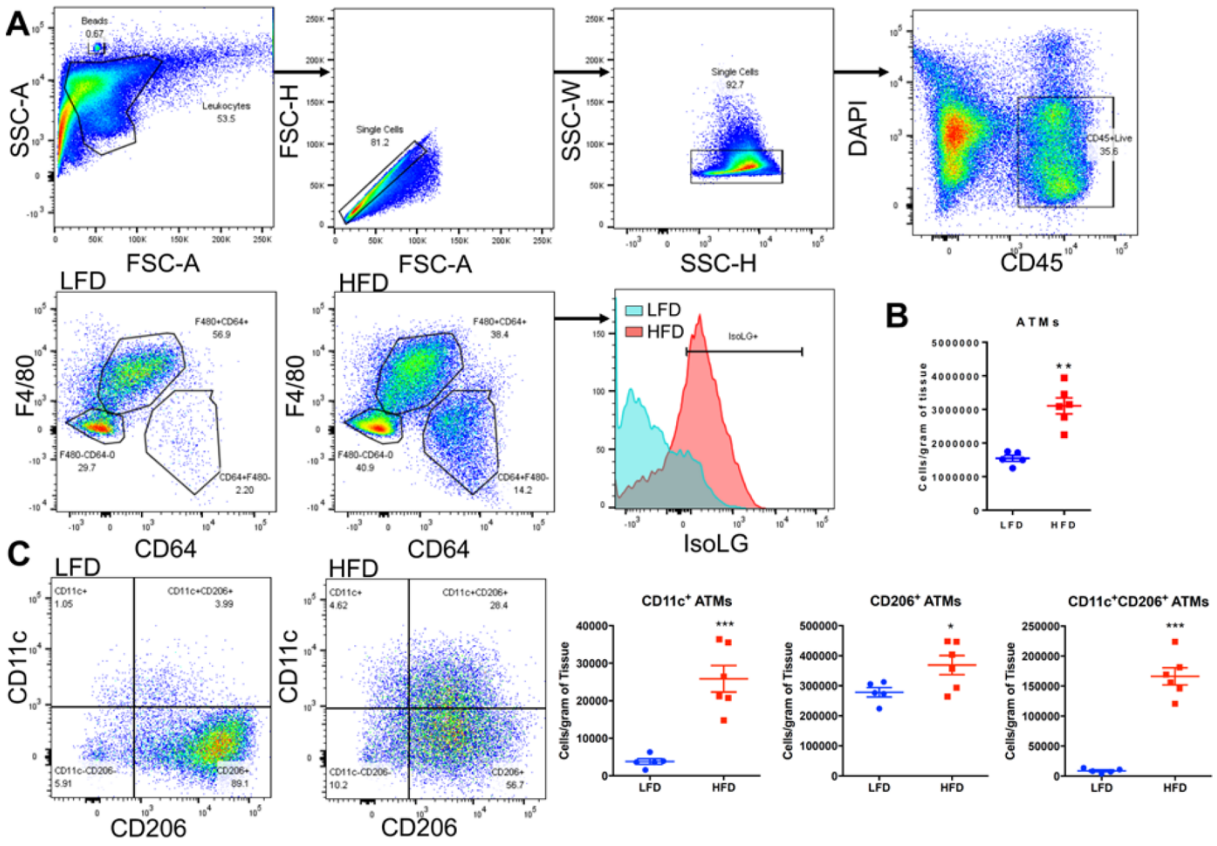


Figure 39. Gating strategy and quantification of ATMs. Gating strategy for total ATMs and (B) quantification of ATMs in LFD and HFD mice. (B) ATMs are significantly higher in diet-induced obesity. (C) Gating strategy for CD11c and CD206 expressing ATMs and quantification of flow analysis.

Because HFD led to elevations in isoLGs in ATMs, we sought to determine whether isoLGs in macrophages induce T-cell activation. As a control and to verify that isoLG-containing macrophages activate T cells, we treated BMDMs with 1 mmol/L TBHP, which induces oxidative stress. Treatment with TBHP led to

elevations in isoLGs in BMDMs (**Figure 40**) and as previously reported.[288]

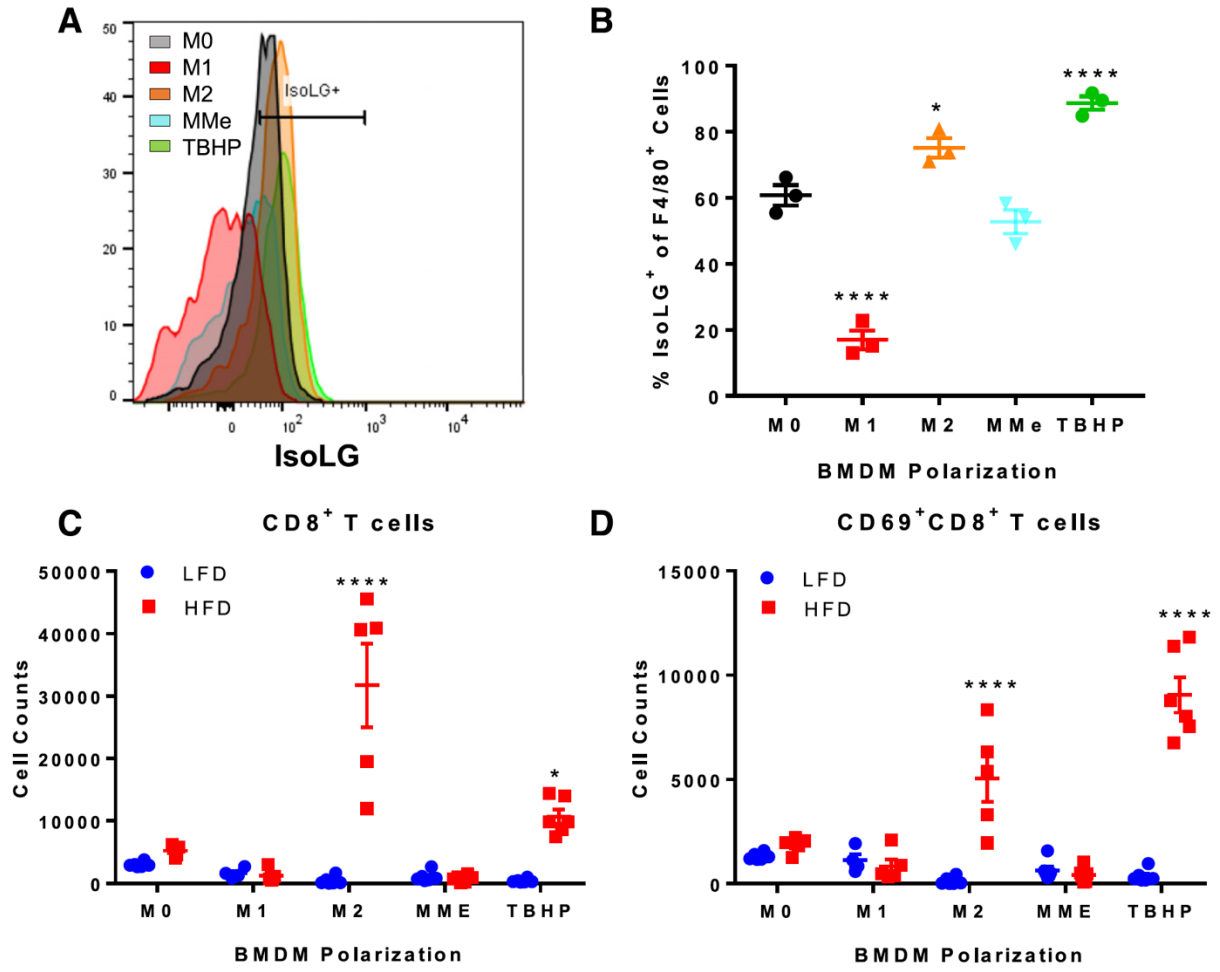


Figure 40. M2 polarization increases isoLG levels and promotes activation of CD8+ T cells. BMDMs were polarized to M0, M1, M2, or MMe or treated with TBHP. After polarization, BMDMs were cocultured with isolated T cells from the spleen of mice fed an LFD or HFD for 3 days. IsoLG content was quantified in polarized BMDMs by flow cytometry. A and B: Flow analysis (A) and quantification (B) of IsoLG+ F4/80+ cells (one-way ANOVA). Data are mean 6 SEM (n = 3 wells/group). T cells were collected from cocultures, and flow analysis was performed to quantify CD8+ T cells (C) and CD8+CD69+ T cells (D) (two-way ANOVA). Data are mean ± SEM (n = 5 mice/LFD or HFD group) and mean ± SEM (n = 5 wells/group). *P < 0.05, ****P < 0.0001.

With use of an in vitro coculture system, we polarized BMDMs to M1, M2, and obesogenic MMes to represent the various subsets present in AT during HFD. Of note, isoLG levels were elevated only in M2-polarized macro- phages compared with M0 ($P < 0.05$), were lower in M1 macrophages ($P < 0.001$), and were not

different in MMes compared with M0 (**Figure 40A and 40B**). TBHP-treated and polarized BMDMs were subsequently cultured with pan-T cells isolated from spleens of LFD- and HFD-fed mice. As expected, coculture with TBHP-treated macrophages led to elevations of total CD8+ T cells (**Figure 40C**) and more specifically in activated cells as indicated by expression of the early activation marker CD69 (**Figure 40C**). Furthermore, only the isoLG-containing M2 macrophages promoted an increase in CD8+ T cells and their activation (**Figure 40C and 40D**), and this was only when cocultured with CD8+ T cells from the obese HFD-fed mice. Together, these findings demonstrate that ATMs express isoLGs under obese conditions and that isoLG-containing M2 macrophages are capable of driving CD8+ T-cell activation in vitro.

Discussion

We demonstrate that HFD-fed mice have a higher AT CD8+ T-cell density and TCR clonality. CD8+ cells from AT of HFD-fed mice also have a distinctive TCR repertoire that differs in both charge and polarity within the CDR3 region compared with LFD-fed mice. These changes are not observed within the CD4+ TCR repertoire, which uses different V-segment genes than CD8+ T cells, regardless of diet (**Figure 32A**). The repertoires of mice fed an HFD are also more public than of mice fed an LFD, which has been reported previously in autoimmune pathologies but un- commonly

in antipathogen responses (reviewed in Madi *et al*.[304]). Furthermore, we provide evidence that these repertoire properties may emerge in part as a response to an increase in immunogenic isoLG adducts present in the ATMs of mice fed an HFD. We have directly identified and quantitated specific clonally expanded TCRs within the CD8+ and CD4+ T-cell populations that appear to arise in response to DIO. We also demonstrate that the AT TCR repertoire alone can distinguish between LFD- and HFD- fed mice. To that end, we note that HFD induces repertoire-level changes previously described in studies of autoimmunity and enriches for amino acid sequences biased toward a positive charge. In fact, Glanville *et al*.[305] recently described a strong inverse correlation between CDR3 charge and epitope charge in TCRs of both humans and mice. Finally, we document elevated numbers of isoLG-positive macrophages in the AT of HFD-fed mice and that M2 macrophages containing isoLG-adducted protein enhance T-cell survival and activation. These data suggest a coordinated mechanism of inflammation within AT wherein T cells infiltrate tissue in response to neo- antigen presented on ATMs.

To date, other studies have described the infiltration of a possibly clonally expanded T-cell population within AT.[281,284,287,295] In mice, Yang *et al*. observed shifts in the clonality of both the AT CD4+ and CD8+ TCR repertoires in obese versus lean mice using PCR-based spectratyping,[286] where clonal expansion was detected as a deviation from the normally Gaussian frequency distribution of CDR3 length.

Although spectratyping has been advantageous in the detection of unusual T-cell clones in lymphoma, it does not provide the specificity or direct measurement of clonal expansion and cannot distinguish between expanded clones with different amino acid sequences and similar V and J gene usage. Consequently, spectratyping also cannot detect clones with identical amino acid sequences and different V and J gene usage, as we observed in the CD8+ and CD4+ TCR repertoires of mice fed both an LFD and an HFD. In contrast, deep sequencing of the TCR repertoire makes it possible to distinguish between the TCR repertoires of LFD- or HFD-fed mice (Figure 33) and to assess intricate biochemical details of the TCR at the repertoire level (**Figures 34-38**). Similar observations have been made with regard to HFD-induced T-cell infiltration in pancreatic islets,[306–310] skeletal muscle,[311] and liver.[312] We observed similar levels of macrophage and CD8+ T-cell infiltration as reported in prior studies and further identified several highly oligoclonal CD8+ T-cell populations induced by HFD. As previously reported by Nishimura *et al*.,[281] we observed no difference in CD4+ T-cell infiltration into AT and extend this observation to include no significant difference in CD4+ T-cell clonality ($P = 0.18$). Although we do not observe enrichment of Vb gene families 7 and 10b as in Nishimura *et al*., we do observe slight enrichment of Vb families 17, 18, 21, 22, 25, 27, 28, and 29 (**Figure 32**). We observe a pronounced increase in the clonality of CD8+ T cells in AT of HFD-fed mice and note that future studies should specifically

examine the extent to which this is induced by diet alone or entirely depends on obesity. Future dietary studies also would benefit from analysis of age because at least one study indicated that AT mass and CD8+ T cells—but not CD4+ T cells—are elevated independent of diet in aging male and female mice.[313]

Highly clonal T-cell populations have been described in a broad array of pathologies.[314] Our group has demonstrated recently that the TCR repertoire is more clonal in the AT of obese and overweight adults with HIV infection receiving long-term therapy than in the blood and that the V-J gene family pairings appear to differ between tissue compartments,[315] suggesting that similar repertoire contraction also occurs in the AT of humans. Adverse effects of similarly clonal populations have been described in both obesity and diabetes. Highly clonal T-cell populations recognizing an HLA-DR4–restricted epitope of insulin have been documented in patients with type 1 diabetes.[316] Obesity in both mice and humans leads to markedly lower naive T-cell pools and highly clonal T-cell populations, further restricting the availability of diverse TCRs.[317] A recent report by Pham *et al*.[279] indicated that HFD also limits the diversity of the B-cell repertoire, which corresponds with IR. The presence of a clonally expanded T-cell population within tissue can be explained by several immunological processes. One possible scenario is that a T-cell population may infiltrate in response to an antigen within the tissue, as documented in the AT of humans and macaques during chronic simian

immunodeficiency viral and HIV infection.[318] Another is the presence of multiple T-cell populations sharing the same TCR but with drastically different functional roles and properties.[319] Alternatively, dominant clones from the blood could enter the tissue in an antigen-blind fashion.

Our findings raise the question of where and how antigen is presented to drive TCR clonal expansion. Several published reports have indicated that antigen presentation to CD4+ T cells occurs directly in AT. Multiple MHC- II–expressing APC populations in AT can present to CD4+ T cells, including B cells,[280] dendritic cells,[320] macrophages,[321] and even adipocytes.[322] Mice with global deficiency of MHC-II demonstrate protection from AT inflammation and systemic IR when placed on an HFD.[320] A portion of the reduced inflammation in AT was due to attenuation of CD11c+ macrophages and CD4+ T-cell accumulation.[320] Adipokines also can activate these APCs directly, including leptin, adiponectin, and retinol-binding protein 4.[322–324] Thus, AT has all the prerequisite components to activate CD4+ T cells. However, our studies show clonality in the CD8+ T-cell population, begging the question of their activation signals. MHC-I is expressed on all cell types and could allow for presentation of neoantigens or modified proteins. The current data suggest an enrichment of arginine and positively charged amino acids in the TCRs of CD8+ T cells that are clonally expanded in HFD-fed mice. This positive charge led us to consider that protein modifications leading to a negative charge might serve as

neoantigens in AT. Indeed, the data suggest that ATMs display increased levels of isoLG adducts (**Figure 38**). Similar alterations in charge and amino acid content have been reported in the context of renal disease.[288,325] Insulin-reactive TCRs have been reported to be enriched for arginine and for nonpolar amino acids at position 7,[326] which we also observed. We identified TCRs previously associated with obesity and IR in mice (**Table 1**), two of which were significantly enriched in HFD TCR repertoires. One clonotype that was highly prevalent in AT is known to be insulin reactive, and another is known to be enriched in the pancreas of diabetic mice. In contrast, two previously reported diabetogenic clonotypes were detected at elevated levels in LFD-fed mice alone (**Table 1**), which demonstrates that changes to the TCR repertoire extend to multiple tissues in the context of obesity and IR and that insulin reactivity alone may not explain clonal expansion within AT. Although we did not observe significant changes in the TCR repertoire of CD4+ T cells of the AT, future studies also could examine the TCR repertoire of CD4+ T cells within the liver, where fat also accumulates during DIO.

The current findings suggest that isoLGs may act as neoantigens responsible for AT TCR clonal expansion. Generation of these reactive lipid aldehydes primarily depends on oxidative stress, and reactive lipid aldehydes are detected at elevated levels in other autoimmune processes, including multiple sclerosis, glaucoma, allergic inflammation, and cancer (reviewed in Salomon and Bi).[327] Previous studies

examined the role of various lipid adducts on the immunogenicity of APCs and found that only proteins adducted to isoLGs induced APC immunogenicity.[288] A number of reports demonstrated that oxidative stress levels are elevated in AT of obese mice and humans.[328,329] In pulmonary injury, impaired NADPH oxidase reduces production of isoLG-adducted proteins, whereas impairment of the Nrf2-antioxidant response element signaling pathway increases isoLG-modified protein levels.[330] Furthermore, inhibition of NADPH oxidase reduces obesity- induced AT reactive oxygen species production, inflammation, and IR.[328,329] Thus, obesity-induced oxidative stress in ATMs may account for the elevation in isoLGs.

Although our results suggest that M2-like ATMs induce or enhance clonal expansion of AT-infiltrating T cells through the presentation of immunogenic isoLG-adducted peptides, the extent to which ATM-generated isoLGs induce clonal expansion in T cells remains to be explored. We also note that T cells derived from splenocytes in the liver may not precisely reflect those found in AT and that a culture model able to effectively enrich such lymphocytes from the AT would be a valuable contribution to the field. The current results also suggest several new directions for further investigation. The antigen specificity of clonally expanded T cells within AT remains to be determined, and single-cell TCR sequencing can connect surface phenotype information from flow cytometry with transcriptomics and paired TCRa/b sequences.[331] Although we did not identify any clonotypes that

were universally shared at high (>1%) frequencies between or within diet groups, we did observe several TCRs that were shared within the top 100 clonotypes of each dietary group (**Table 1**), particularly with respect to HFD-fed mice. This suggests that the antigens or neoantigens presented in obesity may not be the same from mouse to mouse. These TCRs may be specific for one or more antigens, although we have previously shown that TH1 CD4+ cells, CD8+ T cells, and memory cells accumulate in the AT of weight-cycled mice, suggesting secondary immune responses with repeat exposure to HFD,[332] and our *in vitro* experiments suggest that isoLG-containing macrophages augment T-cell activation and survival in tissue-resident immune populations that are expanded in obesity (**Figure 40**).

In summary, we have demonstrated that T cells infiltrating AT of obese mice are clonally expanded, that the TCR repertoires of lean and obese mice are distinct in their sequences and biochemical properties, that immunogenic isoLGs are elevated in ATMs of obese mice, and that T cells of obese mice accumulate and activate in response to isoLG-containing M2 macrophages. Our observations also suggest that therapy targeting the CD8+ T-cell population and antigens thereof could potentially abrogate inflammation and other consequences of adaptive immune activation in obesity.

Chapter IV

Adipose tissue in persons with HIV is enriched for CD4+ T effector memory and T effector memory RA+ cells, which show higher CD69 expression and CD57, CX_3CR1, GPR56 co-expression with increasing glucose intolerance

N.B. This work was published in its entirety as Wanjalla & McDonnell *et al.* **Frontiers in Immunology** 2019; 10:408. doi: 10.3389/fimmu.2019.00408. In this work, we identify for the first time a novel population of CD4 effector memory T cells that increase in frequency according to diabetic status. These cells are cytotoxic, express a unique vascular homing profile, and could serve as a biomarker or a therapeutic target of incident type II diabetes.

Abstract

Chronic T cell activation and accelerated immune senescence are hallmarks of HIV infection, which may contribute to the increased risk of cardiometabolic diseases in people living with HIV (PLWH). T lymphocytes play a central role in modulating adipose tissue inflammation and, by extension, adipocyte energy storage and release. Here, we assessed the CD4+ and CD8+ T cell profiles in the subcutaneous adipose tissue (SAT) and blood of non-diabetic (n = 9; fasting blood glucose [FBG] < 100 mg/dL), pre-diabetic (n = 8; FBG = 100–125 mg/dL) and diabetic (n = 9; FBG ≥ 126 mg/dL) PLWH, in addition to non- and pre-diabetic, HIV-negative controls (n = 8). SAT was collected by liposuction and T cells were extracted by collagenase digestion. The proportion of naïve (TNai) CD45RO−CCR7+, effector memory (TEM) CD45RO+CCR7−, central memory (TCM) CD45RO+CCR7+, and effector memory revertant RA+ (TemRA) CD45RO− CCR7− CD4+ and CD8+ T cells were measured by

flow cytometry. CD4+ and CD8+ TEM and TemRA were significantly enriched in SAT of PLWH compared to blood. The proportions of SAT CD4+ and CD8+ memory subsets were similar across metabolic status categories in the PLWH, but CD4+ T cell expression of the CD69 early-activation and tissue residence marker, particularly on TEM cells, increased with progressive glucose intolerance. Use of t-distributed Stochastic Neighbor Embedding (t-SNE) identified a separate group of predominantly CD69lo TEM and TemRA cells co-expressing CD57, CX_3CR1, and GPR56, which were significantly greater in diabetics compared to non-diabetics. Expression of the CX_3CR1 and GPR56 markers indicate these TEM and TemRA cells may have anti-viral specificity. Compared to HIV- negative controls, SAT from PLWH had an increased CD8:CD4 ratio, but the distribution of CD4+ and CD8+ memory subsets was similar irrespective of HIV status. Finally, whole adipose tissue from PLWH had significantly higher expression of TLR2, TLR8, and multiple chemokines potentially relevant to immune cell homing compared to HIV-negative controls with similar glucose tolerance.

Introduction

People living with human immunodeficiency virus (HIV) are at an increased risk of developing insulin resistance and overt diabetes mellitus, but the factors contributing to the high prevalence of metabolic disease in the HIV population are

not fully understood.[333–335] Since HIV was identified as the cause of acquired immune deficiency syndrome (AIDS) in the early 1980s, the metabolic consequences of altered adipose tissue function in people living with HIV (PLWH) have been a major research focus.[336] Early studies identified accelerated lipolysis and hepatic lipogenesis as central energy metabolism abnormalities in untreated HIV infection,[337–339] while early- generation nucleoside reverse transcriptase inhibitors (NRTIs) caused adipocyte mitochondrial damage and adipose tissue fibrosis,[340–344] and treatment with early protease inhibitors was accompanied by accumulation of visceral adipose tissue (VAT), hyperlipidemia, and insulin resistance.[345–348] More recently, several studies describe profound changes in adipose tissue T cell populations during chronic HIV and simian immunodeficiency virus (SIV; a non-human primate virus similar to HIV) infections, which may influence adipose tissue metabolic function. These include changes in T cell surface marker phenotypes, cytokine production, antigen receptor repertoire, and capacity for latent infection with HIV or SIV provirus.[315,318,349–352] Notably, several studies found that HIV and SIV were accompanied by a substantial increase in the proportion of adipose CD8+ T cells relative to CD4+ T cells, which is strikingly similar to the enrichment in CD8+ T cells also described in obesity.[281,283,353]

Adipose tissue is a complex and vascularized cellular amalgam, comprised of multipotent adipocyte progenitors, mature adipocytes, fibroblasts, and immune

cells of the adaptive and innate lineages. A diverse group of immune cells collectively identify and eliminate the range of viruses and other pathogens which can infiltrate adipose tissue, and these processes impact local levels of pro-inflammatory and other cytokines with subsequent effects on adipocyte regulation and energy storage and release. T lymphocytes play several beneficial and detrimental roles within this environment. Studies of obese humans and animals demonstrate an increase in adipose tissue CD8+ T cells and CD4+ TH1 cells, a decrease in T regulatory cells, and an increase in M1-phenotype (CD68+, tumor necrosis factor-α [TNF-α], interleukin [IL]-6, IL-12, and IL-23-producing) pro-inflammatory macrophages compared to non-obese controls.[280,281,283,284] In animal models of obesity, the infiltration of CD8+ T cells into adipose tissue precedes the recruitment of macrophages;[281] and the resulting increase in local IL-6, TNF-α and other inflammatory mediators act on adipocyte surface receptors and other mechanisms to inhibit insulin signaling via reduced insulin receptor substrate-1 (IRS-1), phosphoinositide 3-kinase p85α, and glucose transporter type 4 (GLUT4) expression.[14,281,349,354–358]

The increase in the CD8+ to CD4+ T cell ratio observed in HIV infection is like that seen in diet-induced obesity, though the mechanisms underlying the accumulation of CD8+ T cells in the adipose tissue of PLWH are not well-defined. Studies in macaques suggest that the high proportion of CD8+ T cells is not due to

depletion of CD4+ T cells.[318] Although lack of Ki-67 expression in adipose tissue T cells has been interpreted as lack of evidence for in situ proliferation, greater CD8+ TCR clonality in subcutaneous adipose tissue (SAT) implies antigen specificity might drive the increase rather than stochastic recruitment of circulating CD8+ T cells. This is further supported by the finding that CD8+ and CD4+ T cells in adipose tissue predominantly display a memory phenotype with increased levels of CD69 expression compared to those in blood.[318,349]

While prior studies have shown enrichment of CD8+ over CD4+ T cells in adipose tissue after HIV infection, there is a paucity of data on whether a particular subset of cells underlies this change, and whether adipose tissue T cell profiles differ according to insulin sensitivity in PLWH (as might be expected given prior findings in obesity-related insulin resistance). In this study, we hypothesized that the enrichment of CD8+ T cells in the adipose tissue of PLWH could be attributed to an over-representation of one or a few memory cell subtypes, and that greater CD8+ and CD4+ T cell activation would characterize the adipose tissue of diabetic PLWH. We evaluated SAT CD4+ and CD8+ T cell subsets (including naïve cells, activated cells, and central memory [TCM], effector memory [TEM], and effector memory revertant RA+ [TemRA] cells) in PLWH vs. HIV- negative controls, and among diabetic vs. non-diabetic PLWH.

Methods

Study participants

We enrolled 26 PLWH on long-term antiretroviral therapy (ART) with sustained virologic suppression from the Vanderbilt Comprehensive Care Clinic between August 2017 and June 2018. Hemoglobin A1c (HbA1c) and fasting blood glucose (FBG) were used to classify participants as non-diabetic (n = 9; HbA1c < 5.7% and FBG < 100 mg/dL), pre-diabetic (n = 8; HbA1c 5.7–6.5% and/or FBG 100–125 mg/dL), and diabetic (n = 9; HbA1c ≥ 6.5% and/or FBG ≥ 126 mg/dL, and on anti- diabetes medications). A group of 8 HIV-negative, non- and pre-diabetic controls were enrolled from the community. The PLWH were on ART for at least 18 months, had HIV-1 RNA <50 copies/ml for the prior 12 months, CD4+ count >350 cells/μl, and had no known inflammatory or rheumatologic conditions. We excluded persons with self-reported heavy alcohol use (defined as >11 drinks/week), any cocaine/amphetamine use, and those receiving corticosteroids or growth hormone.

All visits occurred in the Vanderbilt Comprehensive Care Clinic research suite or the Vanderbilt Clinical Research Center between 8 and 11 am. Participants fasted for a minimum of 8 h prior to blood collection for laboratory measurements and peripheral blood mononuclear cell (PBMC) separation (PLWH only). Blood glucose, HbA1c, high-sensitivity C-reactive protein (hsCRP), low-density lipoprotein (LDL),

triglycerides, and high- density lipoprotein (HDL) were measured in the fasting blood samples at the Vanderbilt Clinical Chemistry Laboratory.

Adipose tissue biopsy and T cell extraction

SAT biopsies were collected ~3 cm to the right of the umbilicus after anesthetizing the skin with lidocaine and infiltrating 40 ml of sterile saline and lidocaine into the subcutaneous adipose tissue as tumescent fluid. We collected ~5 grams of adipose tissue using a 2.1mm blunt, side-ported liposuction catheter (Tulip CellFriendly™ GEMS system Miller Harvester, Tulip Medical Products) designed for the extraction of viable adipocytes and stromal vascular cells during cosmetic adipose tissue transfer procedures.[359] With this approach, adipose tissue is recovered in droplets generally <3 mm in diameter, limiting the need to mechanically mince the sample, and the tissue is placed in 40– 50 cc of cold saline and mixed to rinse. Any visible clots are removed before the sample is transferred to a 70 μm mesh filter for repeated saline rinses with constant stirring. The adipose tissue is then placed in a gentleMACS™ Dissociator (Miltenyi Biotec) followed by incubation with collagenase (Roche Catalog #11088866001). Mononuclear cells were separated by Ficoll-Paque Plus density gradient and cryopreserved in fetal bovine serum (FBS) with 10% DMSO.

Flow cytometry analysis

PBMC and SAT mononuclear cell aliquots were processed and stained as previously published.[315] In brief, matched cryopreserved PBMCs and SAT cells were quickly thawed at 37◦C and suspended in R10 media (RPMI with 10% FBS). These were then washed once in phosphate buffered saline (PBS) and stained with multiple fluorescent tagged antibodies: CD3- BV786 (Clone SK7; BD Biosciences #563800), CD4-PcPCy5.5 (Clone RPA-T4; BD Biosciences #560650), CD8-A700 (Clone PRA-T8; BD Biosciences #557945), CD57-FITC (Lot 4182924; BD Pharmingen # 555619), CX_3CR1-PE (Clone 2A9-1; BD Biosciences #565796), CD45RO-PECF594 (Clone UCHL1; BD Biosciences #562299), CD14-V500 (Clone M5E2, BD Biosciences #561391), CD19-V500 (Clone HIB19; BD Biosciences #561121), LIVE/DEAD Fixable Aqua (ThermoFisher; L34957), CD69-APC (Clone FN50; BD Biosciences #560711), CCR7-V450 (Clone 150503; BD Biosciences #562555), GPR56-PECy7 (Clone CG4; BioLegend #358205), and HLA-DR APC Cy7 (Clone G46.6; BD Biosciences #561358). CCR7 and CX_3CR1 antibody stains were performed at 37◦C; the remainder was stained at room temperature. Cells were analyzed on a 4-laser FACSAria III (all samples from PLWH) or 5-laser LSRII (HIV-negative SAT samples) (BD Biosciences, San Jose, CA). Bead calibration was used to standardize runs done on different days. Flow cytometry data were analyzed using FlowJo software (version 10.4.1) and Cytobank (version 6.3.1).[360,361]

The T cell memory populations in our study are defined as naive T cells (TNai, CD45RO−CCR7+), T effector memory (TEM , CD45RO+ CCR7−), T central (TCM , CD45RO+CCR7+), and T effector memory revertant/re-expressing CD45-RA (TemRA , CD45RO−CCR7−). Representative gating strategies in **Figure 41** and **Figure 42** show the phenotypic markers used to define the memory subsets (CD45RO and CCR7). Total memory cells are a combined group consisting of TEM, TCM, and TemRA.

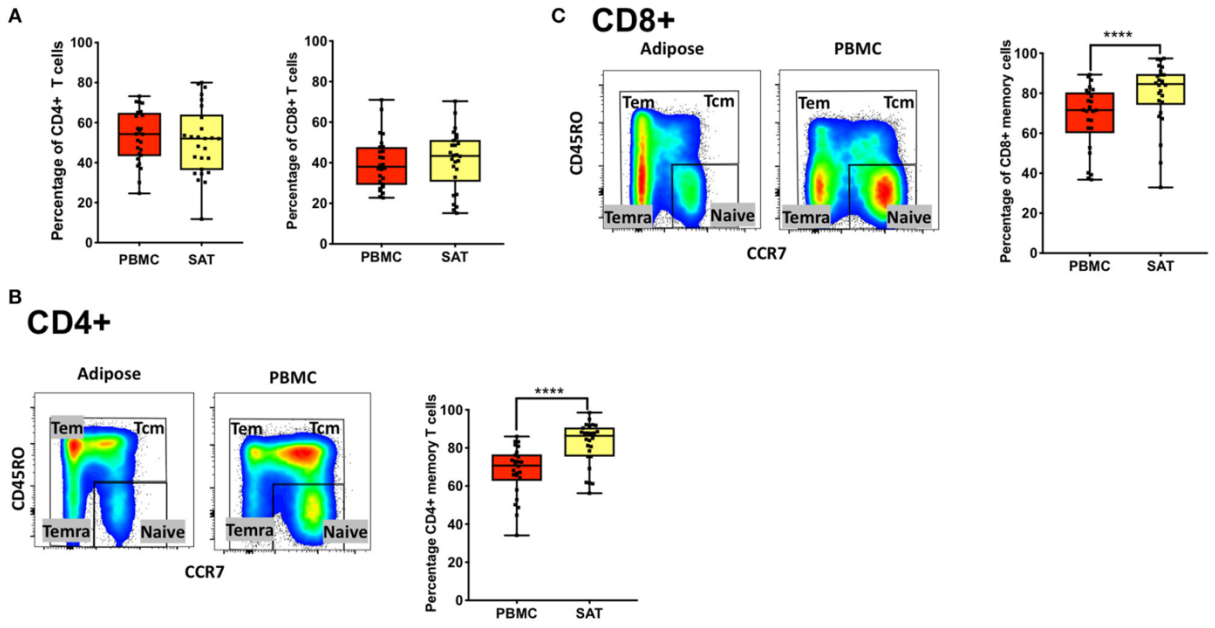


Figure 41. Subcutaneous adipose tissue from PLWH has a higher percentage of CD4+ and CD8+ memory T cells compared to matched blood samples. (A) Frequencies of total CD4+ and CD8+ T cells (percentage of CD3+ cells) in subcutaneous adipose tissue (SAT) and blood (PBMC) from all 26 PLWH. (B) Representative plot showing gating of CD4+ memory T cells (gated on CD3+) in SAT and PBMC depicted by the yellow and red shading, respectively; on the right are individual values and means. (C) Representative plot showing total CD8+ memory T cells in adipose tissue and PBMC. The box and whiskers plot indicate mean ± SD. Wilcoxon matched pair signed test was used to calculate statistics (A-C); ****p < 0.0001.

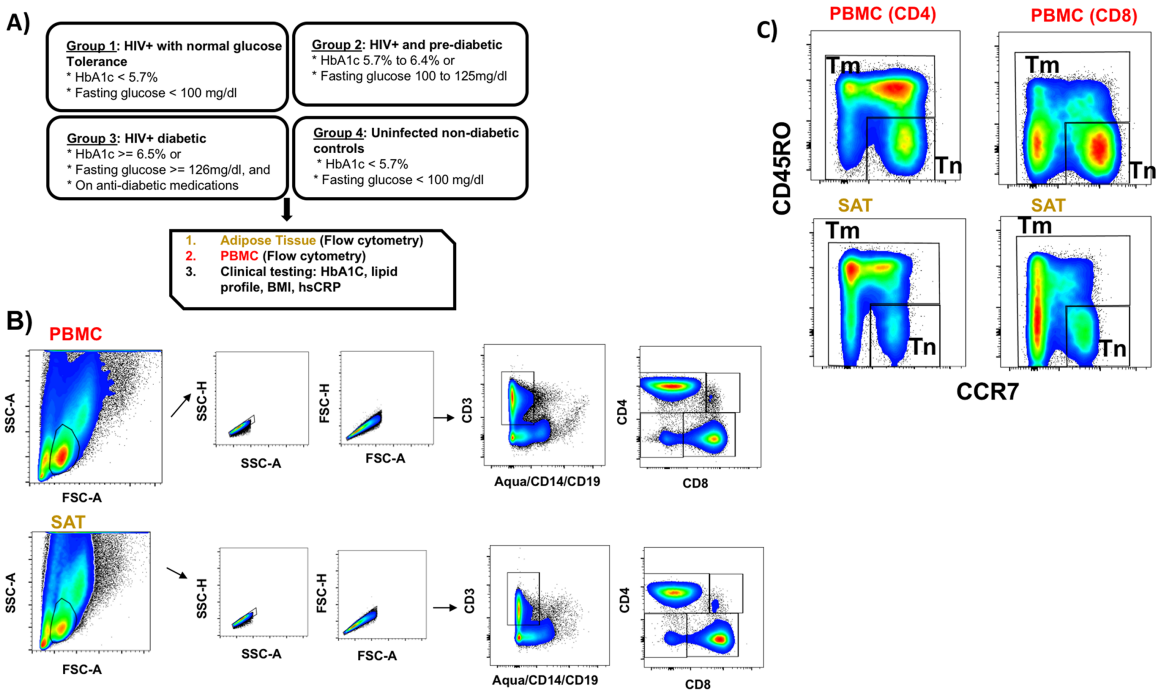


Figure 42. Study groups and flow cytometry gating panels. (A) Study subjects were placed in four groups. Matched cryopreserved PBMC and SAT were analyzed from subjects from Group 1 (n=9), Group 2 (n=8), and Group 3 (n=9). (B) Representative plots showing multi-parametric flow cytometry gating panels used to analyze CD3+, CD4+ and CD8+ T cells in matched PBMC and SAT. (C) Memory CD4 and CD8 T cells were assessed using CCR7 and CD45RO surface markers (TEM: CCR7−CD45RO+; TCM: CCR7+CD45RO+; TemRA: CCR7−CD45RO−).

Adipose tissue gene expression

Adipose tissue was rinsed, placed in cryovials and snap-frozen in liquid nitrogen immediately after collection for subsequent mRNA expression assays. mRNA was extracted after mechanical lysis with the Qiagen RNeasy Lipid Tissue Kit. We used the Nanostring nCounter Plex^2 human inflammation panel to quantify mRNA expression of over 250 genes spanning a broad range of relevant immune pathways including interleukin signaling, Ras, T-cell markers, and Toll-like receptor signaling. Adipose tissue mRNA expression levels were normalized using 14 synthetic spike-ins (6 positive controls and 8 negative controls) and 6 cellular

housekeeping genes included in the assay (GAPDH, GUSB, HPRT1, PGK1, TUBB, and CLTC). We first calculated the coefficient of variation (CV) for the control genes. The CV of the positive controls is proportional to the technical variability introduced by the nCounter platform. The CV for the housekeeping controls is proportional to the confounding biological variation due to sample input. The mean endogenous CV shows the global noise of experimentally observed genes. After evaluating different normalization approaches based on CV values, we developed a normalization strategy including the following steps. First the background count levels were calculated using the mean of negative controls, then subtracted from each sample. The normalization factor for sample/RNA content was calculated using the geometric mean of a set of pre-specified annotated housekeeping genes. The algorithm normalizes for sample or RNA content, i.e., “pipetting” fluctuations, using the geometric mean of pre-specified annotated housekeeping genes. The count data were then divided by the normalization factor to generate counts normalized to the geometric mean of housekeeping genes. None of the housekeeping genes differed significantly in their distribution between study groups.

Statistical analyses

For comparisons, PLWH were grouped according to metabolic status as non-diabetic (n = 9), pre-diabetic (n = 8), and diabetic (n = 9). The 8 HIV-negative controls

were matched to 9 PLWH with similar HbA1c and body mass index (BMI) values to yield a similar comparison group (one control was matched to two PLWH). Percentages of CD4+ and CD8+ subsets were compared between paired blood and adipose tissue samples using Wilcoxon signed-rank tests for paired data. Differences between the PLWH metabolic groups, and between HIV- negative controls and the PLWH, were calculated using Mann-Whitney and Kruskal Wallace tests. When significant between-group comparisons were noted among the PLWH, univariable and multivariable linear regression were used to assess the relationship of cellular populations with progressive glucose intolerance. Statistical analyses and graphs were performed using SPSS (IBM, Armonk, NY), R (www.r-project.org), and GraphPad Prism 7 (GraphPad Software, La Jolla, CA).

Adipose tissue genes were grouped according to immune system pathways specified in the NanoString kit, and normalized mRNA levels were compared between the PLWH and HIV- negative participants. DEseq2 was used to detect differential expression between two groups based on the normalized count data, taking into account technical and biological variability.[362,363] Differences in gene expression were calculated as fold-changes, and p-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure. Analyses were performed using R. Volcano plots displaying the NanoString data were generated using XL-STAT.

Results

Clinical and demographic characteristics of PLWH and HIV-negative subjects

The non-diabetic (*n* = 9), pre-diabetic (*n* = 8), and diabetic (*n* = 9) groups of PLWH are compared in **Table 2**. Age, race and sex were similar across the groups, as were CD4+ nadir, duration of ART, and the proportions receiving integrase inhibitor-based regimens ($p > 0.05$ for all comparisons). BMI and waist circumference increased with progressive glucose intolerance ($p = 0.04$ for both). The HIV-negative controls (*n* = 8) and comparator PLWH (9 out of 26 total) were similar in age, BMI, and HbA1c values ($p > 0.05$ for all), though the controls were more likely to be female and white (**Table 3**, next page). Bold values in **Tables 2, 3** indicate $p < 0.05$.

Variable	PLWH (*n* = 9)	HIV-negative controls (*n* = 8)	*P*-value
Age, median years (IQR)	39 (38, 49)	49 (37, 62)	0.32
Female (%)	1 (10%)	5 (63)	**0.02**
Caucasian race (%)	4 (40%)	8 (100%)	**<0.01**
BMI, kg/m^2	30.8 (27.8, 36.4)	39.35 (33.3, 48.4)	0.89
GLUCOSE METABOLISM MEASUREMENTS			
Hemoglobin A1c, %	5.3 (5.15, 5.4)	5.7 (5.4, 5.9)	0.84

Table 2. Comparison of HIV-negative participants and comparator non- and pre-diabetic PLWH. The bolded values are p-values <0.05, indicating statistical significance.

Variable	HIV (+) Non-diabetic HbA1c < 5.7 *and* FG < 100 mg/dL (*n* = 9)	HIV (+) Pre-diabetic HbA1c 5.7–6.5 *or* FG = 100–125 mg/dL (*n* = 8)	HIV (+) Diabetic on pharmacotherapy (*n* = 9)	*P*-value
Age, median yr (IQR)	39 (38, 49)	59 (33, 63)	54 (45, 61)	0.22
Female (%)	1 (11%)	2 (25%)	4 (44%)	0.23
Caucasian race (%)	4 (40%)	4 (50%)	3 (33%)	0.78
BMI, kg/m^2	30.8 (27.8, 36.4)	32.7 (30.8, 37.7)	40.2 (34.0, 51.9)	**0.04**
Waist: hip ratio	0.97 (0.92, 1.00)	0.97 (0.92, 1.04)	0.98 (0.96, 1.09)	0.60
Waist Circumference, cm	106 (101, 114.5)	107 (101, 111)	116 (110, 142)	**0.04**
GLUCOSE METABOLISM MEASUREMENTS				
Hemoglobin A1c, %	5.3 (5.1, 5.4)	5.6 (4.9, 5.9)	7.2 (6.1, 9.8)	**<0.001**
Fasting glucose, mg/dL	87 (75, 96)	116 (102, 122)	141 (120, 238)	**<0.001**
PLASMA LIPIDS				
Cholesterol, mg/dL	181 (149, 194)	185 (172, 210)	175 (143, 193)	0.28
LDL, mg/dL	95 (77, 109)	116 (96, 131)	82 (56, 107)	**0.05**
Fasting HDL, mg/dL	52 (39, 81)	38 (34, 56)	43 (35, 56)	0.31
Fasting TG, mg/dL	104 (61, 173)	196 (82, 257)	149 (114, 267)	0.37
IMMUNE MARKERS				
High sensitive C-reactive protein, mg/dl	2.7 (1.1, 8.4)	3.1 (0.9, 4.2)	3.1 (1.9, 9.8)	0.67
CD4 at study visit	803 (767, 1065)	702 (643, 942)	1056 (950, 1167)	**0.02**
CD4 at start of therapy	482 (364, 617)	448 (154, 567)	323 (173, 432)	0.12
CD4:CD8 ratio	1.07 (0.88, 1.43)	1.04 (0.73, 1.18)	1.36 (0.94, 1.52)	0.49
ANTIRETROVIRAL THERAPY HISTORY				
Exposure to AZT, *n* (%)	2 (22%)	3 (38%)	5 (56%)	0.35
Duration of ART therapy, yr	7.5 (4.0, 13.2)	11.8 (5.8, 18.9)	11.6 (5.4, 18.6)	0.56
Integrase inhibitor-based therapy	8 (89%)	4 (40%)	7 (78%)	0.18
OTHER				
Smoker (%)	2 (22%)	2 (25%)	2 (22%)	0.99
Hepatitis C (%)	1 (11%)	0 (0%)	2 (22%)	0.36

Table 3. Cohort demographic and clinical characteristics according to glucose tolerance. ART, antiretrovirals; AZT, azidothymidine; BMI, body mass index; FG, fasting glucose; HDL, high-density lipoproteins; IQR, interquartile range; LDL, low-density lipoproteins; TG, triglycerides; yr, year. The bolded values are p-values <0.05, indicating statistical significance.

Adipose tissue from PLWH is enriched in memory CD8+ and CD4+ T cells

We found no difference in the percentage of total CD8+ and CD4+ T cells (gated on CD3+ T cells) between SAT and peripheral blood from PLWH (**Figure 42A**). We also observed, as previously described by other groups, a higher percentage of total CD4+ memory T cells (**Figure 42B**) and CD8+ memory T cells (**Figure 42C**) in SAT compared to peripheral blood ($p < 0.0001$).[349]

SAT from PLWH is enriched in CD4+ and CD8+ TEM and TemRA cells compared to blood

Despite the growing body of literature on adipose tissue immune cells, the distribution of memory T cell subsets in adipose tissue from PLWH has not been characterized, and the role of memory T cell subsets within tissue compartments in general is not well- understood. A prior study assessed the distribution of CD4+ and CD8+ TNai , TEM , TCM , and TemRA subsets in lung, spleen, colon, ileum, jejunum, and lymph nodes, but not adipose tissue, from healthy donors, and found that CD4+ TNai , TEM , and TCM proportions in lung and pulmonary and mesenteric lymph nodes were similar to the proportion in blood, while jejunum, ileum, and colon were roughly 3- to 4-fold enriched in TEM.[364] In all tissues, CD4+ TemRA were relatively sparse (<10%). In contrast, ~30% of CD8+ T cells in blood, spleen, and lung had a TemRA phenotype, while the CD8+ T cells in jejunum, ileum, and colon were overwhelmingly TEM.[364] A subsequent study in the adipose tissue of mice found approximately equal proportions of CD4+ TEM and CD4+ TRM cells (a "resident memory" phenotype defined by the authors as CD69 expression on TEM cells) in adipose tissue samples, and fewer TCM.[365] However, CD8+ memory cells were predominantly TRM in SAT, with similar, lower proportions of TEM and TCM. The same study looked at mesenteric adipose tissue in macaques and again found high

levels of CD8+ TRM, while CD4+ cells were primarily TNai and TCM, with low levels (<10%) of TEM and TRM, potentially highlighting important differences between species.

We first assessed CD4+ and CD8+ memory subsets in SAT and blood from all 26 PLWH. Multiparameter gates were used to quantify memory CD4+ and CD8+ T cells within each sample as shown in **Figure 41**. Each sample had >100,000 ungated cells (**Figure 43**).

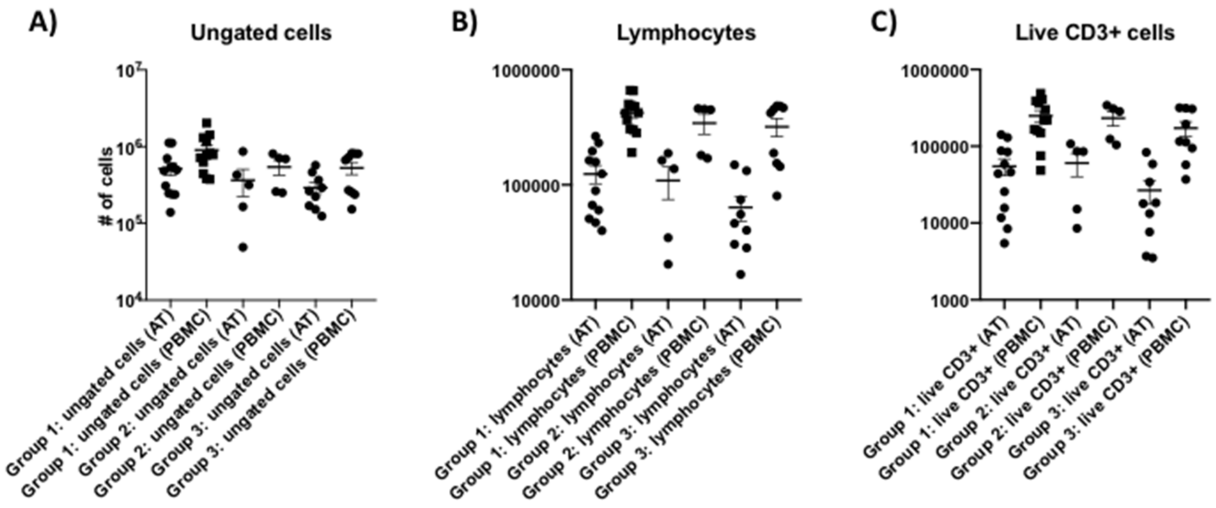


Figure 43. Event counts of individual SAT and PBMC samples from all twenty-six subjects. Ungated cells B) Lymphocytes and C) Live CD3+ T cells.

SAT was significantly enriched in CD4+ and CD8+ TEM and TemRA cells (**Figure 44**) compared to blood but had fewer TNai and TCM cells. We then compared the average TNai, TCM, TEM, and TemRA fractions within SAT and blood among the 26 PLWH (Supplementary Figure 2). We found that CD4+ T cells in SAT can be ranked by frequency as TEM > TCM > TNai > TemRA, whereas peripheral

blood contains CD4+ TCM > TNai > TEM > TemRA. In contrast, SAT CD8+ T cells are primarily TemRA followed by TEM.

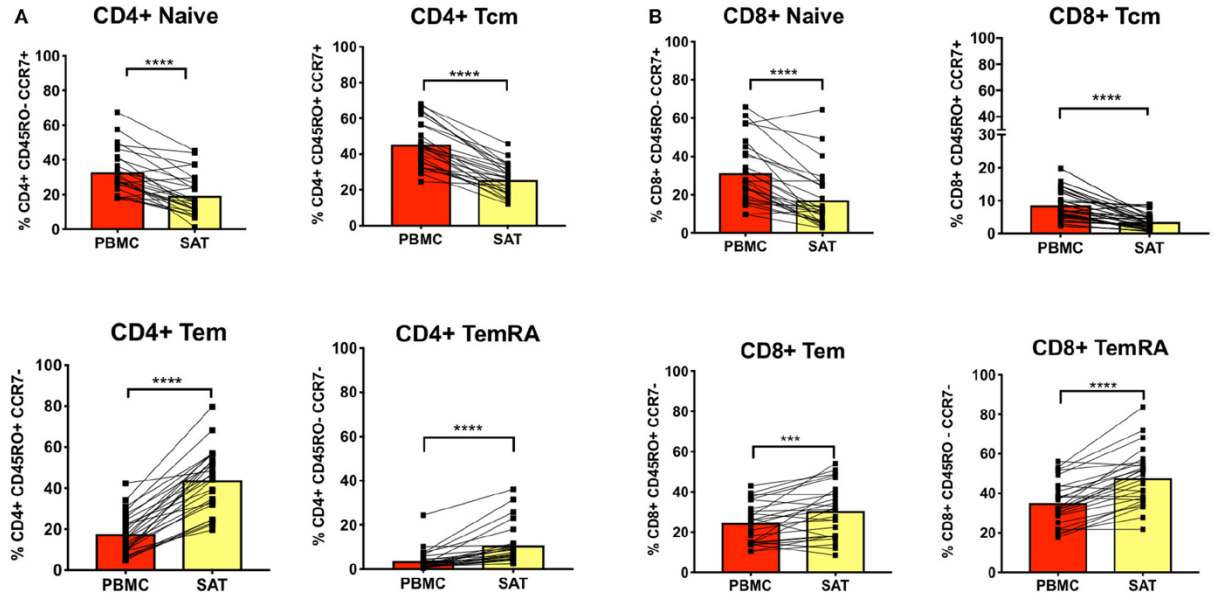


Figure 44. Subcutaneous adipose tissue has a higher percentage of TEM (CD45RO+CCR7-) and TemRA (CD45RO-CCR7-) cells compared to blood in PLWH. (A) The bar graphs show frequencies of CD4+ TNai, TEM, TCM, and TemRA cells in subcutaneous adipose tissue (SAT) and blood (PBMC) from all twenty-six subjects. (B) Frequencies of CD8+ TNai, TEM, TCM, and TEMRA cells in SAT and PBMC. The numbers in the bar graphs on the left indicate mean and diagonal lines indicate matched pairs of SAT and PBMCs. Wilcoxon matched pair signed test was used to calculate statistics ****$p < 0.0001$; ***$p < 0.001$.

The relative distribution of SAT memory T cell subsets does not differ with metabolic status in PLWH

The study of adipose tissue T cells in mice by Han et al. also found that recall responses of adipose tissue memory T cells were enhanced after antigen re-challenge, with downregulation of several metabolic pathways in whole adipose tissue (including lipid biosynthesis and cholesterol and long-chain fatty-acyl- CoA metabolic processes) and a detectable reduction in serum levels of adiponectin and

cholesterol.[365] These findings suggest a possible mechanism by which the accumulation, and subsequent stimulation, of memory cells in adipose may disrupt metabolic homeostasis.

In light of prior animal studies indicating a potential contribution of effector memory T cells to inflammation and impaired metabolic function in adipose tissue, we sought to characterize the distribution of SAT memory T cell subsets in PLWH to determine if there were clear differences between non-diabetics, pre-diabetics, and diabetics in the relative proportions of TNai, TCM, TEM, and TemRA cells. We utilized t- distributed Stochastic Neighbor Embedding (t-SNE) to visualize groups of adipose tissue and blood CD4+ and CD8+ T cell populations based on 12-color flow cytometry staining.

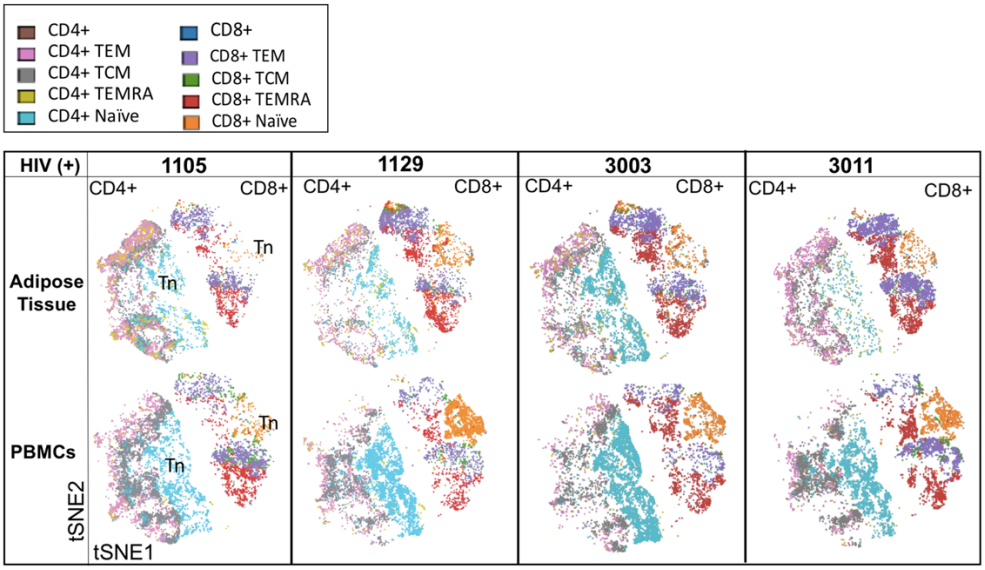


Figure 45. Unsupervised analysis showing distribution of CD4+ and CD8+ subsets compared to matched PBMCs. Representative plots showing analysis performed on gated CD3+ T cells in PBMCs and SAT cells. viSNE analysis illustrates the distribution of memory (TCM, TEM, TemRA) and naïve CD8+ and CD4+ T cell subsets within gates CD3+ T cells.

Figure 46 shows the distribution of TNai, TCM, TEM, and TemRA cells in CD4+ and CD8+ T cells in the adipose tissue and blood from four representative non-diabetic and diabetic PLWH. The plots showed fewer TNai cells and enriched CD4+ and CD8+ TEM and TemRA cell fractions in the SAT compared to the matched peripheral blood.

The t-SNE findings were congruous with our results from 2-dimensional flow cytometry gating. For CD4+ T cells, the proportion of TNai and TCM cells were significantly lower compared to blood in all three groups, while the proportion of TEM and TemRA cells was significantly higher (**Figure 45**).

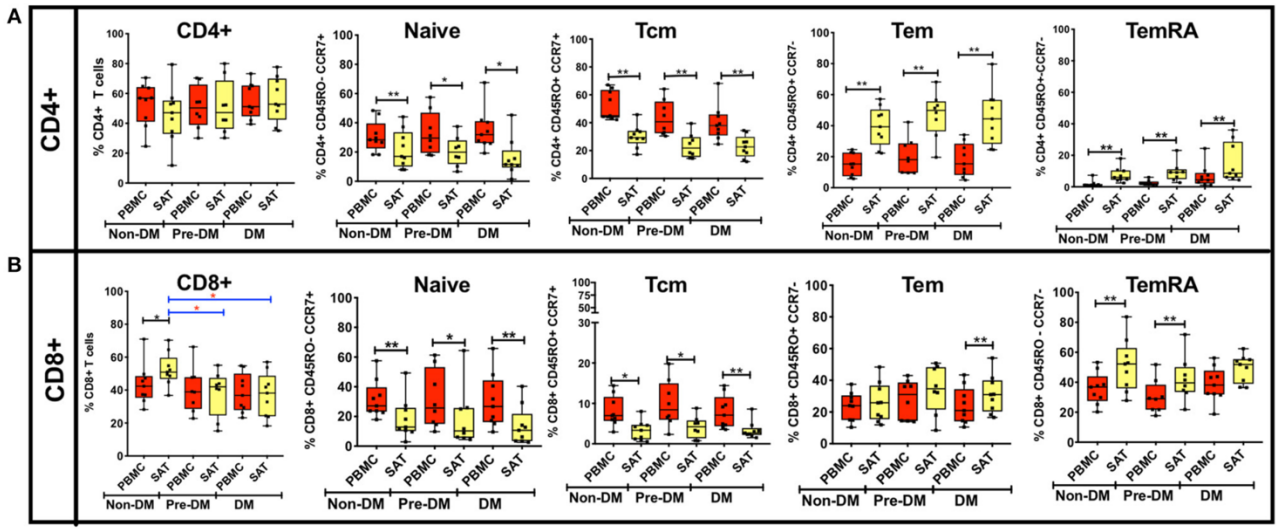


Figure 46. Analysis of CD4+ and CD8+ memory subsets by metabolic status in PLWH.

However, none of the adipose memory subsets in the pre- diabetics or diabetics were significantly different from the non- diabetics in pairwise comparisons. There

was less consistency in the relative proportions of CD8+ memory cells in the blood and adipose tissue. While CD8+ TNai and TCM cells were significantly lower in the SAT in all three groups (**Figure 46**), SAT CD8+ TEM cells were only significantly higher compared to blood in the diabetics, and SAT TemRA cells were only significantly higher in the non- and pre-diabetic groups. In summary, these findings suggest that while adipose tissue is enriched in CD4+ and CD8+ TEM and TemRA cells compared to blood, the relative distribution of naïve and memory T cells within SAT does not markedly vary by metabolic status in PLWH.

Next, we assessed SAT memory T cells according to BMI category and age (**Figure 47, Figure 48**). We observed no significant differences in SAT CD4+ and CD8+ memory T cell populations according to BMI status.

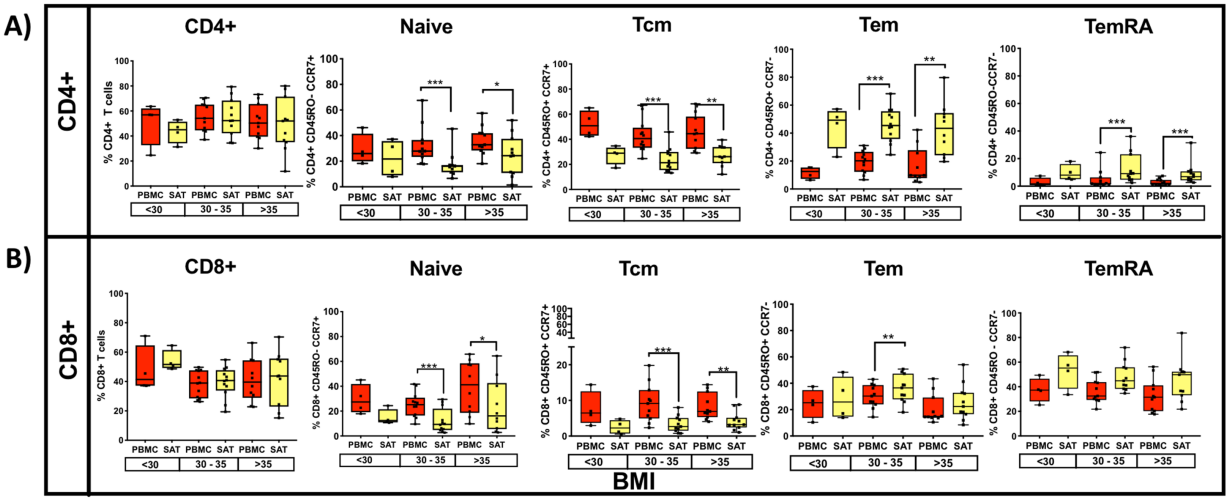


Figure 47. Analysis of CD4+ and CD8+ memory subsets and BMI. The graphs on the top row show frequencies of TNai, TEM, TCM and TemRA CD4+ cells in SAT and PBMC from all twenty-six subjects. Plots separated based on BMI (<30 [n=4], 30-35 [n=12], > 35 kg/m2 [n=10]. B) The bottom row shows frequencies of CD8+ TNai, TEM, TCM and TEMRA cells in adipose tissue and PBMC according to BMI. The box and whiskers plot indicate mean + SD. Wilcoxon matched pair signed test was used to calculate statistics between matched PBMC

and adipose pairs; Mann-Whitney test was used to analyze differences in SAT and PBMC T cell subsets between metabolic groups; ***< 0.001, **< 0.01, * p < 0.05.

When we compared naïve and memory T cell subsets by age, we found the proportion of CD4+ TNai cells in SAT from PLWH <35 years (34% of total CD4+ cells) was significantly higher than in older subjects 35– 55 years (12%, p < 0.05) and those >55 years (16%, p < 0.01). The relative reduction in CD4+ TNai cells appeared principally due to an increase in the proportions of TEM cells at higher ages; TEM cells constituted 29% of total CD4+ cells in participants <35 years, but rose to 50% in those 35–55 years (p = 0.06) and 47% in those >55 years (p < 0.01). These differences were less pronounced for CD8+ T cells; the TEM proportion in PLWH >55 years was significantly higher than those <35 years (38% vs. 22%, p < 0.05), but the difference was not significant for those ages 35– 55.

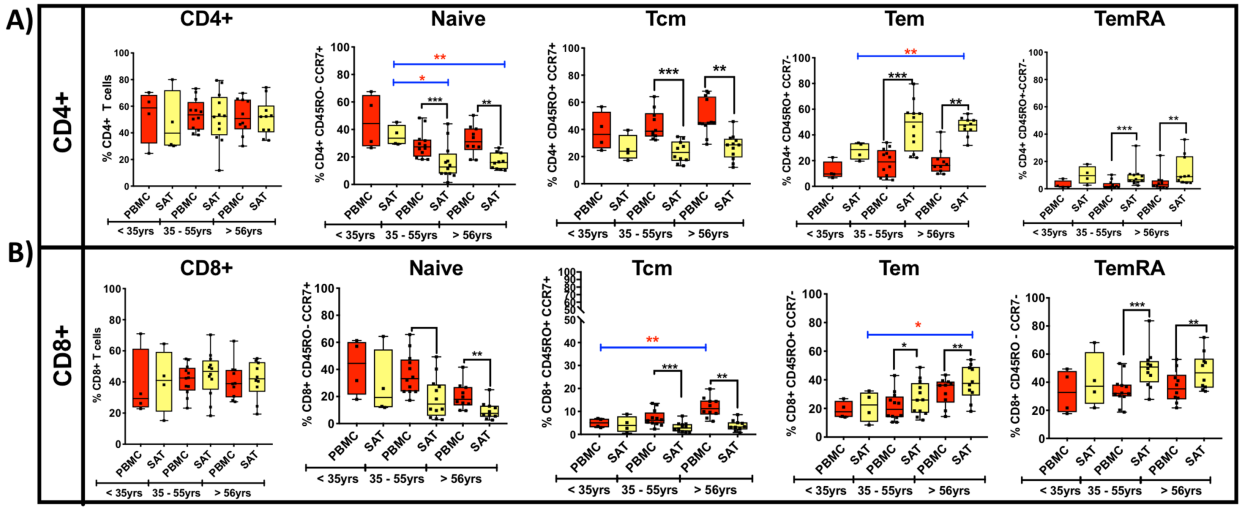


Figure 48. CD4+ and CD8+ TEM proportions in SAT increase with age while naive T cells decrease with age. Frequencies of CD4+ TNai, TEM, TCM and TemRA cells in adipose tissue and PBMC from all twenty-six subjects. Plots separated based on Age (<35 [n=4], 35-55 [n=12], > 56 years [n=11]). B) The bottom row shows frequencies of CD8+ TNai, TEM, TCM and TemRA cells in SAT and PBMC. The box and whiskers plot indicate mean + SD. Wilcoxon matched pair signed test was used to calculate statistics between matched PBMC and adipose pairs; Mann-Whitney test used to analyze differences in SAT and PBMC T cell subsets between metabolic groups; ** p< 0.01, * p < 0.05.

Our findings were similar to a prior study of health donors, which found the proportion of total CD4+ memory (CD45RO+) T cells in lung and mesenteric lymph nodes, spleen, ileum and colon also rose with increasing age.[364]

CD4+ T cell CD69 expression increases with progressive glucose intolerance in PLWH

CD69 is an inducible, early-activation indicator which also serves as a putative tissue-resident marker on memory T cells in human, as well as in animal, mucosal and lymphoid tissues,[364,366,367] but is largely absent on memory T cells in blood.[364] CD69 has been used as a marker of adipose tissue resident memory cells in animals,[365] including in SIV-infected macaques,[318] as well as in prior studies of

PLWH.[349,350] At present, there are few studies on CD69 expression on adipose tissue T cells in PLWH, none of which has looked at their link with metabolic status.

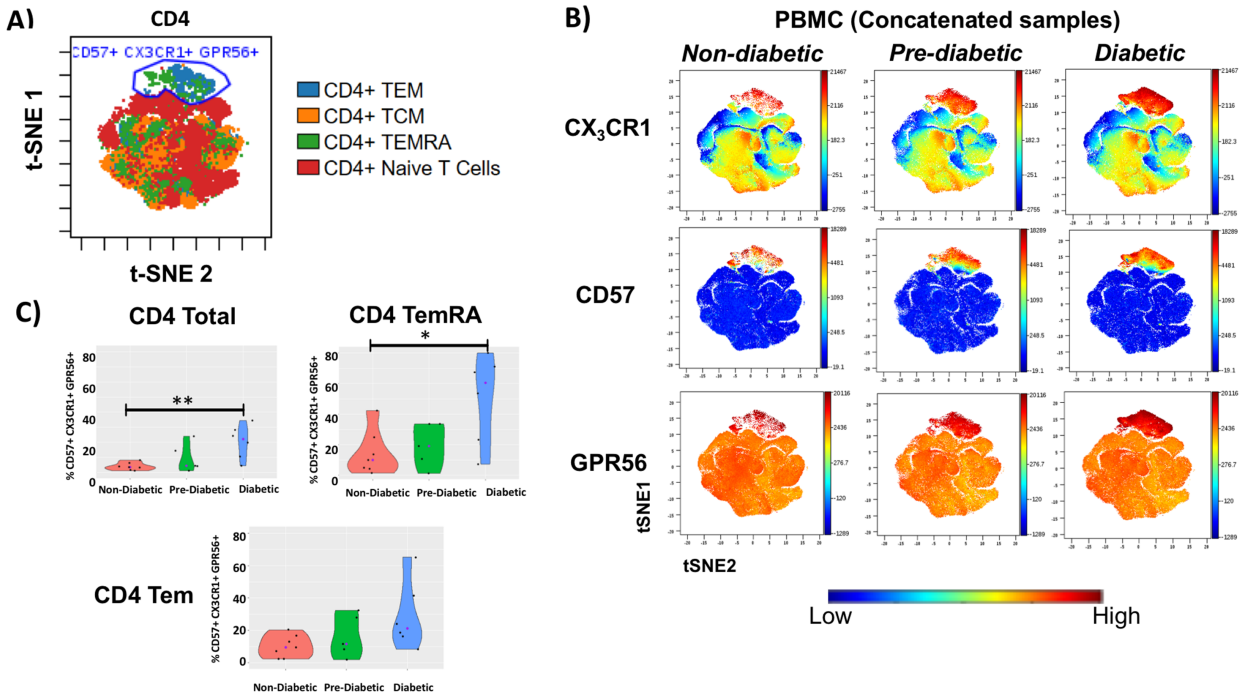


Figure 49. CD4+ T cells co-expressing CD57, CX_3CR1, and GPR56 in PBMCs are higher in pre-diabetic and diabetic PLWH. A) Visualization of groups of CD4+ TNai, TEM, TCM and TemRA cells from peripheral blood B) Concatenated viSNE plots of non-diabetic (n=7), pre-diabetic (n= 5) and Diabetic (n=6) PLWH showing clusters of cells expressing CX_3CR1, CD57, GPR56 and CD69 C) Violin plots showing percentage of CD4+ total, TEM and + TemRA cells co-expressing CD57, CX_3CR1 and GPR56. Mann-Whitney test used to analyze differences between unpaired samples; ** $P < 0.01$, * $P < 0.05$.

We measured CD69 expression on memory T cell subsets in SAT and blood. As reported in prior studies of PLWH, CD69 expression was present on CD4+ T cell subsets from SAT but almost absent in peripheral blood (**Figure 50**),[349,350] whereas CD69 expression on CD8+ T cells from SAT and blood was similar (data not shown). While we observed the relative proportions of SAT CD4+ and CD8+ TCM, TEM, and TemRA cells to be similar regardless of metabolic status in PLWH (**Figure 50**), expression of CD69 on CD4+ T cells rose with progressive glucose intolerance in a

step-wise progression from non-diabetic, to pre-diabetic, to diabetic. Compared to non-diabetics, diabetic PLWH had significantly higher CD69 expression on total CD4+ T cells, TCM, and TEM ($p < 0.01$ for all three), as well as TNai and TemRA ($p < 0.05$ for both, **Figure 50**).

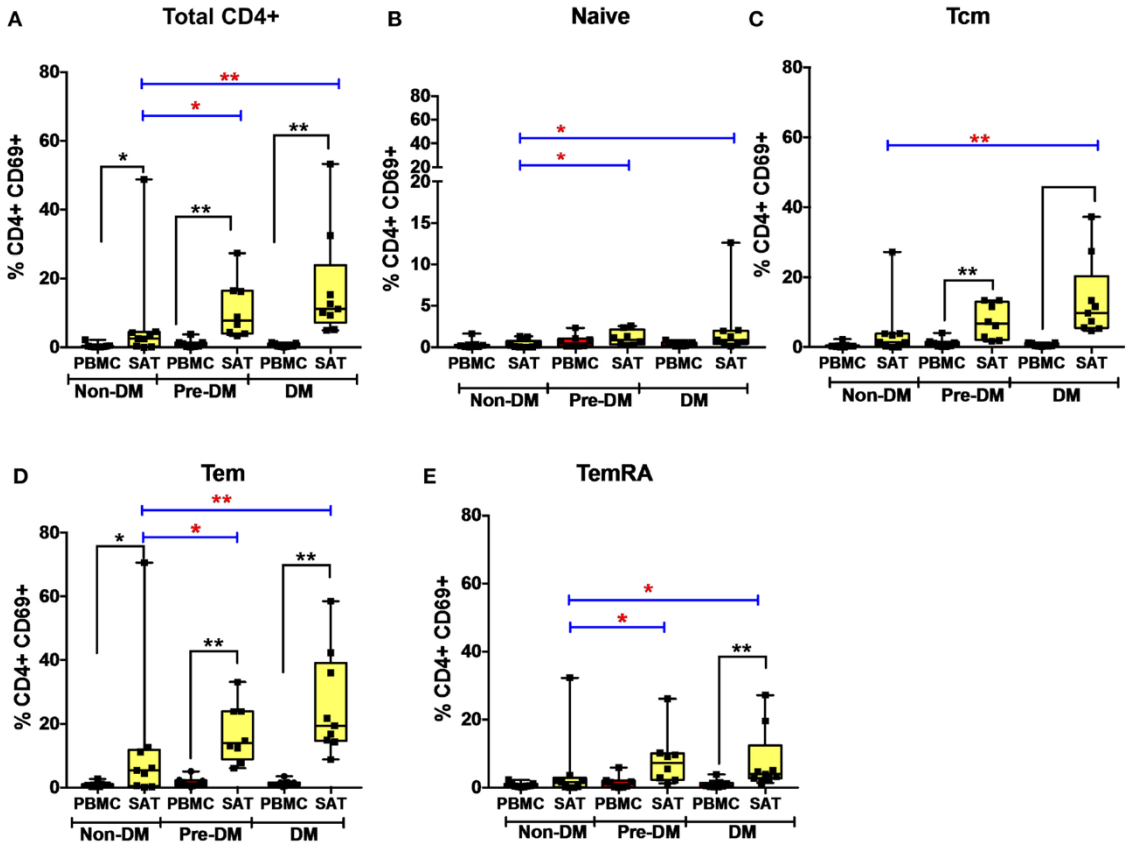


Figure 50. CD69 expression on subcutaneous adipose tissue CD4+ T cells increases with progressive glucose intolerance. (A–E) Frequencies of CD4+ total T cells, TNai, TCM, TEM, and TemRA cells expressing the CD69 activation and putative tissue-residence marker in subcutaneous adipose tissue (SAT) and blood (PBMC). The box and whiskers plot indicate mean ± SD. Wilcoxon matched-pair rank test was used to calculate differences between PBMC and SAT. Mann-Whitney test used to calculate differences between groups; blue lines and red * depict differences between groups **p < 0.01, *p < 0.05.

Similarly, the pre-diabetics also had higher CD69 expression on SAT total CD4+ T cells and TNai, TEM, and TemRA ($p < 0.05$ for all) compared to non-diabetics, but not on TCM ($p < 0.07$). In contrast, we did not observe any significant differences in CD69 expression on CD8+ T cells according to metabolic status.

Given the stepwise progression of CD69 expression on CD4+ T cell subsets with rising glucose intolerance, and the significant pairwise comparisons, a linear regression model was used to assess CD69 expression according to metabolic status. Mean CD4+ T cell CD69 expression for non-diabetics, pre- diabetics, and diabetics is shown in **Table 4**.

T cell subset	Non-diabetic $n = 9$	Pre-diabetic $n = 8$	Diabetic $n = 9$	*p*-Value*
Total $CD4^+$ T cells	7.5%	10.9%	17.2%	**0.004**
CD4 T_{NAIVE}	0.6%	1.1%	2.0%	**0.03**
CD4 $T_{CENTRAL\ MEMORY}$	5.9%	7.8%	11.9%	**0.02**
CD4 $T_{EFFECTOR\ MEMORY}$	14.7%	18.2%	22.3%	**0.04**
CD4 $T_{EFFECTOR\ MEMORY\ RA^+}$	4.9%	8.9%	7.3%	**0.04**
Total $CD8^+$ T cells	5.4%	6.2%	10.0%	0.09
T_{NAIVE}	1.5%	0.6%	4.2%	0.28
$T_{CENTRAL\ MEMORY}$	6.9%	5.5%	12.8%	0.45
$T_{EFFECTOR\ MEMORY}$	11.9%	9.3%	17.0%	0.19
$T_{EFFECTOR\ MEMORY\ RA^+}$	5.0%	5.7%	9.0%	0.12

Table 4. Mean SAT CD4+ and CD8+ T cell CD69 expression according to glucose tolerance. *Univariable linear regression. The association of total CD4+ T cell CD69 expression with progressive glucose intolerance was robust to adjustment for BMI (p = 0.03) and to age (p = 0.01) in separate models. The bolded values are p-values <0.05, indicating statistical significance.

Frequencies of CD69+ cells were natural-log transformed to improve normality (Shapiro-Wilk $p < 0.01$ for all untransformed CD4+ and CD8+ T cell subsets). Expression of CD69 on total CD4+ T cells rose with progressive glucose intolerance (**Table 4**, $p = 0.004$), which was robust to adjustment for BMI ($p = 0.03$ for metabolic status) and to age ($p = 0.01$ for metabolic status) in separate models (data not shown). Among CD4+ T cell subsets, progression from non-diabetic to diabetic groups was accompanied by increased CD69 expression on TCM ($p = 0.02$), TEM ($p = 0.04$), and TemRA ($p = 0.04$) cells.

CD57 expression is higher on SAT CD4+ and CD8+ T cells, but does not vary with metabolic status in PLWH

We previously reported a higher proportion of late- differentiated, CD57+ CD8+ T cells in the SAT of non-diabetic PLWH compared to blood (37 vs. 23%, $p < 0.01$).[295] CD57 is a terminally-sulfated glycan carbohydrate epitope found on T cells and natural killer (NK) cells which serves as a marker of late differentiation, though there is limited consensus as to whether CD57 is a marker of an inability to proliferate in response to antigen stimulation, signifies reduced replicative capacity, or represents an increased susceptibility to activation- induced apoptosis.[368–370] Prior studies have shown that CD57 expression on CD4+ and CD8+ T cells is higher in the blood of PLWH compared to that of HIV-negative controls.[370–372] CD8+ T cells

expressing CD57 produce more interferon-γ and TNF-α after TCR stimulation than CD57− T cells, and CD57+ CD8+ T cells have a distinct gene expression profile characterized by greater cytotoxic effector potential (e.g., production of perforin, granzymes, and granulysin).[373,374] Additionally, a higher percentage of CD57+ expression on T cells has been implicated in other inflammatory diseases, such as rheumatoid arthritis[375] and beryllium-induced disease.[376]

Given the potential pro-inflammatory effects of CD57+ T cells in adipose tissue, we compared CD4+ and CD8+ T cell expression of CD57 in SAT vs. blood, and according to metabolic status (**Figure 51**).

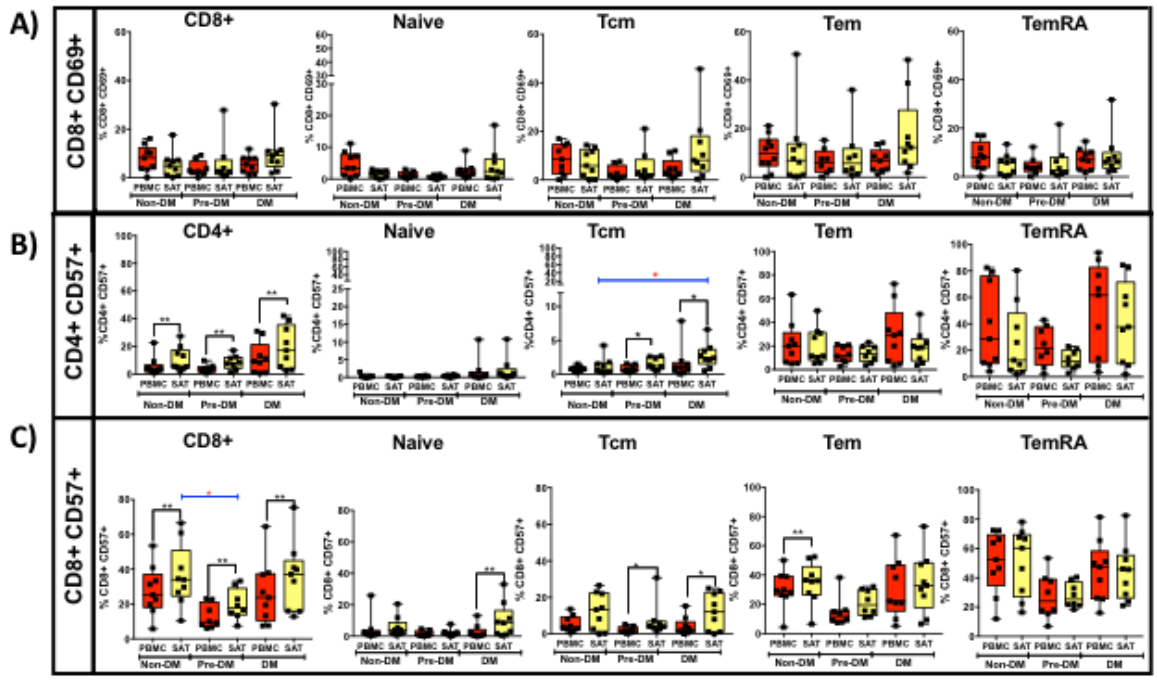


Figure 51. CD57 expression on CD4, CD8 T cells and CD69 expression on CD8 T cells. Total CD8+ T, TNai , TEM, TCM and TemRA cells expressing the CD69 marker, and B) Frequencies of total CD4+ cells and subsets expressing CD57 and C) total CD8+ cells and subsets expressing CD57. The box and whiskers plot indicate mean + SD. Wilcoxon matched-pair rank test was used to calculate differences between PBMC and SAT. Mann-Whitney test used to calculate differences between groups; blue lines and red * depict differences between groups** $p < 0.01$, * $p < 0.05$.

In all three metabolic groups, expression of CD57 on SAT CD8+ T cells was higher compared to blood, confirming our prior study findings.[295] Among the memory subsets, CD57 expression was highest on CD4+ TemRA and CD8+ TEM and TemRA cells. However, we did not observe an increase in CD57 expression on either CD4+ or CD8+ memory cell subsets with progressive glucose intolerance, with the exception of higher CD57 expression on CD4+ TCM in diabetic individuals compared to those without diabetes.

CD4+CD69lo cells co-expressing CD57, CX_3CR1, and GPR56 are associated with increasing glucose intolerance

Adipose tissue serves as a reservoir for both latently HIV-infected CD4+ T cells and free HIV RNA virus, and in the Genotype-Tissue Expression (GTEx) project adipose tissue contained one of the highest levels of CMV transcripts.[295,349,354,377] These findings suggest adipose tissue may also serve as a site for antiviral immune activity. Recent studies of CD4+ TemRA cells have identified major subsets based on G protein-coupled receptor GPR56 expression, with virus-specific cells more frequently GPR56+ and more clonally expanded compared to GPR56− cells.[378] Increased expression of GPR56 and killer-like receptors (KLR) has been linked to higher cytokine expression by memory CD4+ T cells, including T cells obtained from

liver tissue.[379] CD4+ and CD8+ TemRA cytotoxic T cells expressing GPR56 have also been associated with increased co-expression of CX_3CR1.[378,380,381] CX_3CR1 receptor is expressed on terminally differentiated T cells, gamma-delta T cells and NK cells, and has also been identified as a marker of anti-CMV T cells.[111,381–383]

To assess the presence of GPR56+ CX_3CR1+ T cells, and the parent memory cell population(s), we used t-SNE and viSNE to identify surface marker clusters that differed between non- diabetic, pre-diabetic, and diabetic individuals. We identified a group of cells that expressed CD57, lacked CD69, and also co-expressed GPR56 and CX_3CR1. A representative plot of t- SNE maps generated from CD4+ gated T cells showed that the CD57+ CX_3CR1+ GPR56+ co-expression was mainly on TEM and TemRA cells (**Figure 52A, see next page**).

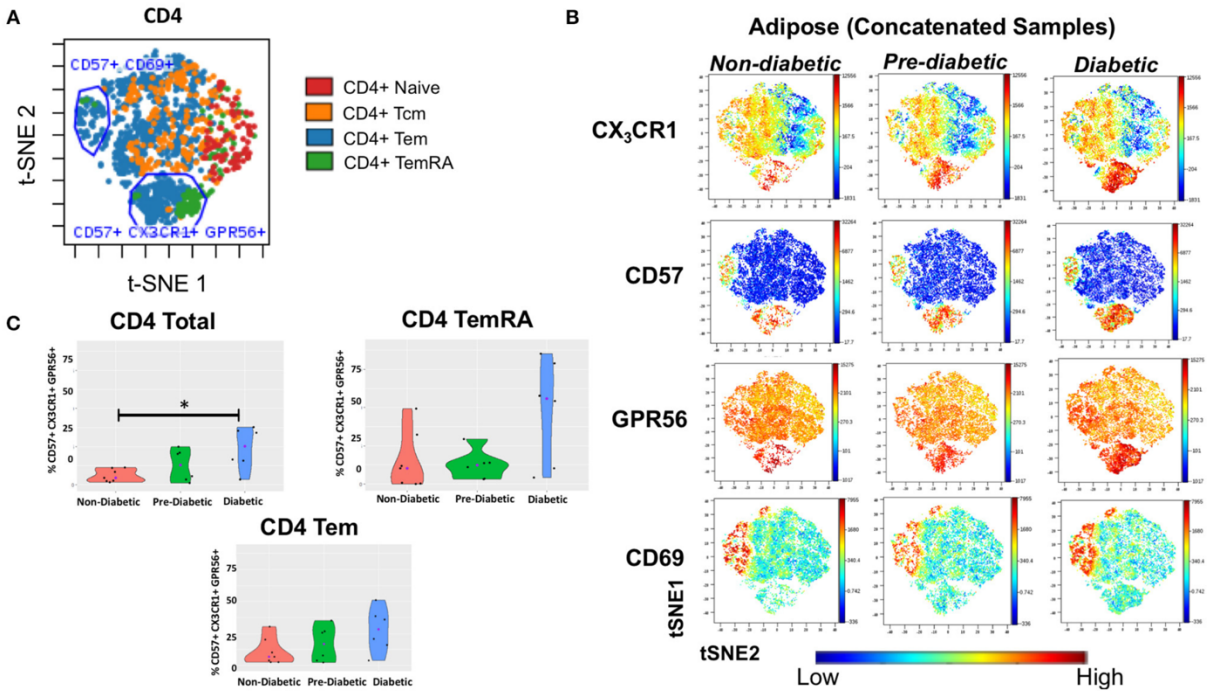


Figure 52. CD4+ T cells co-expressing CD57, CX_3CR1, and GPR56 in subcutaneous adipose tissue increase with progressive glucose intolerance. viSNE plots were generated in Cytobank to identify groups of cells that differ between non-diabetic, pre-diabetic and diabetic categories. (A) Clusters of CD4+, TNai, TEM, TCM, and TemRA cells from subcutaneous adipose tissue (SAT). (B) Concatenated viSNE plots of non-diabetic, pre-diabetic, and diabetic PLWH showing clusters of cells expressing CX_3CR1, CD57, GPR56, and CD69. (C) Violin plots showing percentage of total CD4+, TEM, and TemRA cells co-expressing CD57, CX_3CR1, and GPR56. Mann-Whitney test used to analyze differences between metabolic groups; *P < 0.05.

Concatenated viSNE plots of non-diabetic, pre-diabetic and diabetic PLWH demonstrated two distinct clusters of cells: CD57−/+ CD69+ CX_3CR1−/+ GPR56−/+ and CD57+ CD69lo CX_3CR1+ GPR56+ (**Figure 52B**). A significantly larger proportion of total CD4+ and CD4+ TemRA cells in SAT from diabetics were CD57+ CD69lo CX_3CR1+ GPR56+ compared to SAT from non-diabetics (p = 0.051), and approached significance for CD4+ TEM cells (p = 0.07) (**Figure 52C**). Of note, we observed a similar population of CD4+ TemRA and TEM cells co-expressing CD57+ CX_3CR1+ GPR56+ in matched PBMC samples which, again, increased with glucose intolerance

and were significantly different between diabetic and non-diabetic PLWH (total CD4+ [$p < 0.01$], CD4+ TEM [$p = 0.05$] and CD4+ TemRA [$p < 0.05$]) (data not shown). Similar plots of CD8+ T cells were generated for adipose tissue and PBMC (data not shown) from PLWH. As with the CD4+ T cells, CD57+CX_3CR1+GPR56+ co-expressing CD8+ T cells were predominantly TEM and TemRA, though there were no significant differences between non-diabetics and diabetics.

Lastly, we assessed CD57+ CX_3CR1+ GPR56+ co-expression on SAT CD4+ and CD8+ T cells from our HIV-negative controls (**Figure 53**). We identified two distinct clusters of CD4+ and CD8+ TEM and TemRA cells co-expressing CD57, CX_3CR1, and GPR56 though the expression was more diffuse compared to SAT cells from PLWH, particularly for GPR56.

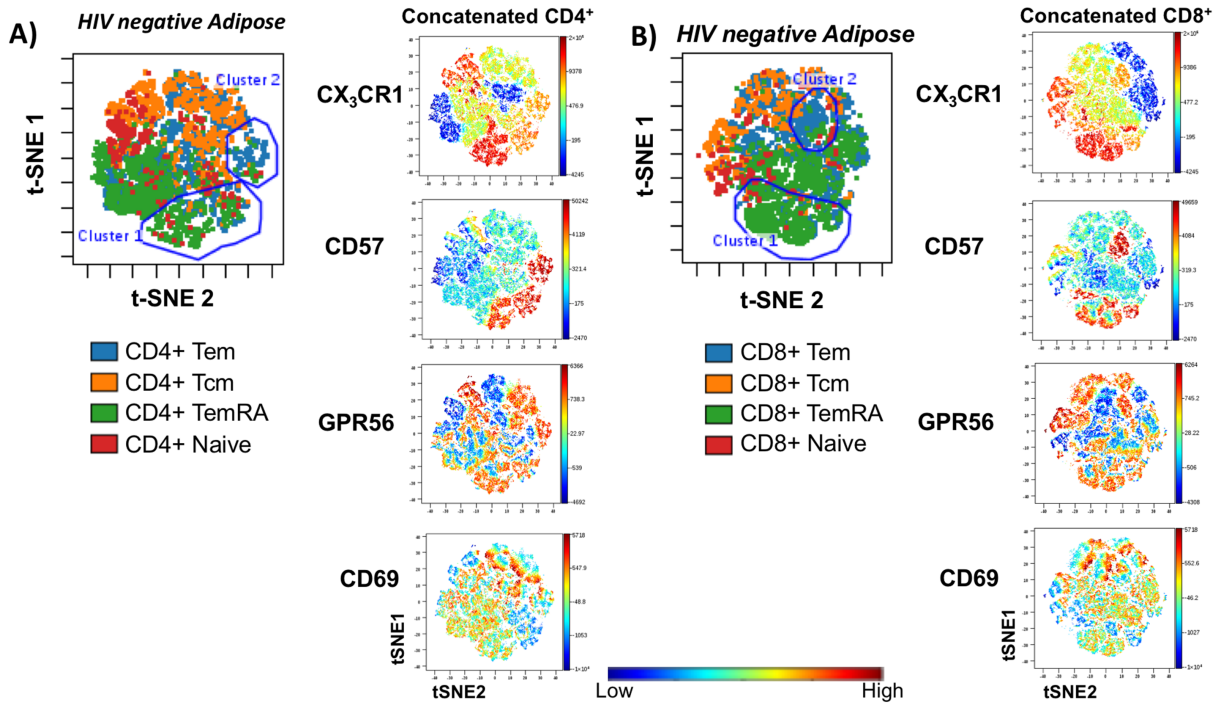


Figure 53. CD4+ and CD8+ T cells co-expressing CD57, CX_3CR1, and GPR56 in SAT of HIV-negative persons. Representative plot showing clusters of naïve, TEM, TCM and TemRA CD4+ T from SAT and Concatenated viSNE plots (n=8) showing clusters of cells expressing CX_3CR1, CD57, GPR56 and CD69. B) Representative plot showing clusters of naïve, TEM, TCM and TemRA CD8+ T from SAT and concatenated viSNE plots (n=8) showing clusters of cells expressing CX_3CR1, CD57, GPR56 and CD69.

Taken together, our results demonstrate a population of CD57+CX_3CR1+GPR56+ co-expressing predominantly CD4+ TemRA cells in both blood and SAT of PLWH which appear to be associated with diabetes. The expression of the GPR56 and CX_3CR1 markers may indicate these cells are virus-specific TemRA cells, a population that could contribute to inflammation in adipose tissue.

SAT CD4+ and CD8+ T cell memory subsets compared by HIV status

Couturier et al. identified a major shift in the CD4:CD8 T cell ratio in PLWH compared to HIV-negative controls,[349] a finding that has been replicated in subsequent HIV and SIV studies.[295,318,350] Specifically, adipose tissue stromal vascular fraction (SVF) CD3+ T cells from individuals without HIV were predominantly memory CD4+ CD45RO+ T cells rather than memory CD8+ T cells. In contrast, this distribution was reversed in PLWH, with more memory CD8+ T cells in the adipose tissue, which represented a ~50% enrichment in memory CD8+ T cells over the peripheral blood, and could not be attributed to differences in peripheral blood T cell subsets.

Phenotypic analysis of SAT memory T cell subsets in PLWH could provide insight on possible mechanisms contributing to the profound shift in the CD4:CD8 ratio that accompanies HIV infection. Therefore, we compared the proportion of CD4+ and CD8+ naive, TCM, TEM, and TemRA cells between 8 non- and pre-diabetic HIV-negative persons and 9 PLWH; these subjects were selected from the 26 PLWH in our cohort based on similar age, HbA1c and BMI values (**Table 3**). As in prior studies, we found that SAT from PLWH is enriched in CD8+ T over CD4+ T cells (51 vs. 47%, respectively) compared to HIV-negative persons (21 vs. 66%; **Figure 54A,D**). Although PLWH have a much higher proportion of CD8+ T cells in the SAT, we found no significant difference in the overall proportions of SAT CD4+ and CD8+ memory

cells between PLWH and HIV-negative persons (**Figure 54C,F**), and the distribution of CD4+ and CD8+ memory T cell subsets was remarkably similar (**Figure 54B,E**). However, three notable differences were present: compared to PLWH, the HIV-negative persons had a significantly higher percentage of CD4+ TEM (58 vs. 39%, p = 0.02), a lower percentage of CD4+ TCM (15 vs. 29%, p = 0.07) in their SAT (Figure 6B), and a significantly higher percentage of CD8+ TCM compared to PLWH (6.4 vs. 3.4%, p = 0.03) (**Figure 54E**).

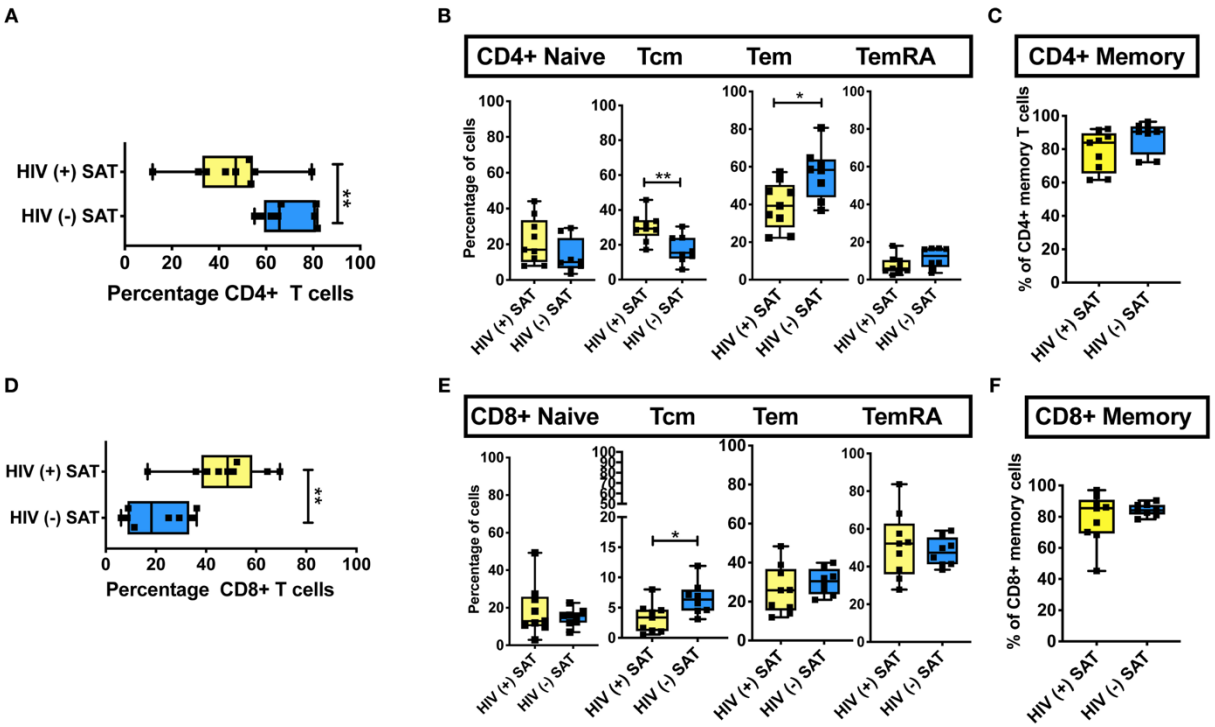


Figure 54. Comparison of CD4+ and CD8+ T cell subsets in subcutaneous adipose tissue of PLWH and match HIV-negative comparators. Cells recovered from subcutaneous adipose tissue (SAT) of PLWH (n = 9) and HIV-negative controls (n = 8) with similar BMI and hemoglobin A1c were compared. (A) CD4+ T cells as a percentage of total live CD3+ T cells. (B) Frequencies of TNai, TEM, TCM, and TemRA cells in SAT of PLWH and HIV-negative controls. (C) Total CD4+ memory T cells (TCM + TEM + TemRA). (D) Similar analysis also done on CD8+ T cells, which are displayed as percentage of total live CD3+ T cells. (E) Frequencies of CD8+ TNai, TEM, TCM, and TemRA cells. (F) Total CD8+ memory T cells. The box and whiskers plot indicate mean ± SD. Mann-Whitney test used to analyze differences between unpaired PLWH and (–) subjects; **p < 0.01, *p < 0.05.

These findings suggest the profound change in the SAT CD4:CD8 T cell ratio observed in PLWH is not driven by the enrichment or depletion of a single memory T cell phenotype. Rather, it is phenotype agnostic and involves shifts in disparate naïve and memory T cell phenotypes from both the CD4+ and CD8+ lineages. This raises the possibility that a chemotactic signal from SAT in PLWH is recruiting CD8+ T cells more robustly in the context of infection with HIV. However, the similar subset distributions suggest this signal, if present, may represent the amplification of a normal physiologic pathway as opposed to an "HIV-specific" process.

Comparison of SAT gene expression in PLWH and HIV-negative people with similar glucose tolerance

Given the increased proportion of CD8+ T cells in SAT from PLWH, we performed a sub-study to investigate potentially relevant adipose tissue immune signaling pathways upregulated in the context of HIV infection. Extracted SAT mRNA from 6 PLWH and 7 HIV-negative individuals was assayed using a NanoString human inflammation panel to quantitate expression of over 250 genes representing a broad range of immune pathways (**Figure 55A**). The overall fold change in genes expressed from SAT of PLWH over HIV-negative is shown in the volcano plot (**Figure 55B**). In general, we found increased expression of chemokine receptors (CXCR2, CXCR1, CXCR4) and ligands (CCL5, CXCL5) in those with HIV. Additional genes

including ALOX12 and IL12A were also elevated. We further analyzed inflammatory gene pathways defined by NanoString (**Figure 55C**). CXCR2, CXCR1, CXCR4, TLR2, and TLR8 gene expression were significantly higher in PLWH compared to HIV-negative (**Figure 55C**) whereas CD4 and CXCL9 were higher in HIV- negative individuals. Future studies are needed to identify the cells expressing these genes, and whether receptor and ligand expression might account for differences in the cells that traffic to SAT in PLWH.

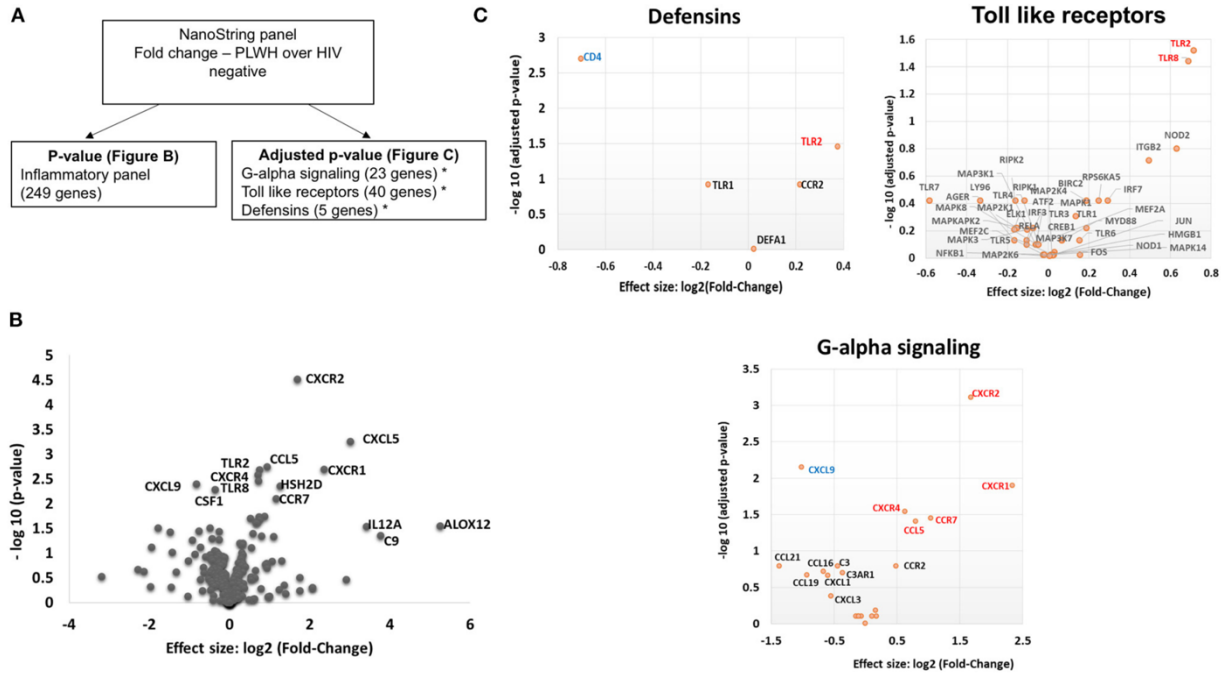


Figure 55. PLWH have a higher CXCR1, CXCR2, CXCR4, CCL5, CXCL5, TLR2, and TLR8 RNA expression within the subcutaneous adipose tissue compared to HIV-negative individuals with similar glucose tolerance. RNA was extracted from subcutaneous adipose tissue of PLWH (n = 6) and HIV negative individuals (n = 7). (A) Schematic showing data analysis. (B) Volcano plot showing log2 fold change (PLWH vs. HIV negative) vs. log10 p-value. Relevant genes out of 249 genes are labeled. (C) Sub-groups of inflammatory gene panels available through NanoString were compared and adjusted p-values calculated within these groups. Genes in G-alpha signaling, defensins and Toll-like receptors with significant findings are shown. Red denotes genes that are higher in PLWH and blue denotes higher gene expression in HIV negative.

Discussion

Several recent studies have uncovered a profound change in the balance of adipose tissue CD8+ and CD4+ T cells in PLWH, and similar studies in SIV-infected macaques indicate that the relative enrichment of CD8+ T cells stems from viral infection rather than from ART treatment or CD4+ cell depletion. However, the characteristics and consequences of this CD8+ cell enrichment are generally unknown. Here, we demonstrate that the adipose tissue of PLWH with viral suppression on long-term ART is a reservoir of CD4+ and CD8+ TEM and TemRA cells; two cell types with high pro-inflammatory potential when stimulated. Furthermore, we show that expression of CD69, a putative marker of TCR-linked activation and tissue residency, increased on CD4+ TEM and TemRA cells in a stepwise manner from non-diabetic, to pre-diabetic, to diabetic individuals. We also identify a phenotypically unique population of CD4+ T cells co-expressing CX_3CR1, GPR56, and CD57 that is specifically enriched in the SAT of diabetic PLWH. While the significance of these cells is currently unclear, CX_3CR1 and GPR56 have previously been identified on cells with antiviral functions.[111,381–384] Finally, our results demonstrate that the enrichment of CD8+ T cells over CD4+ T cells in PLWH as compared to HIV- negative individuals is relatively non-specific in the sense that the change is not attributable to a profound increase or reduction in a single memory cell subset.

Our findings contribute to the growing body of literature on both adipose tissue T cells in general, as well as the immune phenotype of adipose tissue in the context of metabolic disease among PLWH. Han *et al*. recently demonstrated that white adipose tissue in mice is a major reservoir for memory T cells with potent proliferative, effector, and protective potential.[365] The adipose tissue T cells predominantly expressed CD44, a marker of antigen experience, and were CD62L negative, which is consistent with TEM and TRM populations in mice. Furthermore, approximately half the CD44+CD62L− CD8+ and CD4+ cells expressed CD69. The study also assessed recall responses of adipose tissue T cells in mice previously challenged with antigen, demonstrating an influx of monocytes and neutrophils, as well as highly reactive memory T cells. Specifically, these memory T cells indicated rapid and enhanced effector potential with upregulation of several genes involved in antimicrobial defenses. Furthermore, antigen re-challenge led to downregulation of several metabolic pathways (including lipid biosynthesis and cholesterol and long-chain fatty-acyl-CoA metabolic processes) as well as to a detectable reduction in serum levels of adiponectin and cholesterol, further highlighting the potential role memory T cells play in metabolic disease.

These animal studies suggest the accumulation of TEM and TemRA cells in the adipose tissue of PLWH may represent a potent source of inflammation in the setting of antigen stimulation. Viral pathogens, including HIV and CMV, are found in

adipose tissue and could serve as a chronic stimulus for TEM and TemRA cells, with downstream effects on metabolic function (17, 21, 39, 51).[295,349,365,377] Our observation of increasing TEM and TemRA CD69 expression with declining glucose tolerance in PLWH may indicate the enrichment of a resident memory phenotype in diabetic individuals. Our study design precludes an assessment of whether the presence of increased CD69+ TEM and TemRA cells preceded or followed the development of glucose intolerance, though future longitudinal studies in PLWH with early indications of metabolic disease could address this question. Furthermore, future studies to identify the receptor specificity of the CX_3CR1, GPR56, and CD57-expressing CD4+ cells, and experiments to co-culture these cells with adipocytes, may help characterize the role of these immune cells in adipose tissue, identify their cytokine expression patterns, and explore the potential effects of these cells on adipocyte function.

An early study of SAT and VAT from PLWH by Couturier *et al*. identified major differences in CD4+ and CD8+ T cell populations compared to HIV-negative controls,[349] which were subsequently reported in other HIV and SIV studies.[295,318,350] In the HIV-negative controls, adipose tissue SVF CD3+ T cells in SAT were predominantly memory CD4+ CD45RO+ T cells (61%) with fewer memory CD8+ T cells (15%), while this distribution was reversed in PLWH, with more memory CD8+ T cells (46%) compared to memory CD4+ T cells (35%). This represented an ~50%

enrichment in memory CD8+ T cells over the blood in the subjects with HIV and was not reflective of differences in peripheral blood T cell subsets between the two groups.

Notably, the Couturier et al. study found significant differences in the rates of CD69 expression on memory CD4+ and CD8+ T cells in adipose tissue vs. blood: <5% of these cells expressed CD69 in the blood, compared to 60–67% in adipose from PLWH and 61–72% in the adipose from HIV negative persons. A similar study comparing activation markers on adipose tissue T cells in HIV-negative lean, overweight and obese individuals using fine needle aspiration found ~5–10% CD69 expression on SAT CD4+ T cells and ~25% expression on CD8+ T cells.[385] We observed the highest CD69 expression on TEM cells in diabetic PLWH (mean 22%). This was significantly higher than CD69 expression on TEM cells from non-diabetics (mean 15%) and pre-diabetics (mean 18%), and 20-fold higher compared to TEM cells in blood. The reason for the lower CD69 expression in our cohort compared to Couturier et al. is unclear and may reflect differences in CD69 expression in adipose tissue samples collected after death or by surgical resection as opposed to liposuction aspirates processed within 30–60 min of collection in our study, differences in tissue processing to extract T cells, or could be explained by residual peripheral blood in samples obtained via liposuction.

Prior studies have also demonstrated a significantly higher proportion of SAT Treg cells (defined as CD25+ FOXP3+ CD4+ T cells; a cell type thought to exert an anti-inflammatory effect and reported to be depleted in obese adipose tissue) in PLWH compared to HIV-negative persons, with no major differences in TH 1 and TH 17 pro-inflammatory subsets.[283,350] The proportion of TH1 CD4+ T cells (expressing intermediate or high levels of T-bet) did not differ according to HIV status, while TH2 CD4+ T cells (expressing GATA-3) were barely detected in the adipose tissue from both PLWH and HIV-negative persons. While CD4+ T cell expression of HLA-DR (24%) in SAT was higher compared to blood in the HIV-negative, similar levels of SAT HLA-DR expression were observed in the PLWH. Furthermore, SAT CD4+ T cell expression of PD-1 expression was much higher compared to PBMCs (45 vs. 3%), but again there were no significant differences in SAT according to HIV status.[350]

Two studies of SIV-infected cynomolgus macaques confirm the adipose tissue CD8+ T cell enrichment observed in PLWH is a viral phenomenon, rather than related to ART treatment.[354,386] In both studies, SIV infection was associated with a higher percentage of CD8+ T cells in both the SAT and VAT compared to non-infected animals. One study also demonstrated the inverted CD8:CD4 ratio was not driven by a reduction in the total number of CD4+ T cells in infected animals; rather, SIV- infected animals had significantly higher density of CD8+ T cells in VAT and a somewhat higher density in SAT.[354] For both non-infected healthy and SIV-infected

monkeys, the majority of the adipose tissue CD4+ and CD8+ T cells were memory T cells (>94% CD95+), with a large fraction of activated cells marked by expression of CD69+ (62–84%) and CD25+ (3–13%).

A central question that remains unanswered is whether the increase in the proportion of adipose tissue CD8+ T cells in PLWH and macaques with SIV results from in situ proliferation vs. increased infiltration from the circulation. The high expression of CD69 on SAT memory cells suggests a “tissue resident” phenotype, but does not answer the question of whether these cells are expanded clones or prior migrants. On the one hand, a study by Damouche et al. found no significant differences in the proportion of CD4+ or CD8+ T cells expressing Ki-67, indicative of cycling and recently divided cells, between PLWH vs. HIV-negative controls.[350] The authors suggested the low percentage of Ki-67+ cells in SAT (<2%) from PLWH and HIV- negative subjects reflected minimal T cell proliferation within the tissue. Similarly, a study in macaques found no differences in proportions of Ki-67-expressing CD4+ or CD8+ T cells in animals with and without SIV, suggesting the higher density of CD8+ T cells in adipose tissue does not result from recent proliferation.[354] However, proliferation of memory T cells within adipose tissue is clearly demonstrated in other studies. Han et al. injected mice with pseudotuberculosis and evaluated memory T cells 4 weeks later. They showed

proliferation of TEM and TRM by Ki67 expression, bromodeoxyuridine labeling, and cell cycle stage analysis of memory T cells.[365]

The use of T cell receptor (TCR) sequencing of adipose tissue T cell subsets in future studies may provide insight into the clonal lineage of adipose tissue T cells. Previously, our group demonstrated that CD8+ TCR clonality is higher in SAT compared to blood in PLWH using bulk TCRβ CDR3 deep sequencing, where bias-adjusted V and J gene primers are used to amplify rearranged V(D)J segments.[290,295] In that study, the 10 most prevalent TCRβ clones comprised a significantly larger percentage of total clones in SAT (25%) compared to paired blood (16%), and the Shannon's Entropy index, a measure of overall repertoire diversity, was lower in adipose tissue compared to blood (4.39 vs. 4.46, respectively). Additionally, V-J gene pairing and gene usage differed between blood and adipose tissue, albeit not statistically significant, potentially due to the small sample size. While these findings are intriguing, the lower proportion of CD4+ and CD8+ TNai cells in SAT demonstrated in the current analysis may have had a role in the higher SAT clonality scores reported in the prior study. In addition to clonal lineage, TCR sequencing may also inform our understanding of potential antigen targets for adipose tissue T cells in PLWH. A recent study in mice demonstrated that diet-induced obesity is characterized by increased adipose tissue CD8+ T cell density, and the TCR repertoire of these CD8+ T cells is more clonal and positively charged

(in respect to amino acids).[353] This work was also the first to demonstrate that isolevuglandins (a group of negatively charged reactive gamma-ketoaldehydes generated by free radical oxidation) presented on adipose tissue macrophages from obese fat can independently activate T cells, potentially highlighting a mechanism contributing to inflammation.[353] Examining TCR charge and polarity, in addition to clonality, and investigating antigen presenting cells in the adipose tissue may be an important approach for further characterizing the adipose tissue immune milieu in PLWH vs. HIV-negative persons.

Profiling of adipose tissue mRNA expression in SIV-infected macaques showed significantly higher levels of IL-2, IL-7, and CCL19 in acute SIV compared to uninfected controls, which may contribute to the homing and survival of T cells.[386] In our gene expression sub-study on whole adipose tissue of PLWH and HIV-negative controls with similar glucose tolerance, we found >2- fold expression of IL12A, CXCR1, CXCL5, ALOX12, and C9 in the PLWH compared to HIV-negative. Analysis of subgroups of inflammatory gene pathways revealed significantly higher levels of TLR2, TLR8, CXCR4, CCR7, CCL5, CXCR1, and CXCR2 and lower levels of CXCL9 and CD4 in the PLWH.

Previous studies have linked inducible and increased expression of TLR2 in PBMC and SAT of obese individuals with type 2 diabetes.[387–390] TLR8 is an endosomal receptor that binds HIV-1 single-stranded RNA and is similar to TLR2 in

its MyD88-dependent activation of NF-kB. Previous studies of TLR8 expression indicate that it is most often absent or present at low levels in adipose tissue,[387] but that a positive correlation exists between plasma CRP levels and adipose tissue TLR8 expression. In our study, we are unable to discriminate whether the differences in TLR expression are due to alterations on T cells, macrophages, adipocytes or pre-adipocytes. However, we likely have identified a contribution of TLR8 to SAT inflammation that is more pronounced in PLWH compared to HIV-negative individuals with similar glucose tolerance and BMI.

Regarding chemokine ligands and receptors, M1 macrophages express higher levels of CCR7, CCL9, and CCL5 and lower levels of CXCR4 in comparison to M2 macrophages.[391] In the context of HIV infection, CXCR1 and CXCR2 are of interest because HIV-1 matrix protein p17 has been shown to mimic IL-8 and binds CXCR2 with high affinity, stimulating pro-angiogenic ERK downstream.[392] Angiogenesis has been linked to obesity via neovascularization-driven migration of adipocytes.[393] Further studies of chemokine receptors on CD4+ and CD8+ T cells may provide insight into whether HIV infection results in signals that preferentially recruit CD8+ T cells over CD4+ T cells independent of obesity.

Limitations

Our cohort study was cross-sectional and enrolled a total of 34 participants. We classified participants as non-diabetic and pre-diabetic based on fasting blood glucose and HbA1c values as recommended for clinical practice by the American Diabetes Association (diabetics were classified based on medication usage). In the future, the measurement of fasting insulin and use of the homeostatic model assessment (HOMA) may provide more nuanced stratification of participants. While a strength of our study was the collection of SAT via liposuction and rapid processing to extract T cells, samples may have contained residual peripheral blood, which would bias our comparisons of SAT and blood toward the null hypothesis. We were limited in the number of markers we could examine via flow cytometry and did not examine additional markers of tissue resident cells, as well as T regulatory cells and TH1/TH2 subsets. Our gene expression analysis of SAT from PLWH and HIV-negative persons was on whole adipose tissue as opposed to sorted SVF cells. Therefore, differences in genes cannot be directly linked to specific cell types.

Conclusions

Although PLWH can survive decades on effective ART, this success is offset by the rising burden of metabolic diseases affecting the HIV population.[333–335] Here, we

demonstrate that the adipose tissue of PLWH is a reservoir of CD4+ and CD8+ TEM and TemRA cells; two cell types with high pro-inflammatory potential when stimulated. Furthermore, we show that expression of CD69, a putative marker of TCR-linked activation and tissue residence, on CD4+ T cells increases in a stepwise manner from non-diabetic, to pre-diabetic, to diabetic individuals, which may reflect an interaction between cells of the adaptive immune system and adipocytes in the context of metabolic disease. We also identify a population of CD57+CX_3CR1+GPR56+ co- expressing CD4+ T cells that is specifically enriched in the SAT of diabetics. This could represent a group of virus-specific cells that contribute to inflammation in adipose tissue and potentially pre-dispose PLWH to metabolic disease, though further studies to assess the effects of these cells on adipocytes are needed. Finally, our results demonstrate that the enrichment of CD8+ T cells over CD4+ T cells in PLWH as compared to HIV- negative individuals is relatively non-specific in the sense that the change is not attributable to a profound increase or reduction in a single memory cell subset. Further studies are needed to understand the clonal lineage, TCR characteristics, antigen targets, chemokine receptors and the functional phenotype of CD57+CX_3CR1+GPR56+ T cells in the adipose tissue of PLWH, and to identify potential therapeutic targets.

Chapter V

Clonal EBV-like memory CD4+ T cell activation in fatal checkpoint inhibitor-induced encephalitis

N.B. This work in the entirety of the version below was in revision at *Nature Medicine*/under embargo when this dissertation was completed. Here, we outline the first known molecular profiling of immune-mediated neurotoxicity accompanying checkpoint inhibitor therapy. We describe a pathologic, cytotoxic CD4+ T cell population found in the brain of a checkpoint inhibitor-treated melanoma patient and identify a possible immune trigger originating from EBV reactivation in the brain. In addition to applying novel immunoinformatic approaches, we demonstrate proof-of-concept for performing single-molecule phenotyping of immune cells found in FFPE samples.

Summary

Immune checkpoint inhibitors induce durable anti-tumor responses across cancer types but may simultaneously unleash autoimmune-like toxicities affecting nearly all organs. The molecular underpinnings of these toxicities have not been well characterized. Here, we report a case of fatal encephalitis associated with immune checkpoint inhibitor therapy, with in-depth molecular evaluation of the affected brain tissue. We observed robust T cell infiltration of affected tissue, high degree of compensatory immune checkpoint expression, and a unique oligoclonal population of cytotoxic CD4+ T cells. Further analyses revealed homology of this clonal subset with known Epstein-Barr virus (EBV)-recognizing T cells, in addition to high-frequency EBV-specific T cells and latent EBV+ B lymphocytes. This is the first deep molecular description of immune checkpoint inhibitor neurotoxicity. Further, we

show a potential interaction between immune checkpoint inhibitor toxicity and a T cell response to EBV viral proteins.

Abstract

Checkpoint inhibitors produce durable responses in numerous metastatic cancers, but immune-related adverse events (irAEs) complicate and limit their benefit. IrAEs may affect any organ system in an idiosyncratic fashion; presentations range from mild and self-limited to fulminant and fatal. The molecular mechanisms underlying these irAEs are not well understood. Here, we report a case of fatal encephalitis arising during anti-PD-1 therapy. Histopathologic analysis of affected brain tissue revealed a robust T-cell infiltration and prominent PD-L1 expression. To assess the scope and impact of checkpoint inhibitor-associated encephalitis, we searched global pharmacovigilance databases and identified 209 reported cases of encephalitis associated with anti-PD-1 and/or anti-CTLA-4 regimens. Cases occurred across cancer types and had a 19% fatality rate, demonstrating the recurrent and often fulminant nature of these irAEs. Further analyses were performed from the index case and two additional cases.

Multidimensional protein and transcriptomic analyses pinpointed memory activated CD4+ T cell phenotypes as highly enriched in the inflamed region. T cell receptor (TCR) sequencing identified a highly oligoclonal T cell repertoire, with one

clone representing nearly 20% of detected T cells. The clonal TCR bore striking resemblance to a known CD8+ TCR with Epstein-Barr virus (EBV) reactivity, and several other highly represented clones were identical to known TCRβs recognizing EBV. Dual immunohistochemistry (IHC)/RNA in situ analysis localized this clone to memory activated cytotoxic (CD45RO+GZMB+) CD4 cells. Staining of neural tissue for EBER transcripts identified EBV+ lymphocytes in the cortex and meninges of the affected region. Collectively, this study suggests that latent EBV+ lymphocytes infiltrating the CNS may contribute to neural inflammation after release from peripheral tolerance by anti-PD-1 therapy and identifies cytotoxic CD4+ and CD8+ T cells as culprits of checkpoint inhibitor-associated immune encephalitis.

Introduction

Monoclonal antibodies blocking key negative regulators of T cell function have transformed the management of numerous cancers.[394] While these immune checkpoint inhibitors may induce durable responses in a subset of patients, aberrant immune-mediated phenomena, termed immune-related adverse events (irAEs) may be unleashed in an unpredictable fashion.[395] Most commonly, irAEs are self-limited or resolve with glucocorticoid therapy,[396] although highly morbid and even fatal events can occur (4-6).[397–399] The effect of treatment cessation on symptom resolution is unclear as the half-life of immunoglobulin at 12 mg/ml is

approximately 21 days.[400] At a cellular level, inflamed tissues affected by irAE are characterized by robust infiltration by T lymphocytes and myeloid-lineage cells.[398,399,401,402] Two cases of immune-mediated myocarditis also suggested that such T cell infiltrates are clonal, display an activated phenotype, and induce reactive immune checkpoint expression (e.g. PD-L1) in damaged tissues.[398] IrAEs involving the nervous system have been described in individual reports and case series. These have most commonly affected the central nervous system (encephalitis), peripheral nervous system (Guillain-Barre, peripheral sensory neuropathy), and neuromuscular junction (myasthenia gravis).[403] Despite these early findings, the molecular basis of irAEs, including neurologic irAEs, remains poorly understood.

We performed deep molecular profiling of a case of fulminant and fatal anti-PD-1-induced encephalitis to provide insight into a severe irAE affecting the central nervous system (CNS). Using a global pharmacovigilance database, we show that immune checkpoint inhibitor-induced encephalitis can occur across cancer settings and can be severe and even fatal in a significant number of patients. Here, we suggest a role for latent viral infection contributing to irAE pathogenesis. We demonstrate a highly clonal CD4+ memory activated cytotoxic T cell population with striking similarity to a known Epstein-Barr Virus (EBV)-specific CD8+ T cell, several other high-frequency known EBV-specific clones, and latently-infected EBV+

lymphocytes using next-generation sequencing and immunohistochemistry (IHC) analyses of inflamed neural tissue.

Methods – Wet lab

Index patient and autopsy case

The index patient was consented to protocols that permitted de-identified research use of biospecimens, genetic testing results, and clinical data (Vanderbilt IRB numbers 100178, 181685, and 030220) and his family consented to a post-mortem histopathologic evaluation; cases from outside institutions were approved with institution-specific protocols. Clinical data was obtained through chart review. During the autopsy, evaluation for melanoma using standard melanoma stains (MITF, MelanA, SOX10, S100), and histopathologic evaluation of infiltrating immune cell subsets using CD3, CD20, and CD68 was performed in accordance with institutional protocols.

Additional cases of checkpoint inhibitor-associated encephalitis

Additional case 1 was seen at the University Hospital, Zurich Switzerland and has been previously described.[404] Briefly, she was a 60-year-old (2012) patient with an isolated lesion in the right gyrus frontalis medius as well as bilateral pulmonary metastases. After neurosurgical resection followed by adjuvant radiotherapy (with a

total of 30 Gray) the patient was treated with ipilimumab (3 mg/kg IV) for a total of 4 doses. Three months after the last dose, the patient reported generalized tiredness, weakness, and loss of appetite. However, serum endocrinological work-up revealed a non-significant cortisol shift without osmolality changes, which was not specific for hypopituitarism or adrenal insufficiency. No new metastases were found in the MRI of the brain. A new PET-CT scan showed bilateral pulmonary progression. Subsequent therapy with pembrolizumab was initiated (1 infusion). The patient passed away without any antecedent evidence of altered physical status within 2 weeks. At autopsy, gross inspection revealed a 3 cm defect in the right frontal lobe. No other lesions were seen on further sectioning. Microscopic examination was remarkable for scattered perivascular and parenchymal collections of mature lymphocytes, mostly CD8-positive T-cells, in the hemispheres and brainstem. In addition, sections of brainstem showed CD68-positive microglial cells, occasionally organized in loose nodules. IHC preparations for cytomegalovirus, herpes simplex virus I and II, simian virus 40, toxoplasmosis and varicella zoster virus were negative. There was evidence of melanoma metastasis, even after additional sampling. A diagnosis of encephalitis was rendered based on the presence of intraparenchymal and perivascular CD8-positive lymphocytes accompanied by microglial proliferation.

Additional case 2 was seen at Dana-Farber Cancer Institute in Boston, MA and initially presented in 2013 with a T1a melanoma of the left lower extremity. He was followed until 2015 when he noted increased dyspnea on exertion and a CT scan showed numerous enlarged lymph nodes throughout the chest, abdomen and pelvis. The largest of these was a 7.4 × 5.3 cm left pelvic sidewall node with numerous other large nodes noted. He underwent an ultrasound guided biopsy which demonstrated metastatic melanoma. He was started on single agent pembrolizumab and after the first dose started reporting increased tremors and leg weakness. Due to the subacute nature of these symptoms he received two more treatments with pembrolizumab with progressively worsening symptoms. He developed increased weakness and memory issues and the pembrolizumab was held and he was admitted for work-up of this. Lumbar puncture X3, MRI of the brain and spine, neurologic evaluation and extensive work-up for other causes of his symptoms was performed without an etiology. He received a short course of high dose steroids followed by IVIG and a longer steroid taper without benefit. His functional status continued to decline without a clear etiology despite these interventions and brain biopsy was performed which showed T-cell inflammatory infiltrate and gliosis suspicious for pembrolizumab-related encephalopathy. He subsequently died from progressive neurologic decline.

Additional cases of encephalitis of other etiologies

Additional cases of encephalitis not associated with checkpoint inhibitor use were collected under an IRB-approved protocol to study cellular heterogeneity in normal brain and brain neoplasms (IRB#180238). Cases were stained with *EBNA1* RNA-ISH as detailed below, and with anti-PD-L1 IHC as detailed below. Details of the cases are listed in **Table 5**. Positive and negative control probes were used in all cases to verify specificity and presence of intact RNA for analysis.

Diagnosis	Age/Sex	Pathology notes	EBNA1+ (% cells)	PD-L1+ (% cells)
Chronic meningoencephalitis consistent with clinical history of atypical Rasmussen's encephalitis	50/F	Relatively mild inflammation	0	0
Meningoencephalitis with anti-calcium channel antibodies. CMV identified reactive cells present[c]	40/M		0	5
Toxoplasmosis	34/M	Small sample	0	0
Progressive Multifocal Leukoencephalopathy (JC virus infection)	62/F	Very small sample	0	2
Demyelinating process (multiple sclerosis)	25/F		0	0
Rasmussen's encephalitis[b]	3/M		0	0
Autoimmune encephalitis of unknown etiology[a]	50/M	Relatively mild inflammation	0	1
Acute hemorrhagic leukoencephalitis (variant of acute disseminated encephalomyelitis)	63/M		0	0
Rasmussen's encephalitis	5/F		0	5

[a] Non-ICI-encephalitis (1) in **Figure 2**
[b] Non-ICI-encephalitis (2) in **Figure 2**
[c] Non-ICI-encephalitis (3) in **Figure 2**

Table 5. Cases of non-immunotherapy-induced encephalitis assessed for EBNA1 staining, NanoString DSP, and RNAseq.

Immunohistochemistry

Archival tissue from the autopsy was obtained and IHC evaluation was performed. Antibodies utilized were anti-CD45RO (Thermo Scientific, cat# MA5-11532, 1:1,600),

anti-GZMB (Granzyme B, Biocare cat# ACI 3202 AA, prediluted), anti-Ki67 (Dako, cat# M7240, 1:200), anti-CD4 (StatLab, cat# RM27-10, prediluted) , anti-CD8 (StatLab, cat#MM39-10, prediluted), anti-PD-L1 (Cat.#PA5-28115 ThermoFisher, Grand Island, NY; 1:7500), anti-CD68 (PA0191, Leica, Buffalo Grove, IL; prediluted), anti-PD-1 (HPA035981, Sigma-Aldrich Co., St. Louis, MO; 1:75), anti-CD20 (PA0906, Leica, Buffalo Grove, IL, prediluted), anti-LAG3 (Cell Signaling, Catalog# 15372, dilution 1:200), CD244 (Proteintech, Catalog# 16677-1-AP, dilution 1:200), CD160 (Abcam, Catalog# ab202845, dilution 1:600). The Bond Polymer Refine detection system or Envision system was used for visualization. Slides were then dehydrated, cleared, and coverslipped.

RNA in situ hybridization (RNA-ISH) analysis

For RNAscope RNA-ISH (Advanced Cell Diagnostics) analysis of *EBNA1*, standard RNAscope manufacturer's protocols were followed using the RNAscope H_2O_2 and Protease pretreatment kit (ACD, reference# 322381), RNAscope Target retrieval buffer (ACD, reference# 322000), and appropriate positive and negative RNA probes for controls. For custom BaseScope RNA-ISH (Advanced Cell Diagnostics) TCRβ analysis (ACD, reference# 712111), standard BaseScope manufacturer's protocols were followed using BaseScope™ Detection Reagents– RED (ACD, reference# 322910). Dual ISH-IHC was performed using the custom TCRβ target

probe(s) followed by incubation with primary antibodies (CD4, CD8, Ki67, CD45RO, and GZMB) as described above, using the Envision (Dako) system, DAB (Dako), and hematoxylin counterstain. For EBER staining, slides were placed on the Leica Bond Max IHC stainer, and all steps, besides dehydration, clearing and coverslipping, were performed on the Leica Bond Max. Slides were deparaffinized, and enzyme retrieval was performed using Proteinase K (Ref# S3020, Dako, Santa Clara, CA) for 5 minutes. Slides were put through ISH Hybridization with the Ready-To-Use EBER probe (Ref# ISH5687-A, Leica, Buffalo Grove, IL) for two hours. Slides were placed in an anti-Fluorescein antibody (Ref# AR0222, Leica, Buffalo Grove, IL) for 15 minutes. The Bond Polymer Refine detection system was used for visualization. Slides were then dehydrated, cleared and coverslipped. For dual staining, data were quantified by colocalization of the signal as a fraction of the TCR+ cells, by a licensed pathologist (PIE-G).

Methods – Dry Lab

RNA sequencing

Total RNA quality was assessed using the 2200 TapeStation (Agilent). At least 20 ng of DNase-treated total RNA having at least 30% of the RNA fragments with a size >200 nt (DV200) was used to generate RNA Access libraries (Illumina) following manufacturer's recommendations. Library quality was assessed using the 2100

Bioanalyzer (Agilent) and libraries were quantitated using KAPA Library Quantification Kits (KAPA Biosystems). Pooled libraries were subjected to 75 bp paired-end sequencing according to the manufacturer's protocol (Illumina HiSeq3000). Bcl2fastq2 Conversion Software (Illumina) was used to generate de-multiplexed Fastq files. Quality control for the paired-end raw sequencing reads of all samples were performed using FastQC for the analysis of sequence quality, GC content, the presence of adapters, overrepresented k-mers and duplicated reads. Sequencing reads were mapped to human reference genome GRCh38 (Release-85, Ensembl55) using STAR 2.2.1 with 2-pass mapping.[405] QC for read alignment and mapping was evaluated with RSeQC[406] for sequencing saturation, mapped reads clipping profile, mapped read distribution, and coverage uniformity. The TPM (transcripts per million) values were calculated using RSEM[407] and used to assess the global quality and reproducibility of the RNA-seq dataset and exported for downstream data analyses.

CIBERSORT and gene set analysis

CIBERSORT analysis was performed on RNAseq TPM level data generated above, using both the 'relative' and 'absolute' methods available on the analysis website (https://cibersort.stanford.edu/). The CIBERSORT method used is explained in full detail elsewhere.[408] To complete this analysis, we used the LM22 dataset which was

carefully developed and validated in the manuscript. The genes utilized, along with computational methods and development/validation approach are provided as supplemental data in the original report[408] and also freely available on the CIBERSORT website above. Analysis was performed as previously described.[408] For interferon gene set analysis, the HALLMARK_INTERFERON_GAMMA_RESPONSE (M5913; 200 genes, 177 genes in overlap/matching) signature was downloaded from the Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb). For each gene, the log 2 fold change was calculated between the inflamed and unaffected region, or across additional tissues as indicated.

Digital Spatial Profiling (DSP) of Protein

Single FFPE slides from inflamed and non-inflamed brain tissue were selected, as well as from non-inflammatory 'healthy' brain (epileptic) and 3 additional cases of non-ICI-induced encephalitis. Ten regions of interest (ROIs) within each slide were selected by pathologists. Multiplexed protein profiles were generated with the NanoString® GeoMX™ digital spatial profiling platform (RUO – Research Use Only). Formalin-fixed paraffin-embedded tissue sections of 5um thickness were subjected to antigen retrieval (citrate buffer pH 6) and stained with a cocktail of antibodies labeled with photocleavable DNA-indexing oligos (to generate quantitative protein

profiles), fluorescent anti-CD3 and anti-GFAP (visualization markers) and SYTO 83 nuclear dye (Thermo Fisher, S11364). Tissues were imaged by fluorescence microscopy on the GeoMX platform and regions of interest (ROI) were chosen for molecular profiling. ROIs were selected as circles 200 μm in diameter from inflamed or heathy tissue or by custom masking using the fluorescent CD3 signal to specifically identify T cells and a digital micromirror device to precisely control the pattern of UV illumination. ROIs were exposed to UV light (365 nm) to release oligos which were captured via microfluidics and stored in individual wells of a microtiter plate. Following collection from all ROIs, oligos were hybridized to unique 4-color, 6-spot optical barcodes and enumerated on the nCounter® platform. Data were normalized to ERCC-sequence specific probes to control for technical variation in hybridization efficiency, followed by area normalization to control for ROI size and control IgG (rabbit and mouse) to normalize for background. Data were visualized by unsupervised hierarchical clustering or grouped dot plots.

T cell receptor (TCR) sequencing

TCR sequencing and clonality quantification was assessed in neuropathologist-selected FFPE samples of highly inflamed and non-inflamed brain parenchyma using survey level immunoSEQ™ and the Immunoverse™ assay, as previously described (Adaptive Biotechnologies, ArcherDX).[409–411] Sequencing results were

evaluated using the immunoSEQ analyzer version 3.0. Shannon entropy, a measure of sample diversity, was calculated on the clonal abundance of all productive TCR sequences in the data set. Shannon entropy was normalized by dividing Shannon entropy by the logarithm of the number of unique productive TCR sequences. This normalized entropy value was then inverted (1 – normalized entropy) to produce the clonality metric.

HLA genotyping

High-resolution HLA-A, -B, -C, and DP/DQ/DR HLA typing was performed at the Institute for Immunology and Infectious Disease (Perth, Western Australia). Allele-specific primers targeting exons 2 and 3 of the HLA loci were used to perform PCR amplification. Unequivocal HLA genotypes were assigned at four-digit resolution using proprietary commercial software developed by the Institute for Immunology and Infectious Disease.

TCR repertoire analysis

Data processing

We processed raw immunoSEQ data for downstream analysis using VDJTools.[293] We excluded clonotypes if a TRBV or TRBJ gene was not assigned or if they were non-productive. We then recalculated the frequency of each remaining productive

sequence and collapsed the resulting TCR repertoire dataset by CDR3 amino acid sequence (i.e. the counts for two clonotypes with different nucleotide sequences encoding the same CDR3 AA sequence would be summed). Clonotype annotation and known antigen assignment. We first retrieved continuous and discontinuous CDR3β motifs significantly enriched at least 3-fold above a naïve repertoire of 200,000 unselected CD4 and CD8 T cells at a resampling depth of 10,000 using the GLIPH algorithm.[305] We then annotated each TCR repertoire for known and possible antigen-specificity and pathogenic associations using the greedy VDJMatch algorithm in VDJDB[412] against each CDR3β and significant motifs from GLIPH, and a Levenshtein distance of 1 for sequences stored in McPAS-TCR.[413] We also searched 2,460 HLA-A2 restricted previously published[414] melanoma-specific TCRs for matches and homology to our TCR repertoire data.

In silico epitope prediction

Finally, we annotated each clonotype sequence for potential recognition of 38 viral and 5 cancer T cell epitopes using CDR3β sequence, TRBV gene, and TRBJ gene as input to the random forest TCRex model[415,416] at a false positive rate of 1 in 10,000 (0.01%).

Pharmacovigilance database analysis

We queried VigiBase, the WHO database of individual case safety reports on January 23, 2019. VigiBase is managed by Uppsala Monitoring Centre (UMC) and contains more than 18 million ICSRs submitted by national pharmacovigilance centers since 1968. These reports originate from different sources such as healthcare professionals, patients, and pharmaceutical companies and are generally notified post-marketing. We used the following Medical Dictionary for Regulatory Activities (MedDRA) preferred terms for encephalitis associated with ICI: encephalitis, encephalitis autoimmune, limbic encephalitis, meningoencephalitis, and cerebellitis for the following drugs: nivolumab, pembrolizumab, ipilimumab, atezolizumab, durvalumab, avelumab, and tremelimumab. We collected date of report, therapy, indication (cancer type), age, gender, concurrent adverse events, time of onset, and outcome for all patients when available. We used descriptive statistics to quantify medians and range for continuous variables and percentages for categorical variables. We used the R statistical language (3.5.1) to perform the $\chi 2$ test for contingency tables.

Results

Index case report

A man in his 70s with metastatic melanoma developed meningoencephalitis 18 months after start of pembrolizumab. He was initially diagnosed with early stage melanoma on his back, then developed regional lymph node recurrence 9 years later that was surgically resected and treated with adjuvant ipilimumab (complicated by colitis requiring steroids and infliximab). He then developed a brain metastasis two years later that required resection and postoperative radiation followed by emergence of omental metastases treated with BRAF and MEK inhibitor therapy followed by progression (**Figure 56A, see next page**).

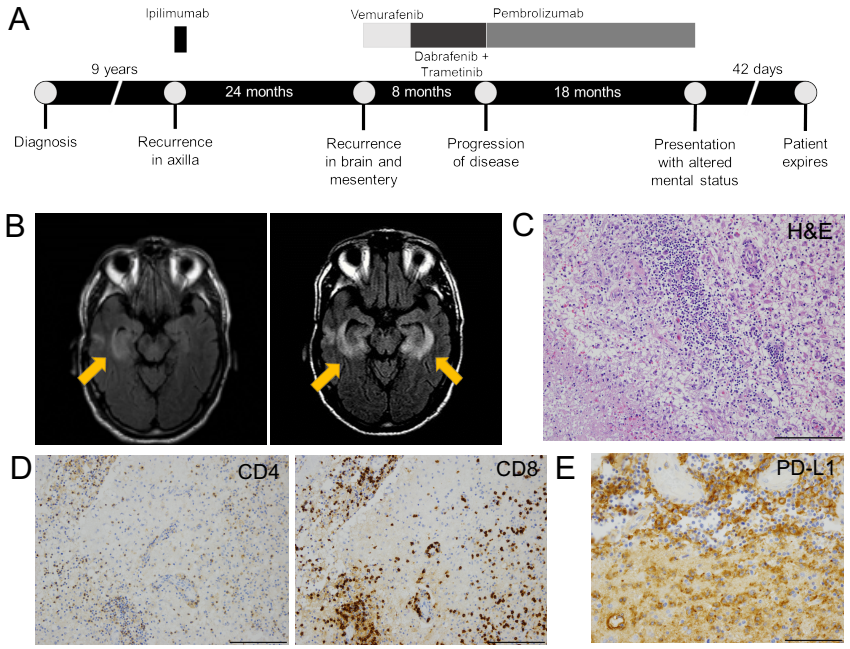


Figure 56. Clinical course of anti-PD-1-induced encephalitis and histologic findings at autopsy. Timeline of index patient diagnosis, prior therapies and anti-PD-1 therapy. B) (left) Magnetic resonance image (MRI) demonstrating restricted diffusion and hyper-intensity in the right temporal lobe and left basal ganglia from the day of hospital admission. (right) MRI showing progressive bilateral medial temporal encephalitis with bilateral putamen involvement. C) H&E stain shows intense perivascular lymphocytic infiltrate extending to adjacent brain parenchyma, next to a region of infarction (lower left); 20x. D) CD4 and CD8 IHC showing the lymphocytic infiltrate is composed of approximately equal ratios of CD4+ and CD8+ T cells; 20x. E) Diffuse PD-L1 positivity staining in cells with macrophage morphology surrounding the region of infarction; 20x. *Of particular note, tissues collected and assayed in this study anteceded treatment with steroids and anti-TNFa, which could impact the immunologic status of the patient.

His subsequent treatment with pembrolizumab resulted in a near complete response without toxicity. Approximately 18 months into pembrolizumab therapy, he developed nausea, fevers, confusion, and aphasia progressing over 2 days. Brain MRI showed restricted diffusion and enhancement in the basal ganglia and right temporal lobe consistent with inflammatory or herpes simplex virus (HSV) encephalitis (**Figure 56B**). Lumbar puncture (LP) showed a neutrophilic pleocytosis (590 nucleated cells/μL; neutrophils 80%, lymphocytes 6%), normal glucose

(67mg/dl; normal range 45-75) and elevated protein (135mg/dl; normal range 15-40). Broad spectrum antibiotics and acyclovir were initiated. Bacterial cultures, HSV-1/2 polymerase chain reaction (PCR), cytology, and paraneoplastic panels were negative; Epstein Barr Virus (EBV) PCR later returned positive with viral load of 1200 copies/mL and was also persistently detected in the blood (1888-4321 copies). PCR for EBV was persistently positive across multiple CSF samples. He continued to have neurologic deterioration requiring intubation and mechanical ventilation; MRI showed progressive encephalitis involving both temporal lobes (**Figure 56B**). Five days into the hospitalization, methylprednisolone 1mg/kg twice daily was initiated, leading to dramatic and rapid clinical improvement.

Two weeks later, he developed recurrent neurologic symptoms while tapering steroids; MRI showed worsening of encephalitis; subsequent CSF evaluation showed evolution to a lymphocyte predominance (83 nucleated cells, 8% neutrophils, 91% lymphocytes). Over the next several weeks, he had brief episodes of clinical improvement with high-dose steroids and intravenous immunoglobulin, followed by progressive deterioration of his mental status and motor strength. He expired 42 days after symptom onset.

Pharmacovigilance database analysis

To assess whether encephalitis is a recurrent event among patients with treated with immune checkpoint inhibitors, we queried VigiBase. VigiBase is a WHO database of individual case safety reports and contains more than 18 million ICSRs submitted by national pharmacovigilance centers since 1967. These reports originate from sources including healthcare professionals, patients, and pharmaceutical companies and are generally reported post-marketing. Of >18,000,000 individual reports, 47,240 were from patients receiving nivolumab, pembrolizumab, ipilimumab, atezolizumab, durvalumab, avelumab, or tremelimumab. Of these, 209 reports of encephalitis were identified. The median age was 61 years (range 7−85) and 59% had either melanoma or lung cancer (**Table 6, see next page**).

Table 6: Characteristics of patients with immune checkpoint inhibitor associated encephalitis (n=209)	
Characteristic	**Number (%)**
Male	113 (54)
Female	69 (33)
Not listed	27 (13)
Age (years; median, range)	61, 16 – 85
Cancer	
Melanoma	44 (21)
Lung cancer (NSCLC and other)	80 (38)
Hodgkin's lymphoma	14 (7)
Non-Hodgkin's lymphoma	3 (1.5)
Renal	9 (4.5)
Other/not listed	59 (28)
Region reporting	
Americas	117 (56)
Europe	67 (32)
Asia	19 (9)
Oceania	6 (3)
Regimen	
Anti-PD-1 monotherapy	137 (66)
Combination anti-PD-1 + anti-CTLA-4	31 (15)
Other combinations*	14 (7)
Anti-PD-L1 monotherapy	15 (7)
Anti-CTLA-4 (ipilimumab) monotherapy	12 (6)
Timing, days (median, range)	67 (5 – 456)
Reporting term	
Encephalitis	145 (69)
Encephalitis autoimmune	36 (17)
Limbic encephalitis	15 (7)
Meningoencephalitis	12 (6)
Cerebellitis	1 (1)
Concurrent irAEs	
Colitis	4 (2)
Pneumonitis	6 (3)
Thyroiditis	7 (3)
Myocarditis	2 (1)
Vasculitis	2 (1)
Adrenal insufficiency	4 (2)
Dermatitis	4 (2)
Other‡	7 (3)
Fatal outcome	39 (19)
Reporting year	
2012 – 2014	5 (2.5)
2015	5 (2.5)
2016	36 (17)
2017	77 (37)
2018 (through Sep 5)	86 (41)

Table 6. Characteristics of patients with immune checkpoint inhibitor-associated encephalitis (n=209).

66% of patients received anti-PD-1 monotherapy and 88% of patients had no other concurrent irAEs reported. The median time to symptom onset was 67 days (range 5 – 456 days) and 39 patients (19%) had a fatal outcome. In contrast to other toxicities (e.g. myocarditis),[397] the fatality rate was similar with encephalitis from anti-PD-1 monotherapy vs. combined anti-PD-1 + anti-CTLA-4 (20% vs. 26%, chi-square $P = 0.50$). No patients treated with ipilimumab monotherapy experienced fatal encephalitis. To confirm these findings in clinically-diagnosed cases of encephalitis (as opposed to de-identified pharmacovigilance reports), we queried 2501 patients from 4 large academic centers treated with immune checkpoint inhibitors, and identified 22 cases of meningoencephalitis (0.88% incidence). These findings (**Table 7**) were similar to the pharmacovigilance database, with a median age of 65 years and median onset of 80 days (range 4-684).

Age (years; median, range)	65 (26-83)
Female gender	6 (27)
Primary cancer melanoma	20 (91)
Agent	
Anti-PD-1/PD-L1 monotherapy	6 (27)
Ipilimumab monotherapy	6 (27)
Anti-PD-1 + ipilimumab	10 (44)
Prior brain metastases?	5 (23)
Active brain metastases?	2 (9)
Prior brain surgery?	2 (9)
Prior brain radiation	2 (9)
Neurologic co-morbidities	7 (32)*
Prior autoimmune disease	3 (14)
Time to onset (days, median, range)	80 (4-684)
Anti-PD-1/PD-L1 monotherapy	86 (5-684)
Ipilimumab monotherapy	58 (39-138)
Anti-PD-1 + ipilimumab	93 (4-517)
Symptoms	
Confusion	11 (50)
Headaches	8 (36)
Seizures	5 (23)
Fevers	4 (18)
Focal neurologic deficits	5 (23)
Concurrent irAEs	
None	10 (44)
Colitis/enteritis	3 (14)
Thyroiditis	3 (14)
Rash	3 (14)
Other	3 (14)[#]
MRI findings	
Unremarkable	16 (73)
Findings present	6 (27)[**]
Cerebrospinal fluid analysis	
Nucleated cells/mL (median, range)	37 (2-590)
Hospitalized	20 (91)
Multiple hospitalizations	5 (23)
Length of stay (days; median, range)	6.5 (2-120)
Treatments	
High dose steroids	19 (86)
Pulse dose (e.g. 1g/daily)	3 (14)
1-2mg/kg	16 (73)
Intravenous immunoglobulin	4 (18)
Other (Plasma exchange, rituximab)	2 (9)
Outcome	
Resolved	15 (68)
Persistent symptoms	5 (22)
Fatal	2 (9)[##]

Table 7. Characteristics of 22 cases of ICI-induced meningoencephalitis. *Includes seizure disorder (n=3), cerebrovascular accident (n=2), migraines, cerebral arteriovenous formation. #Hypophysitis, hepatitis, and epididymitis. **Enhancement of temporal lobes (n=3), enhancement of occipital lobes (n=2), perivascular enhancement in caudate/putamen/internal capsule (n=1), cranial nerve enhancement (n=1). ##Both in patients with MRI findings

Confusion was the most frequent presenting symptom (50%), followed by headaches (36%), seizures (23%), and fevers (18%). The most common imaging finding was temporal or occipital lobe enhancement in 5 patients (22%). Most patients recovered completely (68%) although 5 (22%) had persistent neurologic symptoms and 2 (9%) died; median hospitalization time was 6.5 days (range 2-120 days). Thus, encephalitis is a rare but recurrent and potentially fatal toxicity of checkpoint inhibitors that occurs across cancer types.

Autopsy and histologic findings

An autopsy of the index case was performed. No evidence of melanoma was observed in the central nervous system or elsewhere using pan-melanoma stains S100, SOX10, melanoma antigen recognized by T cells 1 (MelanA/MART-1), and MITF (data not shown). Analysis of T cell receptors (reported subsequently in this manuscript) also did not identify any known or likely MelanA/MART-1 specific HLA-A*02:01 TCRs.

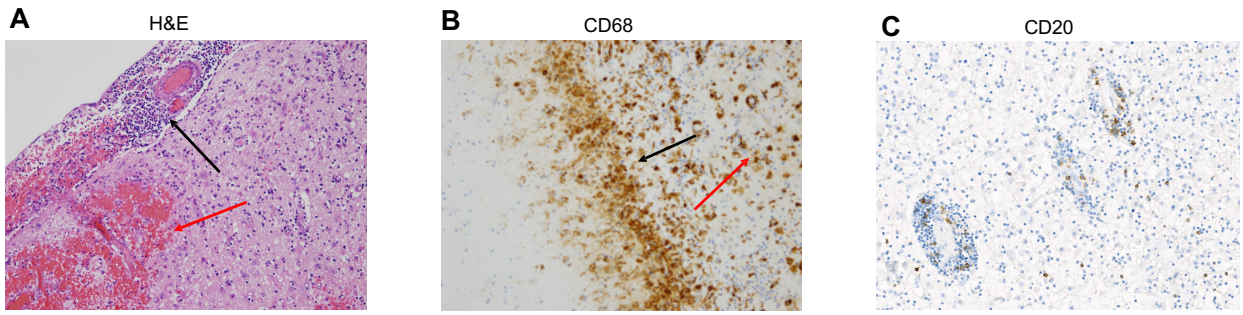


Figure 57. Inflammatory and myeloid/microglial infiltrate in infarct regions of brain. A) H&E stain shows a dense chronic inflammatory infiltrate in the meningees (upper left; black arrow), focal hemorrhage in the underlining brain parenchyma (lower left; red arrow), and gliotic gray matter; 20x). B) CD68+ cells at the interface of the necrotic area (black arrow) and the adjacent brain parenchyma (red arrow) 20x. C) Limited expression of CD20 staining in perivascular and parenchymal regions of brain; 20x.

Vigorous inflammation, multiple cerebral infarctions with necrosis, and sporadic hemorrhagic transformation were noted in the bilateral temporal lobes and striatum with surrounding gliosis and numerous macrophages (**Figure 56C, Figure 57A**). Exuberant perivascular inflammation extending to adjacent gray matter characterized by approximately equal ratios of CD4+ and CD8+ T cells was observed (**Figure 56D**), although it was noted that CD4+ cells were present in localized patterns while CD8+ cells were diffuse and consistently present. Frequent CD68+ cells, a macrophage marker (which also stains microglial cells), were also observed (**Figure 57B**). Minimal CD20+ infiltrates or immunoglobulin deposition were observed, suggesting that B cells did not constitute a major component of the inflammatory process (**Figure 57C**). Similar patterns of T cell and macrophage infiltration were observed in inflamed meninges and adjacent gray matter,

including perivascular spaces (**Figure 58**).

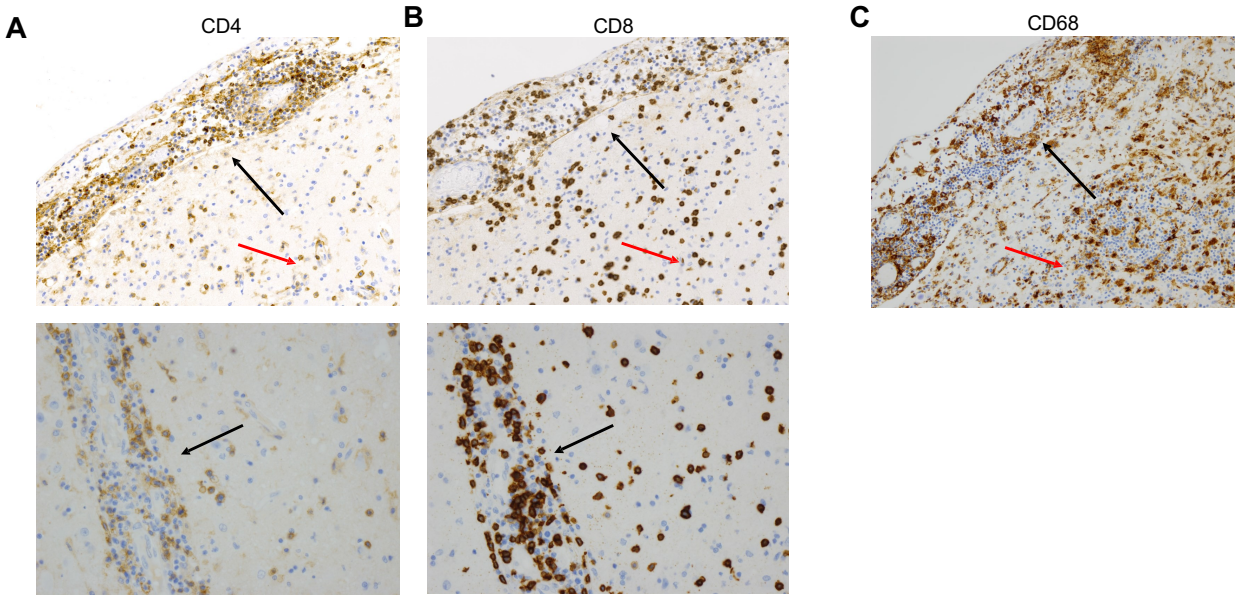


Figure 58. Lymphocytic and myeloid infiltrate in meninges, perivascular, and parenchymal regions. The lymphocytic infiltrate involving the meninges (upper image; upper left of image, black arrow), parenchyma and brain perivascular regions (upper image; upper left of image, black arrow), parenchyma and brain perivascular regions (upper image; lower right of image, red arrow and lower image) includes CD4+ and (B) CD8+ T cells; 20x. C) CD68+ cells accompanying lymphocytic infiltrate in meninges (upper left of image, black arrow) and brain parenchyma (lower right, red arrow); 20x.

By contrast, non-affected brain tissue obtained from radiographically and macroscopically normal brain was histologically unremarkable with no inflammatory infiltrate (**Figure 59, see next page**).

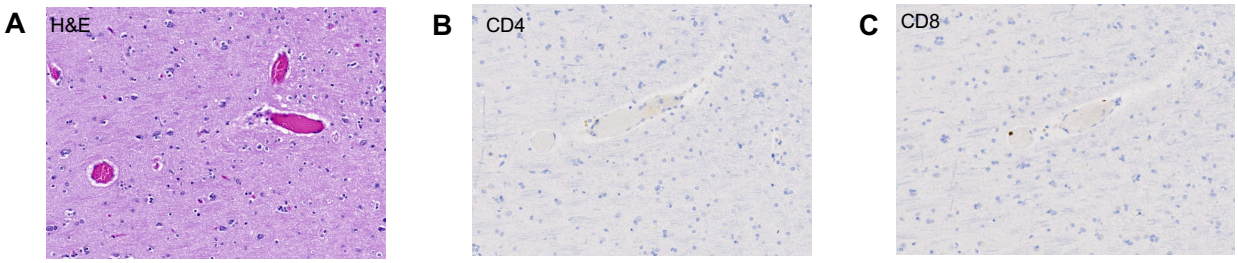


Figure 59. Absence of substantial inflammatory infiltrate in radiologically and macroscopically non-affected area. A) H&E corresponding to radiologically and macroscopically non-affected area with preserved architecture and no apparent inflammatory infiltrate; 20x. B) CD4 and (C) CD8 stains demonstrating sparse presence or absence of T cell infiltrates in radiologically and macroscopically non-affected area; 20x.

We then profiled the T cell exhaustion markers PD-1 and PD-L1. PD-L1 was diffusely expressed by cells with macrophage morphology (**Figure 56E**). We also observed few-to-no cells expressing PD-L1 in samples from patients with encephalitis of other etiologies (**Table 6**); these samples also showed minimal to no lymphocytic infiltration, suggesting that PD-L1 expression in the brain may represent an attempt to dampen lymphocyte-derived inflammation and injury. PD-1 was expressed at lower levels by infiltrating lymphocytes as well as by pericytes, cells which surround the capillary and venular endothelium and play a critical role in blood-brain barrier maintenance (**Figure 60, next page**).[417]

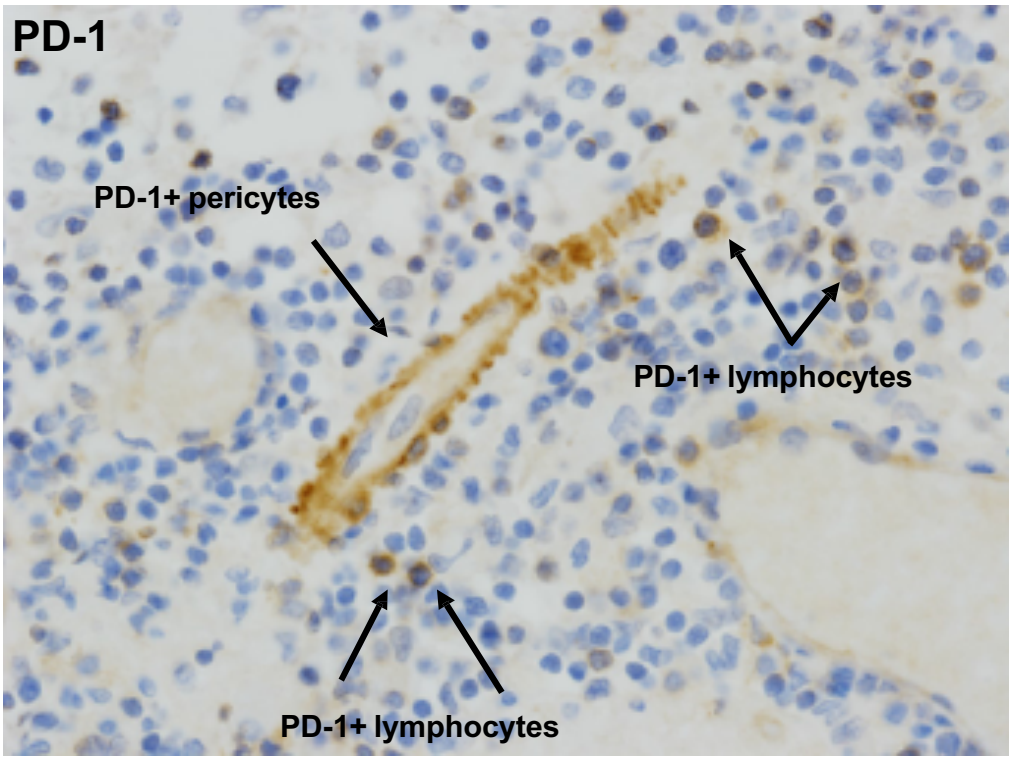


Figure 60. PD-1 expression in inflamed region of brain. PD-1+ perivascular lymphocytes and pericytes; 20x.

Taken together, the pattern of immune cell infiltration and exhaustion markers were reminiscent of that reported in two cases of fulminant PD-1/CTLA-4 blockade-induced myocarditis.[398]

Additionally, we profiled other markers of immune cell exhaustion including LAG-3 (T cells), CD244 (NK cells), and CD160 (CD8+ T cells and NK cells) in the inflammatory microenvironment and histologically normal regions (**Figure 61**). LAG-3 was highly expressed in the inflamed region, and absent in the uninflamed region. CD244+ cells were also present although less dramatically than with LAG-3.

CD160+ cells were identified in both tissues, although present at a higher degree in the inflamed region. These results suggest that repeated antigen exposure likely resulted in the infiltrating T cell population receiving negative co-stimulation, as demonstrated here.

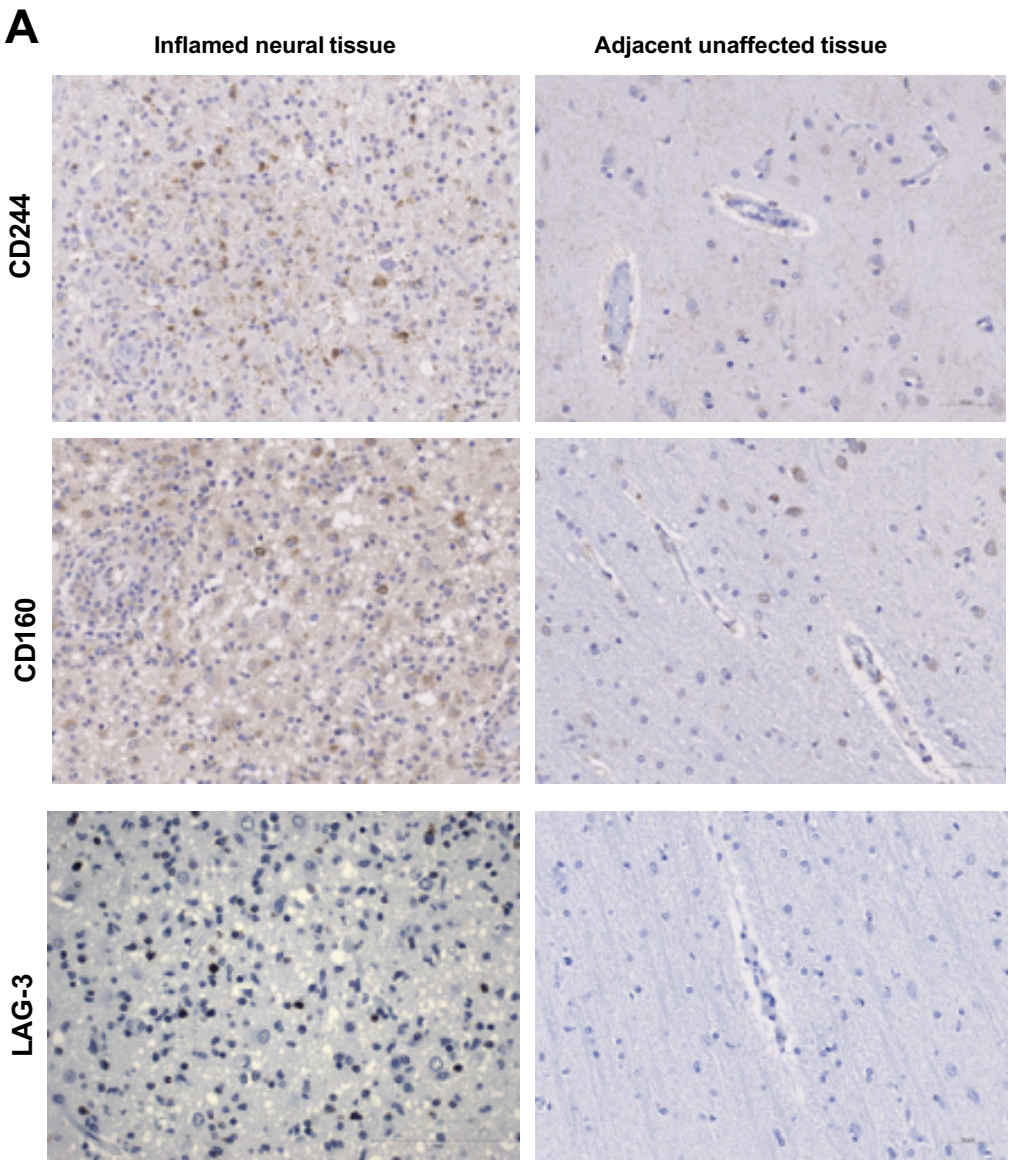


Figure 61. Expression of T cell and NK cell markers of immune cell exhaustion. Inflamed and non-inflamed adjacent regions of neural tissue were immunostained for CD244, CD160, and LAG-3; 20x.

Digital spatial profiling of encephalitic immune microenvironment

As only formalin–fixed autopsy tissue was available, we characterized the immune microenvironment using NanoString digital spatial profiling. This methodology utilizes UV-photocleavable oligonucleotide barcodes conjugated to monoclonal and polyclonal antibodies to quantitatively detect over 30 immune-related protein targets simultaneously from a single tissue section. Tissue architecture is visualized by fluorescent-labeled antibodies—*e.g.* in this case glial fibrillary acidic protein (GFAP) to identify neural tissue and CD3 to identify T lymphocytes. Micro-capillaries collect barcodes released by the focusing of UV light on specific regions of interest (ROIs) selected spatially or based on specific cell populations (e.g. CD3+ cells only). NanoString nCounter analysis directly and digitally quantifies the numbers of each barcode present on each ROI or cell population selected (**Figure 62A, next page**).

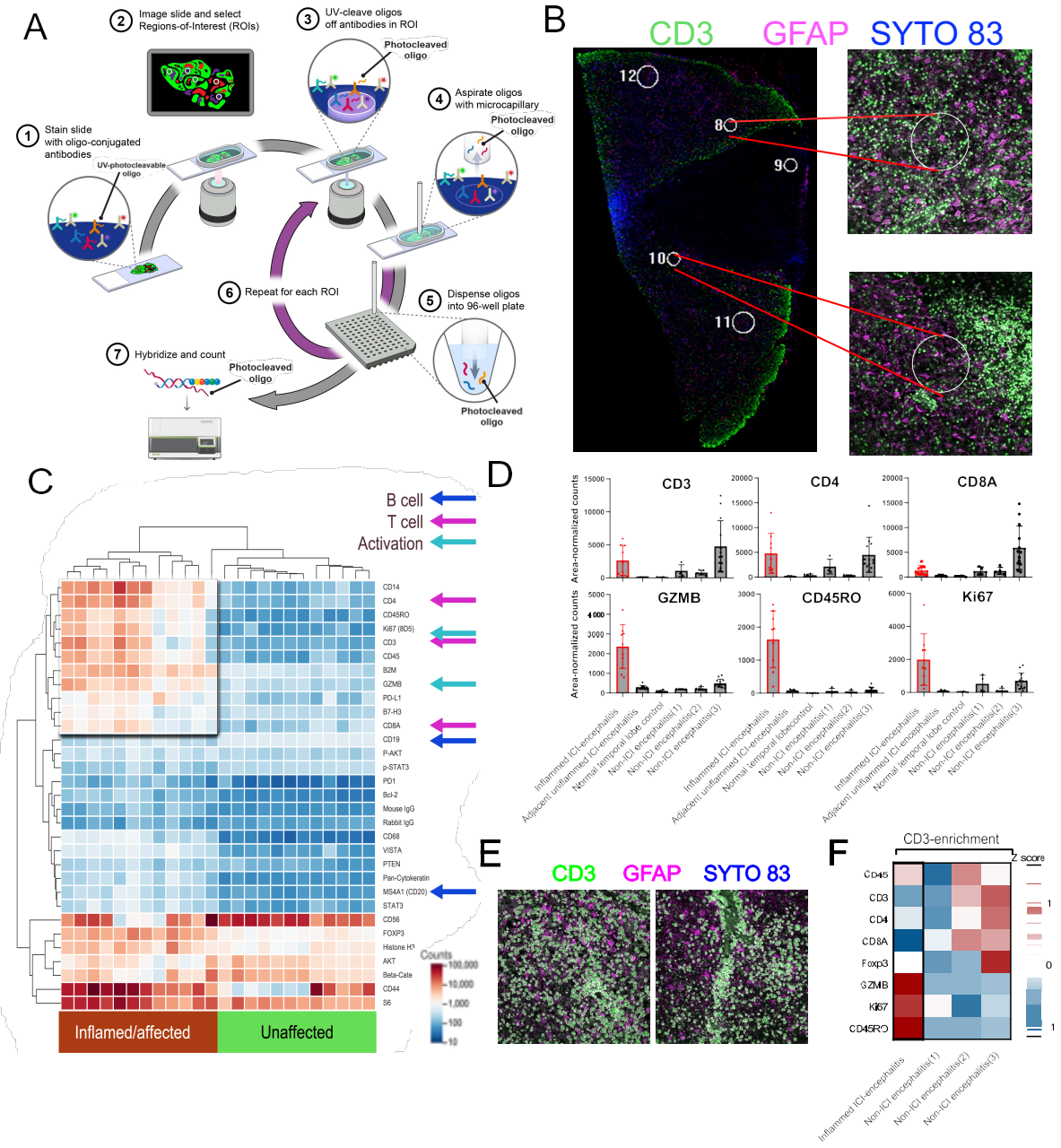


Figure 62. Digital spatial profiling of immune-related protein markers across inflamed and non-inflamed neural tissue. A) Schematic describing DSP workflow. B). Representative region-of-interest selection under immunofluorescence for analysis. Tissue was stained with anti-GFAP (pink) and anti-CD3 (green) in addition to SYTO 83 nuclear staining for ROI visualization and selection. Circular regions 200 µm in diameter were excited with UV light to release barcodes for collection and analysis. C) Heatmap of area-normalized barcode counts from inflamed and unaffected tissue sections across 12 ROIs. Key antigens representing activation, T cell, and B cell markers are identified by arrows. D) Spatio-regional protein expression patterns across 8-10 ROIs from brain tissues demonstrating enrichment for T cell markers in encephalitis cases, but only memory activated markers in (Ki67^{HI} CD45RO^{HI} GZMB^{HI}) in ICI-encephalitis E) Representative images of CD3+ T cells gated for barcode collection across 2 spatial ROIs in the inflamed region. F) Area-normalized counts for key antigens demonstrating enrichment of memory activated phenotypes (Ki67^{HI} CD45RO$^{H\ I}$GZMB^{HI}) in CD3+ population ROIs from ICI-induced encephalitis versus non-ICI-encephalitis.

We selected 12 ROIs each from inflamed and adjacent unaffected tissue sections (**Figure 62B**). This analysis demonstrated a high degree of T cell and cytotoxic activation, with large increases in CD4, CD8, CD3, Ki67, PD-L1, GZMB, and CD45RO in the affected region and minimal increase in the B cell markers CD19 and CD20 (**Figure 62C**). We extended this analysis across additional tissue controls, including a normal temporal lobe specimen and 3 cases of non-ICI-associated encephalitis (**Table 7**). While T cell infiltrate (CD3/4/8) was variable but similar across affected in both the present case as well as the non-ICI encephalitic cases, there was a preponderance for markers of memory activated phenotypes (granzyme B expression, CD45RO, and Ki67) in the present case (**Figure 62D**). Selecting two CD3+ T cell-specific ROIs on the inflamed region of brain, as well as non-ICI encephalitic cases (**Figure 62E**), we next asked what markers were specifically enriched on the infiltrating T cell population. While CD3 is expected to be enriched in these ROIs based on its use of fluorescently-labeled selection markers, this analysis further localized Ki67, CD45RO, and GZMB specifically to T cells (**Figure 62F**). Altogether, we reasoned that memory activated T cells were likely playing an important role in the pathophysiology.

Transcriptomic sequencing identifies memory-activated CD4 T cell phenotypes

We performed mRNA sequencing on tissue from both the inflamed and unaffected brain of the patient, as well as tissue controls utilized above. Using the CIBERSORT method,[408] we characterized the predicted composition of the immune infiltrate. Absolute fractional analysis demonstrated, as expected, an increase in nearly all populations of immune cells in the inflamed region.

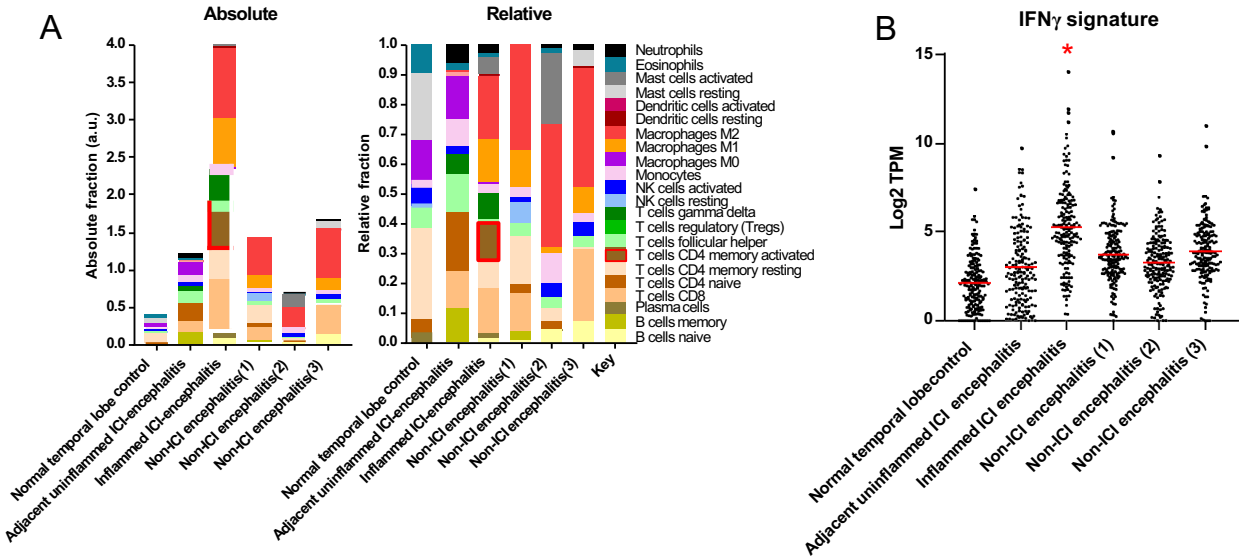


Figure 63. RNAseq analysis of encephalitic and unaffected tissue. Absolute and relative quantification of immune subsets by CIBERSORT. B) Interferon-γ-inducible genes (HALLMARK_INTERFERON_GAMMA_RESPONSE; M5913; 177 overlapping genes) quantified in inflamed and unaffected regions by RNAseq, as well as additional cases identified in Supplementary Table 1. *adjusted P value <0.0001 versus all other samples via ANOVA with Tukey's post-hoc test.

However, adjusting for immune cell input through the relative fraction analysis, we observed several key changes: 1) conversion of naïve CD4 cells to CD4 memory-activated T cell population in the inflamed tissue, 2) conversion from resting to activated mast cells, 3) a loss of memory B cells, and 4) differentiation of M0 to

M1/M2 macrophage phenotypes (**Figure 63A**). As response to anti-PD-1 therapy has been linked to type-II interferon responses, we asked whether IFNγ-inducible genes were also over-represented in the inflamed region. Such genes were markedly upregulated (**Figure 63B**).

T cell receptor profiling identifies oligoclonal cytotoxic CD4 memory T cells

To examine the clonality of infiltrating T cells, we performed TCRβ CDR3 sequencing using the immunoSEQ and ArcherDX platforms on: 1) the inflamed and unaffected regions of the brain, 2) prior resections of the tumor-adjacent lymph node, mesentery, and metastasis to the brain, and 3) the recurrence scar and spleen. The inflamed section demonstrated a high T cell fraction (0.71 of all nucleated cells; **Figure 64A**). An extremely high degree of productive T cell clonality was also demonstrated (0.369; **Figure 64B, 64C**). Notably, a single TCRβ sequence comprised 19.6% of all infiltrating T cells (CDR3 amino acid CASSFPSGSYEQYF; **Figure 64D**), with two other high frequency clones making up 5.98% and 5.91% of infiltrating T cells, respectively. As a comparison, the most highly prevalent clone infiltrating the heart muscles among two analyzed patients with myocarditis[398] was approximately 9% compared to 19.6% in this case of encephalitis. We also detected this dominant clone at 1.08% in the spleen and 6.44% in the recurrence scar, though analyses of

the TCR repertoire of additional tissues revealed mostly unique repertoires in comparison to the brain (**Figure 64G**).

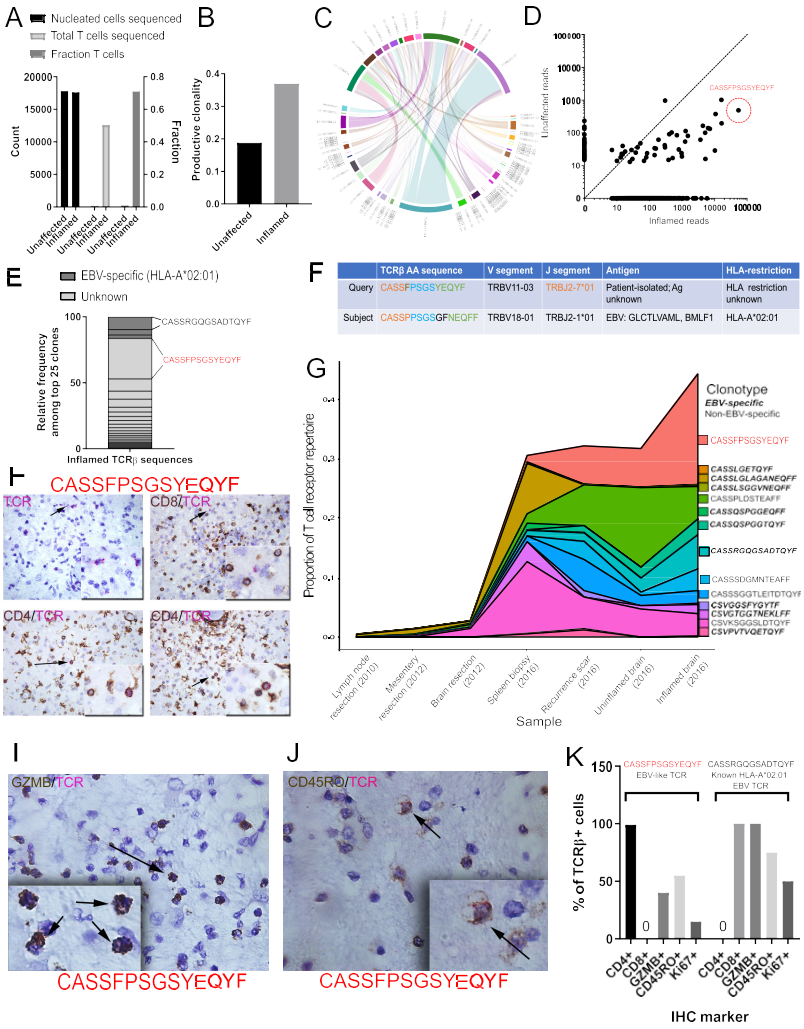


Figure 64. TCR sequencing identification of oligoclonal CD4+ cytotoxic T cells in inflamed encephalitic tissue. A) ImmunoSEQ statistics of total cellular content, composition (fraction of T cells) and total T cells sequenced in inflamed and unaffected neural tissue. B) Enrichment of productive clonality in inflamed region of encephalitic tissue. C) Circos plot depicting the V(D)J rearrangements in TCRβ in the repertoire. D) Biplot demonstrating enrichment of shared TCRs among the unaffected (likely regional presence) and inflamed tissues. The amino acid sequence predicted from the most highly represented DNA sequence is identified (CASSFPSGSYEQYF). E) Representation of top 25 clones according to known EBV-specificity based on patient HLA haplotype. F) Sequence similarity in TCRβ amino acid sequence to a known EBV clone with specificity for the immediate-early HLA-A2-restricted epitopes GLC (BMLF1) G) Clonotype tracking over time and across tissue samples of EBV-specific and non-EBV-specific dominant clones found in the inflamed brain tissue. H) Dual IHC/RNA-ISH analysis for CD4/CD8 and EBV-like CASSFPSGSYEQYF RNA sequence. Upper left: only the TCR probe (pink); upper right: exclusion of TCR probe with CD8+ T cells, lower: Co-expression of CD4 with TCR probe. I) Co-expression of granzyme B with TCR probe J) Co-expression of CD45RO with TCR probe. K) Quantification for Ki67, CD45RO, and granzyme B (GZMB) expressed as percent of all CASSFPSGSYEQYF TCRβ+ cells, or the comparator CASSRGQGSADTQYF TCRβ sequence, a known HLA-A*02:01 restricted EBV clone.

We next performed 6-digit HLA-haplotyping on the patient in order to have a greater contextual understanding of the TCR repertoire and to leverage existing knowledge bases of TCR-antigen specificity. The patient was homozygous for HLA-A*02:01:01, a common allele in the European/Caucasian population. The high degree of clonality suggested that a single antigenic epitope might have triggered this aberrant T cell response. We used several previously described algorithms to identify: 1) significantly enriched amino acid motifs in the TCR repertoire and 2) the antigen specificity of TCRs involved in the immune response of the inflamed section. These annotations revealed a [PSGS] motif expanded in the TCR repertoire in the inflamed section of the brain. This same [PSGS] motif was solely found in the dominant clone and in the junctional region where CDR3 diversity is highest (i.e. outside of the germline CASS and YEQYF regions encoded by the V and J genes). When we queried available public TCR databases, we also found this motif in two known EBV-specific TCRs recognizing the immediate-early epitopes GLCTLVAML and RAKFKQLL. This enrichment of the [PSGS] motif in the inflamed section of the brain indicates that TCRs containing this sequence are likely involved in an antigen-specific response not found in the naïve T cell compartment, where these motifs are highly infrequent. Similarly, TCRs not containing significantly enriched motifs can be inferred to be less important to, or involved in, the antigen-specific response

of a given T cell repertoire. We queried public TCR databases[412,413] and found that the [PSGS] motif in reported TCR sequencing data to-date is exclusively found in TCRs recognizing viral proteins; this motif was not identified in any known autoimmune-associated or autoimmune-specific TCRs (**Table 8**).

CDR3	TRBV gene	TRBJ gene	HLA restriction	Epitope	Epitope origin	PMID
CASSVAPSGSESPLHF	V9*01	J1-6*01	HLA-DRA*01	PKYVKQNTLKLAT	InfluenzaA	12466894
CASRPSGSELIYEQYF	V5-1*01	J2-7*01	HLA-A*02	GILGFVFTL	InfluenzaA	28423320
CASSPPSGSAYNEQFF	V9*01	J2-1*01	HLA-B*07:02	LPRRSGAAGA	InfluenzaA	28636589
CAISDPSGSSYNEQFF	V10-3*01	J2-1*01	HLA-B*07:02	RPRGEVRFL	HSV-2	20139278
CASSPPSGSYEQYF	V7-9*01	J2-7*01	HLA-B*08	FLKEKGGL	HIV-1	17287271
CASSQEPSGSWGEQYF	V3-1*01	J2-7*01	HLA-A*02	CINGVCWTV	HCV	28146579
CASSQEPSGSWGEQYF	V3-1*01	J2-7*01	HLA-A*02	CINGVCWTV	HCV	28146579
CASSQEPSGSWGEQYF	V3-1*01	J2-7*01	HLA-A*02	CINGVCWTV	HCV	28146579
CASSPPSGSGFNEQFF	V18*01	J2-1*01	HLA-A*02:01	GLCTLVAML	EBV	12504586
CASSSDPSGSIAYEQYF	V5-1*01	J2-7*01	HLA-B*08:01	RAKFKQLL	EBV	24512815
CASSPPSGSYEQYF	V7-9*01	J2-7*01	HLA-A*02	NLVPMVATV	CMV	28423320
CSAPSGSSYEQYF	V20-1*01	J2-7*01	HLA-A*02	NLVPMVATV	CMV	28423320
CATSRVPSGSYEQYF	V15*01	J2-7*01	HLA-A*02	NLVPMVATV	CMV	28423320
CASSLVPSGSTDTQYF	V5-1*01	J2-3*01	HLA-A*02:01	NLVPMVATV	CMV	28636589

Table 8. TCRs containing the [PSGS] motif are viral-specific and not associated with autoimmune reactivity.

Further analyses including query of publicly available databases[412,413] revealed 23 additional EBV-specific TCRs, 7 of which were exact matches to known HLA-A2-restricted EBV-specific TCRs that were previously described.[305,418–423] These

sequences represented 10.9% of the infiltrating T cells in the infarct, including the second most dominant TCR in the inflamed region of the brain (**Figure 64E, 64F**), with specificity for the immediate-early HLA-A2-restricted epitopes GLC (BMLF1) and YVL (BRLF1). We observed these sequences at frequencies higher than those found in acute phase infectious mononucleosis patients and in the synovial fluid of rheumatoid arthritis patients, in whom these TCRs were first detected at frequencies of 5% of CD8+ T cells.[418] While the frequency of these EBV-specific sequences fell to less than 1% of CD8+ T cells over time in acutely infected patients in those studies, we observed a distinct increase in the frequency of EBV-specific sequences in multiple tissues of the index case over time (**Figure 64G**). These sequences also occupied 16.1% of sequences detected in the spleen biopsy of this patient and 5% of the brain recurrence scar. Notably, we did not detect any previously described HLA-A2-restricted or HLA-restriction-unknown TCRs specific to MelanA epitopes (see Methods).

Given consistent findings in our earlier data suggesting a predominance of CD4 memory activated T cells in the inflamed region, we performed dual RNA *in situ* analysis (using an RNA probe specific to the oligoclonal RNA sequence) and IHC to characterize the cells. This analysis surprisingly yielded a complete localization of CASSFPSGSYEQYF+ cells to CD4+ T cells (**Figure 64H**). Furthermore, these CD4+ cells were also GZMB+ (40%), CD45RO+ (55%) and Ki67+ (10-15%) (**Figure 64I-K**). Thus,

the oligoclonal cells detected by TCRβ sequencing are likely the same cells identified by digital spatial analysis and CIBERSORT, together corroborating that CD4 memory activated T cells played a key role in this syndrome. Importantly, the second dominant TCR sequence (CASSRGQGSADTQYF), recognizing immediate-early HLA-A2-restricted epitopes GLC (BMLF1) and YVL (BRLF1), was found specifically in CD8+ T cells as expected, and also colocalized with markers of activation (**Figure 64J** and **Figure 65**).

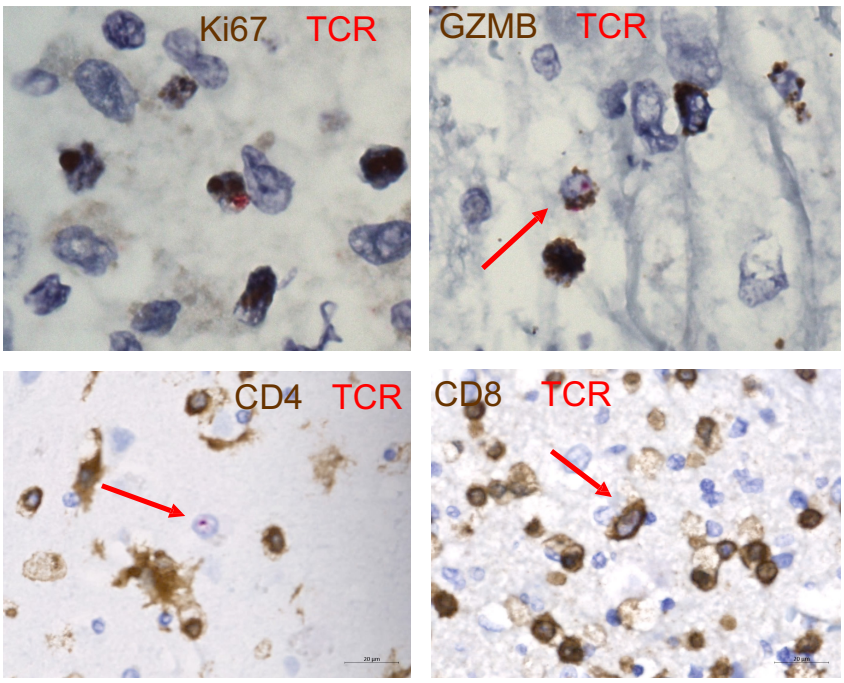


Figure 65. Overlap of HLA-A*02:01-restricted known EBV-specific TCR with CD8+ Ki67+ and GZMB+ phenotypes. Representative images of the TCRβ RNA-ISH probe overlaid with IHC markers.

Latent EBV+ lymphocytes in neural tissue of anti-PD-1-induced encephalitis

Given the findings that the patient had persistently positive EBV PCR from multiple CSF samples and that there were known HLA-A*02:01:01 EBV-specific TCRβ sequences detected at high clonality in the inflamed tissue, we asked whether EBV+ lymphocytes could be found in the local immune microenvironment. A substantial percentage (~90%) of the population is known to be infected with EBV,[424] which persists indefinitely in latently infected B lymphocytes in peripheral lymphatic structures.[425] These cells, however, are rarely found in the CNS, and EBV-induced encephalitic syndromes are rare, but can occur (25-28).[426–429] It should be noted that the number of EBV+ B cells found in circulation in asymptomatic donors is less than 0.01% of all B cells (or approximately 38 out of 10^6 B cells).[430] Notably, other herpesviruses can also establish latency in the central nervous system, including varicella zoster (VZV), herpes simplex (HSV), and human herpes virus 6 (HHV-6). Using a diagnostic RNA *in situ* hybridization (RNA-ISH) stain against EBV-encoded RNAs (EBER1/2), we identified rare but reproducible EBER+ lymphocytes in both the cortex and meninges (**Figure 66A and B**, respectively) of the affected region of the patient's brain. EBNA1-specific RNA-ISH confirmed the presence of EBV-infected cells (Figure **66C and D**). Staining of a lymph node dissection anteceding administration of anti-PD-1 therapy demonstrated EBER+ cells in the node, suggesting that the encephalitis did not arise from acute infection (Figure 66E).

Surprisingly, EBER staining of the patient's resected brain metastasis, also predating immunotherapy and symptoms, identified rare melanoma cells with EBER reactivity (**Figure 66F**).

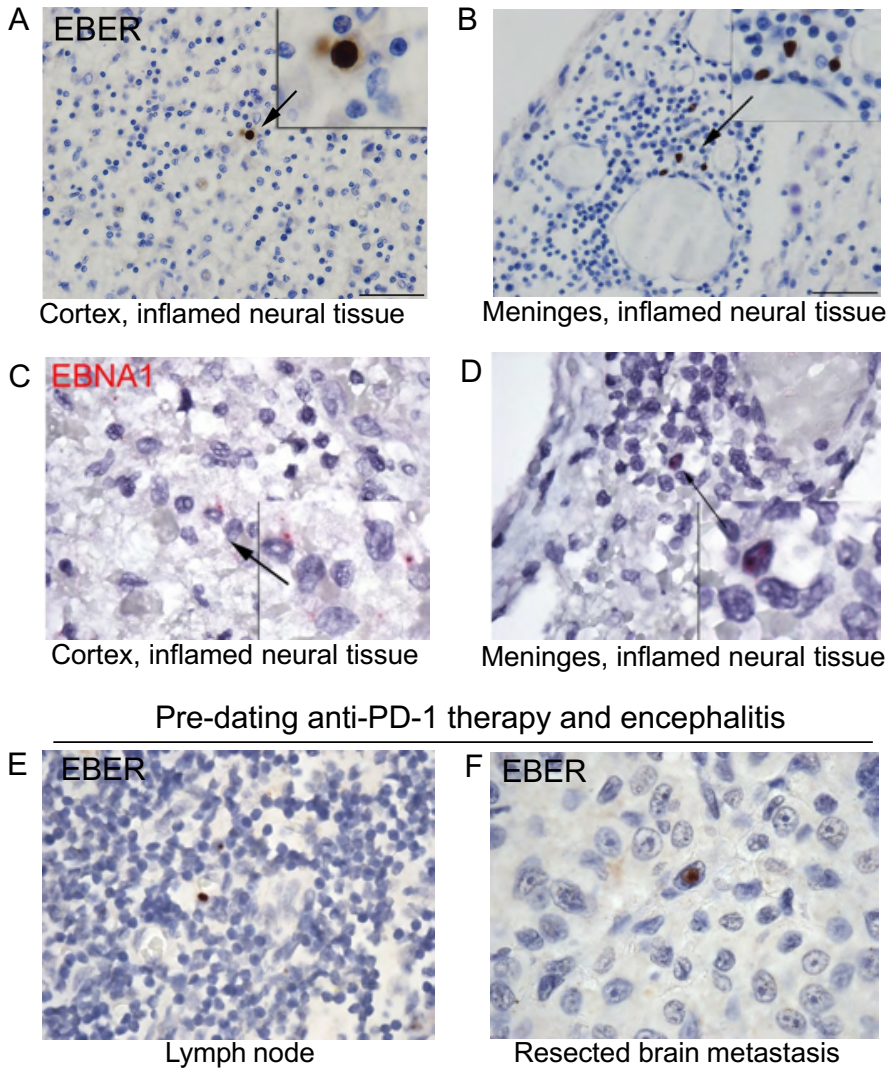


Figure 66. Evidence of latent EBV infection at the site of encephalitic inflammation. A) EBER(1/2) staining of lymphocytes by RNA in situ hybridization in the cortex and meninges (B) of the encephalitic infarct. C) EBNA1 staining of lymphocytes by RNA in situ hybridization in the cortex and meninges (D) of the encephalitic infarct. E) Positive EBER staining of rare lymphocytes in lymph node resection pre-dating anti-PD-1 therapy, suggesting historic EBV infection. F) Positive EBER staining of rare tumor cells in brain metastasis resection pre-dating anti-PD-1 therapy, also suggesting historic EBV infection.

In contrast, EBNA1 staining of 9 additional non-immunotherapy related encephalitis cases demonstrated complete absence of EBER positive lymphocytes in all cases, suggesting that EBV+ lymphocyte presence is an uncommon finding in cases of encephalitis caused by other etiologies (**Table 7**). Notably, archival neural tissue from two other cases of immune checkpoint inhibitor encephalitis, which were characterized by substantially less robust inflammation, had negative EBER staining (**Figure 67**; case details in methods), suggesting that other mechanisms may also contribute in these cases.

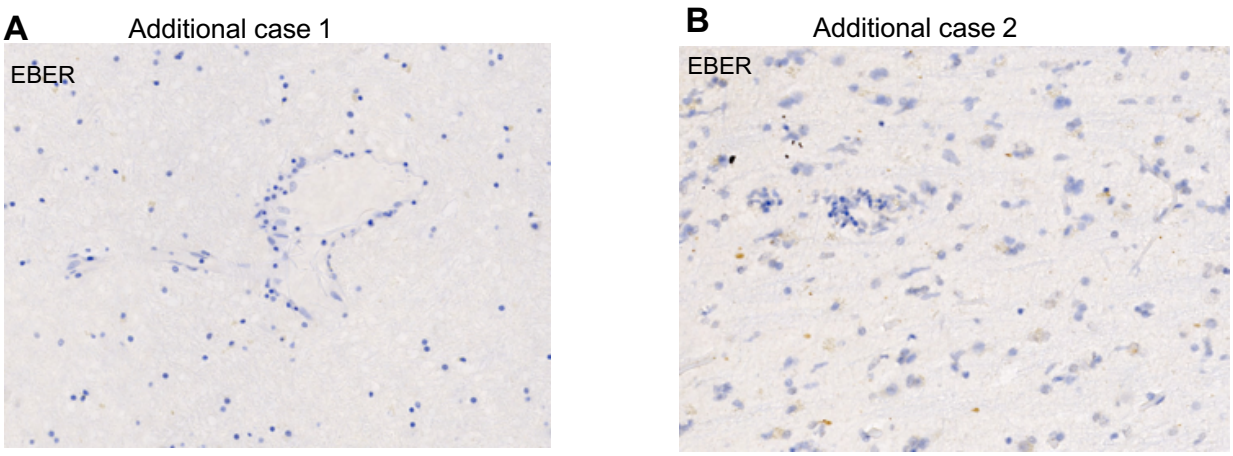


Figure 67. Lack of detection of EBER+ cells in 2 additional cases of checkpoint-inhibitor encephalitis. A) EBER stain of neural tissue from additional case 1 and B) additional case 2 (details of case in methods).

Discussion

The molecular basis of immune checkpoint inhibitor toxicities is unclear, and difficult to analyze as such cases seem to arise stochastically in the clinic. While these toxicities can affect nearly all organ systems, encephalitis is the most

common neurological toxicity associated with checkpoint inhibitor therapy.[431] Herein, we report for the first time the clinical and molecular features of a fulminant and fatal case of anti-PD-1 induced encephalitis. It is important to note that tissues collected in this study anteceded treatment with steroids and anti-TNFα, which could impact the immunologic status of the patient. Characterization of affected brain tissue demonstrated robust and clonal infiltration of cytotoxic CD4+ activated memory T cells, reactive immune checkpoint (PD-L1) expression, and evidence of concurrent EBV infection. To our knowledge, this is the first molecular description of immune-therapy induced neurotoxicity and the first suggestion of interplay between viral infection and immune checkpoint inhibitor toxicity.

Recent studies have suggested that neurologic irAEs affect approximately 1-3% of patients treated with anti-PD-1 therapy, with higher proportions of ipilimumab-treated patients affected (up to 14%) (10,31-33).[403,432–434] Encephalitis/meningitis, neuropathy, and myasthenia gravis were the most common adverse events noted in these studies, showing these events are recurrent but uncommon. Although these studies are important findings, our pharmacovigilance database characterized 209 cases of immune checkpoint inhibitor encephalitis (compared with a total number of 9 cases combined in the

largest series published to date), which is augmented by 22 cases identified and curated at large academic centers (**Table 7**).

In this case, the patterns of immune cell infiltration and TCR clonality were also reminiscent of our recent observations in fulminant immune checkpoint inhibitor-mediated myocarditis. Given this recurrent presentation, we posit that the canonical features of severe and fulminant checkpoint inhibitor toxicity include 1) T cell and macrophage infiltration into inflamed tissue, 2) clonal expansion of infiltrating T cells, and 3) reactive upregulation of PD-L1as a compensatory mechanism to limit tissue injury. We speculate that PD-L1 upregulation fails to abrogate tissue damage in the presence of anti-PD-1, thus potentiating inflammation and injury.

The cellular mechanisms of immune checkpoint inhibitor toxicity remain poorly characterized. We and others have suggested that causes or correlations could include shared antigens present in tumor and affected tissue,[398] pre-existing subclinical autoimmune responses (e.g. autoantibodies),[435,436] early B cell changes,[437] and specific intestinal microbiota.[438] We observed an extremely clonal population of T cells present in this sample. Intriguingly, bioinformatic analyses suggested that this represented an unusual activated memory cytotoxic CD4+ T cell population. Interestingly, cytotoxic CD4+ T cells have demonstrated roles in autoimmunity (including multiple sclerosis),[439] antitumor immunity,[440,441] and both

protective and pathologic[442] roles in antiviral defense (including against EBV).[443] Cytotoxic CD4+ T cells also play an essential protective role by directly lysing EBV-infected B cells while providing cytokine help; this highlights not only their abilities to recognize and kill virally infected cells but also their enhancement of the adaptive and innate immune response against EBV (reviewed by Martorelli *et al*. and Hislop and Graham).[444–447]

Though our ability to assign causality to EBV specifically is limited by tissue availability and lack of in-depth longitudinal molecular monitoring of the primary patient, several features led us to hypothesize that EBV in the setting of concurrent anti-PD-1 may have served as a potential trigger of this encephalitis. First, the patient had persistent albeit modest levels of EBV in the CSF and blood in conjunction with imaging and clinical findings reminiscent of viral encephalitis. Second, the extremely high degree of T cell clonality, which far exceeded our analyses of other irAEs (myocarditis, myositis, and colitis), suggested that a single epitope drove inflammation. Third, >10% of TCRs in the inflamed brain were identical to known CD8+ TCRs specific to EBV, a frequency much higher than previously described even with acute EBV infection or in autoimmune disease where EBV-specific T cells were identified. In addition, the dominant clone bore a striking degree of homology to a known CD8+ EBV-specific TCR in the form of a PSGS motif that was significantly enriched; this motif has been described only in

antiviral rather than autoimmune responses to date. However, this sequence localized to the aforementioned cytotoxic CD4+ T cell population, limiting our ability to more precisely specify the culprit antigen. Additionally, while we identified TCR sequences known to recognize EBV antigens in a class I-restricted fashion, we cannot exclude the possibility that the dominant CD4+ clone we describe above is restricted by HLA-A2, as both CD4+ T cells restricted by HLA class I and CD8+ T cells restricted by HLA class II have been described in a variety of healthy and disease states.[448–451] The frequency and cytotoxic memory phenotype of the dominant CD4+ clone we describe above suggests a novel role for cytotoxic memory CD4+ T cells if the clone is tumor-reactive, both in the context of tumor control and potential danger of autoreactivity as shown by others.[440,441] Indeed, others have observed that select TCRs are restricted by peptides presented individually on both HLA class I and class II, implying that such TCRs possess multiple specificities.[452] Finally, latent B cells present in the brain expressed EBER and EBNA1 mRNAs, which encode EBV-proteins. Although these cells were infrequent, nine additional encephalitis cases completely lacked these cells. We speculate that latent EBV+ lymphocytes may have leaked into the CSF, perhaps due to the prior CNS metastatectomy or radiation, setting the stage for the fatal immune response. Recent work by Harley *et al*. implicates EBNA2-dependent transcriptional programs as potential causal agents of multiple immunopathies,[453] though in a fashion not

necessarily mediated by T cell inflammation and subsequent pathology as suggested by the data we present here. The presence and frequency of both CD4+ and CD8+ cytotoxic T cell populations in the brain and tumor microenvironment merits further study in additional cancer types and in autoimmune models, especially given that the additional cases we examined both showed marked inflammation-associated T cell infiltration into the brain.

In conclusion, this is the first molecular analysis of neurotoxicity due to immune checkpoint inhibitors to our knowledge. In addition, we believe this is the first evidence linking concurrent viral infection with irAEs, and the first detailed identification and phenotyping of culprit immune cell populations. Further studies are needed to determine the best interventions for prevention and management of these rare but fulminant toxicities.

Chapter VI

Select vignettes of the application of computational methods in the analysis of single-cell RNA sequencing data.

N.B. The methods described in this chapter are methods that I have applied to select or multiple scRNA-seq data sets derived from samples of human disease and healthy human samples. They comprise a computational toolbox that can be used to understand heterogeneity of the immune system, identify disease-specific transcriptomic features, dissect single cells using multi-modal/multi-omic approaches, and will be critical to future studies flowing from the work described in this thesis.

As I describe in Chapter I of this dissertation, current analytical approaches to scRNA-seq and similar single cell measurements typically fall into one of four classes, ranked from more to less common: 1) clustering and analyses, 2) trajectory and differentiation analyses, 3) gene and transcription factor mechanic analyses, and 4) network inference analyses. In this chapter, I provide several brief examples and applications of these different types of analyses to scRNA-seq data derived from different examples and types of human disease.

Consensus and non-consensus clustering of single immune cells

Allopurinol can induce an HLA-B*58:01-associated severe blistering disease of the skin known as Stevens-Johnson syndrome/toxic epidermal necrolysis (SJS/TEN). The major cytolytic mediator of this disease is granulysin (gene symbol GNLY), which is known to be produced by T lymphocytes and NK cells in the blister fluid of patients

suffering from SJS/TEN. Consensus clustering is a clustering approach that uses multiple types of clustering algorithms—each with distinct loss functions, simply described as “goals”—with the aim of locating stably identifiable groups of observations (single cells). Here, I apply consensus clustering implemented by the SC3 algorithm to show that there are distinctive genes that clearly separate activated CD8+CD137+ T cells from the blister fluid of an allopurinol SJS/TEN patient from non-activated CD8+CD137− T cells from the same blister fluid (**Figure 68**).

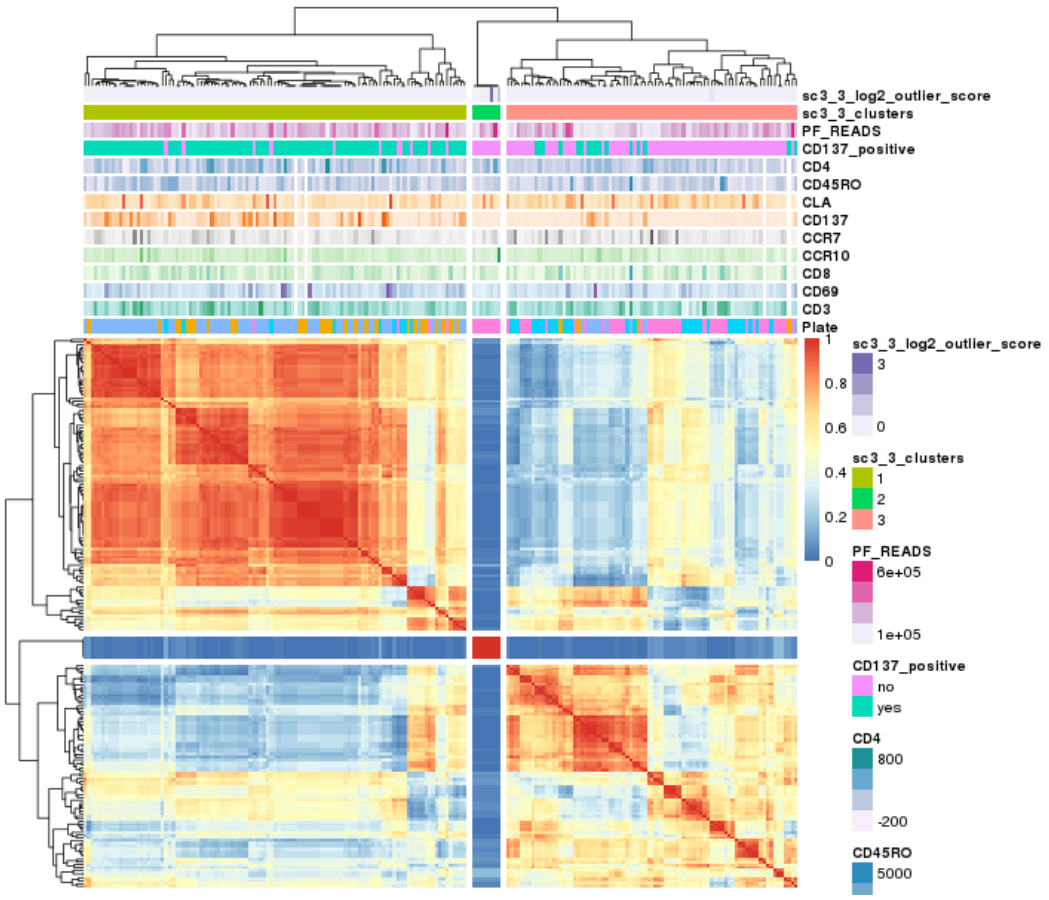


Figure 68. SC3 consensus clustering separates activated and non-activated single T cells from the blister fluid of an allopurinol SJS/TEN patient. In this diagram, rows represent genes and columns represent individual cells from multiple plates of index-sorted CD8+CD137+ or CD8+CD137− T cells from an allopurinol SJS/TEN patient. The indexed protein measurements from flow cytometry are represented as rows above the columns, as well as the read alignment ratio (PF_READS), the plate of origin, and each cell’s CD137 expression

status (±). Three distinct clusters are observable: 1) a mostly CD8+CD137+ population, 2) a mostly CD8+CD137− population, and 3) an outlier cluster of CD8+CD137− cells that are enriched in mitochondrial genes and which express fewer genes than their peers in clusters 1 and 2.

This demonstrates several of the advantages of consensus clustering, namely: 1) the ability to inspect multiple pieces of metadata in relation to the degree of heterogeneity within a cellular population, 2) the automatic identification of signatures and "metagenes" that best explain the variation between groups of single cells, and 3) the identification of groups of cells that multiple algorithms vote on as being biologically similar, and therefore more likely to exist (validated in this analysis by the clear separation of ground-truth CD8+CD137+ and CD8+CD137− cells). Further inspection of the signature distinguishing between these populations using a mixed-effect Poisson regression test for differential expression with a random-effect to account for technical variation across plates clearly reveals genes of immunologic importance enriched in the activated population, including: 1) the identification of the *TRBV* and *TRAV* genes encoding the dominant and pathologic clonotype in this patient, 2) a gene signature with enrichment of genes that negatively regulate IL-12 production (p = 2.78 x 10^{-6}, FDR q-value 5.86 x 10^{-3}), and 3) selective expression of *KLRC1*, a negative inhibitor of NK and T cell cytotoxicity, in the pathologic T cell population (**Figure 69**). A subset of these genes have at least 80% power to classify T cells from this blister fluid as belonging to one of the 3 main clusters using a simple one-vs-one multiclass classification function, demonstrating

the power of this approach to identify from hundreds to thousands of differentially expressed genes a more limited set that may be more relevant (**Figure 70, next page**).

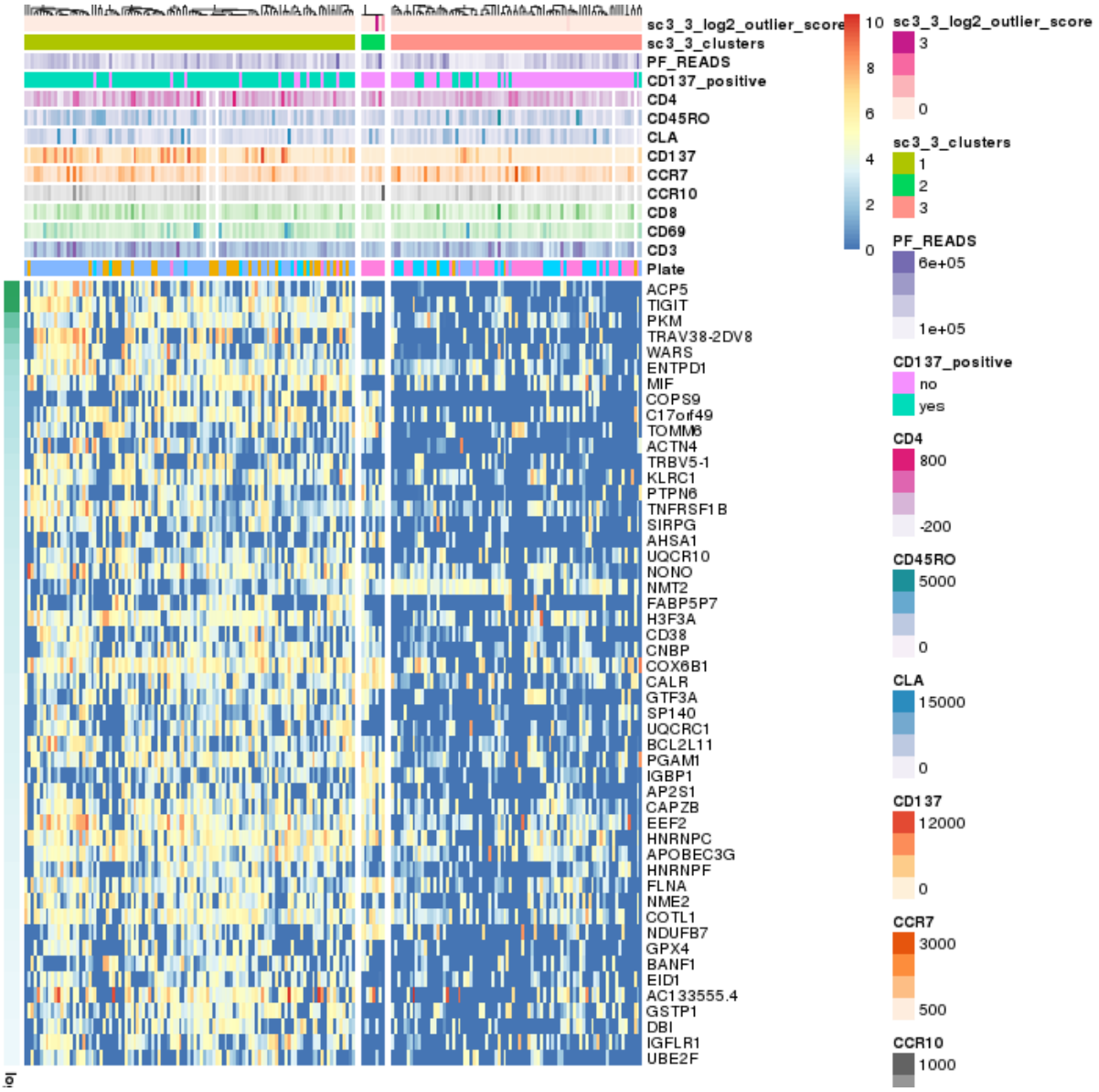


Figure 69. Consensus clustering reveals a subset of strongly differentially expressed genes between activated CD8+CD137+ and CD8+CD137− blister fluid T cells. The 50 most-differentially expressed genes between the 3 consensus clusters in this analysis reveals enrichment of co-stimulatory and co-inhibitory genes (TIGIT, KLRC1, CALR, TNFRSF1B, WARS), metabolic genes (PKM, WARS, HNRNPC, H3F3A, EID1), genes with a 3' UTR NFAT-binding transcription factor binding site (H3F3A, NONO, WARS, COTL1, TNFRSF1B, PGAM1), and genes with a 3' UTR PAX4-binding transcription factor binding site (HNRNPC, HNRNPF, PTPN6, UBE2F, ENTPD1, IGFLR1, CD38).

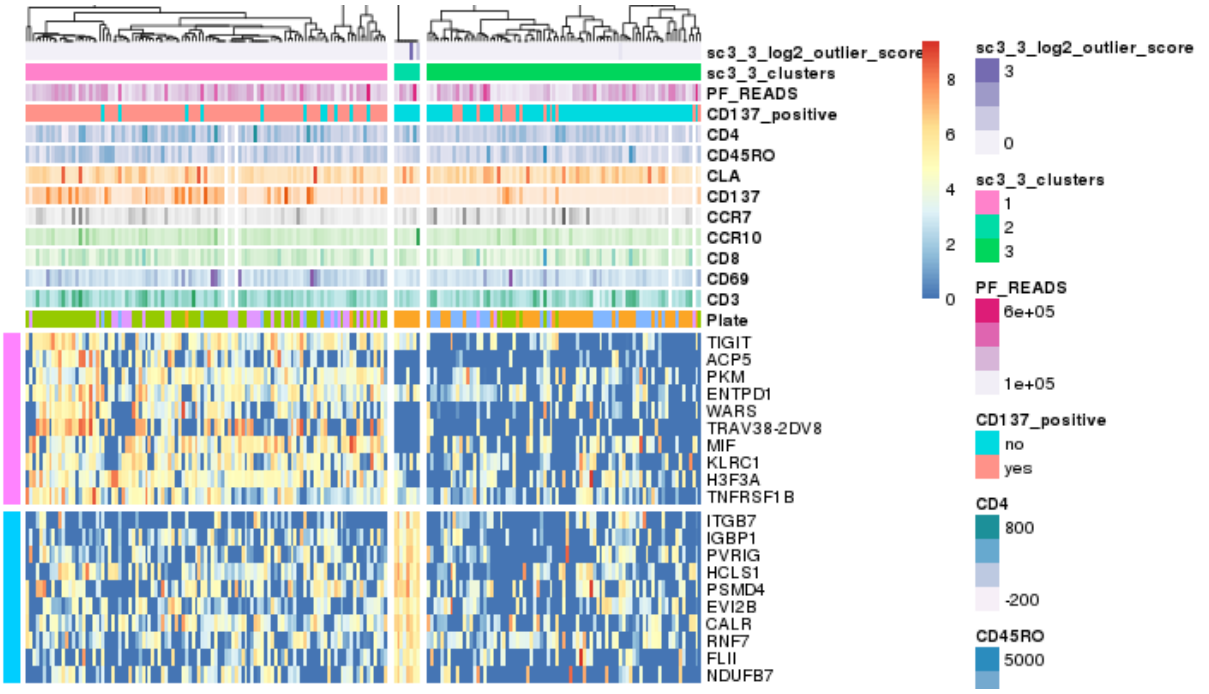


Figure 70. A subset of genes have strong classification power (AUROC > 0.8) between CD8+CD137+ and CD8+CD137− T cells from the blister fluid of an allopurinol SJS/TEN patient. Genes displayed here in rows have strong classification/discriminatory power (at least 80% area under the receiver operating characteristic curve) between activated and non-activated single T cells from the blister fluid.

Quality analysis of these results indicates that the outlier/low-quality cluster is a group of cells that are commonly unassigned to the CD8+CD137+ or CD8+CD137− clusters; plotting the silhouette of how far cells within these clusters are from cells in the other clusters reveals that these cells are frequently consensus clustered "between" the two main clusters (i.e. they are difficult or impossible to stably identify as belonging to one of the main clusters, which are strong communities in multidimensional space) (**Figure 71, next page**). This suggests that these cells should be treated as their own analytical subgroup/cluster, excluded from further

analysis downstream, or evaluated with a different clustering scheme.

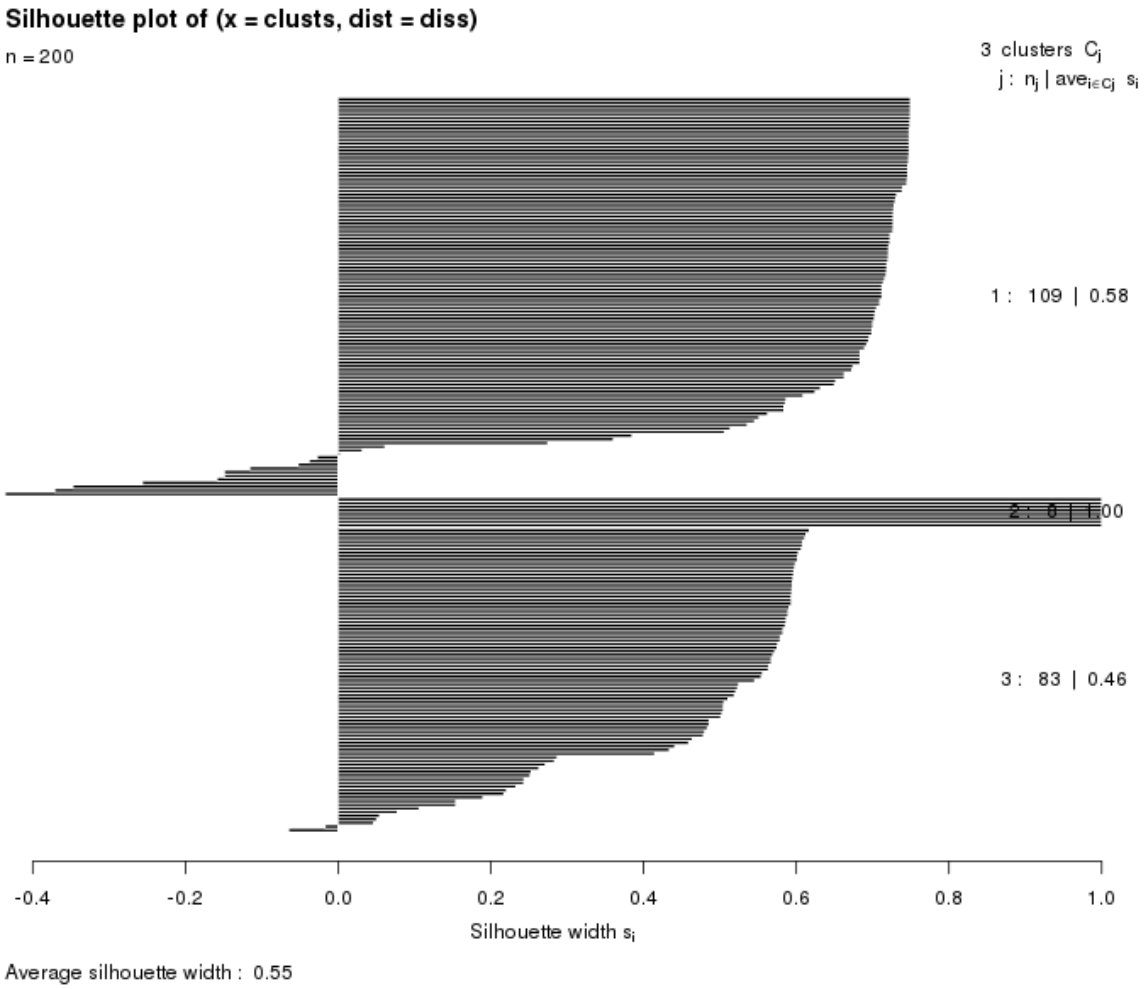


Figure 71. Evaluation of SC3 consensus clustering quality via silhouette analysis identifies low-quality single T cells from the blister fluid. Each line is a single T cell as shown in the figures above. Negative silhouette widths represent cells that are unstably clustered into a given group, while positive silhouette widths represent single cells that are stably and repeatedly assigned to a cluster by multiple algorithms.

Elimination of persistent batch effects from SMART-seq2 scRNAseq data and 10X Chromium scRNAseq data

Idiopathic subglottic stenosis (iSGS) is an upper airway disease that afflicts a population of Caucasian women of western European descent who are mostly over

the age of 40 years old.[454] This disease is characterized by accrual of fibrotic scar tissue enriched in CD8+ resident memory and effector memory T cells. Due to the stochastic nature of sample availability (i.e. fresh single-cell suspensions are only available for processing, sorting, and sequencing when a subglottic scar is removed from one of these patients), cells from the patients analyzed here had to be sequenced across two sequencing runs. Several samples failed the first sequencing run (run #124), and had to be re-sequenced in a second run (run #124).

Initial analyses of these data revealed strong separation in multidimensional space that was almost completely explainable by which sequencing run the data corresponding to each cell were generated on (**Figure 72**). Attempts to correct these data using the canonical correlation analysis (CCA) algorithm of Seurat[189] were unsuccessful, and clustering analyses merely separated cells into clusters that grouped by sequencing run—making downstream analysis unreliable as the sequencing run was a perfectly confounding variable. This is a frequent challenge faced by many groups, with potential to endanger published results if the conversation between sequencing cores, wet lab biologists, and computational biologists does not identify such batch effects. While there are many batch-correction algorithms,[188,243,455–460] most of which function by identifying "nearest neighbors" among cells in multidimensional space or by using a form of deep machine learning called autoencoding, there are few-to-no benchmark datasets

available to test existing and new correction tools, and existing methods frequently compare their performance to only one or two other methods; hence, a tool that works well for one dataset may not always work well for another.

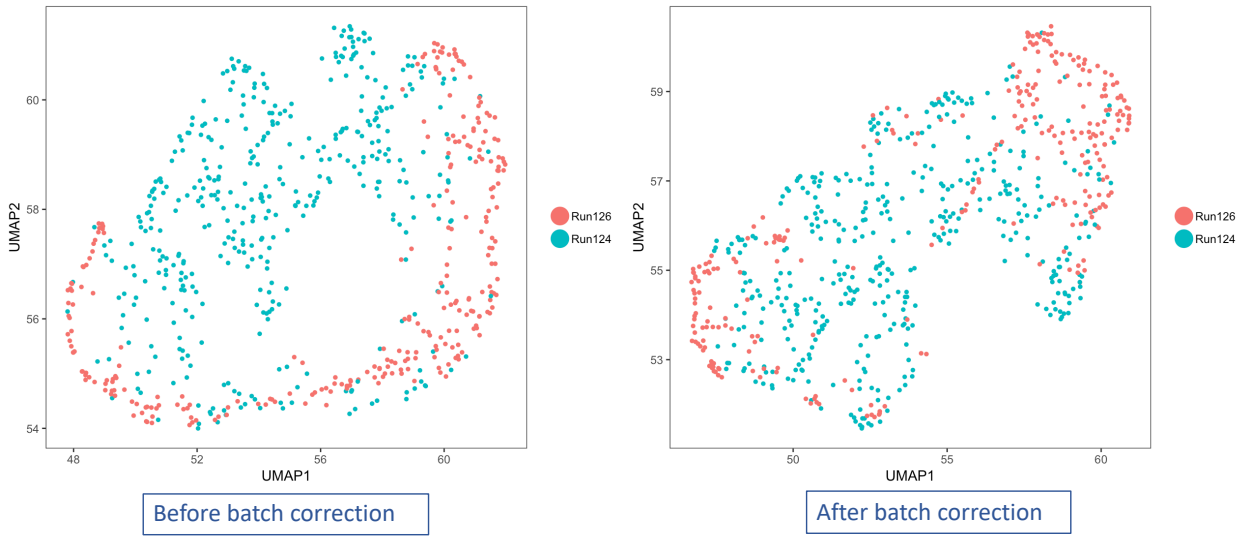


Figure 72. Removal of batch effects from SMART-seq2 scRNA-seq data. Left: Efforts to correct the data for sequencing run using Seurat's CCA algorithm were unsuccessful, and showed a clear separation and clustering in multidimensional space based on sequencing run (shown via UMAP dimensional reduction). Right: Application of the Harmony correction algorithm, which uses local linear regression after performing k-clustering across a wide range of *k*, successfully eliminates this batch effect.

Given that our attempts to correct the data with the CCA algorithm failed, we instead applied the Harmony algorithm to the principal components of these data and re-ran Leiden clustering in Seurat. After this successful correction, 3 distinct clusters of cells appear that are evenly and unevenly represented in cells from 3 patients with idiopathic subglottic stenosis, one with ANCA-associated vasculitis, and a sixth with post-intubation stenosis (**Figure 73**). Thus, performing differential expression analysis of cells within these 3 clusters enables the identification of

genes that are uniquely expressed by T cells of iSGS.

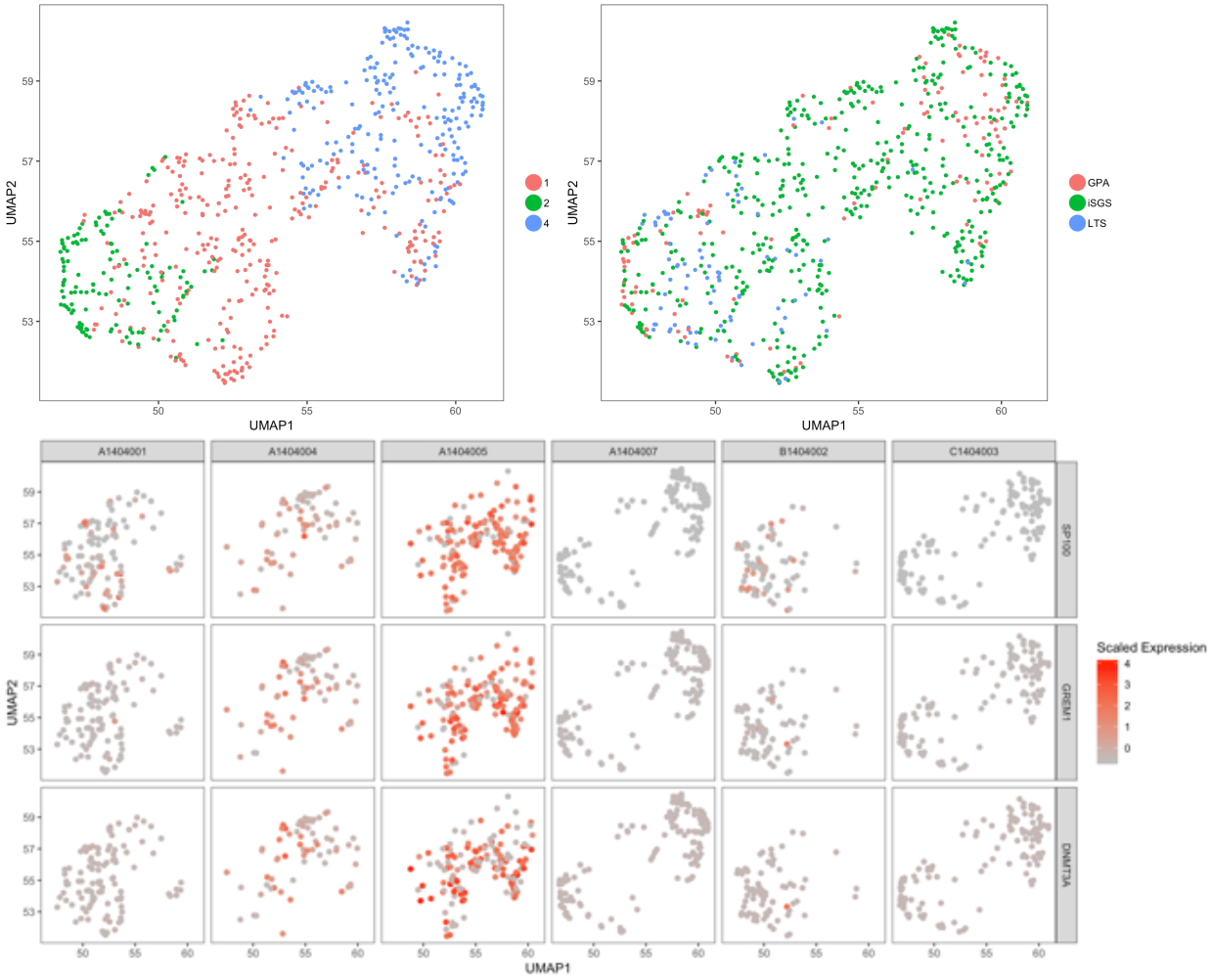


Figure 73. Harmony correction enables joint clustering and batch effect removal, leading to the identification of a select set of marker genes unique to CD8+ T cells infiltrating the scar tissue of iSGS patients. Top left: Three clusters of cells are readily identifiable by Leiden clustering using the Harmony-corrected principal components as input. Top right: The resulting clusters are comprised of T cells from iSGS patients (iSGS, green, 4 left-most columns bottom panel), an ANCA-associated vasculitis patient (GPA, red, second-from-right column bottom panel), and a post-intubation patient (LTS, blue, right-most column bottom panel). Bottom: Differential expression between these three clusters enables the identification of *SP100*, *GREM1*, and *DNMT3A* as iSGS-specific marker genes ($p < 0.0001$, all comparisons).

Harmony and similar approaches are also excellent for the correction of effects introduced by biological effects (as I show in chapter II), but also for technical effects introduced by different single-cell sequencing technologies. Below,

I show that application of Harmony to data from the same allopurinol blister fluid samples sequenced on both the 10X Chromium 3' platform and the 10X Chromium 5' platform effectively removes batch effects present in the data due to differences in the sequencing technology and also enables the identification of shared transcriptomic features and immune populations in two separate donors (**Figure 74**).

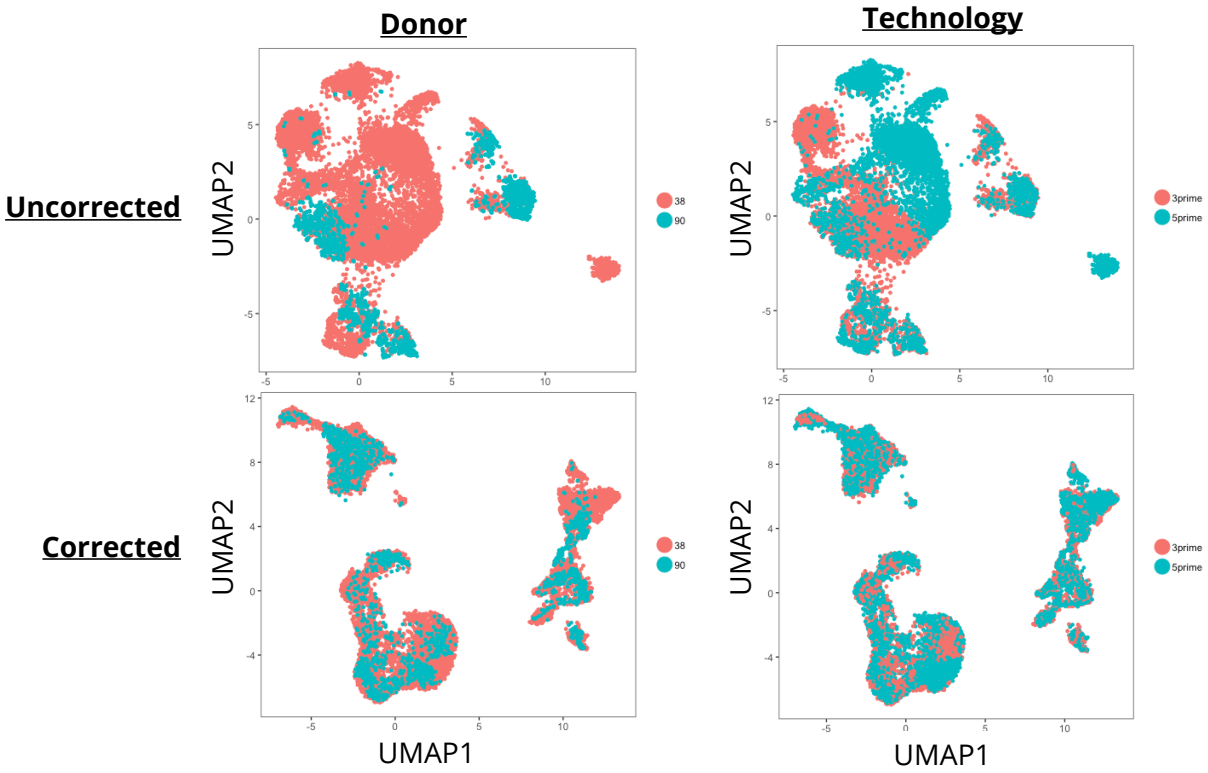


Figure 74. Harmony correction enables the identification of shared populations of immune cells found in two different allopurinol SJS/TEN patients and correction of sequencing platform batch effects. Top panels: UMAP dimensional reduction utilizing uncorrected principal components; plots showing individual cells as dots colored by donor of origin (left) and sequencing technology (right). Bottom panels: UMAP dimensional reduction using Harmony-corrected principal components; cells are colored as in the top panels.

This analytical approach also proves useful in learning the strongest features that are pathology-specific or cell-type-specific. In an extension of this experiment, after

stimulating separate aliquots of PBMCs from these patients with staphylococcal enterotoxin B (SEB), oxypurinol (the drug metabolite to which allopurinol SJS/TEN patients experience a reaction), or unstimulated PBMCs (though it is possible that T cells within the PBMCs were being stimulated *in vivo* by oxypurinol as the patient was experiencing severe SJS/TEN at the time of blood draw), we applied Harmony correction to these data to demonstrate that the cytotoxic and cytolytic T cells found in the blister fluid of these patients (CD3+CD8+GNLY+GZMB+) are enriched in CXCL13, a chemotactic molecule used to recruit B cells into germinal centers (**Figure 75, see next page**). These results also suggest that the blister fluid of allopurinol patients shares aspects of tertiary lymphoid structures found in patients whose tumors contain CD8+CXCL13+ T cells.[461,462]

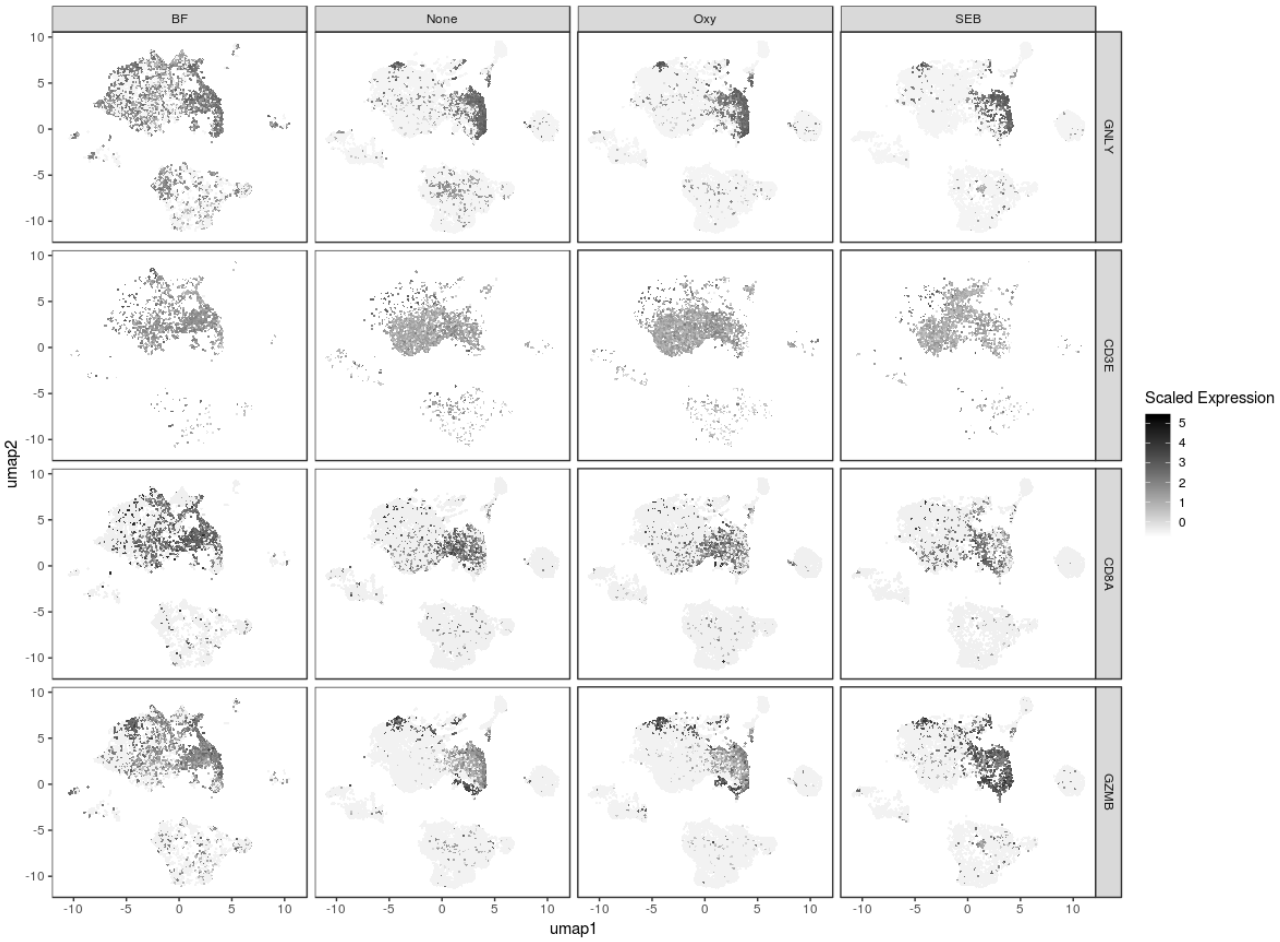


Figure 75. Harmony analysis enables the identification of blister-fluid-specific transcriptomic features that differ from artificially stimulated T cells and T cells found in the peripheral blood. Expression is colored on each cell from lowest (white) to highest (dark). Each dot represents a single cell, and cells are separated by their sample types in columns. This shows that the cytolytic granulysin-granzyme B-T cell signature is not unique to blister fluid. Rather, the expression of granulysin within a distinct population of T cells in the blister fluid, many of which also express the chemotactic molecule CXCL13, is unique—as well as the widespread expression of these cytotoxic molecules in other T cell populations in the blister fluid.

These analyses show the power and flexibility of so-called joint analyses, whose strengths include the benefit of added statistical power (by pooling cells from multiple samples, the number of cells that can be compared between clusters is increased, and the probability of detecting rare populations of cells is increased), the correction of batch effects within data, and the ability to identify common

transcriptomic features within heterogenous populations of single cells from different individuals.

Normalization of SMART-seq2 data using 'sequins'

Normalization of scRNA-seq data is a problem for which many researchers have devised approaches. One of the most recent approaches includes the use of synthetic External RNA Control Consortium (ERCC) spike-in synthetic RNAs,[463] which are pooled together and vary in their length and GC base content. However, ERCC sequences are derived from *in vitro* transcription of synthetic DNA sequences originating from *Bacillus subtilis* and *Methanocaldococcus jannaschii*.[463] As a result, these sequences are useful for assessing the quantitative accuracy of scRNA-seq and RNA-seq experiments, but they are limited in their usefulness to assess differential expression of individual isoforms, identify isoform fusions, and represent single-exon transcripts. In contrast, 'sequins' (so named as a portmanteau of "sequencing" and "spike-ins"), are synthetic RNAs designed by sampling regions of the human genome and 1) inverting the sequences, 2) mutating the sequences to remove homology to known human genes while maintaining GC content, sequence uniqueness, repetition, and original nucleotide composition.[464] The *in silico* chromosome containing these sequins is comprised of 78 loci, with 164 alternatively spliced isoforms made of 869 exons and 754 unique

introns. These sequins also vary significantly in the number of exons in each sequin, in length, and in isoform complexity.

Recent innovations in scRNA-seq analyses such as isoform-level differential expression,[186] reconstruction of the TCR and BCR,[183,184] and detection of single-cell-specific gene fusion events would all benefit from wet-lab and dry-lab approaches enhancing study design. Sequins represent a promising addition to the single-cell toolkit. In a proof-of-principle study, we spiked Sequins over a range of concentrations into 96-well plates containing the necessary reagents to perform SMART-seq2 scRNA-seq.

Using the Anaquin pipeline,[465] we aligned and quantified Sequin-derived sequencing reads, and used the scater package[466] to normalize read counts in relation to 1) the number of Sequin genes detected, 2) the abundance of each Sequin gene, and 3) known sequencing quality control metrics. We found that the acceptable concentration of Sequin per well likely ranges between 0.005 picogram (pg)/μL and 0.09 pg/μL (**Figure 76, see next page**). This ensures that Sequins only occupy 1-10% of the sequencing library for each cell.

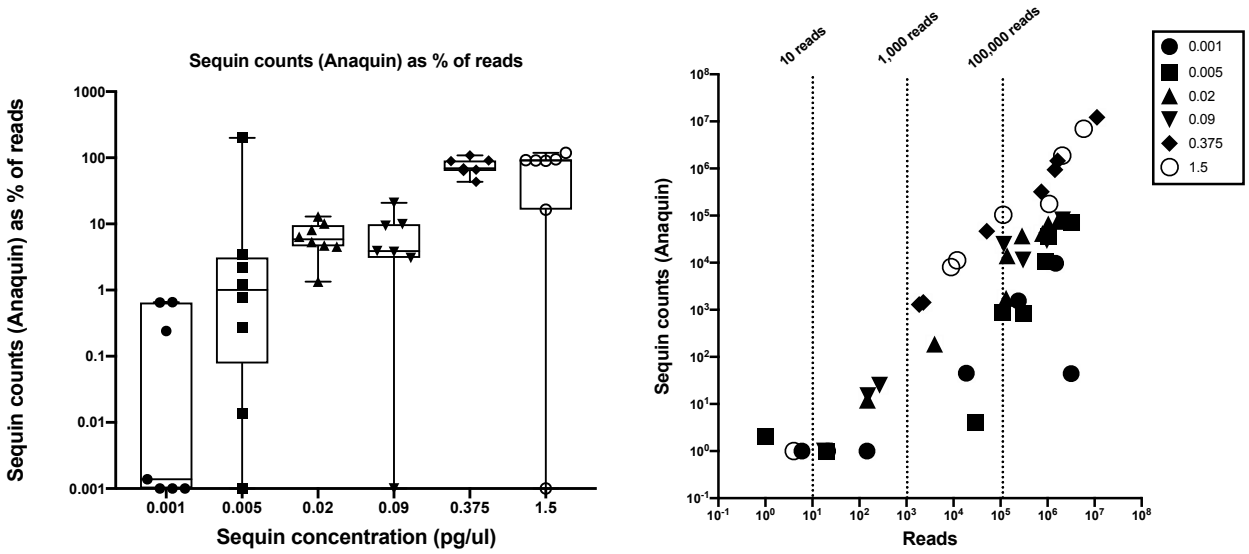


Figure 76. Sequins can be spiked-in to plate-based scRNA-seq samples without compromising library integrity. Left: The optimal concentration of Sequin per well is approximately 0.005 to 0.09 pg/µL. Beyond this concentration, Sequins take over the sequencing library (i.e. cell-derived RNA is drowned out by the Sequin-derived signal). Right: Relationship between sequencing depth (reads/cell) and Sequin counts as quantified by the Anaquin pipeline. As expected, the number of Sequins detected per cell increases with sequencing depth, demonstrating that Sequins are a reliable synthetic spike-in.

Analysis of the Sequin-corrected count data revealed successful reconstruction of correlations between known quality-control genes (*ACTB* and *GAPDH*), the relationship between mitochondrial content per cell and Sequin diversity (suggesting that Sequin diversity within a given single cell library is an intrinsically valuable quality control measurement), and known cell-size correlations (*CD3E* and *ACTB*) (**Figure 77**). These analyses provide evidence that Sequins are a suitable substitute for ERCC spike-ins; importantly, Sequins are free of charge for research use, suggesting that well-designed scRNA-seq experiments can benefit from the

addition of Sequins without adding reagent cost.

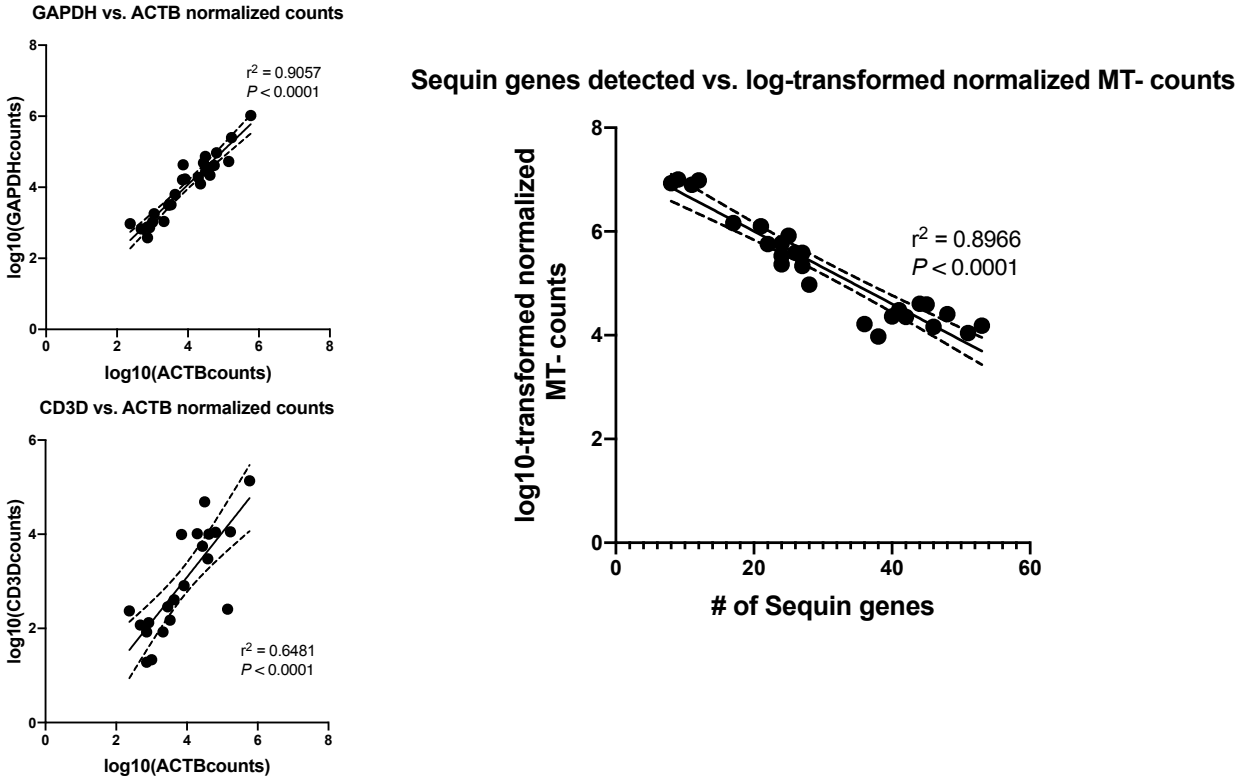
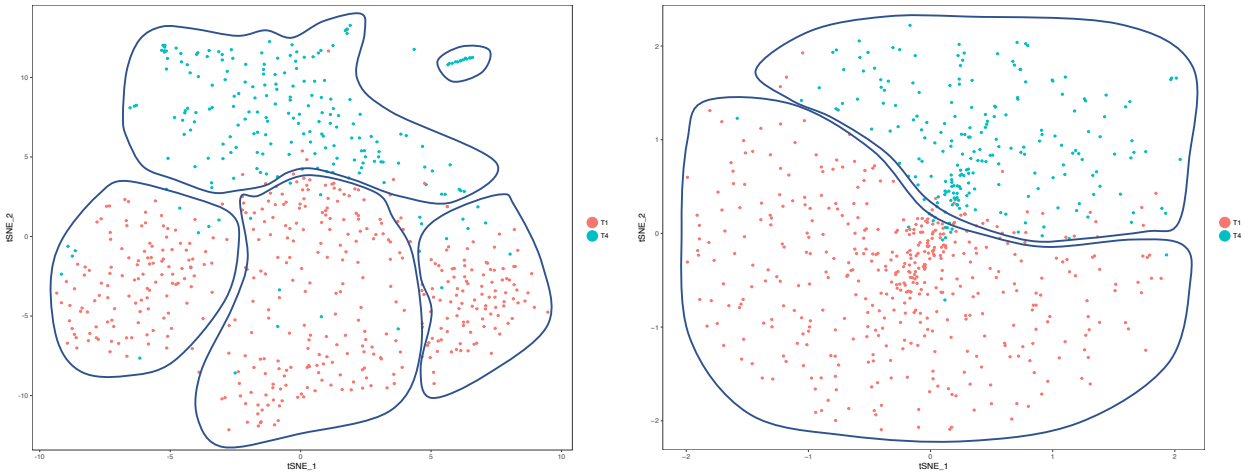


Figure 77. Sequins recover known correlations of high-strength, quality control, and cell size. Top left: The correlation between *GAPDH* and *ACTB* expression is commonly >0.9 in bulk RNA-seq experiments, but due to dropout and technical variation this correlation is lost in scRNA-seq data. Simple normalization using Sequin counts recovers this relationship. Bottom left: The correlation between *ACTB* and *CD3D* (and other CD3 variants, data not shown) suggests, as would be expected, that increased levels of CD3 complex members correspond to cell size. Right: The mitochondrial content (the percent of counts corresponding to mitochondrial genes) strongly negatively correlates with the number of Sequin genes detected in each single cell, suggesting that Sequins can automatically identify cells that are information-poor.

Rebuilding the TCR repertoire of cells responding to an IL-12 adjuvanted HIV vaccine

HIV Vaccine Trials Network trial 087 (HVTN-087) evaluated the immune response elicited by priming over the course of 3 months with a DNA plasmid vaccine containing *gag*, *pol*, and *env* and boosting 3 months afterwards with a pseudotyped

vesicular stomatitis virus expressing *gag*. Trial participants also received 1 of 4 concentrations of IL-12 adjuvant in parallel: 0, 250, 1000, or 1500 μg of IL-12. Evaluation by intracellular staining of CD4+ T cells from individuals in this study in response to HIV Gag peptide pools revealed that increased dosage of this vaccine did not increase the size of the HIV-specific CD4+ T cell response, but that it did correlate with increased HIV-specific CD8+ T cell responses in a dose-dependent fashion.[467] We sorted activated CD8+CD137+ T cells after overnight HIV Gag peptide pool stimulation of study participant PBMCs from 4 no-IL-12 recipients and 2 1500-μg-IL-12 recipients, and performed SMART-seq2 to analyze the transcriptomes of these potentially vaccine-reactive CD8+ T cells. After correcting the raw data for cell cycle (as in McDavid *et al.* [468]) and sequencing depth using Seurat, we visualized these single cells using the t-SNE dimensional reduction algorithm (**Figure 78, next page**).



Standard t-SNE embedding

t-SNE embedding of ML-derived features

Figure 78. t-SNE and Leiden clustering reveal distinctive groups of CD8+CD137+ T cells in response to HIV peptide stimulation and an IL-12-adjuvanted transcriptomic signature that is identifiable with supervised learning.

We then used the TPOT genetic programming suite[469,470] to build a support-vector machine capable of classifying cells as originating from no-IL-12 or 1500-µg-IL-12 donors using a limited set of genes; this supervised learning approach allowed us to build a potential "vaccine response" signature that could be applied to single cells from other vaccinees in the future (**Figure 78**). In order to gain a better understanding of the high-level differences between these T cells, we performed two main analyses: differential expression analysis using the MAST and bimod tests[223,471] and reconstruction of the TCR using TraCeR.[183] We found that IL-12-

receiving individuals produced CD8+CD137+ T cells high in inflammatory molecules such as MIP1α, MIP1β, and STING (all $p < 0.0001$, **Figure 79**).

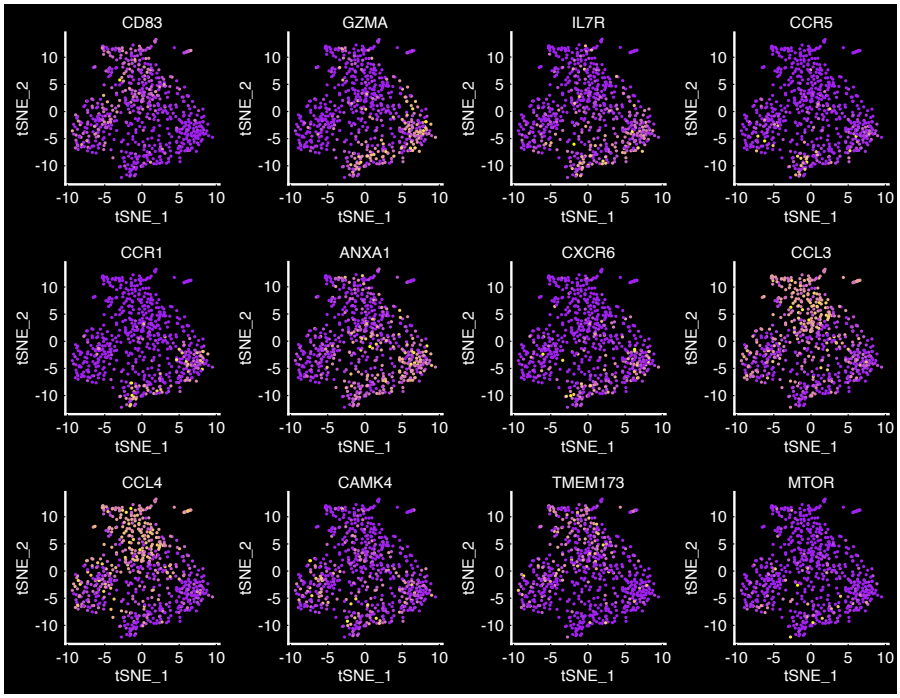


Figure 79. Identification of differentially expressed genes between IL-12-receiving and IL-12-non-receiving individuals. CCL3 and CCL4 are two examples of genes distinguishing between the single T cells of IL-12-receiving and IL-12-non-receiving cells shown in the previous **Figure** (where CCL3 and CCL4 are more highly expressed in the peptide-reactive T cells of IL-12-receiving vaccinees).

We also identified a population of mucosal-associated invariant T cells found in all vaccinees; whether or not these MAIT cells were responsive to HIV peptide representative of noise due to activation marker selection is still unclear, and no reports in the literature have identified these cells using similar approaches or single-cell sequencing. We recovered the majority of TCR sequences from each sorted cell using TraCeR, and identified dominant clonotypes within each patient that corresponded with bulk TCR sequencing frequencies of the same activated

CD8+CD137+ population (**Figure 80; Table 9, next page**).

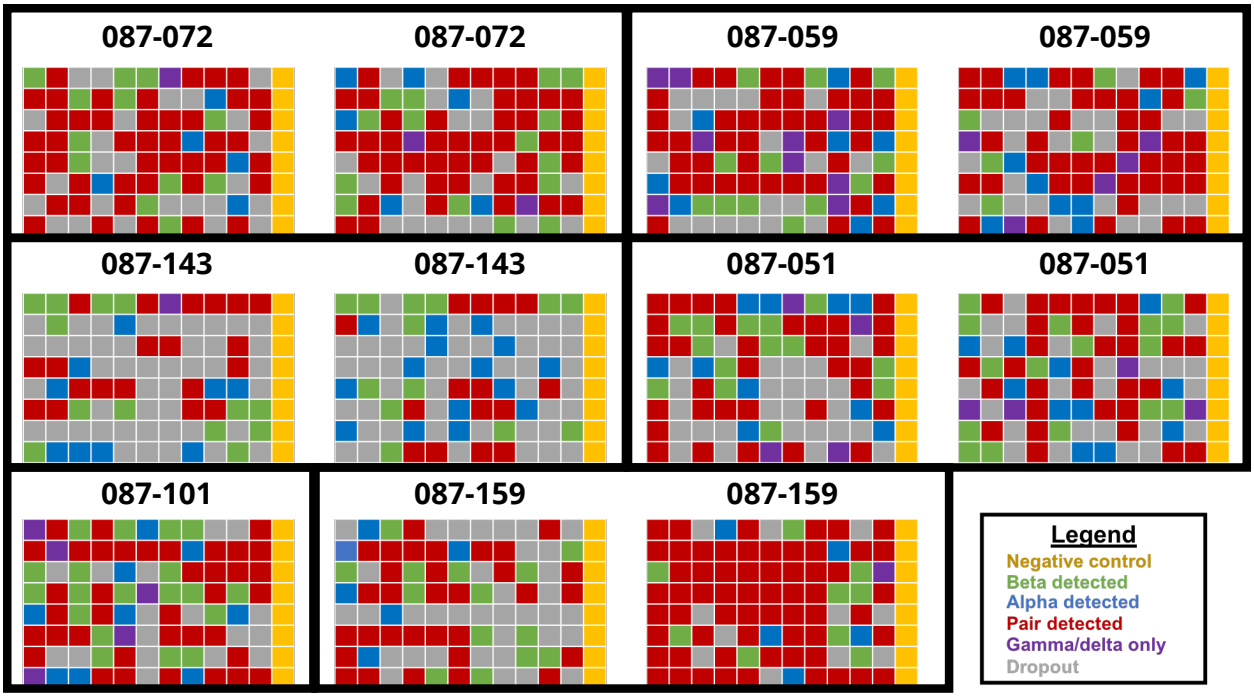


Figure 80. Successful reconstruction of alpha, beta, gamma, and delta TCRs from SMART-seq2 data of CD8+CD137+ T cells. Each square represents a single cell in a single well of a 96-well plate. We successfully recovered paired TCR$\alpha\beta$ sequences from the majority of cells, but we also recovered TCR$\gamma\delta$ sequences from multiple cells. Further analyses (data not shown) revealed that approximately 30% of single T cells in this experiment express 2 TCRα sequences, indicating that allelic exclusion is detectable at the single-cell level.

087-159	Frequency								
TCR	**Single cell**	**Bulk**	**TRBV**	**CDR3**	**TRBJ**		**TRAV**	**CDR3**	**TRAJ**
1	56.3%	28.5%	*TRBV-9*	CASSVDWAEYNEQFF	*TRBJ 2-1*		*TRAV 13-1*	CATYNNDMRF	*TRAJ 43*
2	9.4%	15.3%	*TRBV 4-1*	CASSQDRLTGGYTF	*TRBJ 1-2*		*TRAV 24*	CASSSSNTGKLIF	*TRAJ 37*
3	7.0%	6.7%	*TRBV-19*	CASFWTTGDGNTIYF	*TRBJ 1-3*		*TRAV 38-2*	CALGGGAQKLVF	*TRAJ 54*
4	1.6%	1.9%	*TRBV 5-6*	CASSPGWGTYEQYF	*TRBJ 2-7*		*TRAV 12-3*	CAMSGDTGNQFYF	*TRAJ 49*

087-101	Frequency								
TCR	**Single cell**	**Bulk**	**TRBV**	**CDR3**	**TRBJ**		**TRAV**	**CDR3**	**TRAJ**
1	26.2%	27.0%	*TRBV 13*	CASSLQRDGEQYF	*TRBJ 2-7*		*TRAV 20*	CAVPVRDSGYALNF	*TRAJ 41*
2	15.4%	3.2%	*TRBV 13*	CASSLSRAWEDTQYF	*TRBJ 2-3*		*TRAV 41*	CAVVPSTGTASKLTF	*TRAJ 44*
3	1.5%	1.7%	*TRBV 29*	CSVDYWDSLRDGYTF	*TRBJ 1-2*		*TRAV 39*	CAVDGGTSGTYKYIF	*TRAJ 40*

087-072	Frequency								
TCR	**Single cell**	**Bulk**	**TRBV**	**CDR3**	**TRBJ**		**TRAV**	**CDR3**	**TRAJ**
1	9.2%	49.8%	*TRBV 5-6*	CASSLGWGQNQPQHF	*TRBJ 1-5*		*TRAV 12-2*	CAVNMGGDNFNKFYF	*TRAJ 21*
2	2.5%	5.9%	*TRBV 7-6*	CASSLGWSTDTQYF	*TRBJ 2-3*		*TRAV 8-1*	CAVHFGNEKLTF	*TRAJ 48*

087-059	Frequency								
TCR	**Single cell**	**Bulk**	**TRBV**	**CDR3**	**TRBJ**		**TRAV**	**CDR3**	**TRAJ**
1	8.5%	43.0%	*TRBV 27*	CASSFDGPMNTEAFF	*TRBJ 1-1*		*TRAV 27*	CAGGTYKYIF	*TRAJ 40*
2	1.6%	1.4%	*TRBV 30*	CAWTYEQYF	*TRBJ 2-7*		*TRAV 1-2*	CAVRDSNYQLIW	*TRAJ 33*
3	0.8%	0.6%	*TRBV 6-4*	CASSQEPSGANVLTF	*TRBJ 2-6*		*TRAV 1-2*	CAAMDSSYKLIF	*TRAJ 12*

087-143	Frequency								
TCR	**Single cell**	**Bulk**	**TRBV**	**CDR3**	**TRBJ**		**TRAV**	**CDR3**	**TRAJ**
1	50.6%	87.8%	*TRBV 5-6*	CASSLGWGNEKLFF	*TRBJ 1-4*		*TRAV 13-1*	CAASNEGNTPLVF	*TRAJ 29*
2	2.4%	2.4%	*TRBV 5-6*	CASSLGFGGGDTQYF	*TRBJ 2-3*		*TRAV 38-2*	CASNTGNQFYF	*TRAJ 49*

087-51	Frequency								
TCR	**Single cell**	**Bulk**	**TRBV**	**CDR3**	**TRBJ**		**TRAV**	**CDR3**	**TRAJ**
1	64.1%	32.3%	*TRBV 5-1*	CASRLAERKFF	*TRBJ 2-1*		*TRAV 8-2*	CVVSETGNQFYF	*TRAJ 49*
2	4.3%	11.0%	*TRBV 5-1*	CASSLVYEQFF	*TRBJ 2-1*		*TRAV 13-2*	CADLKEGGSYIPTF	*TRAJ 1-2*
3	0.9%	10.9%	*TRBV 5-1*	CASSLGGKAFF	*TRBJ 1-1*				
4	0%	2.3%	*TRBV 30*	CAWMGPAGANVLTF	*TRBJ 2-6*				

Table 9. Dominant TCR clonotypes of CD8+CD137+ T cells from 6 HVTN-087 donors reconstructed using TraCeR.

These data demonstrate that it is possible to link surface protein expression, transcriptome, and TCR clonotype together in a fashion that enables the identification of features indicative of differences in the vaccine response to HIV, and potentially other pathogens.

Summary and Future Directions: Expanding the Immunologist's Toolbox

In the previous chapters, I have outlined a population-to-single-individual, and population-to-single-cell approach to studying immune-mediated human disease. In chapter II, I provided evidence that alteration of the CD4 T cell compartment is associated with mortality using flow cytometry measurements from a cohort of veterans. In chapters III and IV, I show how using an increased number of immunomic approaches (flow cytometry, TCR sequencing, RNA sequencing and quantification) can shed light on how the immune system is associated with glucose intolerance and obesity in mice and in humans (these studies are excellent examples of utilizing the contrast provided by the perturbation framework I describe in **Chapter I**, **Figure 2**. I conclude the dissertation in **Chapters V and VI** by demonstrating various ways in which single-molecule and immunomic approaches, including single-cell RNA sequencing, can clarify complex immunopathologic processes in multiple disease settings.

While randomized studies and prospective longitudinal studies are gold standards and keystones of evidence-based medicine, these approaches are less common in immunologic—and notably, single-cell—studies. There are several factors that explain this, including 1) the difficulty in collecting diseased tissue and cellular samples from humans while complying with ethical research standards and logistic challenges of collecting these samples, 2) the fiscal cost and analytical

expertise required to apply immunomic approaches to large cohorts, and 3) a surprising lack of tools that enable researchers to leverage the power of these study designs in existing computational tools.

In addition to existing proteoimmunogenomic approaches such as CITE-seq, further developments with respect to single-cell technologies would do well to elucidate the specificity of the B cell receptor repertoire; efforts are well underway in this regard in T cell immunology. While recent research has established approaches to identify spatial niches using single-cell sequencing and imaging, and even to identify T- and B-cell/APC interacting pairs,[472] bioinformatic and wet-lab approaches to achieve this in humans will need to be developed to extract this valuable information from rare human disease samples; one promising approach recently established a way of estimating these interactions by scoring receptor-ligand interaction via scRNA-seq.[473]

Recent developments in spatial transcriptomics[474] represent significant advancements in the single-cell field; these approaches enable the construction of spatial transcriptomic signatures that can be used to estimate the spatial proximity of single cells in samples where no spatial information could be preserved, and more importantly allow for the identification of multiple immune cell subsets “at the site of the crime.” I show one such example of this in Chapter V with a limited

number of transcriptomic features, single-molecule imaging, and traditional ISH/IHC.

From a more broad perspective, I have investigated the influence of immune perturbations introduced by experiments of nature (infection with CMV and/or HIV, **Chapter II**) and experiments of man (provision of excess energy to the immune system, the effect of ART on immune cell subsets, and the effects of checkpoint inhibitor therapy, **Chapters III-IV**) as I describe in **Chapter I** (**Figure 2**). Key points and future studies supported by these data include the following:

1. ***Chapter II***:
    a. Key point 1: Co-infection with CMV in the setting of HIV infection is beneficial, however,
    b. Key point 2: when CMV-seropositive individuals walk too far along an immune trajectory associated with CD4 differentiation (and likely T cell memory inflation), the resulting immunologic composition is associated with higher mortality.
    c. Key point 3: Chronic HIV infection is associated with higher levels of CD4 Tem and TemRA cells (which in turn correlate negatively with CD4 Tcm and CD4 Treg populations); this suggests that immunologic senescence is a stronger predictor of mortality than chronologic age.

d. Future studies should include: the validation of select immune cell subsets as predictors of all-cause and cause-specific mortality in other cohorts, inspection of the relationships between genotype and these immune cell subsets, and randomized controlled trials of anti-CMV drugs such as letermovir to reduce immune-associated mortality.

2. ***Chapter III***:

   a. Key point 1: Diet-induced obesity via a high-fat diet causes a post-translational modification of protein presented in the adipose tissue by M2-polarized macrophages.

   b. Key point 2: This post-translational modification (increased isoLG expression) is associated with clonal expansion of a CD8+ T cell population in the adipose tissue which specifically expands in response to isoLG-presenting APCs.

   c. Future studies should: assess the ability of scavengers such as 2-HOBA to prevent or reduce the formation and presentation of isoLGs on adipose tissue APCs, and use lineage-tracing mice such as the CRISPR-barcoded MARC1 mouse to assess whether or not the expansion of the CD8+ T cells in adipose tissue is a naïve response, a memory/restimulated response, and if this response

can also seed oxidative damage in other adipose tissue reserves throughout the body.

3. ***Chapter IV***:
    a. Key point 1: Increases in CD4+ and CD8+ Tem and TemRA T cells in the adipose tissue of PLWH may contribute to or be associated with protection from type 2 diabetes.
    b. Key point 2: CD4+CX_3CR1+GPR56+CD57+ T cells are a biomarker associated with glucose intolerance in the peripheral blood and adipose tissue. These cells could match the phenotypes of both HIV- and CMV-specific T cells (among other viral-specific T cells) based on literature reports.
    c. Future studies should: examine the lymphocyte landscape of human adipose tissue in PLWH and HIV-negative persons using single-cell transcriptomics, and study the effects of anti-CMV and other T cell metabolic drugs on the phenotype and infiltration of T cells into human adipose tissue.
4. ***Chapter V***:
    a. Key point 1: Fatal checkpoint inhibitor-associated neurotoxicity is associated with clonal expansion of, and pathology induced by, cytotoxic CD4 and CD8 T cells.

b. Key point 2: *In situ* techniques are critically important to understanding and dissecting these rare immune-mediated diseases; spatial transcriptomics will become increasingly important to replicate and further dissect these pathologies.

c. Future studies should: monitor the presence and association of human herpesviruses and other human infections in multiple types of checkpoint inhibitor toxicity (colitis, myocarditis, encephalitis, myasthenia gravis, and type I diabetes), monitor the association of specific HLA alleles with these toxicities to enable preventive screening, and prospectively study the differentiation trajectories of peripheral lymphocytes from checkpoint inhibitor-treated patients using scRNA-seq in order to identify the functional and cellular bases of how these toxicities develop.

5. Additional future studies should:

a. Use autologous patient-derived T cells to confirm and study the antigen specificity of single T cells isolated by altering *ex vivo* T lymphocytes in order to retain the genetic features and phenotypes of the patient's own cells;

b. Use high-throughput screening technologies (HLA-gnostic combining HLA typing and TCR sequencing, and HLA-agnostic

combining peptide pools and TCR stimulation screening) with human herpesvirus (HHV) and other holobiont peptide pools to examine the extent to which response to these viruses and organisms is associated with checkpoint inhibitor toxicity, adverse drug reactions, and other immune-mediated adverse events.

This thesis is a product of new and powerful methods of our time, which represent a set of emergent properties from convergent advances in molecular biology, liquid handling and miniaturization, optical and sequencing technology, mathematics and computation, and computing hardware innovation. Coming advances in spatial and single-cell transcriptomic approaches will include: Slide-seq and DART-seq and others[240,475–484]; the application of deep learning and transfer learning models to multi-modal spatial and scRNA-seq data[187,190,191,476,485–489]; extension of scRNA-seq-based predictive models of mortality; automation of processes such as cell types and subtypes identification[490–493]; scRNA-seq experimental design for maximal recovery of new cell types and rare cells[494,495]; implementation of graph-based reference genomes and transcriptomes[496,497]; and discovery of antibody-free flow sorting schemes from scRNA-seq data[498] will begin to comprise robust modes of discovery. Collectively, many of these approaches will become part of the “immunologist’s toolbox.”

References

1. Tomes, N. J. & Warner, J. H. Introduction to Special Issue on Rethinking the Reception of the Germ Theory of Disease: Comparative Perspectives. *Journal of the History of Medicine and Allied Sciences* **52**, 7–16 (1997).
2. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. L. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747–1757 (2006).
3. Omran, A. R. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Mem Fund Q* **49**, 509–538 (1971).
4. Omran, A. R. The epidemiologic transition: a theory of the epidemiology of population change. 1971. *Milbank Q* **83**, 731–757 (2005).
5. Carter, K. C. The germ theory, beriberi, and the deficiency theory of disease. *Med Hist* **21**, 119–136 (1977).
6. de Maat, M. P. M. *et al.* Genetic Influence on Inflammation Variables in the Elderly. *Arteriosclerosis, Thrombosis, and Vascular Biology* **24**, 2168–2173 (2004).
7. Keaney, J. F. *et al.* Heritability and correlates of intercellular adhesion molecule-1 in the Framingham Offspring Study. *Journal of the American College of Cardiology* **44**, 168–173 (2004).

8. de Craen, A. J. M. *et al.* Heritability estimates of innate immunity: an extended twin study. *Genes & Immunity* **6**, 167–170 (2005).

9. Berg, A. H. & Scherer, P. E. Adipose Tissue, Inflammation, and Cardiovascular Disease. *Circulation Research* **96**, 939–949 (2005).

10. Fernández-Real, J. M. & Ricart, W. Insulin Resistance and Chronic Cardiovascular Inflammatory Syndrome. *Endocrine Reviews* **24**, 278–301 (2003).

11. Scrivo, R., Vasile, M., Bartosiewicz, I. & Valesini, G. Inflammation as “common soil” of the multifactorial diseases. *Autoimmunity Reviews* **10**, 369–374 (2011).

12. Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860–867 (2006).

13. Hotamisligil, G. S., Shargill, N. S. & Spiegelman, B. M. Adipose expression of tumor necrosis factor-alpha: direct role in obesity-linked insulin resistance. *Science* **259**, 87–91 (1993).

14. Hotamisligil, G. S. *et al.* IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF-alpha- and obesity-induced insulin resistance. *Science* **271**, 665–668 (1996).

15. Tarkowski, E. *et al.* Early Intrathecal Production of Interleukin-6 Predicts the Size of Brain Lesion in Stroke. *Stroke* **26**, 1393–1398 (1995).

16. Kostulas, N., Pelidou, S. H., Kivisäkk, P., Kostulas, V. & Link, H. Increased IL-1β, IL-8, and IL-17 mRNA Expression in Blood Mononuclear Cells Observed in a Prospective Ischemic Stroke Study. *Stroke* **30**, 2174–2179 (1999).

17. Levin, B. R. Population Biology, Evolution, and Infectious Disease: Convergence and Synthesis. *Science* **283**, 806–809 (1999).

18. Tsatsoulis, A., Mantzaris, M. D., Bellou, S. & Andrikoula, M. Insulin resistance: An adaptive mechanism becomes maladaptive in the current environment — An evolutionary perspective. *Metabolism* **62**, 622–633 (2013).

19. Søndergaard, L. Homology between the mammalian liver and the Drosophila fat body. *Trends Genet.* **9**, 193 (1993).

20. Sarri, K. O., Tzanakis, N. E., Linardakis, M. K., Mamalakis, G. D. & Kafatos, A. G. Effects of Greek orthodox christian church fasting on serum lipids and obesity. *BMC Public Health* **3**, (2003).

21. Sarri, K. O., Linardakis, M. K., Bervanaki, F. N., Tzanakis, N. E. & Kafatos, A. G. Greek Orthodox fasting rituals: a hidden characteristic of the Mediterranean diet of Crete. *British Journal of Nutrition* **92**, 277–284 (2004).

22. Trepanowski, J. F., Canale, R. E., Marshall, K. E., Kabir, M. M. & Bloomer, R. J. Impact of caloric and dietary restriction regimens on markers of health and longevity in humans and animals: a summary of available findings. *Nutrition Journal* **10**, (2011).

23. Vardavas, C. I., Linardakis, M. K., Hatzis, C. M., Saris, W. H. M. & Kafatos, A. G. Prevalence of obesity and physical inactivity among farmers from Crete (Greece), four decades after the Seven Countries Study. *Nutr Metab Cardiovasc Dis* **19**, 156–162 (2009).

24. Vardavas, C. I., Linardakis, M. K., Hatzis, C. M., Saris, W. H. M. & Kafatos, A. G. Cardiovascular disease risk factors and dietary habits of farmers from Crete 45 years after the first description of the Mediterranean diet. *Eur J Cardiovasc Prev Rehabil* **17**, 440–446 (2010).

25. Noguchi, M. Interleukin-2 receptor γ chain mutation results in X-linked severe combined immunodeficiency in humans. *Cell* **73**, 147–157 (1993).

26. Buck, D. *et al.* Cernunnos, a Novel Nonhomologous End-Joining Factor, Is Mutated in Human Immunodeficiency with Microcephaly. *Cell* **124**, 287–299 (2006).

27. Hambleton, S. *et al. IRF8* Mutations and Human Dendritic-Cell Immunodeficiency. *New England Journal of Medicine* **365**, 127–138 (2011).

28. Ottenhoff, T. H. M., Kumararatne, D. & Casanova, J.-L. Novel human immunodeficiencies reveal the essential role of type-1 cytokines in immunity to intracellular bacteria. *Immunology Today* **19**, 491–494 (1998).

29. Salzer, U. *et al.* Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nature Genetics* **37**, 820–828 (2005).

30. Chun, H. J. *et al.* Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature* **419**, 395–399 (2002).

31. Orange, J. S. Human natural killer cell deficiencies. *Curr Opin Allergy Clin Immunol* **6**, 399–409 (2006).

32. Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* **348**, 62–68 (2015).

33. Miller, J. S. Successful adoptive transfer and in vivo expansion of human haploidentical NK cells in patients with cancer. *Blood* **105**, 3051–3057 (2005).

34. Thomas, E. D., Lochte, H. L., Lu, W. C. & Ferrebee, J. W. Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy. *New England Journal of Medicine* **257**, 491–496 (1957).

35. Yarchoan, R. *et al.* Administration of 3'-azido-3'-deoxythymidine, an inhibitor of HTLV-III/LAV replication, to patients with AIDS or AIDS-related complex. *Lancet* **1**, 575–580 (1986).

36. Mitsuya, H., Yarchoan, R. & Broder, S. Molecular targets for AIDS therapy. *Science* **249**, 1533–1544 (1990).

37. Ho, D. D. Perspectives series: host/pathogen interactions. Dynamics of HIV-1 replication in vivo. *J. Clin. Invest.* **99**, 2565–2567 (1997).

38. Carcelain, G., Debré, P. & Autran, B. Reconstitution of CD4+ T lymphocytes in HIV-infected individuals following antiretroviral therapy. *Curr. Opin. Immunol.* **13**, 483–488 (2001).

39. Perelson, A. S. *et al.* Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188–191 (1997).

40. Deeks, S. G., Lewin, S. R. & Havlir, D. V. The end of AIDS: HIV infection as a chronic disease. *The Lancet* **382**, 1525–1533 (2013).

41. Deeks, S. G., Tracy, R. & Douek, D. C. Systemic Effects of Inflammation on Health during Chronic HIV Infection. *Immunity* **39**, 633–645 (2013).

42. Melchior, J.-C. *et al.* Malnutrition and wasting, immunodepression, and chronic inflammation as independent predictors of survival in HIV-infected patients. *Nutrition* **15**, 865–869 (1999).

43. Dinh, D. M. *et al.* Intestinal Microbiota, Microbial Translocation, and Systemic Inflammation in Chronic HIV Infection. *Journal of Infectious Diseases* **211**, 19–27 (2015).

44. Deeks, S. G. HIV Infection, Inflammation, Immunosenescence, and Aging. *Annual Review of Medicine* **62**, 141–155 (2011).

45. Hunt, P. W. HIV and Inflammation: Mechanisms and Consequences. *Curr Hiv Aids Reports* **9**, 139–147 (2012).

46. Tsai, A. C., Bangsberg, D. R. & Weiser, S. D. Harnessing Poverty Alleviation to Reduce the Stigma of HIV in Sub-Saharan Africa. *PLoS Medicine* **10**, e1001557 (2013).

47. Pascoe, S. J. S. *et al.* Poverty, Food Insufficiency and HIV Infection and Sexual Behaviour among Young Rural Zimbabwean Women. *PLOS ONE* **10**, e0115290 (2015).

48. The 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

49. Carapito, R., Radosavljevic, M. & Bahram, S. Next-Generation Sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies. *Human Immunology* **77**, 1016–1023 (2016).

50. Laydon, D. J., Bangham, C. R. M. & Asquith, B. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, 20140291 (2015).

51. Soto, C. *et al.* High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* (2019). doi:10.1038/s41586-019-0934-8

52. Gray, L.-A. *et al.* Genome-Wide Analysis of Genetic Risk Factors for Rheumatic Heart Disease in Aboriginal Australians Provides Support for Pathogenic Molecular Mimicry. *The Journal of Infectious Diseases* **216**, 1460–1470 (2017).

53. Manczinger, M. *et al.* Pathogen diversity drives the evolution of generalist MHC-II alleles in human populations. *PLOS Biology* **17**, e3000131 (2019).

54. Middleton, D. & Gonzelez, F. The extensive polymorphism of KIR genes. *Immunology* **129**, 8–19 (2010).

55. Ljunggren, H. G. & Kärre, K. In search of the 'missing self': MHC molecules and NK cell recognition. *Immunol. Today* **11**, 237–244 (1990).

56. Harel-Bellan, A. *et al.* Natural killer susceptibility of human cells may be regulated by genes in the HLA region on chromosome 6. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5688–5692 (1986).

57. Gardiner, C. M. *et al.* Different NK cell surface phenotypes defined by the DX9 antibody are due to KIR3DL1 gene polymorphism. *J. Immunol.* **166**, 2992–3001 (2001).

58. Boyington, J. C., Motyka, S. A., Schuck, P., Brooks, A. G. & Sun, P. D. Crystal structure of an NK cell immunoglobulin-like receptor in complex with its class I MHC ligand. *Nature* **405**, 537–543 (2000).

59. Long, E. O. & Rajagopalan, S. HLA class I recognition by killer cell Ig-like receptors. *Semin. Immunol.* **12**, 101–108 (2000).

60. Ott, J. A. *et al.* Somatic hypermutation of T cell receptor α chain contributes to selection in nurse shark thymus. *eLife* **7**, (2018).

61. Charmley, P., Wang, K., Hood, L. & Nickerson, D. A. Identification and physical mapping of a polymorphic human T cell receptor V beta gene with a frequent null allele. *J. Exp. Med.* **177**, 135–143 (1993).

62. Day, C. E., Zhao, T. & Robinson, M. A. Silent allelic variants of a T-cell receptor V beta 12 gene are present in diverse human populations. *Hum. Immunol.* **34**, 196–202 (1992).

63. Barron, K. S. & Robinson, M. A. The human T-cell receptor variable gene segment TCRBV6S1 has two null alleles. *Hum. Immunol.* **40**, 17–19 (1994).

64. Barozzi, P. *et al.* Circulating functional T cells specific to human herpes virus 6 (HHV6) antigens in individuals with chromosomally integrated HHV6. *Clin. Microbiol. Infect.* **22**, 893–895 (2016).

65. Robbins, P. F. *et al.* Single and dual amino acid substitutions in TCR CDRs can enhance antigen-specific T cell functions. *J. Immunol.* **180**, 6116–6131 (2008).

66. Robbins, P. F. *et al.* Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J. Clin. Oncol.* **29**, 917–924 (2011).

67. Geldmacher, C. *et al.* Minor viral and host genetic polymorphisms can dramatically impact the biologic outcome of an epitope-specific CD8 T-cell response. *Blood* **114**, 1553–1562 (2009).

68. Looney, T. J. *et al.* Haplotype Analysis of the TRB Locus by TCRB Repertoire Sequencing. *bioRxiv* (2018). doi:10.1101/406157

69. Luo, G. *et al.* Autoimmunity to hypocretin and molecular mimicry to flu in type 1 narcolepsy. *Proceedings of the National Academy of Sciences* **115**, E12323–E12332 (2018).

70. Han, F. *et al.* Genome Wide Analysis of Narcolepsy in China Implicates Novel Immune Loci and Reveals Changes in Association Prior to Versus After the 2009 H1N1 Influenza Pandemic. *PLoS Genetics* **9**, e1003880 (2013).

71. Ollila, H. *et al.* Narcolepsy risk loci are enriched in immune cells and suggest autoimmune modulation of the T cell receptor repertoire. *bioRxiv* (2018). doi:10.1101/373555

72. Pathiraja, V. *et al.* Proinsulin-specific, HLA-DQ8, and HLA-DQ8-transdimer-restricted CD4+ T cells infiltrate islets in type 1 diabetes. *Diabetes* **64**, 172–182 (2015).

73. Michels, A. W. *et al.* Islet-Derived CD4 T Cells Targeting Proinsulin in Human Autoimmune Diabetes. *Diabetes* **66**, 722–734 (2017).

74. Watson, C. T. *et al.* Revisiting the T-cell receptor alpha/delta locus and possible associations with multiple sclerosis. *Genes Immun.* **12**, 59–66 (2011).

75. Planas, R., Metz, I., Martin, R. & Sospedra, M. Detailed Characterization of T Cell Receptor Repertoires in Multiple Sclerosis Brain Lesions. *Frontiers in Immunology* **9**, (2018).

76. Sakurai, K. *et al.* HLA-DRB1 Shared Epitope Alleles and Disease Activity Are Correlated with Reduced T Cell Receptor Repertoire Diversity in CD4+ T Cells in Rheumatoid Arthritis. *The Journal of Rheumatology* **45**, 905–914 (2018).

77. Manabe, K. *et al.* T-cell receptor constant beta germline gene polymorphisms and susceptibility to autoimmune hepatitis. *Gastroenterology* **106**, 1321–1325 (1994).

78. Yu, Y., Ceredig, R. & Seoighe, C. A Database of Human Immune Receptor Alleles Recovered from Population Sequencing Data. *The Journal of Immunology* **198**, 2202–2210 (2017).

79. Luo, S., Yu, J. A., Li, H. & Song, Y. S. Worldwide genetic variation of the IGHV and TRBV immune receptor gene families in humans. *bioRxiv* (2017). doi:10.1101/155440

80. DeWitt, W. S. *et al.* Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife* **7**, (2018).

81. Becker, A. J., McCULLOCH, E. A. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–454 (1963).

82. Siminovitch, L., Mcculloch, E. A. & Till, J. E. THE DISTRIBUTION OF COLONY-FORMING CELLS AMONG SPLEEN COLONIES. *J Cell Comp Physiol* **62**, 327–336 (1963).

83. Till, J. E. & McCulloch, E. A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. 1961. *Radiat. Res.* **178**, AV3-7 (2012).

84. Wu, A. M., Till, J. E., Siminovitch, L. & McCulloch, E. A. A cytological study of the capacity for differentiation of normal hemopoietic colony-forming cells. *J. Cell. Physiol.* **69**, 177–184 (1967).

85. Wu, A. M., Till, J. E., Siminovitch, L. & McCulloch, E. A. Cytological evidence for a relationship between normal hemotopoietic colony-forming cells and cells of the lymphoid system. *J. Exp. Med.* **127**, 455–464 (1968).

86. Vermijlen, D. *et al.* Human cytomegalovirus elicits fetal gammadelta T cell responses in utero. *J. Exp. Med.* **207**, 807–821 (2010).

87. Chan, W. Y., Kohsaka, S. & Rezaie, P. The origin and cell lineage of microglia: new concepts. *Brain Res Rev* **53**, 344–354 (2007).

88. Pui, J. C. *et al.* Notch1 expression in early lymphopoiesis influences B versus T lineage determination. *Immunity* **11**, 299–308 (1999).

89. Carlyle, J. R. & Zúñiga-Pflücker, J. C. Requirement for the thymus in alphabeta T lymphocyte lineage commitment. *Immunity* **9**, 187–197 (1998).

90. Gordon, J. *et al.* Functional evidence for a single endodermal origin for the thymic epithelium. *Nat. Immunol.* **5**, 546–553 (2004).

91. Radtke, F. *et al.* Deficient T cell fate specification in mice with an induced inactivation of Notch1. *Immunity* **10**, 547–558 (1999).

92. Surh, C. D. & Sprent, J. T-cell apoptosis detected in situ during positive and negative selection in the thymus. *Nature* **372**, 100–103 (1994).

93. Saint-Ruf, C. *et al.* Analysis and expression of a cloned pre-T cell receptor gene. *Science* **266**, 1208–1212 (1994).

94. Pang, S. S. *et al.* The structural basis for autonomous dimerization of the pre-T-cell antigen receptor. *Nature* **467**, 844–848 (2010).

95. Livak, F., Petrie, H. T., Crispe, I. N. & Schatz, D. G. In-frame TCR delta gene rearrangements play a critical role in the alpha beta/gamma delta T cell lineage decision. *Immunity* **2**, 617–627 (1995).

96. Lewis, J. M. *et al.* Selection of the cutaneous intraepithelial gammadelta+ T cell repertoire by a thymic stromal determinant. *Nat. Immunol.* **7**, 843–850 (2006).

97. Kreslavsky, T., Garbe, A. I., Krueger, A. & von Boehmer, H. T cell receptor-instructed alphabeta versus gammadelta lineage commitment revealed by single-cell analysis. *J. Exp. Med.* **205**, 1173–1186 (2008).

98. Ciofani, M., Knowles, G. C., Wiest, D. L., von Boehmer, H. & Zúñiga-Pflücker, J. C. Stage-specific and differential notch dependency at the alphabeta and gammadelta T lineage bifurcation. *Immunity* **25**, 105–116 (2006).

99. Borowski, C., Li, X., Aifantis, I., Gounari, F. & von Boehmer, H. Pre-TCRalpha and TCRalpha are not interchangeable partners of TCRbeta during T lymphocyte development. *J. Exp. Med.* **199**, 607–615 (2004).

100. Hardardottir, F., Baron, J. L. & Janeway, C. A. T cells with two functional antigen-specific receptors. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 354–358 (1995).

101. Huang, C.-Y., Sleckman, B. P. & Kanagawa, O. Revision of T cell receptor {alpha} chain genes is required for normal T lymphocyte development. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14356–14361 (2005).

102. Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L. & Kappler, J. W. Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).

103. Scott-Browne, J. P., White, J., Kappler, J. W., Gapin, L. & Marrack, P. Germline-encoded amino acids in the alphabeta T-cell receptor control thymic selection. *Nature* **458**, 1043–1046 (2009).

104. Zerrahn, J., Held, W. & Raulet, D. H. The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* **88**, 627–636 (1997).

105. Kishimoto, H. & Sprent, J. Negative selection in the thymus includes semimature T cells. *J. Exp. Med.* **185**, 263–271 (1997).

106. Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298**, 1395–1401 (2002).

107. McCaughtry, T. M., Baldwin, T. A., Wilken, M. S. & Hogquist, K. A. Clonal deletion of thymocytes can occur in the cortex with no involvement of the medulla. *J. Exp. Med.* **205**, 2575–2584 (2008).

108. Steinman, R. M. & Nussenzweig, M. C. Avoiding horror autotoxicus: the importance of dendritic cells in peripheral T cell tolerance. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 351–358 (2002).

109. Patin, E. *et al.* Natural variation in the parameters of innate immune cells is preferentially driven by genetic factors. *Nat. Immunol.* **19**, 302–314 (2018).

110. Brodin, P. *et al.* Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47 (2015).

111. Abana, C. O. *et al.* Cytomegalovirus (CMV) Epitope-Specific CD4+ T Cells Are Inflated in HIV+ CMV+ Subjects. *J. Immunol.* **199**, 3187–3201 (2017).

112. Schlott, F. *et al.* Characterization and clinical enrichment of HLA-C*07:02-restricted Cytomegalovirus-specific CD8+ T cells. *PLOS ONE* **13**, e0193554 (2018).

113. Hosie, L. *et al.* Cytomegalovirus-Specific T Cells Restricted by HLA-Cw*0702 Increase Markedly with Age and Dominate the CD8+ T-Cell Repertoire in Older People. *Front Immunol* **8**, 1776 (2017).

114. Alejenef, A. *et al.* Cytomegalovirus drives Vδ2neg γδ T cell inflation in many healthy virus carriers with increasing age. *Clin Exp Immunol* **176**, 418–428 (2014).

115. Crompton, L., Khan, N., Khanna, R., Nayak, L. & Moss, P. A. CD4+ T cells specific for glycoprotein B from cytomegalovirus exhibit extreme conservation of T-cell receptor usage between different individuals. **111**, 2053–2061 (2008).

116. Khan, N., Cobbold, M., Cummerson, J. & Moss, P. A. Persistent viral infection in humans can drive high frequency low-affinity T-cell expansions. *Immunology* **131**, 537–548 (2010).

117. Klenerman, P. & Oxenius, A. T cell responses to cytomegalovirus. **16**, 367–377 (2016).

118. Adler, S. P. & Reddehase, M. J. Pediatric roots of cytomegalovirus recurrence and memory inflation in the elderly. *Med. Microbiol. Immunol.* (2019). doi:10.1007/s00430-019-00609-6

119. Baumann, N. S. *et al.* Early primed KLRG1- CMV-specific T cells determine the size of the inflationary T cell pool. *PLoS Pathog.* **15**, e1007785 (2019).

120. Renzaho, A. *et al.* Transcripts expressed in cytomegalovirus latency coding for an antigenic IE/E phase peptide that drives 'memory inflation'. *Med. Microbiol. Immunol.* (2019). doi:10.1007/s00430-019-00615-8

121. Adams, N. M. *et al.* Cytomegalovirus Infection Drives Avidity Selection of Natural Killer Cells. *Immunity* (2019). doi:10.1016/j.immuni.2019.04.009

122. Nikzad, R. *et al.* Human natural killer cells mediate adaptive immunity to viral antigens. *Science Immunology* **4**, eaat8116 (2019).

123. Furman, D. *et al.* Cytomegalovirus infection enhances the immune response to influenza. *Sci Transl Med* **7**, 281ra43-281ra43 (2015).

124. Freeman, M. L., Lederman, M. M. & Gianella, S. Partners in Crime: The Role of CMV in Immune Dysregulation and Clinical Outcome During HIV Infection. **13**, 10–19 (2016).

125. Spyridopoulos, I. *et al.* CMV seropositivity and T-cell senescence predict increased cardiovascular mortality in octogenarians: results from the Newcastle 85+ study. **15**, 389–392 (2016).

126. Chanouzas, D. *et al.* CD4+CD28– T-cell expansions in ANCA-associated vasculitis and association with arterial stiffness: baseline data from a randomised controlled trial. **385**, S30 (2015).

127. Li, H. *et al.* Chronic CMV infection in older women: Longitudinal comparisons of CMV DNA in peripheral monocytes, anti-CMV IgG titers, serum IL-6 levels, and

CMV pp65 (NLV)-specific CD8+ T-cell frequencies with twelve year follow-up. **54**, 84–89 (2014).

128. Margolick, J. B. *et al.* Relationship Between T-Cell Responses to CMV, Markers of Inflammation, and Frailty in HIV-uninfected and HIV-infected Men in the Multicenter AIDS Cohort Study. *J. Infect. Dis.* **218**, 249–258 (2018).

129. McElhaney, J. E. *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* **30**, 2060–2067 (2012).

130. Jones, R. B. *et al.* Tim-3 expression defines a novel population of dysfunctional T cells with highly elevated frequencies in progressive HIV-1 infection. *J. Exp. Med.* **205**, 2763–2779 (2008).

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

132. Blackburn, S. D. *et al.* Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat. Immunol.* **10**, 29–37 (2009).

133. Yamamoto, T. *et al.* Surface expression patterns of negative regulatory molecules identify determinants of virus-specific CD8+ T-cell exhaustion in HIV infection. *Blood* **117**, 4805–4815 (2011).

134. Kaufmann, D. E. *et al.* Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. *Nat. Immunol.* **8**, 1246–1254 (2007).

135. Banga, R. *et al.* PD-1(+) and follicular helper T cells are responsible for persistent HIV-1 transcription in treated aviremic individuals. *Nat. Med.* **22**, 754–761 (2016).

136. Chomont, N. *et al.* HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009).

137. Hunt, P. W. *et al.* T Cell Activation Is Associated with Lower CD4 $^{+}$ T Cell Gains in Human Immunodeficiency Virus–Infected Patients with Sustained Viral Suppression during Antiretroviral Therapy. *The Journal of Infectious Diseases* **187**, 1534–1543 (2003).

138. Hunt, P. W. *et al.* Relationship between T Cell Activation and CD4 $^{+}$ T Cell Count in HIV-Seropositive Individuals with Undetectable Plasma HIV RNA Levels in the Absence of Therapy. *The Journal of Infectious Diseases* **197**, 126–133 (2008).

139. Hunt, P. W. *et al.* Valganciclovir Reduces T Cell Activation in HIV-Infected Individuals With Incomplete CD4+ T Cell Recovery on Antiretroviral Therapy. **203**, 1474–1483 (2011).

140. Lee, S. A. *et al.* Impact of HIV on CD8+ T Cell CD57 Expression Is Distinct from That of CMV and Aging. *PLoS ONE* **9**, e89444 (2014).

141. Kovacs, J. A. *et al.* Identification of dynamically distinct subpopulations of T lymphocytes that are differentially affected by HIV. *J. Exp. Med.* **194**, 1731–1741 (2001).

142. Hong, H. S. *et al.* HIV Infection Is Associated with a Preferential Decline in Less-Differentiated CD56dim CD16+ NK Cells. *Journal of Virology* **84**, 1183–1188 (2010).

143. Moir, S. *et al.* HIV-1 induces phenotypic and functional perturbations of B cells in chronically infected individuals. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10362–10367 (2001).

144. Julius, S. *et al.* Feasibility of treating prehypertension with an angiotensin-receptor blocker. *N. Engl. J. Med.* **354**, 1685–1697 (2006).

145. Chobanian, A. V. *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* **289**, 2560–2572 (2003).

146. Narayan, K. M. V., Ali, M. K. & Koplan, J. P. Global noncommunicable diseases--where worlds meet. *N. Engl. J. Med.* **363**, 1196–1198 (2010).

147. Guzik, T. J. *et al.* Role of the T cell in the genesis of angiotensin II induced hypertension and vascular dysfunction. *J. Exp. Med.* **204**, 2449–2460 (2007).

148. Marvar, P. J. *et al.* Central and peripheral mechanisms of T-lymphocyte activation and vascular inflammation produced by angiotensin II-induced hypertension. *Circ. Res.* **107**, 263–270 (2010).

149. Ehret, G. B., O'Connor, A. A., Weder, A., Cooper, R. S. & Chakravarti, A. Follow-up of a major linkage peak on chromosome 1 reveals suggestive QTLs associated with essential hypertension: GenNet study. *Eur. J. Hum. Genet.* **17**, 1650–1657 (2009).

150. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).

151. Tragante, V. *et al.* Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci. *Am. J. Hum. Genet.* **94**, 349–360 (2014).

152. Ammirati, E. *et al.* Effector Memory T cells Are Associated With Atherosclerosis in Humans and Animal Models. *J Am Heart Assoc* **1**, 27–41 (2012).

153. Olson, N. C. *et al.* Decreased Naive and Increased Memory CD4+ T Cells Are Associated with Subclinical Atherosclerosis: The Multi-Ethnic Study of Atherosclerosis. **8**, e71498 (2013).

154. Itani, H. A. *et al.* Activation of Human T Cells in Hypertension: Studies of Humanized Mice and Hypertensive Humans. *Hypertension* **68**, 123–132 (2016).

155. Youn, J.-C. *et al.* Immunosenescent CD8 $^{+}$ T Cells and C-X-C Chemokine Receptor Type 3 Chemokines Are Increased in Human Hypertension. *Hypertension* **62**, 126–133 (2013).

156. Silverman, J. D., Roche, K., Mukherjee, S. & David, L. A. Naught all zeros in sequence count data are the same. *bioRxiv* (2018). doi:10.1101/477794

157. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols* **13**, 599–604 (2018).

158. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).

159. Li, N. & Unger, M. Technology for cell transdifferentiation. (2015).

160. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

161. Klein, A. M. *et al.* Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187–1201 (2015).

162. Zhang, X. *et al.* Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *bioRxiv* (2018). doi:10.1101/313130

163. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).

164. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature Methods* **10**, 1096–1098 (2013).

165. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171–181 (2014).

166. Islam, S. *et al.* Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res.* **21**, 1160–1167 (2011).

167. Hochgerner, H. *et al.* STRT-seq-2i: dual-index 5′ single cell and nucleus RNA-seq on an addressable microwell array. *Sci Rep* **7**, 16327 (2017).

168. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* **2**, 666–673 (2012).

169. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology* **17**, (2016).

170. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).

171. Soumillon, M., Cacchiarelli, D., Semrau, S., van Oudenaarden, A. & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* (2014). doi:10.1101/003236

172. Bagnoli, J. W. *et al.* Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nature Communications* **9**, (2018).

173. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).

174. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).

175. Ding, J. *et al.* Systematic comparative analysis of single cell RNA-sequencing methods. *bioRxiv* (2019). doi:10.1101/632216

176. Wang, Y. J. *et al.* Comparative analysis of commercially available single-cell RNA sequencing platforms for their performance in complex human tissues. *bioRxiv* (2019). doi:10.1101/541433

177. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).

178. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).

179. Peterson, V. M. *et al.* Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).

180. Mimitou, E. P. *et al.* Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).

181. Singh, M. *et al.* High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *bioRxiv* (2018). doi:10.1101/424945

182. Gupta, I. *et al.* Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nature Biotechnology* **36**, 1197–1202 (2018).

183. Stubbington, M. J. T. *et al.* T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).

184. Lindeman, I. *et al.* BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods* **15**, 563–565 (2018).

185. Lindeman, I. & Stubbington, M. J. T. Antigen Receptor Sequence Reconstruction and Clonality Inference from scRNA-Seq Data. *Methods Mol. Biol.* **1935**, 223–249 (2019).

186. Ntranos, V., Yi, L., Melsted, P. & Pachter, L. A discriminative learning approach to differential expression analysis for single-cell RNA-seq. *Nature Methods* **16**, 163–166 (2019).

187. Stein-O'Brien, G. L. *et al.* Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. *bioRxiv* (2018). doi:10.1101/395004

188. Wang, T., Johnson, T. S., Shao, W., Zhang, J. & Huang, K. BURMUDA: A novel deep transfer learning method for single-cell RNA sequencing batch correction reveals hidden high-resolution cellular subtypes. *bioRxiv* (2019). doi:10.1101/641191

189. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

190. Stuart, T. *et al.* Comprehensive integration of single cell data. *bioRxiv* (2018). doi:10.1101/460147

191. Deng, Y., Bao, F., Dai, Q., Wu, L. F. & Altschuler, S. J. Scalable analysis of cell-type composition from single-cell transcriptomics using deep recurrent learning. *Nature Methods* **16**, 311–314 (2019).

192. Liu, M. *et al.* Unsupervised learning techniques reveal heterogeneity in memory CD8+ T cell differentiation following acute, chronic and latent viral infections. **509**, 266–279 (2017).

193. Tian, T., Wan, J., Song, Q. & Wei, Z. Clustering single-cell RNA-seq data with a model-based deep learning approach. *Nature Machine Intelligence* **1**, 191–198 (2019).

194. Zhang, M. J., Ntranos, V. & Tse, D. One read per cell per gene is optimal for single-cell RNA-Seq. *bioRxiv* (2018). doi:10.1101/389296

195. Riemondy, K. A. *et al.* Recovery and analysis of transcriptome subsets from pooled single-cell RNA-seq libraries. *Nucleic Acids Research* **47**, e20–e20 (2019).

196. LifeLines Cohort Study *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nature Genetics* **50**, 493–497 (2018).

197. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *bioRxiv* (2019). doi:10.1101/598748

198. Larsson, A. J. M. *et al.* Genomic encoding of transcriptional burst kinetics. *Nature* **565**, 251–254 (2019).

199. Waddington, C. H. *The strategy of the genes*. (Routledge, 2014).

200. Peel, L., Larremore, D. B. & Clauset, A. The ground truth about metadata and community detection in networks. *Sci Adv* **3**, e1602548 (2017).

201. Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S. & Kluger, Y. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **16**, 243–245 (2019).

202. Roweis, S. T. & Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**, 2323–2326 (2000).

203. Balasubramanian, M. & Schwartz, E. L. The isomap algorithm and topological stability. *Science* **295**, 7 (2002).

204. Moon, K. R. *et al.* Visualizing Structure and Transitions for Biological Data Exploration. *bioRxiv* (2019). doi:10.1101/120378

205. Angerer, P. *et al. destiny* : diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).

206. Skinnider, M. A., Squair, J. W. & Foster, L. J. Evaluating measures of association for single-cell transcriptomics. *Nature Methods* **16**, 381–386 (2019).

207. McGee, W. A., Pimentel, H., Pachter, L. & Wu, J. Y. Compositional Data Analysis is necessary for simulating and analyzing RNA-Seq data. *bioRxiv* (2019). doi:10.1101/564955

208. Duò, A., Robinson, M. D. & Soneson, C. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Research* **7**, 1141 (2018).

209. Severson, D. T., Owen, R. P., White, M. J., Lu, X. & Schuster-Böckler, B. BEARscc determines robustness of single-cell clusters using simulated technical replicates. *Nature Communications* **9**, (2018).

210. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).

211. Wilson, N. K. *et al.* Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16**, 712–724 (2015).

212. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360.e4 (2016).

213. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).

214. Guo, M., Wang, H., Potter, S. S., Whitsett, J. A. & Xu, Y. SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. *PLoS Comput. Biol.* **11**, e1004575 (2015).

215. Žurauskienė, J. & Yau, C. pcaReduce: hierarchical clustering of single cell transcriptional profiles. *BMC Bioinformatics* **17**, 140 (2016).

216. Lin, P., Troup, M. & Ho, J. W. K. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* **18**, 59 (2017).

217. Yang, Y. *et al.* SAFE-clustering: Single-cell Aggregated (From Ensemble) Clustering for Single-cell RNA-seq Data. *Bioinformatics* (2018). doi:10.1093/bioinformatics/bty793

218. Sengupta, D., Aich, I. & Bandyopadhyay, S. Feature selection using feature dissimilarity measure and density-based clustering: application to biological data. *J. Biosci.* **40**, 721–730 (2015).

219. Dinger, S. C., Van Wyk, M. A., Carmona, S. & Rubin, D. M. Clustering gene expression data using a diffraction-inspired framework. *Biomed Eng Online* **11**, 85 (2012).

220. Jiang, L., Chen, H., Pinello, L. & Yuan, G.-C. GiniClust: detecting rare cell types from single-cell gene expression data with Gini index. *Genome Biol.* **17**, 144 (2016).

221. Lancichinetti, A. & Fortunato, S. Community detection algorithms: A comparative analysis. *Physical Review E* **80**, (2009).

222. Subelj, L. & Bajec, M. Unfolding communities in large complex networks: combining defensive and offensive label propagation for core extraction. *Phys Rev E Stat Nonlin Soft Matter Phys* **83**, 036103 (2011).

223. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

224. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184–197 (2015).

225. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

226. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nature Methods* **13**, 241–244 (2016).

227. Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14**, 483–486 (2017).

228. Björklund, Å. K. *et al.* The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. *Nat. Immunol.* **17**, 451–460 (2016).

229. Gury-BenAri, M. *et al.* The Spectrum and Regulatory Landscape of Intestinal Innate Lymphoid Cells Are Shaped by the Microbiome. *Cell* **166**, 1231-1246.e13 (2016).

230. Croote, D., Darmanis, S., Nadeau, K. C. & Quake, S. R. High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes. *Science* **362**, 1306–1309 (2018).

231. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

232. Schlitzer, A. *et al.* Identification of cDC1- and cDC2-committed DC progenitors reveals early lineage priming at the common DC progenitor stage in the bone marrow. *Nature Immunology* **16**, 718–728 (2015).

233. Shin, J. *et al.* Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. *Cell Stem Cell* **17**, 360–372 (2015).

234. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nature Methods* **13**, 845–848 (2016).

235. Setty, M. *et al.* Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature Biotechnology* **34**, 637–645 (2016).

236. Herman, J. S., Sagar & Grün, D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nature Methods* **15**, 379–386 (2018).

237. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nature Biotechnology* **37**, 547–554 (2019).

238. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology* **20**, (2019).

239. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* (2017). doi:10.1038/nmeth.4402

240. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

241. Ludwig, L. S. *et al.* Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics. *Cell* **176**, 1325-1339.e22 (2019).

242. Xu, J. *et al.* Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. *eLife* **8**, (2019).

243. Korsunsky, I. *et al.* Fast, sensitive, and accurate integration of single cell data with Harmony. *bioRxiv* (2018). doi:10.1101/461954

244. Dorshkind, K., Montecino-Rodriguez, E. & Signer, R. A. J. The ageing immune system: is it ever too old to become young again? *Nat. Rev. Immunol.* **9**, 57–62 (2009).

245. Carr, E. J. *et al.* The cellular composition of the human immune system is shaped by age and cohabitation. *Nat. Immunol.* **17**, 461–468 (2016).

246. Fulop, T. *et al.* Immunosenescence and Inflamm-Aging As Two Sides of the Same Coin: Friends or Foes? *Front Immunol* **8**, 1960 (2017).

247. Shanley, D. P., Aw, D., Manley, N. R. & Palmer, D. B. An evolutionary perspective on the mechanisms of immunosenescence. *Trends Immunol.* **30**, 374–381 (2009).

248. Orrù, V. *et al.* Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013).

249. Roederer, M. *et al.* The genetic architecture of the human immune system: a bioresource for autoimmunity and disease pathogenesis. *Cell* **161**, 387–403 (2015).

250. Scepanovic, P. *et al.* Human genetic variants and age are the strongest predictors of humoral immune responses to common pathogens and vaccines. *Genome Med* **10**, 59 (2018).

251. Hammer, C. *et al.* Amino Acid Variation in HLA Class II Proteins Is a Major Determinant of Humoral Response to Common Viruses. *Am. J. Hum. Genet.* **97**, 738–743 (2015).

252. Rubicz, R. *et al.* Genome-wide genetic investigation of serological measures of common infections. *Eur. J. Hum. Genet.* **23**, 1544–1548 (2015).

253. Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun* **8**, 599 (2017).

254. Berger, M. *et al.* Lymphocyte subsets recovery following allogeneic bone marrow transplantation (BMT): CD4+ cell count and transplant-related mortality. *Bone Marrow Transplantation* **41**, 55–62 (2008).

255. Hutchinson, P., Chadban, S. J., Atkins, R. C. & Holdsworth, S. R. Laboratory assessment of immune function in renal transplant patients. *Nephrology Dialysis Transplantation* **18**, 983–989 (2003).

256. Pai, S.-Y. *et al.* Transplantation Outcomes for Severe Combined Immunodeficiency, 2000–2009. *New England Journal of Medicine* **371**, 434–446 (2014).

257. Ducloux, D. *et al.* CD4 lymphocytopenia in long-term renal transplant recipients. *Transplant. Proc.* **30**, 2859–2860 (1998).

258. Boeckh, M. *et al.* Late cytomegalovirus disease and mortality in recipients of allogeneic hematopoietic stem cell transplants: importance of viral load and T-cell immunity. *Blood* **101**, 407–414 (2003).

259. Luque, Y. *et al.* Long-term CD4 lymphopenia is associated with accelerated decline of kidney allograft function. *Nephrol. Dial. Transplant.* **31**, 487–495 (2016).

260. Ducloux, D., Challier, B., Saas, P., Tiberghien, P. & Chalopin, J. CD4 Cell Lymphopenia and Atherosclerosis in Renal Transplant Recipients. *Journal of the American Society of Nephrology* **14**, 767–772 (2003).

261. Duncan, R. A. *et al.* Idiopathic CD4+ T-lymphocytopenia--four patients with opportunistic infections and no evidence of HIV infection. *N. Engl. J. Med.* **328**, 393–398 (1993).

262. Sheikh, V. *et al.* Administration of interleukin-7 increases CD4 T cells in idiopathic CD4 lymphocytopenia. *Blood* **127**, 977–988 (2016).

263. Katlama, C. *et al.* Treatment intensification followed by interleukin-7 reactivates HIV without reducing total HIV DNA: a randomized trial. *AIDS* **30**, 221–230 (2016).

264. Baglama, J., Reichel, L. & Lewis, B. W. *irlba: Fast partial SVD by implicitly-restarted Lanczos bidiagonalization.* (2019).

265. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology* **37**, 38–44 (2018).

266. Cannoodt, R. *et al.* SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. *bioRxiv* (2016). doi:10.1101/079509

267. Alpert, A. *et al.* A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nature Medicine* **25**, 487–495 (2019).

268. Chiu, D. S. & Talhouk, A. diceR: an R package for class discovery using an ensemble driven approach. *BMC Bioinformatics* **19**, 11 (2018).

269. Huang, H., Liu, Y., Yuan, M. & Marron, J. S. Statistical Significance of Clustering using Soft Thresholding. *J Comput Graph Stat* **24**, 975–993 (2015).

270. Colbert, R. W., Kundu, S., Duncan, M. S., So-Armah, K. & Freiberg, M. S. MORTALITY RISK AFTER HEART FAILURE DIAGNOSIS IN HIV PATIENTS. *Journal of the American College of Cardiology* **73**, 792 (2019).

271. Freiberg, M. S. *et al.* Association Between HIV Infection and the Risk of Heart Failure With Reduced Ejection Fraction and Preserved Ejection Fraction in the Antiretroviral Therapy Era: Results From the Veterans Aging Cohort Study. *Jama Cardiol* (2017). doi:10.1001/jamacardio.2017.0264

272. Meira-Machado, L., Cadarso-Suárez, C., Gude, F. & Araújo, A. smoothHR: An R Package for Pointwise Nonparametric Estimation of Hazard Ratio Curves of Continuous Predictors. *Computational and Mathematical Methods in Medicine* **2013**, 1–11 (2013).

273. Lohse, N. & Obel, N. Update of Survival for Persons With HIV Infection in Denmark. *Annals of Internal Medicine* **165**, 749 (2016).

274. Nakagawa, F., May, M. & Phillips, A. Life expectancy living with HIV: recent estimates and future implications. *Current Opinion in Infectious Diseases* **26**, 17–25 (2013).

275. Sabin, C. A. Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy? *BMC Medicine* **11**, (2013).

276. Nsanzimana, S. *et al.* Life expectancy among HIV-positive patients in Rwanda: a retrospective observational cohort study. *The Lancet Global Health* **3**, e169–e177 (2015).

277. Oswald-Richter, K. *et al.* Identification of a CCR5-expressing T cell subset that is resistant to R5-tropic HIV infection. *PLoS Pathog.* **3**, e58 (2007).

278. Kohlgruber, A. C., LaMarche, N. M. & Lynch, L. Adipose tissue at the nexus of systemic and cellular immunometabolism. *Semin. Immunol.* **28**, 431–440 (2016).

279. Pham, T. D. *et al.* High-fat diet induces systemic B-cell repertoire changes associated with insulin resistance. *Mucosal Immunol* (2017). doi:10.1038/mi.2017.25

280. Winer, D. A. *et al.* B cells promote insulin resistance through modulation of T cells and production of pathogenic IgG antibodies. *Nat. Med.* **17**, 610–617 (2011).

281. Nishimura, S. *et al.* CD8+ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nature Medicine* **15**, 914–920 (2009).

282. Duffaut, C., Galitzky, J., Lafontan, M. & Bouloumié, A. Unexpected trafficking of immune cells within the adipose tissue during the onset of obesity. *Biochem. Biophys. Res. Commun.* **384**, 482–485 (2009).

283. Feuerer, M. *et al.* Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nat. Med.* **15**, 930–939 (2009).

284. Winer, S. *et al.* Normalization of obesity-associated insulin resistance through immunotherapy. *Nat. Med.* **15**, 921–929 (2009).

285. Stolarczyk, E. *et al.* Improved insulin sensitivity despite increased visceral adiposity in mice deficient for the immune cell transcription factor T-bet. *Cell Metab.* **17**, 520–533 (2013).

286. Yang, H. *et al.* Obesity increases the production of proinflammatory mediators from adipose tissue T cells and compromises TCR repertoire diversity: implications for systemic inflammation and insulin resistance. *J. Immunol.* **185**, 1836–1845 (2010).

287. Correia-Neves, M., Waltzinger, C., Mathis, D. & Benoist, C. The shaping of the T cell repertoire. *Immunity* **14**, 21–32 (2001).

288. Kirabo, A. *et al.* DC isoketal-modified proteins activate T cells and promote hypertension. *J. Clin. Invest.* **124**, 4642–4656 (2014).

289. Orr, J. S., Kennedy, A. J. & Hasty, A. H. Isolation of Adipose Tissue Immune Cells. *Journal of Visualized Experiments* (2013). doi:10.3791/50707

290. Robins, H. *et al.* Ultra-sensitive detection of rare T cell clones. *J. Immunol. Methods* **375**, 14–19 (2012).

291. Trouplin, V. *et al.* Bone marrow-derived macrophage production. *J Vis Exp* e50966 (2013). doi:10.3791/50966

292. Hill, A. A. *et al.* Activation of NF-κB drives the enhanced survival of adipose tissue macrophages in an obesogenic environment. *Mol Metab* **4**, 665–677 (2015).

293. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).

294. Nazarov, V. I. *et al.* tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics* **16**, 175 (2015).

295. Koethe, J. R. *et al.* Adipose Tissue is Enriched for Activated and Late-Differentiated CD8+ T Cells and Shows Distinct CD8+ Receptor Usage, Compared With Blood in HIV-Infected Persons. *J. Acquir. Immune Defic. Syndr.* **77**, e14–e21 (2018).

296. Suzuki, R. & Shimodaira, H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).

297. Feoli, E. & Orlóci, L. Analysis of concentration and detection of underlying factors in structured tables. *Vegetatio* **40**, 49–54 (1979).

298. Williams, E. J. Use of Scores for the Analysis of Association in Contingency Tables. *Biometrika* **39**, 274 (1952).

299. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including

sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281-287 (2012).

300. Candéias, S., Waltzinger, C., Benoist, C. & Mathis, D. The V beta 17+ T cell repertoire: skewed J beta usage after thymic selection; dissimilar CDR3s in CD4+ versus CD8+ cells. *J. Exp. Med.* **174**, 989–1000 (1991).

301. Emerson, R. *et al.* Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. *J. Immunol. Methods* **391**, 14–21 (2013).

302. Nawaz, A. *et al.* CD206+ M2-like macrophages regulate systemic glucose metabolism by inhibiting proliferation of adipocyte progenitors. *Nat Commun* **8**, 286 (2017).

303. Li, P. *et al.* Functional heterogeneity of CD11c-positive adipose tissue macrophages in diet-induced obese mice. *J. Biol. Chem.* **285**, 15333–15345 (2010).

304. Madi, A. *et al.* T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Research* **24**, 1603–1612 (2014).

305. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98 (2017).

306. Lennon, G. P. *et al.* T cell islet accumulation in type 1 diabetes is a tightly regulated, cell-autonomous event. *Immunity* **31**, 643–653 (2009).

307. Butcher, M. J. *et al.* Association of proinflammatory cytokines and islet resident leucocytes with islet dysfunction in type 2 diabetes. *Diabetologia* **57**, 491–501 (2014).

308. Brooks-Worrell, B., Narla, R. & Palmer, J. P. Islet autoimmunity in phenotypic type 2 diabetes patients. *Diabetes Obes Metab* **15 Suppl 3**, 137–140 (2013).

309. Sarikonda, G. *et al.* CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes. *J. Autoimmun.* **50**, 77–82 (2014).

310. Wang, J. *et al.* In situ recognition of autoantigen as an essential gatekeeper in autoimmune CD8+ T cell inflammation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9317–9322 (2010).

311. Khan, I. M. *et al.* Attenuated adipose tissue and skeletal muscle inflammation in obese mice with combined CD4+ and CD8+ T cell deficiency. *Atherosclerosis* **233**, 419–428 (2014).

312. Wolf, M. J. *et al.* Metabolic activation of intrahepatic CD8+ T cells and NKT cells causes nonalcoholic steatohepatitis and liver cancer via cross-talk with hepatocytes. *Cancer Cell* **26**, 549–564 (2014).

313. Ahnstedt, H. *et al.* Sex Differences in Adipose Tissue CD8+ T Cells and Regulatory T Cells in Middle-Aged Mice. *Front Immunol* **9**, 659 (2018).

314. Miles, J. J., Douek, D. C. & Price, D. A. Bias in the αβ T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol. Cell Biol.* **89**, 375–87 (2011).

315. Koethe, J. R. *et al.* Adipose Tissue is Enriched for Activated and Late-differentiated CD8+ T cells, and Shows Distinct CD8+ Receptor Usage, Compared to Blood in HIV-infected Persons: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 1 (2017). doi:10.1097/QAI.0000000000001573

316. Kent, S. C. *et al.* Expanded T cells from pancreatic lymph nodes of type 1 diabetic subjects recognize an insulin epitope. *Nature* **435**, 224–228 (2005).

317. Yang, H. *et al.* Obesity accelerates thymic aging. *Blood* **114**, 3803–3812 (2009).

318. Damouche, A. *et al.* Adipose Tissue Is a Neglected Viral Reservoir and an Inflammatory Site during Chronic HIV and SIV Infection. *PLoS Pathog.* **11**, e1005153 (2015).

319. Bilate, A. M. *et al.* Tissue-specific emergence of regulatory and intraepithelial T cells from a clonal T cell precursor. *Science Immunology* **1**, eaaf7471–eaaf7471 (2016).

320. Cho, K. W. *et al.* An MHC II-dependent activation loop between adipose tissue macrophages and CD4+ T cells controls obesity-induced inflammation. *Cell Rep* **9**, 605–617 (2014).

321. Cho, K. W. *et al.* Adipose Tissue Dendritic Cells Are Independent Contributors to Obesity-Induced Inflammation and Insulin Resistance. *J. Immunol.* **197**, 3650–3661 (2016).

322. Mattioli, B., Straface, E., Quaranta, M. G., Giordani, L. & Viora, M. Leptin promotes differentiation and survival of human dendritic cells and licenses them for Th1 priming. *J. Immunol.* **174**, 6820–6828 (2005).

323. Jung, M. Y., Kim, H.-S., Hong, H.-J., Youn, B.-S. & Kim, T. S. Adiponectin induces dendritic cell activation via PLCγ/JNK/NF-κB pathways, leading to Th1 and Th17 polarization. *J. Immunol.* **188**, 2592–2601 (2012).

324. Moraes-Vieira, P. M. *et al.* RBP4 activates antigen-presenting cells, leading to adipose tissue inflammation and systemic insulin resistance. *Cell Metab.* **19**, 512–526 (2014).

325. Wu, H., Zhang, G. Y. & Knight, J. F. T cell receptor BV gene usage in interstitial cellular infiltrates in active Heymann nephritis. *Nephrol. Dial. Transplant.* **16**, 1374–1381 (2001).

326. Pearson, J. A. *et al.* Proinsulin Expression Shapes the TCR Repertoire but Fails to Control the Development of Low-Avidity Insulin-Reactive CD8+ T Cells. *Diabetes* **65**, 1679–1689 (2016).

327. Salomon, R. G. & Bi, W. Isolevuglandin Adducts in Disease. *Antioxidants & Redox Signaling* **22**, 1703–1718 (2015).

328. Den Hartigh, L. J. *et al.* Adipocyte-Specific Deficiency of NADPH Oxidase 4 Delays the Onset of Insulin Resistance and Attenuates Adipose Tissue Inflammation in ObesityHighlights. *Arteriosclerosis, Thrombosis, and Vascular Biology* **37**, 466–475 (2017).

329. Coats, B. R. *et al.* Metabolically Activated Adipose Tissue Macrophages Perform Detrimental and Beneficial Functions during Diet-Induced Obesity. *Cell Reports* **20**, 3149–3161 (2017).

330. Mont, S. *et al.* Accumulation of isolevuglandin-modified protein in normal and fibrotic lung. *Scientific Reports* **6**, (2016).

331. Schober, K. & Busch, D. H. A synergistic combination: using RNAseq to decipher both T-cell receptor sequence and transcriptional profile of individual T cells. *Immunol. Cell Biol.* **94**, 529–530 (2016).

332. Anderson, E. K., Gutierrez, D. A., Kennedy, A. & Hasty, A. H. Weight cycling increases T-cell accumulation in adipose tissue and impairs systemic glucose tolerance. *Diabetes* **62**, 3180–3188 (2013).

333. Brown, T. T. Antiretroviral Therapy and the Prevalence and Incidence of Diabetes Mellitus in the Multicenter AIDS Cohort Study. *Archives of Internal Medicine* **165**, 1179 (2005).

334. De Wit, S. *et al.* Incidence and Risk Factors for New-Onset Diabetes in HIV-Infected Patients: The Data Collection on Adverse Events of Anti-HIV Drugs (D:A:D) Study. *Diabetes Care* **31**, 1224–1229 (2008).

335. Capeau, J. *et al.* Ten-year diabetes incidence in 1046 HIV-infected patients started on a combination antiretroviral treatment: *AIDS* **26**, 303–314 (2012).

336. Barre-Sinoussi, F. *et al.* Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**, 868–871 (1983).

337. Hellerstein, M. K. *et al.* Increased de novo hepatic lipogenesis in human immunodeficiency virus infection. *The Journal of Clinical Endocrinology & Metabolism* **76**, 559–565 (1993).

338. Sekhar, R. V. *et al.* Metabolic basis of HIV-lipodystrophy syndrome. *American Journal of Physiology-Endocrinology and Metabolism* **283**, E332–E337 (2002).

339. Reeds, D. N. *et al.* Alterations in lipid kinetics in men with HIV-dyslipidemia. *American Journal of Physiology-Endocrinology and Metabolism* **285**, E490–E497 (2003).

340. McComsey, G. A. *et al.* Peripheral and Central Fat Changes in Subjects Randomized to Abacavir-Lamivudine or Tenofovir-Emtricitabine With Atazanavir-Ritonavir or Efavirenz: ACTG Study A5224s. *Clinical Infectious Diseases* **53**, 185–196 (2011).

341. Nolan, D. *et al.* Mitochondrial DNA depletion and morphologic changes in adipocytes associated with nucleoside reverse transcriptase inhibitor therapy. *AIDS* **17**, 1329–1338 (2003).

342. Hammond, E., Nolan, D., James, I., Metcalf, C. & Mallal, S. Reduction of mitochondrial DNA content and respiratory chain activity occurs in adipocytes within 6–12 months of commencing nucleoside reverse transcriptase inhibitor therapy: *AIDS* **18**, 815–817 (2004).

343. Nolan, D., Hammond, E., James, I., McKinnon, E. & Mallal, S. Contribution of nucleoside-analogue reverse transcriptase inhibitor therapy to lipoatrophy from the population to the cellular level. *Antivir. Ther. (Lond.)* **8**, 617–626 (2003).

344. Mallal, S. A., John, M., Moore, C. B., James, I. R. & McKinnon, E. J. Contribution of nucleoside analogue reverse transcriptase inhibitors to subcutaneous fat wasting in patients with HIV infection: *AIDS* **14**, 1309–1316 (2000).

345. Carr, A. *et al.* A syndrome of peripheral lipodystrophy, hyperlipidaemia and insulin resistance in patients receiving HIV protease inhibitors. *AIDS* **12**, F51-58 (1998).

346. Kim, R. J., Wilson, C. G., Wabitsch, M., Lazar, M. A. & Steppan, C. M. HIV Protease Inhibitor-Specific Alterations in Human Adipocyte Differentiation and Metabolism*. *Obesity* **14**, 994–1002 (2006).

347. Lagathu, C. *et al.* Some HIV antiretrovirals increase oxidative stress and alter chemokine, cytokine or adiponectin production in human adipocytes and macrophages. *Antivir. Ther. (Lond.)* **12**, 489–500 (2007).

348. Kudlow, B. A., Jameson, S. A. & Kennedy, B. K. HIV protease inhibitors block adipocyte differentiation independently of lamin A/C: *AIDS* **19**, 1565–1573 (2005).

349. Couturier, J. *et al.* Human adipose tissue as a reservoir for memory CD4+ T cells and HIV: *AIDS* **29**, 667–674 (2015).

350. Damouche, A. *et al.* High proportion of PD-1-expressing $CD4^+$ T cells in adipose tissue constitutes an immunomodulatory microenvironment that may support HIV persistence. *European Journal of Immunology* **47**, 2113–2123 (2017).

351. Couturier, J. *et al.* Adipocytes impair efficacy of antiretroviral therapy. *Antiviral Research* **154**, 140–148 (2018).

352. Hsu, D. C. *et al.* CD4+ Cell infiltration into subcutaneous adipose tissue is not indicative of productively infected cells during acute SHIV infection. *Journal of Medical Primatology* **46**, 154–157 (2017).

353. McDonnell, W. J. *et al.* High CD8 T-Cell Receptor Clonality and Altered CDR3 Properties Are Associated With Elevated Isolevuglandins in Adipose Tissue During Diet-Induced Obesity. *Diabetes* **67**, 2361–2376 (2018).

354. Damouche, A. *et al.* Adipose Tissue Is a Neglected Viral Reservoir and an Inflammatory Site during Chronic HIV and SIV Infection. *PLOS Pathogens* **11**, e1005153 (2015).

355. Lumeng, C. N., Bodzin, J. L. & Saltiel, A. R. Obesity induces a phenotypic switch in adipose tissue macrophage polarization. *J. Clin. Invest.* **117**, 175–184 (2007).

356. Weisberg, S. P. *et al.* Obesity is associated with macrophage accumulation in adipose tissue. *J. Clin. Invest.* **112**, 1796–1808 (2003).

357. Gao, D. *et al.* Interleukin-1β mediates macrophage-induced impairment of insulin signaling in human primary adipocytes. *American Journal of Physiology-Endocrinology and Metabolism* **307**, E289–E304 (2014).

358. Lumeng, C. N., Deyoung, S. M. & Saltiel, A. R. Macrophages block insulin action in adipocytes by altering expression of signaling and glucose transport proteins. *American Journal of Physiology-Endocrinology and Metabolism* **292**, E166–E174 (2007).

359. Alexander, R. & Harrell, D. Autologous fat grafting: use of closed syringe microcannula system for enhanced autologous structural grafting. *Clinical, Cosmetic and Investigational Dermatology* 91 (2013). doi:10.2147/CCID.S40575

360. Kotecha, N., Krutzik, P. O. & Irish, J. M. Web-Based Analysis and Publication of Flow Cytometry Experiments. *Current Protocols in Cytometry* **53**, 10.17.1-10.17.24 (2010).

361. Diggins, K. E., Ferrell, P. B. & Irish, J. M. Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data. *Methods* **82**, 55–63 (2015).

362. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biology* **11**, R106 (2010).

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

364. Sathaliyawala, T. *et al.* Distribution and Compartmentalization of Human Circulating and Tissue-Resident Memory T Cell Subsets. *Immunity* **38**, 187–197 (2013).

365. Han, S.-J. *et al.* White Adipose Tissue Is a Reservoir for Memory T Cells and Promotes Protective Memory Responses to Infection. *Immunity* **47**, 1154-1168.e6 (2017).

366. Casey, K. A. *et al.* Antigen-Independent Differentiation and Maintenance of Effector-like Resident Memory T Cells in Tissues. *The Journal of Immunology* **188**, 4866–4875 (2012).

367. Farber, D. L., Yudanin, N. A. & Restifo, N. P. Human memory T cells: generation, compartmentalization and homeostasis. *Nature Reviews Immunology* **14**, 24–35 (2014).

368. Focosi, D., Bestagno, M., Burrone, O. & Petrini, M. CD57 $^{+}$ T lymphocytes and functional immune deficiency. *Journal of Leukocyte Biology* **87**, 107–116 (2010).

369. Brenchley, J. M. Expression of CD57 defines replicative senescence and antigen-induced apoptotic death of CD8+ T cells. *Blood* **101**, 2711–2720 (2003).

370. Palmer, B. E., Blyveis, N., Fontenot, A. P. & Wilson, C. C. Functional and Phenotypic Characterization of CD57+CD4+ T Cells and Their Association with HIV-1-Induced T Cell Dysfunction. *The Journal of Immunology* **175**, 8415–8423 (2005).

371. Papagno, L. *et al.* Immune Activation and CD8+ T-Cell Differentiation towards Senescence in HIV-1 Infection. *PLoS Biology* **2**, e20 (2004).

372. Lieberman, J. *et al.* Expansion of CD57 and CD62L-CD45RA+ CD8 T lymphocytes correlates with reduced viral plasma RNA after primary HIV infection. *AIDS* **13**, 891–899 (1999).

373. Le Priol, Y. *et al.* High Cytotoxic and Specific Migratory Potencies of Senescent CD8+CD57+ Cells in HIV-Infected and Uninfected Individuals. *The Journal of Immunology* **177**, 5145–5154 (2006).

374. Tae Yu, H. *et al.* Characterization of CD8+CD57+ T cells in patients with acute myocardial infarction. *Cellular & Molecular Immunology* **12**, 466–473 (2015).

375. Maeda, T. *et al.* Involvement of CD4+,CD57+ T cells in the disease activity of rheumatoid arthritis. *Arthritis & Rheumatism* **46**, 379–384 (2002).

376. Palmer, B. E., Mack, D. G., Martin, A. K., Maier, L. A. & Fontenot, A. P. CD57 expression correlates with alveolitis severity in subjects with beryllium-induced disease. *Journal of Allergy and Clinical Immunology* **120**, 184–191 (2007).

377. Shnayder, M. *et al.* Defining the Transcriptional Landscape during Cytomegalovirus Latency with Single-Cell RNA Sequencing. *mBio* **9**, (2018).

378. Tian, Y. *et al.* Unique phenotypes and clonal expansions of human CD4 effector memory T cells re-expressing CD45RA. *Nature Communications* **8**, (2017).

379. Truong, K.-L. *et al.* Progressive change in killer-like receptor and GPR56 expression defines cytokine production of human CD4 T memory cells. *bioRxiv* (2017). doi:10.1101/191007

380. Böttcher, J. P. *et al.* Functional classification of memory CD8+ T cells by CX3CR1 expression. *Nature Communications* **6**, (2015).

381. Nishimura, M. *et al.* Dual Functions of Fractalkine/CX3C Ligand 1 in Trafficking of Perforin+/Granzyme B+ Cytotoxic Effector Lymphocytes That Are Defined by CX3CR1 Expression. *The Journal of Immunology* **168**, 6173–6180 (2002).

382. Pachnio, A. *et al.* Cytomegalovirus Infection Leads to Development of High Frequencies of Cytotoxic Virus-Specific CD4+ T Cells Targeted to Vascular Endothelium. *PLoS Pathog.* **12**, e1005832 (2016).

383. Gordon, C. L. *et al.* Induction and Maintenance of CX3CR1-Intermediate Peripheral Memory CD8 + T Cells by Persistent Viruses and Vaccines. *Cell Reports* **23**, 768–782 (2018).

384. Peng, Y.-M. *et al.* Specific expression of GPR56 by human cytotoxic lymphocytes. *Journal of Leukocyte Biology* **90**, 735–740 (2011).

385. Travers, R. L., Motta, A. C., Betts, J. A., Bouloumié, A. & Thompson, D. The impact of adiposity on adipose tissue-resident lymphocyte activation in humans. *International Journal of Obesity* **39**, 762–769 (2015).

386. Couturier, J. *et al.* Infectious SIV resides in adipose tissue and induces metabolic defects in chronically infected rhesus macaques. *Retrovirology* **13**, (2016).

387. Vitseva, O. I. *et al.* Inducible Toll-like Receptor and NF-κB Regulatory Pathway Expression in Human Adipose Tissue. *Obesity* **16**, 932–937 (2008).

388. Ahmad, R. *et al.* Elevated expression of the toll like receptors 2 and 4 in obese individuals: its significance for obesity-induced inflammation. *Journal of Inflammation* **9**, 48 (2012).

389. Creely, S. J. *et al.* Lipopolysaccharide activates an innate immune system response in human adipose tissue in obesity and type 2 diabetes. *American Journal of Physiology-Endocrinology and Metabolism* **292**, E740–E747 (2007).

390. Dasu, M. R., Devaraj, S., Park, S. & Jialal, I. Increased Toll-Like Receptor (TLR) Activation and TLR Ligands in Recently Diagnosed Type 2 Diabetic Subjects. *Diabetes Care* **33**, 861–868 (2010).

391. Martinez, F. O., Gordon, S., Locati, M. & Mantovani, A. Transcriptional Profiling of the Human Monocyte-to-Macrophage Differentiation and Polarization: New Molecules and Patterns of Gene Expression. *The Journal of Immunology* **177**, 7303–7311 (2006).

392. Caccuri, F. *et al.* HIV-1 matrix protein p17 promotes angiogenesis via chemokine receptors CXCR1 and CXCR2. *Proceedings of the National Academy of Sciences* **109**, 14580–14585 (2012).

393. Sierra-Honigmann, M. R. *et al.* Biological action of leptin as an angiogenic factor. *Science* **281**, 1683–1686 (1998).

394. Wolchok, J. D. PD-1 Blockers. *Cell* **162**, 937 (2015).

395. Johnson, D. B., Chandra, S. & Sosman, J. A. Immune Checkpoint Inhibitor Toxicity in 2018. *JAMA* **320**, 1702–1703 (2018).

396. Postow, M. A., Sidlow, R. & Hellmann, M. D. Immune-Related Adverse Events Associated with Immune Checkpoint Blockade. *N. Engl. J. Med.* **378**, 158–168 (2018).

397. Moslehi, J. J., Salem, J.-E., Sosman, J. A., Lebrun-Vignes, B. & Johnson, D. B. Increased reporting of fatal immune checkpoint inhibitor-associated myocarditis. *Lancet* **391**, 933 (2018).

398. Johnson, D. B. *et al.* Fulminant Myocarditis with Combination Immune Checkpoint Blockade. *N. Engl. J. Med.* **375**, 1749–1755 (2016).

399. Naidoo, J. *et al.* Pneumonitis in Patients Treated With Anti-Programmed Death-1/Programmed Death Ligand 1 Therapy. *J. Clin. Oncol.* **35**, 709–717 (2017).

400. Ryman, J. T. & Meibohm, B. Pharmacokinetics of Monoclonal Antibodies. *CPT Pharmacometrics Syst Pharmacol* **6**, 576–588 (2017).

401. Gonzalez, R. S. *et al.* PD-1 inhibitor gastroenterocolitis: case series and appraisal of 'immunomodulatory gastroenterocolitis'. *Histopathology* **70**, 558–567 (2017).

402. Verschuren, E. C. *et al.* Clinical, Endoscopic, and Histologic Characteristics of Ipilimumab-Associated Colitis. *Clin. Gastroenterol. Hepatol.* **14**, 836–842 (2016).

403. Larkin, J. *et al.* Neurologic Serious Adverse Events Associated with Nivolumab Plus Ipilimumab or Nivolumab Alone in Advanced Melanoma, Including a Case Series of Encephalitis. *Oncologist* **22**, 709–718 (2017).

404. Bossart, S. *et al.* Case Report: Encephalitis, with Brainstem Involvement, Following Checkpoint Inhibitor Therapy in Metastatic Melanoma. *Oncologist* **22**, 749–753 (2017).

405. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

406. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).

407. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

408. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

409. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**, 568–571 (2014).

410. Gerlinger, M. *et al.* Ultra-deep T cell receptor sequencing reveals the complexity and intratumour heterogeneity of T cell clones in renal cell carcinomas. *J. Pathol.* **231**, 424–432 (2013).

411. Mathew, J. M. *et al.* Generation and Characterization of Alloantigen-Specific Regulatory T Cells For Clinical Transplant Tolerance. *Sci Rep* **8**, 1136 (2018).

412. Shugay, M. *et al.* VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* **46**, D419–D427 (2018).

413. Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E. & Friedman, N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* **33**, 2924–2929 (2017).

414. Chen, G. *et al.* Sequence and Structural Analyses Reveal Distinct and Highly Diverse Human CD8(+) TCR Repertoires to Immunodominant Viral Antigens. *Cell Rep* **19**, 569–583 (2017).

415. De Neuter, N. *et al.* On the feasibility of mining CD8+ T cell receptor patterns underlying immunogenic peptide recognition. *Immunogenetics* **70**, 159–168 (2018).

416. Gielis, S. *et al.* TCRex: a webtool for the prediction of T-cell receptor sequence epitope specificity. *bioRxiv* (2018). doi:10.1101/373472

417. Rustenhoven, J., Jansson, D., Smyth, L. C. & Dragunow, M. Brain Pericytes As Mediators of Neuroinflammation. *Trends Pharmacol. Sci.* **38**, 291–304 (2017).

418. Annels, N. E., Callan, M. F., Tan, L. & Rickinson, A. B. Changing patterns of dominant TCR usage with maturation of an EBV-specific cytotoxic T cell response. *J. Immunol.* **165**, 4831–4841 (2000).

419. Lim, A. *et al.* Frequent contribution of T cell clonotypes with public TCR features to the chronic response against a dominant EBV-derived epitope: application to direct detection of their molecular imprint on the human peripheral T cell repertoire. *J. Immunol.* **165**, 2001–2011 (2000).

420. Cohen, G. B. *et al.* Clonotype Tracking of TCR Repertoires during Chronic Virus Infections. *Virology* **304**, 474–484 (2002).

421. Koning, D. *et al.* In vitro expansion of antigen-specific CD8(+) T cells distorts the T-cell repertoire. *J. Immunol. Methods* **405**, 199–203 (2014).

422. Grant, E. J. *et al.* Lack of Heterologous Cross-reactivity toward HLA-A*02:01 Restricted Viral Epitopes Is Underpinned by Distinct αβT Cell Receptor Signatures. *J. Biol. Chem.* **291**, 24335–24351 (2016).

423. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).

424. Tzellos, S. & Farrell, P. J. Epstein-barr virus sequence variation-biology and disease. *Pathogens* **1**, 156–174 (2012).

425. Young, L. S. & Rickinson, A. B. Epstein-Barr virus: 40 years on. *Nat. Rev. Cancer* **4**, 757–768 (2004).

426. Jordan, B. *et al.* A study on viral CNS inflammation beyond herpes encephalitis. *J. Neurovirol.* **22**, 763–773 (2016).

427. Zarlasht, F., Salehi, M., Abu-Hishmeh, M. & Khan, M. Encephalitis treatment - a case report with long-term follow-up of EBV PCR in cerebrospinal fluid. *Int J Gen Med* **10**, 371–373 (2017).

428. Wang, X. *et al.* Elevated expression of EBV and TLRs in the brain is associated with Rasmussen's encephalitis. *Virol Sin* **32**, 423–430 (2017).

429. Mashima, K. *et al.* Epstein-Barr Virus-associated Lymphoproliferative Disorder with Encephalitis Following Anti-thymocyte Globulin for Aplastic Anemia Resolved

with Rituximab Therapy: A Case Report and Literature Review. *Intern. Med.* **56**, 701–706 (2017).

430. Babcock, G. J., Decker, L. L., Volk, M. & Thorley-Lawson, D. A. EBV persistence in memory B cells in vivo. *Immunity* **9**, 395–404 (1998).

431. Wang, D. Y. *et al.* Fatal Toxic Effects Associated With Immune Checkpoint Inhibitors: A Systematic Review and Meta-analysis. *JAMA Oncol* **4**, 1721–1728 (2018).

432. Kao, J. C. *et al.* Neurological Complications Associated With Anti-Programmed Death 1 (PD-1) Antibodies. *JAMA Neurol* **74**, 1216–1222 (2017).

433. Abdel-Wahab, N., Shah, M. & Suarez-Almazor, M. E. Adverse Events Associated with Immune Checkpoint Blockade in Patients with Cancer: A Systematic Review of Case Reports. *PLoS ONE* **11**, e0160221 (2016).

434. Spain, L. *et al.* Neurotoxicity from immune-checkpoint inhibition in the treatment of melanoma: a single centre experience and review of the literature. *Ann. Oncol.* **28**, 377–385 (2017).

435. Iwama, S. *et al.* Pituitary expression of CTLA-4 mediates hypophysitis secondary to administration of CTLA-4 blocking antibody. *Sci Transl Med* **6**, 230ra45 (2014).

436. Osorio, J. C. *et al.* Antibody-mediated thyroid dysfunction during T-cell checkpoint blockade in patients with non-small-cell lung cancer. *Ann. Oncol.* **28**, 583–589 (2017).

437. Das, R. *et al.* Early B cell changes predict autoimmunity following combination immune checkpoint blockade. *J. Clin. Invest.* **128**, 715–720 (2018).

438. Dubin, K. *et al.* Intestinal microbiome analyses identify melanoma patients at risk for checkpoint-blockade-induced colitis. *Nat Commun* **7**, 10391 (2016).

439. Peeters, L. M. *et al.* Cytotoxic CD4+ T Cells Drive Multiple Sclerosis Progression. *Front Immunol* **8**, 1160 (2017).

440. Curran, M. A. *et al.* Systemic 4-1BB activation induces a novel T cell phenotype driven by high expression of Eomesodermin. *J. Exp. Med.* **210**, 743–755 (2013).

441. Hirschhorn-Cymerman, D. *et al.* Induction of tumoricidal function in CD4+ T cells is associated with concomitant memory and terminally differentiated phenotype. *J. Exp. Med.* **209**, 2113–2126 (2012).

442. Penaloza-MacMaster, P. *et al.* Vaccine-elicited CD4 T cells induce immunopathology after chronic LCMV infection. *Science* **347**, 278–282 (2015).

443. Takeuchi, A. & Saito, T. CD4 CTL, a Cytotoxic Subset of CD4+ T Cells, Their Differentiation and Function. *Frontiers Immunol* **8**, (2017).

444. Nikiforow, S., Bottomly, K., Miller, G. & Münz, C. Cytolytic CD4(+)-T-cell clones reactive to EBNA1 inhibit Epstein-Barr virus-induced B-cell proliferation. *J. Virol.* **77**, 12088–12104 (2003).

445. Landais, E. *et al.* Direct killing of Epstein-Barr virus (EBV)-infected B cells by CD4 T cells directed against the EBV lytic protein BHRF1. *Blood* **103**, 1408–1416 (2004).

446. Martorelli, D. *et al.* Role of CD4+ cytotoxic T lymphocytes in the control of viral diseases and cancer. *Int. Rev. Immunol.* **29**, 371–402 (2010).

447. Hislop, A. D. & Taylor, G. S. T-Cell Responses to EBV. *Curr. Top. Microbiol. Immunol.* **391**, 325–353 (2015).

448. Ranasinghe, S. *et al.* Antiviral CD8+ T Cells Restricted by Human Leukocyte Antigen Class II Exist during Natural HIV Infection and Exhibit Clonal Expansion. *Immunity* **45**, 917–930 (2016).

449. Boyle, L. H., Goodall, J. C. & Gaston, J. S. H. Major histocompatibility complex class I-restricted alloreactive CD4+ T cells. *Immunology* **112**, 54–63 (2004).

450. Wang, M. *et al.* High-affinity human leucocyte antigen class I binding variola-derived peptides induce CD4+ T cell responses more than 30 years post-vaccinia virus vaccination. *Clin. Exp. Immunol.* **155**, 441–446 (2009).

451. Legoux, F. *et al.* Characterization of the human CD4(+) T-cell repertoire specific for major histocompatibility class I-restricted antigens. *Eur. J. Immunol.* **43**, 3244–3253 (2013).

452. Heemskerk, M. H. *et al.* Dual HLA class I and class II restricted recognition of alloreactive T lymphocytes mediated by a single T cell receptor complex. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6806–6811 (2001).

453. Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat. Genet.* **50**, 699–707 (2018).

454. Gelbard, A. *et al.* Causes and consequences of adult laryngotracheal stenosis. *Laryngoscope* **125**, 1137–1143 (2015).

455. Shaham, U. *et al.* Removal of batch effects using distribution-matching residual networks. *Bioinformatics* **33**, 2539–2546 (2017).

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

457. Barkas, N. *et al.* Wiring together large single-cell RNA-seq sample collections. *bioRxiv* (2018). doi:10.1101/460246

458. Zhang, A. W. *et al.* Probabilistic cell type assignment of single-cell transcriptomic data reveals spatiotemporal microenvironment dynamics in human cancers: Supplementary tables. *bioRxiv* (2019). doi:10.1101/521914

459. Li, X. *et al.* Deep learning enables accurate clustering and batch effect removal in single-cell RNA-seq analysis: Supplementary material. *bioRxiv* (2019). doi:10.1101/530378

460. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nature Methods* **16**, 43–49 (2019).

461. Workel, H. H. *et al.* TGF-β induced CXCL13 in CD8+ T cells is associated with tertiary lymphoid structures in cancer. *bioRxiv* (2018). doi:10.1101/303834

462. Sautès-Fridman, C., Petitprez, F., Calderaro, J. & Fridman, W. H. Tertiary lymphoid structures in the era of cancer immunotherapy. *Nat. Rev. Cancer* (2019). doi:10.1038/s41568-019-0144-6

463. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).

464. Hardwick, S. A. *et al.* Spliced synthetic genes as internal controls in RNA sequencing experiments. *Nature Methods* **13**, 792–798 (2016).

465. Wong, T., Deveson, I. W., Hardwick, S. A. & Mercer, T. R. ANAQUIN: a software toolkit for the analysis of spike-in controls for next generation sequencing. *Bioinformatics* **33**, 1723–1724 (2017).

466. McCarthy, D. J., Campbell, K. R., Lun, A. T. L. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179–1186 (2017).

467. Li, S. S. *et al.* DNA Priming Increases Frequency of T-Cell Responses to a Vesicular Stomatitis Virus HIV Vaccine with Specific Enhancement of CD8+ T-Cell Responses by Interleukin-12 Plasmid DNA. *Clin. Vaccine Immunol.* **24**, (2017).

468. McDavid, A., Finak, G. & Gottardo, R. The contribution of cell cycle to heterogeneity in single-cell RNA-seq data. *Nature Biotechnology* **34**, 591–593 (2016).

469. Olson, R. S. *et al.* Automating biomedical data science through tree-based pipeline optimization. *arXiv:1601.07925 [cs]* (2016).

470. Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A. & Moore, J. H. Automating biomedical data science through tree-based pipeline optimization. in *European Conference on the Applications of Evolutionary Computation* 123–137 (2016).

471. McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461–467 (2013).

472. Pasqual, G. *et al.* Monitoring T cell-dendritic cell interactions in vivo by intercellular enzymatic labelling. *Nature* **553**, 496–500 (2018).

473. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).

474. Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G.-C. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nature Biotechnology* **36**, 1183–1190 (2018).

475. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

476. Saikia, M. *et al.* Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nature Methods* **16**, 59–62 (2019).

477. Zhang, S.-Q. *et al.* High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nature Biotechnology* **36**, 1156–1159 (2018).

478. Nam, A. S. *et al.* High throughput droplet single-cell Genotyping of Transcriptomes (GoT) reveals the cell identity dependency of the impact of somatic mutations. *bioRxiv* (2018). doi:10.1101/444687

479. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive scale single-cell epigenomics. *bioRxiv* (2019). doi:10.1101/612713

480. McCarthy, D. J. *et al.* Cardelino: Integrating whole exomes and single-cell transcriptomes to reveal phenotypic impact of somatic variants. *bioRxiv* (2018). doi:10.1101/413047

481. Cao, J. *et al.* Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).

482. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.e11 (2019).

483. Fischer, D. S. *et al.* Inferring population dynamics from single-cell RNA-sequencing time series data. *Nature Biotechnology* **37**, 461–468 (2019).

484. Le, K. *et al.* A novel mammalian cell line development platform utilizing nanofluidics and optoelectro positioning technology. *Biotechnology Progress* **34**, 1438–1446 (2018).

485. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods* **15**, 1053–1058 (2018).

486. Vallania, F. *et al.* Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. *Nature Communications* **9**, (2018).

487. Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Charting a tissue from single-cell transcriptomes. *bioRxiv* (2018). doi:10.1101/456350

488. Liang, S., Wang, F., Han, J. & Chen, K. Latent periodic process inference from single-cell RNA-seq data. *bioRxiv* (2019). doi:10.1101/625566

489. Linderman, G. C., Zhao, J. & Kluger, Y. Zero-preserving imputation of scRNA-seq data using low-rank approximation. *bioRxiv* (2018). doi:10.1101/397588

490. Tan, Y. & Cahan, P. SingleCellNet: a computational tool to classify single cell RNA-Seq data across platforms and across species. *bioRxiv* (2018). doi:10.1101/508085

491. Ma, F. & Pellegrini, M. Automated identification of Cell Types in Single Cell RNA Sequencing: Supplemental File 1. *bioRxiv* (2019). doi:10.1101/532093

492. Dumitrascu, B., Villar, S., Mixon, D. G. & Engelhardt, B. E. Optimal marker gene selection for cell type discrimination in single cell analyses. *bioRxiv* (2019). doi:10.1101/599654

493. Cao, Z.-J., Wei, L., Lu, S., Yang, D.-C. & Gao, G. Cell BLAST: Searching large-scale scRNA-seq databases via unbiased cell embedding. *bioRxiv* (2019). doi:10.1101/587360

494. Dumitrascu, B., Feng, K. & Engelhardt, B. E. GT-TS: Experimental design for maximizing cell type discovery in single-cell data. *bioRxiv* (2018). doi:10.1101/386540

495. Tian, L. *et al.* Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. *Nature Methods* (2019). doi:10.1038/s41592-019-0425-8

496. Rakocevic, G. *et al.* Fast and accurate genomic analyses using genome graphs. *Nature Genetics* **51**, 354–362 (2019).

497. Garrison, E. *et al.* Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nature Biotechnology* **36**, 875–879 (2018).

498. Baron, C. S. *et al.* Cell type purification by single-cell transcriptome-trained sorting: Supplementary figures and tables. *bioRxiv* (2018). doi:10.1101/502773