Shaping the helper T cell landscape: from naive to effectors

Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Christian-Albrechts-Universität zu Kiel

Valeriia Kriukova

Kiel, 2023

First examiner: Prof. Dr. rer. nat. Andre Franke
Second examiner: Prof. Dr. rer. nat. Thomas Roeder
Date of the oral examination: 13.12.2023
Approved for publication: 13.12.2023

Summary - English

Helper T (Th) lymphocytes play a key role in our immunity, orchestrating responses to various pathogens, vaccines, host proteins, microbiota, environmental antigens, and cancers. The development and function of Th cells depend on their T cell receptors (TCR). Each individual T cell progenitor acquires its TCR-encoding genes through stochastic genome recombination within TCRα and TCRβ loci. Collectively, TCRs of independent T cells contribute to an extremely diverse TCR repertoire. The TCR repertoire serves to recognize a large pool of distinct peptide antigens presented within major histocompatibility complexes (MHC). The interaction with MHC plays a central role in the development and function of T cells. Thus, T cell progenitors' survival and early commitment to a Th lineage in thymus depends on MHC class II (MHC-II) recognition and the corresponding TCR signaling. Later peripheral recognition of antigens presented within MHC-II involves naive and memory Th lymphocytes in immune responses. Activated Th cells acquire specific helper phenotypes, which control the development of immune responses of an appropriate type, facilitating pathogen clearance and tissue homeostasis. Later, the exact antigen specificity and the linked Th phenotype are preserved in the form of Th memory clones, providing stable response patterns to familiar antigens. Studying the thymic and peripheral fate selection by Th cells may help to mechanically understand the successes and failures of our immune system to develop protective responses in infectious diseases, vaccinations, and cancers, as well as the generation of immune tolerance to microbiota-derived and host antigens.

Within the scope of this work, several aspects of Th lymphocyte development and phenotype acquisition are studied through the lens of their TCR repertoire analysis. The introductory chapter covers the key concepts of Th biology with a particular emphasis on Th lymphocyte involvement in antiviral immunity and gives an overview of technologies used in the study.

The results contain three manuscripts produced within this work.

Manuscript 1 presents the analysis of TCR repertoires in naive Th lymphocytes selected on different MHC-II allelic variants. In detail, two mouse strains were used, which carried an identical genetic background but different MHC-II alleles, known as H2-A^{j} and H2-A^{b}. The study aimed to investigate the effects of MHC allelic variation on the TCR repertoires of mature naive T regulatory (CD4+Foxp3+, nTreg) and naive T conventional (CD4+Foxp3-, nTconv) lymphocytes. The results of this study showed that the selection on the MHC-II allelic variants shapes the usage of TRAV and TRBV gene

segments and the CDR3 repertoires. This work confirmed several known key TCR features statistically predisposing T cell progenitors to nTreg or nTconv lineage commitment, such as shortened CDR3α and CDR3β regions with a higher average number of strongly interacting hydrophobic amino acid residues, as well as demonstrated how these features may vary in the different MHC-II contexts.

Manuscript 2 aims to describe the phenotypic diversity of Th cells coexisting in the peripheral blood of healthy individuals and patients with an acute viral SARS-CoV-2 infection. Within the scope of the work, a new approach, Sort-Seq, was introduced to connect classically defined Th phenotypes and their underlying scRNA-Seq transcriptomes. As a result, an integrated scRNA-Seq reference dataset of well-annotated peripheral blood Th lymphocytes was created, providing us with a frame for further analysis of Th cells responding to acute viral infections. Additionally, a large set of SARS-CoV-2-specific TCRβ clonotypes from a single donor was discovered using various bioinformatic and experimental approaches. Ultimately, the developed Th reference scRNA-Seq dataset was used to uncover the phenotypes of the identified antigen-specific Th cells in the context of primary and secondary acute SARS-CoV-2 infections.

Manuscript 3 describes the involvement of Th lymphocytes in a model chronic viral infection. Particularly, the properties of IL-10-producing Th lymphocytes promoting mouse cytomegalovirus (MCMV) persistence in salivary glands were studied in comparison to their IL-10-negative counterparts. Both studied Th subsets developed clusters of similar TCRβ clonotypes, suggesting the existence of an ongoing antigen-driven immune response. The clonal overlap between the TCRβ repertoires indicated a probably shared origin of the IL-10-producing and the IL-10-negative Th lymphocytes in MCMV infection. This study gave us a valuable example of how the phenotypic plasticity of Th clones may result in attenuated pathogen clearance and viral chronicity.

In summary, the results of this thesis cover the influence of host genetics and acute and chronic viral infections on different stages of Th development and function, collectively contributing to the landscapes of Th phenotypes and antigen-specificities.

Zusammenfassung – Deutsch

T-Helfer-Lymphozyten (Th-Lymphozyten) spielen eine Schlüsselrolle in unserer Immunität, indem sie die Reaktionen auf verschiedene Krankheitserreger, Impfstoffe, Wirtsproteine, Mikrobiota, Umweltantigene und Krebserkrankungen steuern. Die Entwicklung und Funktion von Th-Zellen hängt von ihrem T-Zell-Rezeptor (TCR) ab. Jede einzelne T-Helferzelle erwirbt ihre spezifisch für den TCR kodierenden Gene durch stochastische Genomrekombination innerhalb der TCRα- und TCRβ-Loci. Gemeinsam tragen die TCR unabhängiger T-Zellen zu einem äußerst vielfältigen TCR-Repertoire bei. Das TCR-Repertoire dient der Erkennung einer großen Bandbreite verschiedener Peptidantigene, die in Haupthistokompatibilitätskomplexen (MHC) präsentiert werden.

Die Interaktion mit MHC spielt eine zentrale Rolle für die Entwicklung und Funktion von T-Zellen. So hängt das Überleben und die frühe Festlegung von Vorläuferzellen auf eine T-Zelllinie im Thymus, von der Erkennung der MHC-Klasse II (MHC-II) und der entsprechenden TCR Signalübertragung ab. Die spätere, periphere Erkennung von Antigenen, die von MHC-II präsentiert werden, bezieht naive und Th-Gedächtniszellen in die Immunantwort ein. Aktivierte Th-Zellen erwerben spezifische Phänotypen, die geeignete Immunantworten steuern und so die Bekämpfung von Krankheitserregern und die Gewebehomöostase erleichtern. Später bleibt die genaue Antigenspezifität und der damit verbundene Th-Phänotyp in Form von Th-Gedächtniszellen erhalten, die stabile Reaktionsmuster auf bekannte Antigene liefern. Die Untersuchung der Th-Zellen im Thymus und in der Peripherie kann dazu beitragen, die Erfolge und Misserfolge unseres Immunsystems bei der Ausbildung von Immunreaktionen auf Infektionskrankheiten, Impfungen und Krebserkrankungen sowie die Entstehung von Immuntoleranz gegenüber Mikrobiota- und Wirtsantigenen besser zu verstehen.

Im Rahmen dieser Arbeit werden verschiedene Aspekte der Ausbildung von Phänotypen von Th-Lymphozyten durch die Analyse ihres TCR-Repertoires untersucht. Das Einführungskapitel behandelt die Schlüsselkonzepte der Biologie der Th-Lymphozyten mit besonderem Schwerpunkt auf ihrer Beteiligung an der antiviralen Immunität und gibt einen Überblick über die in der Studie verwendete Methodik. Der Ergebnisteil setzt sich aus drei Manuskripten zusammen, die im Rahmen dieser Arbeit entstanden sind.

Manuskript 1 stellt die Analyse des TCR-Repertoires in naiven Th-Lymphozyten vor, die auf verschiedene MHC-II-Allele selektiert wurden. Im Einzelnen wurden zwei Mausstämme verwendet, die einen identischen genetischen Hintergrund haben, aber unterschiedliche MHC-II-Allele, bekannt als H2-Aj und H2-Ab, tragen. Ziel der Studie war es, die Auswirkungen der MHC-Allelvariation auf das

TCR-Repertoire von reifen naiven regulatorischen (CD4+Foxp3+, nTreg) und naiven konventionellen (CD4+Foxp3-, nTconv) T-Lymphozyten zu untersuchen. Die Ergebnisse der Studie zeigten, dass die Selektion auf die MHC-II-Allele nicht nur die Verwendung der TCR TRAV und TRBV Gensegmente, sondern auch das CDR3-Repertoire prägt. Diese Arbeit bestätigte mehrere bekannte TCR-Schlüsselmerkmale, die T-Zell-Vorläufer statistisch für eine nTreg- oder nTconv-Linienbindung prädisponieren, wie z. B. verkürzte CDR3α- und CDR3β-Regionen mit einer höheren durchschnittlichen Anzahl stark interagierender hydrophober Aminosäurereste, und zeigte, wie diese Merkmale in den verschiedenen MHC-II-Kontexten variieren können.

Manuskript 2 zielt darauf ab, die phänotypische Vielfalt von Th-Zellen zu beschreiben, die im peripheren Blut von gesunden Personen und Patienten mit einer akuten viralen SARS-CoV-2-Infektion koexistieren. Im Rahmen der Arbeit wurde ein neuer Ansatz (Sort-Seq) eingeführt, um klassisch definierte Th-Phänotypen und den zugrunde liegenden scRNA-Seq-Transkriptomen zu unterscheiden. Als Ergebnis wurde ein integrierter scRNA-Seq-Referenzdatensatz von gut annotierten Th-Lymphozyten aus peripherem Blut erstellt, der einen Rahmen für die weitere Analyse von Th-Zellen bietet, die auf akute virale Infektionen reagieren. Zusätzlich wurde eine große Anzahl von SARS-CoV-2-spezifischen TCRβ-Klonotypen von einem einzigen Spender mit Hilfe verschiedener bioinformatischer und experimenteller Ansätze charakterisiert. Schließlich wurde der erstellte Th-Referenz-scRNA-Seq-Datensatz verwendet, um die Phänotypen der identifizierten antigenspezifischen Th-Lymphozyten, im Zusammenhang mit primären und sekundären akuten SARS-CoV-2-Infektionen, zu untersuchen.

Manuskript 3 beschreibt die Beteiligung von Th-Lymphozyten in einem Modell für eine chronische Virusinfektion. Hier wurden insbesondere die Eigenschaften von IL-10 produzierenden Th-Lymphozyten, die die Persistenz des Maus-Cytomegalovirus (MCMV) in Mäusespeicheldrüsen fördern, im Vergleich zu ihren IL-10 negativen Gegenspielern untersucht. Beide untersuchten Th-Untergruppen entwickelten Cluster ähnlicher TCRβ-Klonotypen, was auf eine laufende antigengesteuerte Immunantwort hindeutet. Die klonale Überlappung zwischen den TCRβ-Repertoires deutet auf einen gemeinsamen Ursprung der IL-10 produzierenden und der IL-10 negativen Th-Lymphozyten bei chronischer MCMV-Infektion hin. Diese Studie lieferte uns ein wertvolles Beispiel dafür, wie phänotypische Plastizität von Th-Klonen zu einer abgeschwächten Erregerbeseitigung und viraler Chronizität führen kann.

Zusammenfassend lässt sich sagen, dass die Ergebnisse dieser Arbeit den Einfluss der Wirtsgenetik sowie akuter und chronischer Virusinfektionen auf die verschiedenen Stadien der Th-Entwicklung

und -Funktion abdecken, die gemeinsam zu der großen Bandbreite der Th-Phänotypen und Antigen-Spezifitäten beitragen.

Table of Contents

List of Figures

Manuscript 1

Manuscript 2

Manuscript 3

List of Tables

Abbreviations

ANCA	Anti-neutrophil cytoplasmic antibodies
APC	Antigen presenting cell
BCR	B cell receptor
cDNA	Complementary Deoxyribonucleic acid
CDR	Complementarity-determining region
CITE-Seq	Cellular indexing of transcriptomes and epitopes (sequencing)
CMV	Cytomegalovirus
COVID-19	Coronavirus disease 2019
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
cTECs	Cortical thymic epithelial cells
DAMP	Danger associated molecular patterns
DP	Double-positive
EBV	Epstein-Barr virus
FR3	Framework region 3
HBV	Hepatitis B virus
HCV	Hepatitis C virus

HIV	Human immunodeficiency virus
IBD	Inflammatory bowel disease
Ig	Immunoglobulin
LCMV	Lymphocytic choriomeningitis virus
MCMV	Murine cytomegalovirus
MHC	Major histocompatibility complex
MHC-I	Major histocompatibility complex class I
MHC-II	Major histocompatibility complex class II
MMLV	Murine leukemia virus
mRNA	Messenger ribonucleic acid
mTECs	Medullary thymic epithelial cells
NGS	Next-generation sequencing
NK	Natural killer (cells)
nTconv	Naive T conventional (cells)
nTreg	Naive T regulatory (cells)
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
pMHC-II	Peptide - Major histocompatibility complex

R1	(Sequencing) read 1
R2	(Sequencing) read 2
RACE	Rapid amplification of cDNA ends
RLRs	RIG-I-like receptors
ROS	Reactive oxygen species
RTE	Recent thymic emigrants
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
scATAC-Seq	Single-cell Assay for Transposase-Accessible Chromatin using sequencing
scRNA-Seq	Single-cell RNA sequencing
scTCR-Seq	Single-cell TCR sequencing
SP	Single-positive
Tcm	Central memory T (cells)
TCR	T cell receptor
Teff	Effector T (cells)
Temra	Terminally differentiated effector memory cells re-expressing CD45RA
Tfh	Follicular helper T (cells)
Tfr	Follicular regulatory T (cells)
Th	T helpers

TLRs	Toll-like receptors
Tn	Naive T (cells)
TRAJ	T cell receptor alpha joining (segment)
TRAV	T cell receptor alpha variable (gene)
TRBJ	T cell receptor beta joining (segment)
TRBV	T cell receptor beta variable (gene)
Treg	Regulatory T (cells)
Trm	Tissue-resident memory T (cells)
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifier

1. Introduction

1.1. Motivation of the study

Functional programs selected by Th lymphocytes control the course of primary and any further secondary immune responses. The functional differentiation of Th subsets starts already in thymus during the process of MHC-II driven Th cells selection and continues upon the peripheral recognition of cognate peptide-MHC-II complexes (pMHC-II). Later, the acquired phenotype is preserved in memory antigen-specific Th clones, allowing for faster and phenotypically predetermined responses to secondary antigen encounters. In this light, such commitment of Th clones to specific phenotypes may be a double-edged sword. On the one hand, it may be beneficial in recurrent infectious diseases; however, on the other hand, the cross-reactivity to unrelated host and environmental antigens might cause an undesired pathology. We launched this study to shed some light on the decisions made by Th cells during their lineage commitment path, starting from the differentiation of naive Th lymphocytes and ending with the involvement in antigen-specific immune responses. Specifically, the study aimed

a) to identify the TCR features associated with nTreg and nTconv thymic selection in different MHC-II genetic contexts;

b) to describe the existing phenotypic and functional heterogeneity of peripheral blood Th lymphocytes and their involvement in an antigen-specific immune response to an acute viral infection;

c) to characterize the features of Th lymphocytes involved in an immune response to a model chronic viral infection.

1.2. Outline of the thesis

The introductory chapter of the thesis gives an overview of the underlying immunological concepts and methods used in this work.

The results consist of three manuscripts written within the scope of the thesis.

Manuscript 1, published in PNAS in 2020[1], describes the selection of nTreg and nTconv driven by different MHC-II allelic variants in a mouse model with a specific focus on TCR repertoire analysis. Our results independently confirmed the preferential engagement of TCRs bearing distinct germline-encoded TRAV and TRBV gene segments by different MHC-II alleles. Furthermore, we reported several observations related to the CDR3 repertoires shaping by selection on MHC-II. Thus, in both MHC-II allelic contexts shortened CDR3α and CDR3β loops with on average higher abundance of strongly interacting amino acids in the center of CDR3α and CDR3β were associated with the differentiation into nTregs. Nevertheless, TCR repertoires selected on the H2-A^j allelic variant were characterized by an even higher abundance of strongly interacting (mainly hydrophobic and aromatic) amino acids within CDR3α and CDR3β, compared to the selection on H2-A^b. Finally, even TCRs produced by the unfavored TRAV or TRBV gene segments could be rescued in the selection process via the CDR3-mediated interactions with MHC-II. Together, these results provided us with a set of rules dictating the selection of TCR repertoires and, ultimately, the lineage commitment in naive Th cells.

Manuscript 2, currently under review in Nature Communications, presents several new approaches to studying the heterogeneity of Th lymphocytes and their involvement in antiviral immune responses. Namely, we introduced a new Sort-Seq approach aimed to assist in annotation of Th subsets sorted based on surface marker expression within the scRNA-Seq data. Based on this advanced annotation, we created an easy-to-use reference scRNA-Seq dataset of peripheral blood Th lymphocytes from 122 healthy and SARS-CoV-2-infected donors. In parallel, we identified over 80 SARS-CoV-2-specific TCRβ clonotypes from an individual donor utilizing longitudinal TCR profiling, T cell functional assays, and the search within previously published databases. Eventually, we delineated the helper phenotypes of some of the identified SARS-CoV-2-specific T cell clones, thus providing a use case for the incorporation of our reference scRNA-Seq dataset into research.

Manuscript 3, published in eLife in 2023[2], characterizes the properties of IL-10-producing Th lymphocytes, which promote chronic MCMV infection in mouse salivary glands. The major findings of the study were the uncovered T-bet-dependency of these pathological IL-10-producing

lymphocytes, as well as the coexpression of arginase-1. However, only the TCR repertoire profiling of these cell populations was performed within the scope of this thesis. IL-10-producing Th lymphocytes in salivary glands in MCMV infection had a more prominent TCRα clonal structure and a reduced TCRα and TCRβ diversity compared to IL-10-negative counterparts. Based on the statistical overrepresentation of the groups of similar TCRs, we fingerprinted the signs of an ongoing antigen-driven T cell immune response in salivary glands. However, considering the clonal and cluster overlap between IL-10-producing and Il-10-negative Th lymphocytes, we proposed a probably shared origin and the existence of shared specificities in these two Th subsets.

The discussion section summarizes the achievements and challenges faced in the era of single-cell transcriptomics in immunological research. Furthermore, some updates are given based on the recent works published in the field.

The overview of the thesis structure is shown in **Figure 1**.

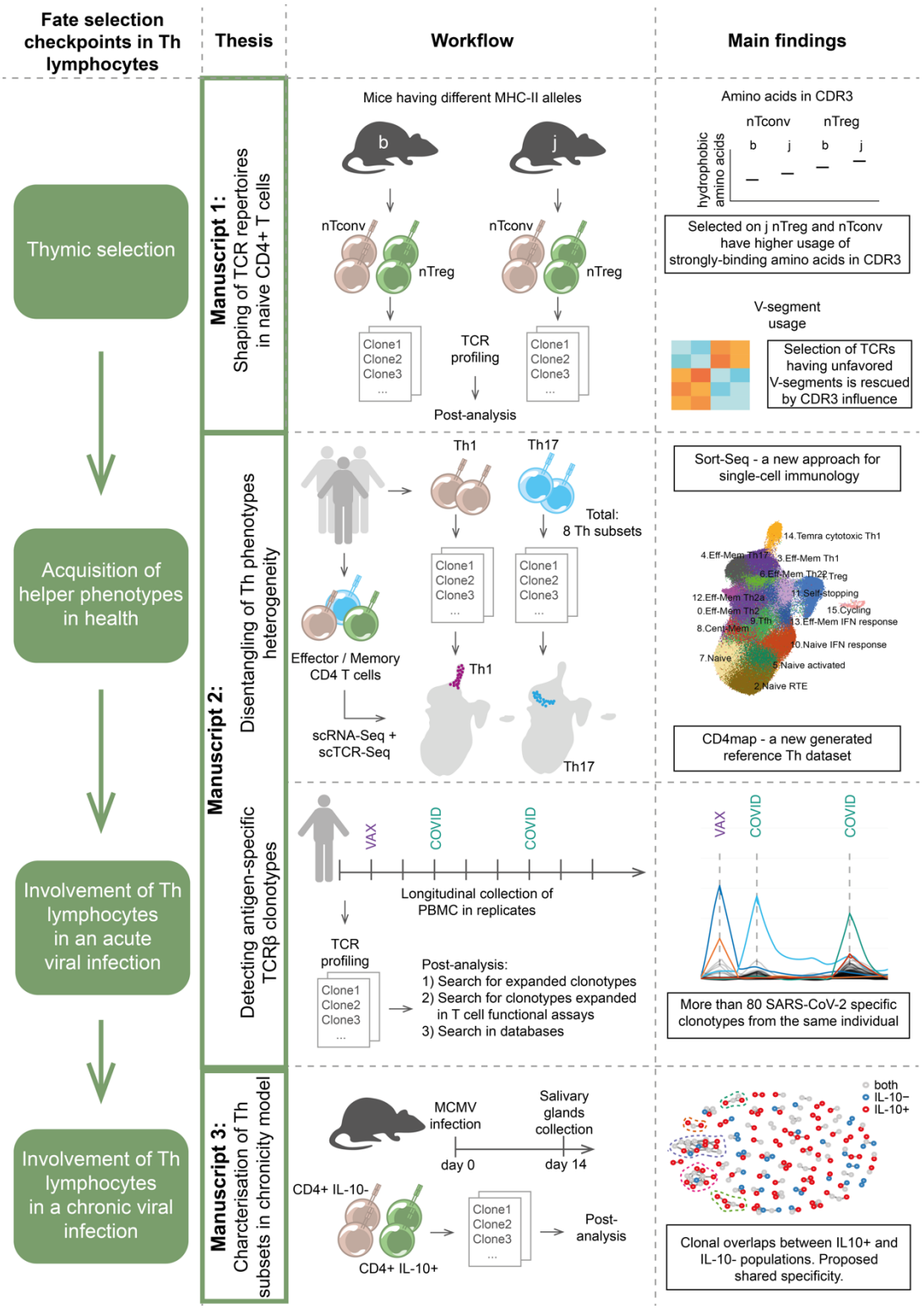

Figure 1. Overview of the thesis. This thesis consists of three manuscripts which describe the helper fate selection at different conditions and stages of Th lymphocyte development.

1.3. Introduction to CD4+ T cell immunobiology

1.3.1. CD4+ T cells as a part of the immune system

Coordinated actions of all cell types maintain our immunity, yet the primary role in defense against pathogens and malignancies is attributed to our immune system. The human immune system has two evolutionary distinct branches known as the innate and adaptive immune systems. The innate immune system is tuned to recognize conservative molecular patterns associated with various groups of pathogens, eliciting various effector programs in response to these threats. However, the innate immune system cannot generate immune memory. Thus, a successful survival in any infection does not arm or prepare the innate immune system for the secondary pathogen encounter in advance. Establishing immune memory is a unique feature of our relatively recently evolved adaptive immune system. This system is comprised of T and B cells, each possessing distinct immune receptors known as TCRs and BCRs, respectively. All mature T and B cells have unique individual TCRs and BCRs, collectively contributing to the diverse TCR and BCR repertoires within the organism. The extensive diversity within the resulting immune repertoires enables the recognition of antigens from various sources, including pathogens, the host organism, dietary compounds, environment, and neoantigens.

B cells recognize their cognate antigens in the form of folded proteins. Once properly activated, B cells may start producing antibodies, which are soluble versions of their BCRs. T cells, in contrast, do not directly recognize intact antigenic proteins but instead require a unique processing and presentation of antigens by other cells. Specifically, TCRs target relatively short antigenic peptides displayed within MHC molecules on the surface of various immune and non-immune cells. All T cells may be divided into two groups based on the surface expression of CD8 or CD4 coreceptors, and the respective restriction to MHC-I or MHC-II complexes. CD8+ T lymphocytes are primarily cytotoxic and govern direct cell killing upon recognition of the cognate peptide-MHC-I (pMHC-I) on the surface of a target cell. This cellular defense mechanism is vital in controlling viral infections and malignant processes.

Conversely, CD4+ T helper (Th) cells usually recognize antigens presented by professional antigen-presenting cells (APCs) and may initiate a plethora of effector programs. This includes the proper priming of innate and adaptive immune cells, coordination of the immune response through cytokine signaling, sustaining high-affinity antibody responses, induction of immune tolerance, and dampening immune reactions. The decisions made in primary immune responses are recorded in the form of memory Th clones, which may be reactivated in the secondary infection and elicit the same

effector function once again. In summary, the multifaceted programs of Th lymphocytes emphasize their central role in complex interactions within the immune system.

1.3.2. Initial TCR repertoire diversity

T cells rely on their TCR repertoire to recognize the vast array of antigens. T cell progenitors develop their TCRs in the thymus through a stochastic yet regulated process known as VDJ recombination. This DNA recombination involves a sequential joining of V-, D- (exclusive to the TCRβ locus), and J-gene segments. In humans, there are (45-47 TRAV x 50 TRAJ) and (40-48 TRBV x 12-13 TRBJ) functional genes[3], which contribute to the combinatorial diversity of TCRs through gene segments rearrangements and α- and β-chain pairing. This diversity is further enhanced through random nucleotide trimming and insertions to the V-D, D-J, and V-J (in the TCRα locus) joints. The resulting mutated joints encode the hypervariable CDR3α and CDR3β loops, pivotal for antigen recognition.

1.3.3. Shaping by MHC

The thymic selection on MHC-I and MHC-II molecules shapes the initial extremely diverse TCR repertoire. Positive selection provides the selection of MHC-restricted TCR repertoires. TCRs primarily recognize MHC molecules by CDR1 and CDR2 loops, which are germline pre-encoded within V-genes[4]. Some MHC allelic variants may exhibit a preference for specific V-segments due to their higher affinity[5]. However, CDR3 also contributes to MHC-II recognition, albeit usually indirectly, by altering the conserved contacts between CDR1, CDR2, and MHC-II[6]. Rescued in the positive selection DP thymocytes on average have higher abundance of hydrophobic amino acids within CDR3 regions of TCRα and TCRβ chains[7], further proving the influence of CDR3 on the positive selection outcome.

The unwanted autoreactivity of the initial TCR repertoire is controlled by negative selection. T cells bearing high-affinity TCRs undergo apoptosis upon recognition of self-antigens, presented by thymic mTECs and B cells[8]. Within the MHC-II-restricted SP T cell population, a subset differentiates into a distinct nTreg lineage already within the thymus. These nTregs are characterized by high-affinity self-reactive TCRs, resulting in heightened signaling strength, slightly lower or comparable to the TCR signaling levels in clonally-deleted SP thymocytes[9] (**Figure 2**). Altogether, these processes collectively establish central immune tolerance.

In the end, post-selection CD4+ T cells exhibit a narrowed TCR repertoire diversity and greater transcriptional and epigenetic heterogeneity compared to the pre-selection thymocytes.

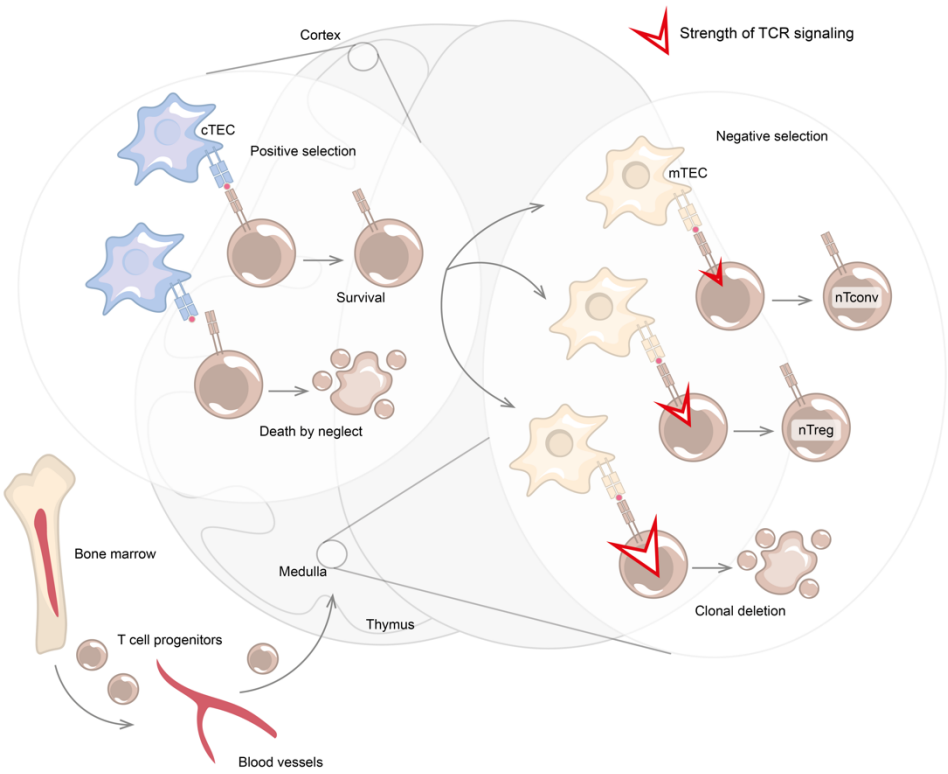

Figure 2. TCR signaling strength in thymic CD4+ T cell selection. T cell progenitors originate in the bone marrow during hematopoiesis. Upon entering the thymus, they undergo recombination of their TCRβ and TCRα loci, start expressing TCR, and probe pMHC complexes presented by cortical thymic epithelial cells (cTECs) during positive selection. At this step, the optimal TCR signaling strength provides survival signals to thymocytes and determines CD4/CD8 lineage commitment. Failure to recognize pMHC-I or pMHC-II induces apoptosis (death by neglect). Survived thymocytes move to the thymic medulla, where they encounter pMHC complexes presented by medullary thymic epithelial cells (mTECs) during negative selection. The strength of TCR signaling in CD4-lineage committed thymocytes leads to either nTconv selection (intermediate signaling strength), nTreg selection (intermediate to high signaling strength), or clonal deletion by apoptosis (signaling strength above a certain threshold).

1.3.4. Naive CD4+ T cells

CD4+ T cells that successfully pass thymic selection exit the thymus to form a pool of naive CD4+ T cells. These cells exhibit distinct traits, including a highly diverse TCR repertoire, a homing to lymphoid tissues (as indicated by the surface expression of CD62L and CCR7), and a dependency on IL-7. Within the population of naive T cells, a specific subset known as recent thymic emigrants (RTEs) exists. Unlike mature naive T cells, RTEs continue to undergo peripheral selection. The

recognition of a cognate antigen by an RTE in the absence of inflammation results in limited RTE proliferation and the induction of clonal anergy[10]. In contrast, low-affinity tonic signals from self-pMHC-II together with IL-7 promote long-term survival[11] and phenotypic shaping[12] of fully mature naive T cells. Collectively, these processes further fine-tune the TCR repertoire of naive T lymphocytes to a particular antigenic landscape of the body, ensuring the establishment and maintenance of the naive T cell pool for many years.

1.3.5. The transition from naive to effectors

The transition from naive to effectors occurs when TCRs recognize cognate pMHC-II complexes presented by properly activated dendritic cells. This transition is marked by rapid proliferation, metabolic switch to glycolysis, and shift in homing receptor expression. Immunologists have traditionally used the latter feature to distinguish effector T cells from naive and central memory T cells (e.g., by CCR7 or CD27 surface staining) (**Table 1**). In addition to these changes, effector CD4+ T cells acquire specific helper phenotypes, determining the mode of host primary and secondary responses to a given antigen. Indeed, the expansion of antigen-specific CD4+ T cell clones with specific Th phenotypes and the subsequent formation of long-living memory clones is a T-cell way to create and preserve immune memory.

1.3.6. Generation of CD4+ T cell immune memory

During the contraction phase of the immune response, the majority of effector antigen-specific T cells undergo apoptosis. However, a subset of these cells acquires a memory phenotype, ensuring their survival. The emergence of memory Th lymphocytes depends on timely development of transcriptionally distinct memory precursors[13].

The Th memory compartment consists of cells exhibiting various levels of effector molecule expression, proliferative potential, and preferences for homing to different tissues (**Table 1**). Effector-memory T cells, for instance, preserve the expression of marker genes associated with different Th programs, circulate through the bloodstream, and essentially patrol our body for potential secondary antigen encounters.

In contrast, central memory T cells are relatively rare in blood and do not express Th-specific effector molecules. We still do not know the precise mechanisms by which human central memory CD4+ T cells retain information about the specific type of immune response to elicit. However, studies in

mouse models[14] and human CD8+ T cell memory[15] hint us that central memory Th cells might still be poised towards a specific helper response fingerprinted in their epigenome.

A distinct subset of the CD4+ T cell memory population is represented by effector-memory T cells re-expressing CD45RA (Temra). Typically arising in response to persistent antigen stimulation, these cells exhibit an extreme effector phenotype, trafficking through blood, and large clonal expansions.

Tissue-resident memory CD4+ T cells (Trm) populate nonlymphoid tissues. These cells share a common origin with peripheral blood effector-memory T cells (Tem), as suggested by clonal overlap. However, once established, Trm possess the capability to self-replenish in a tissue-specific manner, actively participating in controlling secondary infections locally [16].

Collectively, these memory subsets contribute to the long-term antigen-specific Th memory (**Figure 3**).

Table 1. Summarized features of various stages of memory development in CD4+ T cells.

Subset	Homing	Proliferative potential	Effector molecules expression	Longevity	TCR clonality	TCR diversity
Tn	Lymphoid tissue	High	Absent	Long	Extremely low	High
Teff	Blood/ Nonlymphoid tissue	Low	High	Short	High	Intermediate
Tem	Blood	Intermediate	Intermediate	Reasonably long	Intermediate	Intermediate
Tcm	Lymphoid tissue	High	Low/Absent	Long	Low	Intermediate
Trm	Nonlymphoid tissue	Intermediate	Intermediate	Long	High	Intermediate
Temra	Blood	Low/Absent	High	Long	Extremely high	Low

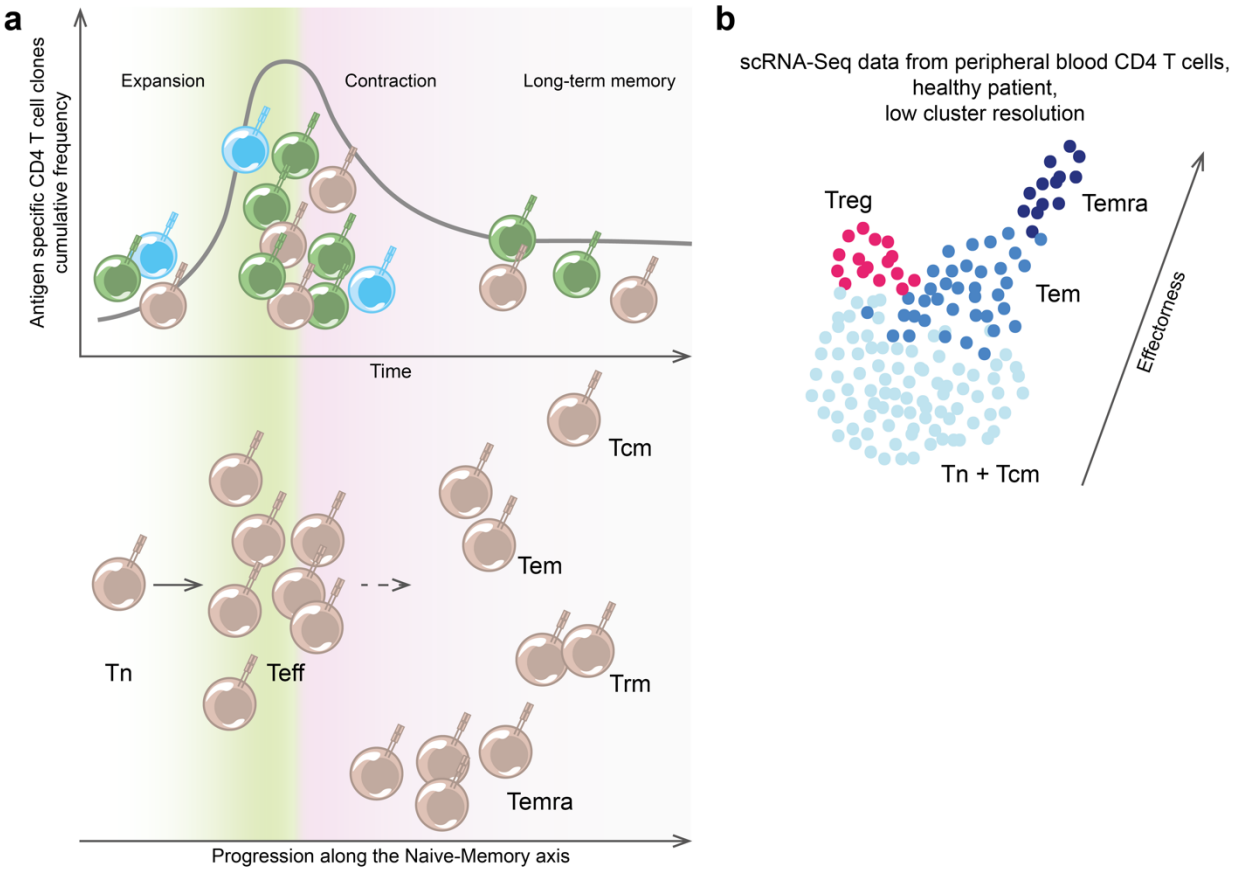

Figure 3. The relationship between the progression of immune response, Th cells differentiation, and scRNA-Seq phenotypes. a) At the moment of a primary immune challenge, naive antigen-specific T cells get primed and clonally proliferate, a stage referred to as expansion. Upon proliferation, antigen-specific T cells differentiate into effector T cells (Teff), and transcriptional and metabolic changes accompany this transition. During the contraction stage, most Teff cells undergo apoptosis and the total number of antigen-specific T cells in peripheral blood drops. The remaining antigen-specific cells in the blood are represented by clones that have successfully established memory phenotypes through transcriptional and metabolic shifts. Peripheral blood memory T cell population consists mainly of effector-memory T cells (Tem), terminally differentiated effector-memory T cells re-expressing CD45RA (Temra), which differentiate in the context of repetitive antigen challenges, and partially from central memory T cells (Tcm), which although preferentially home to lymphoid tissue. A separate population of tissue-resident memory T cells (Trm) forms in tissues. **b)** The majority of gene expression variance in

scRNA-Seq datasets from peripheral Th cells is explained by the effectorness gradient of cells, which is linked to the differentiation from naive to memory states. The UMAP plot is schematically reproduced from Cano-Gamez et al. [17].

1.3.7. Th phenotypes

Cognate antigens for CD4+ T cells are technically just peptides with lengths ranging from 13 to 25 amino acids, presented within host MHC-II complexes. Notably, the exact amino acid sequence of an antigen and the TCR recognition *per se* do not dictate to T cells a proper action mode against particular antigens. In practice, immune responses to antigens originating from pathogens, host proteins, and environmental proteins are expected to differ. Classically, we categorize immune responses into three prototypic types: type I immunity directed against cancers and intracellular pathogens (viruses and intracellular bacteria); type II immunity, aimed at multicellular parasites; and type III response, targeting extracellular microbial pathogens (bacteria and fungi, mainly *Candida albicans*[18]). These responses are primarily promoted by Th1, Th2, and Th17 cells, which express distinct groups of effector molecules associated with the respective types of immune responses. However, the classic paradigm of Th subsets has been further expanded and diversified with the introduction of various novel Th subsets. We will now explore their pivotal contributions to our immunity.

1.3.7.1. Th1

The major effector molecule of Th1 cells is a soluble cytokine IFN-γ. IFN-γ plays a crucial role in enhancing antigen processing and presentation by target cells[19], making stressed tissues more readily detectable by the immune system. In lymph nodes and tertiary lymphoid structures, the influence of IFN-γ directs germinal center antigen-specific B cells predominantly toward the IgG1 class switch[20]. These downstream effects position Th1 cells as significant players in antiviral immunity.

In macrophages, IFN-γ promotes classical M1 activation, facilitating the production of reactive oxygen species (ROS), enabling anti-tumor activity, and the clearance of intracellular bacteria such as *Mycobacteria tuberculosis* and *Listeria monocytogenes*. Additionally, IFN-γ induces the expression of Th1-lineage master transcription factor T-bet in activated T cells. T-bet, in turn, upregulates the

CXCR3 chemokine receptor, its ligand CXCL9, and IFN-γ itself. Thus, IFN-γ – CXCL9 – CXCR3 axis favors the exclusive generation and recruitment of Th1 cells, promoting the type I immune response.

1.3.7.2. Th2 and Th2a

Th2 cells play a prominent role in anti-helminth immunity and defend against biting parasites like ticks. These cells develop under the control of the GATA3 master transcription factor and primarily exert their effects through the expression of IL-4, IL-5, and IL-13. IL-5 recruits eosinophils, which are specialized innate anti-helminth effector cells. In contrast, IL-4 and IL-13 convert macrophages into the M2 activation state, facilitating tissue remodeling and wound repair following macroscopic tissue damage. As a result, the major Th2 cytokines regulate anti-inflammatory and highly specialized inflammatory responses.

In B cells, IL-4 drives the IgE class switch. Mast cells and basophils bind IgE antibodies by FcεR1 and release histamines in an antigen-specific way. In turn, these histamines trigger itching reactions, signaling our central nervous system to take action and physically remove a skin parasite.

Unfortunately, type II immunity drives IgE-dependent hypersensitivities, commonly known as allergies. Such pathological allergen-specific responses are attributed to a novel Th2a subset[21], which is phenotypically distinct from conventional Th2 cells. Th2a cells express high levels of prostaglandin D2 synthase and the prostaglandin D2 receptor CRTH2. Together, classical Th2 and Th2a cytokines fuel a positive feedback loop and promote Type II responses, imposing many challenges for allergen immunotherapy[22].

1.3.7.3. Th17

Extracellular pathogens such as *Candida albicans*[18], *Staphylococcus aureus*[23], and *Bordetella pertussis* are classified as targets for Th17-driven type III immunity. The primary effector molecule of Th17 cells is the proinflammatory cytokine IL-17. Th17 also produce IL-22, which stimulates the production of antimicrobial peptides in epithelial cells. When treated with Th17-derived cytokines, epithelial cells also produce molecules attracting neutrophils to the site of the immune response. Neutrophils, in turn, constrain, trap or kill the extracellular pathogen through phagocytosis, production of ROS, netosis, and the release of granule content. One such molecule, calprotectin, is the major biomarker of gut inflammation and inflammatory bowel disease (IBD)[24]. The effector functions of neutrophils amplify inflammation dramatically by releasing various danger-associated

signals (DAMPs). Unfortunately, similar events may lead to autoimmunity when hardwired to host antigens. Th17 are associated with conditions like rheumatoid arthritis[25], ANCA-glomerulonephritis[26], multiple sclerosis, and experimental autoimmune encephalomyelitis[27].

1.3.7.4. Th22

Th22 cell subset is often considered to be closely related to Th17 cells. Indeed, both subsets may produce IL-22 involved in tissue remodeling response and mucosal immune regulation[28]. This provokes the debates on the origin of Th22 and Th17 cells. The reports in humans[29] and transgenic mice[30,31] show that, at least partially, Th22 cells develop independently.

1.3.7.5. Th1-17

Th1-17 cells phenotypically combine features of Th1 and Th17 cells. Namely, they express T-bet and ROR-gt transcription factors, CCR6 and CXCR3 chemokine receptors, and produce IFN-γ and IL-17 cytokines upon activation. The pathogenic role of skin resident Th1-17 cells is known in psoriasis[32]. The dual nature of Th1-17 cells provokes questions about whether they originate from Th1 cells, Th17 cells, or have a completely separate origin. The clonal relations of peripheral blood Th1-17 cells to Th1 and Th17 cells studied by Kasatskaya et al. proposed a close relationship between Th1 and Th1-17[33].

1.3.7.6. Treg

Treg cells play a pivotal role in establishing immune tolerance broadly. The divergence of nTreg early in thymic selection represents a mechanism for establishing central immune tolerance. Additionally, naive T cells have the capacity to convert into induced Tregs (iTregs) in the periphery under the TGFβ and IL-10 influence. These iTregs are responsible for sustaining symbiotic relationships with microbiota, inducing tolerance to peripheral host antigens, and dampening the immune response prior to pathogen clearance[34] or following it[35], or at the transition to chronicity[36]. Both nTreg and iTreg develop under the control of the FOXP3 master transcription factor and primarily exert their regulatory functions in an antigen-specific CTLA-4 dependent manner, or more broadly in via CD25 and IL-10. The TCR repertoires of both nTreg and effector Tregs are characterized by increased usage of hydrophobic amino acids[37,38]. These hydrophobic amino acids within the CDR3 region are considered to facilitate high-affinity pMHC-II recognition, which may mechanistically explain the stronger TCR signaling observed in Treg cells[39] and, eventually, their ability to suppress antigen-specific immune responses.

1.3.7.7. Tfh and Tfr

In contrast to T cells, antigen-specific B cells can undergo somatic hypermutation of their BCRs, aiming to refine BCR specificity. In the T-cell-dependent responses, the selection of successfully hypermutated high-affinity BCRs is essentially controlled by Tfh cells in the process of BCR affinity maturation in germinal centers. The key signal provided by Tfh cells is IL-21, which promotes the selection of high-affinity BCRs in germinal center reactions[40] and the development of antibody-secreting plasma cells[41]. Currently, various cytokine-skewed Tfh cells are recognized[42]. Among them, Tfr lymphocytes govern the negative selection of autoreactive B cells. Tfr cells are supposed to have a repertoire of specificities to various host proteins, including nuclear antigens[43,44]. Together, Tfh and Tfr cells shape the antibody repertoires in T-cell-dependent B cell responses towards higher specificity and lower self-reactivity.

1.3.7.8. Cytotoxic CD4+

Surprisingly, with the advancement of scRNA-Seq techniques, researchers have shed light on the importance of cytotoxic CD4+ T cells, which were previously underappreciated. Since then, the accumulation of cytotoxic CD4+ T cells in peripheral blood and lymphoid tissue has been associated with human[45] and mouse[46] aging, while skin-resident cytotoxic CD4+ T cells have been directly implicated in senescent cells killing[47]. Many cytotoxic CD4+ T cells show significant clonal expansions, exhibit a prominent effector phenotype, and express combinations of granzymes, perforin, and granulysin, which makes them transcriptionally close to classical cytotoxic CD8+ cells.

1.4. Antiviral immune response

1.4.1. Innate immunity in viral infection

The antiviral immune response largely depends on both innate and adaptive immunity. The first line of defense against respiratory, enteric, and sexually transmitted viruses is represented by epithelial cells. They express many innate sensors for viral recognition (TLRs, RLRs) and even for alarming neighboring cells about viral infection (cGAS-STING pathway[48]). Epithelial borders are constantly monitored by dendritic cells, armed with an even wider array of innate viral-recognition receptors. For example, they express endosomal TLR-7 and TLR-9 in addition to common cytoplasmic TLRs. Any primary adaptive antiviral immune response crucially depends on proper dendritic cell activation.

NK cells are highly specialized warriors in innate antiviral immunity. They distinguish abnormal or infected cells by recognizing surface stress molecules or reduced MHC-I expression. Once reaching

the activation barrier, NK cells are capable of direct killing through the precise delivery of granzymes A and B[49].

Additionally, various mechanisms of cellular antiviral defense, mainly global RNA degradation and translational arrest, are triggered by type I IFNs. These events lead to the block of virus replication and viral protein synthesis[50]. To conclude, the innate immune system is a reliable mediator of antiviral defense; however, it lacks immune memory, which means it cannot recognize the same pathogen more rapidly upon the secondary infection.

1.4.2. Adaptive immunity in viral infection

The generation of immune memory is a classical feature of adaptive immunity. After primary infection, T and B cells may form long-living memory clones or sustain long-lasting antibody production. In the case of a secondary encounter with the same pathogen, serum antibodies may immediately recognize the antigen and memory T and B cells may undergo rapid reactivation and engage in clonal proliferation, effector response, or germinal center reactions.

Adaptive antiviral immune response largely depends on cytotoxic CD8+ T cells capable of direct antigen-specific killing of the infected cells presenting viral antigens within the MHC-I complex. Like NK cells, CD8+ T cells directly deliver granzymes and induce apoptosis in cells. Viral envelope proteins may also cover the surface of cells infected by enveloped viruses and may become targets for antibody recognition[51]. Thus, IgG1 antibodies mark infected cells and navigate NK cells for antibody-mediated cytotoxicity. Antibodies that may bind viral proteins on the surface of virions and prevent viral entry into host cells are called neutralizing antibodies, and their successful generation upon vaccination[52] or timely production in viral infection[53] may be crucial for adequate immune protection. To conclude, both CD8+ cytotoxic and antibody responses are essential in antiviral immune response. However, the proper CD8+ T cells priming[54], successful generation of CD8+ memory clones[55], and the production of high-affinity virus-specific antibodies[56] are impossible without CD4+ T cell help.

1.4.3. Immune response in chronic versus acute viral infection

As a result of the balanced and fine-tuned actions of our immune system, viral infections usually get resolved. In such acute infections, antigens are only available for a short period, and there is a reasonably long window when memory clones survive without antigenic challenge. However, in some cases, viral clearance cannot be achieved. For example, infection with HIV, HBV, HCV, EBV, and CMV in humans, and infection with MCMV and LCMV in mice may develop into chronicity[57]. Chronic

infections are characterized by repetitive and prolonged antigen exposure to immune cells, which inevitably leads to the shaping of the antiviral immune response. In antigen-specific T cells, such conditions lead to phenotypic alterations and the development of T cell exhaustion, marked by the expression of inhibitory molecules PD-1, TIM-3, LAG-3[57]. In chronic CMV, EBV, and LCMV infections, T cells may develop large antigen-specific clones, reaching up to 5-10% of the total TCR repertoire, and may have a cytotoxic terminally differentiated phenotype[58–60]. Besides, virus-specific CD4+ T cells may also acquire an IL-21-producing phenotype, as was shown in the chronic LCMV infection model[61], IL-21 sustains CD8+ T cell responses in LCMV chronicity[62]. In addition, it probably allows for constant help for germinal center B cells and the unstoppable production of virus-specific antibodies, which may cover even escape-variants[63,64].

Although immune responses to chronic viral infections have some common features, they vary greatly depending on a particular virus. In this thesis, we study human adaptive T-cell responses to acute SARS-CoV-2 and chronic MCMV infections in mouse salivary glands. We exploited the natural expansion and contraction of antigen-specific T cell clones in acute infection to identify a large set of SARS-CoV-2 specific TCRβ clonotypes. In chronic MCMV infection, we described TCR repertoires of IL-10-producing CD4+ T cells that promote MCMV persistence in salivary glands for months after infection[65].

In summary, the antiviral immune response is a complex immune system reaction involving many innate and adaptive cells. However, in this thesis, the primary focus is on helper T cells. Th cells not only govern the variety of immune responses but also serve as a tool to memorize the, hopefully, appropriate mode of response to a particular pathogen in a highly antigen-specific way. Throughout our lives, we accurately collect and store this valuable information in our blood and lymphatics to be fully armed upon the secondary antigen encounter.

1.5. NGS in immunology

1.5.1. TCR repertoire sequencing

The invention of NGS made the sequencing of high-diversity libraries truly high-throughput. Researchers in T cell biology adopted NGS to profile TCR repertoires. Unlike B cells, T cells do not experience somatic hypermutations and thus do not accumulate mutations in the V-regions of their immune receptors. Remarkably, germline pre-encoded TCR V-segments have allelic variability in population[66], and sometimes germline variants may even determine the TCR specificity[67]. Nevertheless, the cumulative effect of the naive TCR repertoire shaping by MHC and the altered

immunogenicity of antigenic peptides in various MHC contexts are supposed to contribute much more to the involvement of particular naive clones in the immune response. Thus, if we admit the use of publicly available TCR V-segment and J-segment references, the task of TCR-Seq mainly narrows down to covering CDR3α and CDR3β, and a CDR3-adjacent piece of V-segment for the inference of the used V-gene. Illumina short-read sequencing may easily cover this region, which is preferred among other NGS techniques due to its high accuracy, crucial for errorless coverage of the CDR3 region. Various molecular biology tools like Rapid Amplification of cDNA Ends (RACE) and Unique Molecular Identifiers (UMIs) are utilized in bulk TCR-Seq, in scTCR-Seq, and scRNA-Seq assays and will be discussed together further.

1.5.2. Single-cell technologies

Recent years have become the era of single-cell transcriptomics and, later, other single-cell omics techniques, such as profiling of surface protein expression (CITE-Seq[68]), immune repertoires (scTCR-Seq, scBCR-Seq), chromatin accessibility (scATAC-Seq[69]), histone modifications (scCUT&TAG[70]), and RNA-expression phenotypes in pooled CRISPR-screens[71]. Some of the technologies reach the single-cell level of transcriptome profiling by encapsulating the barcode-ligation reaction within individual cells[72], others use single-cell FACS sorting into individual plate wells[73], which is especially applicable for low-input assays. However, the most popular approach allowing simultaneous profiling of a substantial number of input cells is introduced by 10x genomics and is based on the cDNA synthesis from single cells in droplets. In general, suspensions of fresh viable cells, methanol-fixed cells, or cell nuclei might be used in the assay. Then, cells in a cDNA synthesis reaction mix are loaded through tiny channels into the partitioning oil with oligonucleotide-coated gel beads to generate an emulsion. The conditions favor the generation of droplets containing no more than one cell and one gel bead. Oligos coating the beads include UMIs, cell barcodes, and, depending on the protocol used, a sequence to capture either the poly-A mRNA molecules or the first cDNA strands.

1.5.3. Key features of scRNA-Seq, scTCR-Seq, and bulk TCR-Seq libraries design

1.5.3.1. 3'-RACE versus 5'-RACE

All 10x genomics scRNA-Seq protocols are powered by either 3'-RACE or 5'-RACE technology. 3'-RACE-based assays capture individual mRNA molecules and prime the first strand synthesis with poly-dT containing gel-bead-oligos. Ultimately, 3'-RACE scRNA-Seq libraries consist of 3'-ends of the transcripts, making simultaneous immune repertoire profiling challenging. Though, assuming the successful synthesis of full-length cDNA molecules, immune profiling from 3'-RACE-generated cDNA libraries is indeed possible with the help of V-multiplex primers[74]. However, 5'-RACE technology may

be much easier adopted for immune receptor profiling, and in this study, we have only performed 5'-RACE-based scRNA-Seq assays. In this assay version, the cDNA synthesis reaction is primed by free poly-dT oligos within the reaction mix. The MMLV-derived reverse transcriptase used in the assay has an intrinsic terminal deoxynucleotide transferase activity and adds deoxycytidine stretches to the 3'-end of the first cDNA strand (**Figure 4**). At this point, the gel bead oligos containing three adjacent riboguanosines serve as template-switch oligos for the reverse transcription enzyme and allow the 3'-end complementary cDNA extension. The fully constructed 5'-RACE cDNA library contains a universal primer annealing, which supports the amplification of the CDR3-containing region from a universal primer and a TCR constant region-specific primer. Bulk TCR-Seq may also be based on the 5'-RACE procedure. Unlike V-multiplex-based assay design, 5'-RACE provides a fully unbiased and easy-to-establish solution for TCR-Seq in any model and non-model[75] species. The library design strategies applied in this study are summarized in **Figure 4.**

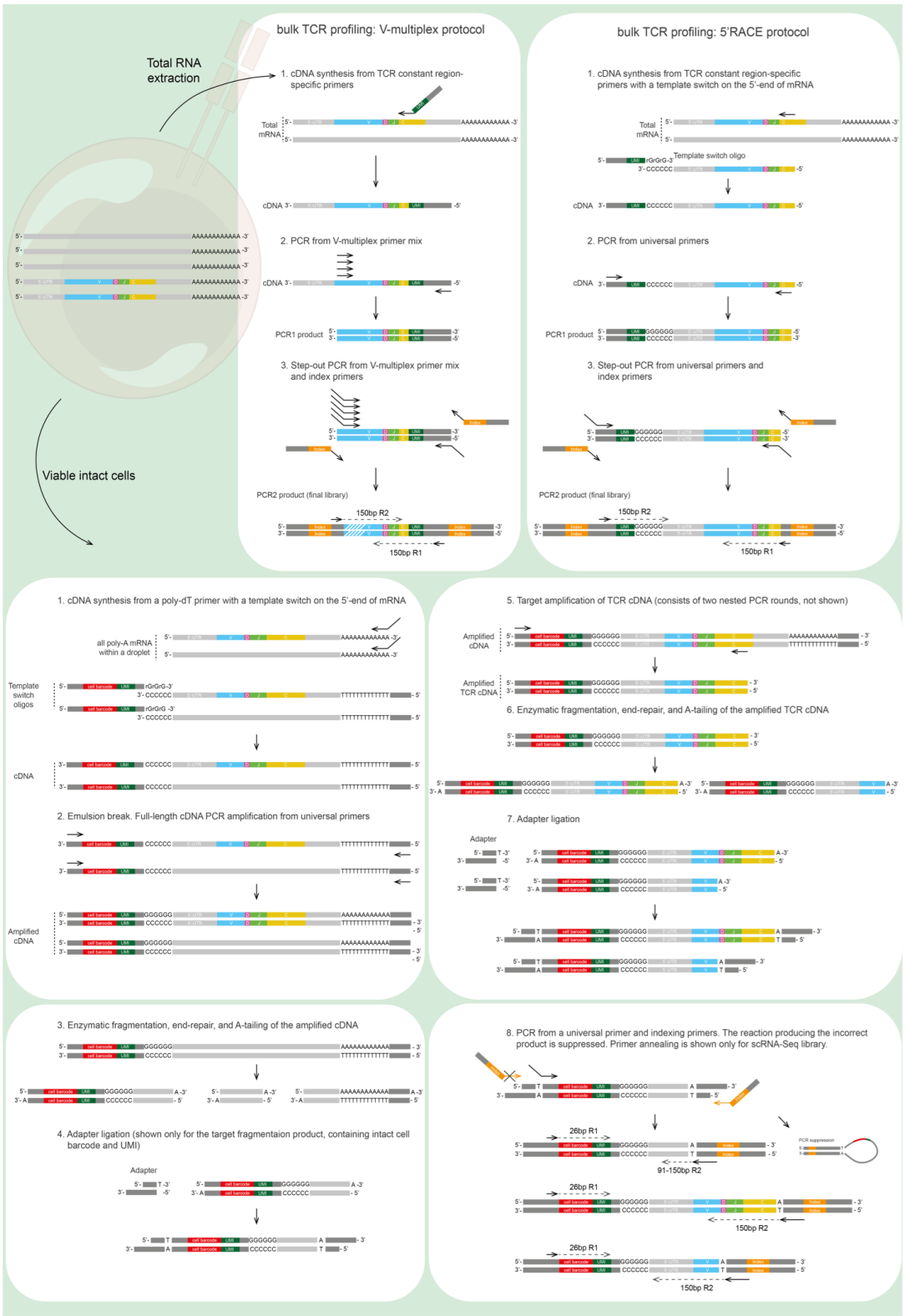

Figure 4. Strategies for bulk TCR-Seq, RNA-Seq, and scRNA-Seq libraries preparation used in this study.

1.5.3.2. UMI

Besides poly-dT or poly-rG required for the 3'- or 5'-RACE, 10x genomics gel beads contain a 10-nucleotide-long UMI and a 16-nucleotide-long cell barcode. UMIs are unique for each oligo and allow an unbiased count of each captured transcript, and further normalization of the gene expression between the cells with various RNA content. UMIs are also utilized in many available bulk TCR-seq protocols. In the TCR repertoire studies UMIs help to calculate the precise clonotype proportions in the repertoire, normalize samples to the same depth, and even to perform the correction of PCR and sequencing errors. The comparison of various TCR-Seq protocols used in this study is given in **Table 2**.

Table 2. Features of TCR-Seq protocols used in this thesis.

	UMI-based 5'-RACE TCR-Seq	**UMI-based multiplex TCR-Seq**	**10x genomics scTCR-Seq**
Covered sequence	CDR3 and a part of FR3	CDR3 and a part of FR3	Full-length
UMI	Yes	Yes	Yes
Starting material	RNA	RNA	RNA in viable cells
Forward PCR primers anneal to	Universal sequence from template switch oligo	V-segments	Universal sequence from template switch oligo
Reverse PCR primers anneal to	C-region	C-region	C-region

1.5.4. Data analysis

Data analysis of scRNA-Seq, scTCR-Seq, and bulk TCR-Seq starts with aligning reads to references. For scRNA-Seq, we use 10x genomics CellRanger software, which performs the alignment to a reference human genome and transcriptome, UMI correction, cell calling, and quality control. CellRanger also processes raw scTCR-Seq data via full-length transcript sequence assembly, alignment to human T cell receptor loci reference genes, and paired TCR clones calling. To analyze raw bulk TCR-Seq reads, we run MiXCR software. MiXCR, similar to CellRanger, initiates a pipeline containing alignment, UMI-correction, UMI-based correction of erroneous bases in reads, clonotype

export, and quality control steps. Later, we perform some additional filtering steps. For example, we select the most abundant TCRα and TCRβ within a cell in scTCR-seq. We also filter and keep only functional clonotypes without frameshifts or stop-codons within CDR3, having the conservative cysteine and phenylalanine in the first and last amino acid position in CDR3. We perform post-analysis using R, owing to the excellent infrastructure of general purpose and specialized bioinformatic packages available. I want to acknowledge the help of the tidyverse collection of R packages[76] for my research. Among all bioinformatic R libraries used, notable credits should be given to Seurat[77], which allows scRNA-Seq post-analysis, unified experiment data and metadata storage, integration of samples, batch-effects correction, and provides a framework for reference mapping of cells. We have created an R package (https://github.com/kriukovav/CD4map) wrapping the latter feature of Seurat for convenient use together with the Th reference dataset generated in **Manuscript 2**. To discover reproducibly expanded in replicates TCR clones, we used the edgeR package[78] available through Bioconductor. To identify overrepresented convergent clonotypes associated with any ongoing immune response, we apply an R package tcrgrapher (https://github.com/KseniaMIPT/tcrgrapher/) created by my colleague and powered by an advanced version of ALICE algorithm[79].

2. Results

2.1. Publications and main findings

The section Results consists of three manuscripts, the content of which is summarized here.

Manuscript 1: MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naive CD4+ T cells (Logunova*, Kriukova* et al*. *PNAS* 2020) *Authors contributed equally to this work.

This study describes the TCR repertoire features of two Th populations, namely naive Treg (CD4+ Foxp3+) and Tconv (CD4+ Foxp3-) lymphocytes selected in different MHC-II genetic contexts. I observed how the T cell selection on different MHC-II allelic variants favors distinct TRAV and TRBV gene segments. Additionally, I observed the shaping of CDR3 repertoires, with the abundance of strongly-binding amino acids in the central region of CDR3 correlated with selection on H2-A^j allelic variant, or the selection of nTregs.

Manuscript 2: Sort-Seq: immune repertoire-based scRNA-Seq systematization (Kriukova et al., submitted to *Nature Communications;* preprint https://doi.org/10.6084/m9.figshare.24428152.v1**)**

In this work I continued to study features of Th lymphocytes, with a greater emphasis on effector-memory Th populations. I performed scRNA-Seq profiling of peripheral blood CD4+ effector-memory T cells, and developed the Sort-Seq approach for the precise annotation of various Th subsets in scRNA-Seq data. Using Sort-Seq, I contributed to the building of an integrated scRNA-Seq reference dataset of peripheral blood CD4+ T lymphocytes. Additionally, I tracked peripheral blood TCR repertoires in a single donor, who experienced anti-SARS-CoV-2 vaccination and two breakthrough infections. I identified more than 80 SARS-CoV-2-specifc TCRβ clonotypes from this single donor. Finally, I applied mapping to our generated reference dataset to reveal the phenotypes of some SARS-CoV-2-specifc Th cells.

Manuscript 3: Inhibitory IL-10-producing CD4+ T cells are T-bet-dependent and facilitate cytomegalovirus persistence via coexpression of arginase-1 (Clement et al. *eLife* 2023)

In this study I described the TCR repertoire features in IL-10-producing Th lymphocytes promoting chronic MCMV infection in mouse salivary glands. I observed an increased clonality and decreased diversity in TCR repertoires of IL-10-producing Th lymphocytes, which could be explained by their involvement in a focused antigen-driven immune response. However, I also found clonal overlaps between IL-10-producing and IL-10-negative subsets, which suggested the existence of shared specificities in these two populations.

2.2. Selection of TCR repertoires in naive CD4+ T lymphocytes

2.2.1. Manuscript 1: MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4+ T cells

MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4$^+$ T cells

Nadezhda N. Logunova[a,1], Valeriia V. Kriukova[b,c,1], Pavel V. Shelyakin[b], Evgeny S. Egorov[b,d], Alina Pereverzeva[b], Nina G. Bozhanova[e,f], Mikhail Shugay[b,c,d], Dmitrii S. Shcherbinin[d], Mikhail V. Pogorelyy[b,d], Ekaterina M. Merzlyak[b,d], Vasiliy N. Zubov[b], Jens Meiler[e,f,g], Dmitriy M. Chudakov[b,c,d], Alexander S. Apt[a,2,3], and Olga V. Britanova[b,2,3]

[a]Laboratory for Immunogenetics, Central Institute for Tuberculosis, Moscow, 107564, Russia; [b]Department of Genomics of Adaptive Immunity, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, 117997, Russia; [c]Center of Life Sciences, Skoltech, Moscow, 143026, Russia; [d]Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, 117997, Russia; [e]Center for Structural Biology, Vanderbilt University, Nashville, TN 37235; [f]Department of Chemistry, Vanderbilt University, Nashville, TN 37235; and [g]Institute for Drug Discovery, Leipzig University, Leipzig, SAC 04103, Germany

Edited by Philippa Marrack, National Jewish Health, Denver, CO, and approved April 27, 2020 (received for review February 21, 2020)

T cell maturation and activation depend upon T cell receptor (TCR) interactions with a wide variety of antigenic peptides displayed in a given major histocompatibility complex (MHC) context. Complementarity-determining region 3 (CDR3) is the most variable part of the TCRα and -β chains, which govern interactions with peptide–MHC complexes. However, it remains unclear how the CDR3 landscape is shaped by individual MHC context during thymic selection of naïve T cells. We established two mouse strains carrying distinct allelic variants of *H2-A* and analyzed thymic and peripheral production and TCR repertoires of naïve conventional CD4$^+$ T (T_{conv}) and naïve regulatory CD4$^+$ T (T_{reg}) cells. Compared with tuberculosis-resistant C57BL/6 (H2-A^b) mice, the tuberculosis-susceptible H2-A^j mice had fewer CD4$^+$ T cells of both subsets in the thymus. In the periphery, this deficiency was only apparent for T_{conv} and was compensated for by peripheral reconstitution for T_{reg}. We show that H2-A^j favors selection of a narrower and more convergent repertoire with more hydrophobic and strongly interacting amino acid residues in the middle of CDR3α and CDR3β, suggesting more stringent selection against a narrower peptide–MHC-II context. H2-A^j and H2-A^b mice have prominent reciprocal differences in CDR3α and CDR3β features, probably reflecting distinct modes of TCR fitting to MHC-II variants. These data reveal the mechanics and extent of how MHC-II shapes the naïve CD4$^+$ T cell CDR3 landscape, which essentially defines adaptive response to infections and self-antigens.

TCR repertoire landscape | MHC-II | naïve CD4$^+$ T cells | regulatory T cells

The interaction of peptide–major histocompatibility complex (p-MHC) with T cell receptors (TCRs) plays a central role in positive and negative selection of T lymphocytes in the thymus as well as subsequent homeostasis of naïve, primed, and effector-memory T cells in the periphery (1). Even subtle shifts in p-MHC–TCR interactions may profoundly affect T cell responses (2–4) and in extreme cases, can result in immunological disorders (5–7). The theoretical diversity of TCRα/β variants initially produced by recombination in the thymus exceeds 10^{15} for mice (8) and 10^{19} for humans [per our current estimate (9)]. However, not all TCRs effectively interact with p-MHC; only ∼5% of T cells successfully pass through positive selection in the thymus, and TCR repertoires are further narrowed by negative selection (reviewed in ref. 10). Selection continues in the periphery, where recent thymic emigrants acquire the functional properties of mature naïve T cells—which are only capable of providing an antigen-specific response—after exhaustive screening against self p-MHCs (11, 12), enforcing MHC restriction. Subsequently, tonic TCR signaling—induced by interaction with self p-MHC—supports long-term survival of mature naïve T cells (13). Thus, the individual repertoire of naïve TCRs is strongly shaped by self p-MHC complexes, which determine the allowed range of affinities and angles of interaction (4, 14, 15). The resulting individual diversity of a functional TCRα/β repertoire gains about 2×10^6 TCRα/β variants per 2×10^7 cells in a mouse spleen (16). For a human, individual naïve TCRα/β diversity may reach 10^8 variants (17).

Binding of TCRα and -β chains to the p-MHC-II complex is largely determined by their complementarity-determining regions (CDRs). CDR1 and CDR2, encoded by a set of germline T cell receptor variable α (*TRAV*) and T cell receptor variable β (*TRBV*) genes, primarily make contact with α-helical domains of MHC-II, while CDR3, which originates from the hypervariable combination of variable, diversity and joining (V-(D)-J) segments, predominantly contacts antigenic peptides docked within the MHC-II grove (18). It is firmly established that the usage of TCRα and TCRβ V genes depends upon *MHC* allelic variations

Significance

The landscape of naïve T cell receptors (TCRs) essentially determines the ability of the host to respond to vaccination, infection, and cancer and to control autoimmunity and inflammation. The diversity of naïve T cells is generated by semirandom recombination, followed by tight shaping during thymic selection against individual pools of self-peptide major histocompatibility complexes (MHCs). We characterize TCR repertoires of naïve conventional and regulatory CD4$^+$ T cells produced in two different MHC-II contexts. We reveal profound differences in the diversity, convergence, and physicochemical properties of their antigen-interacting regions that indicate the distinct mode of TCR interaction with peptide–MHC-II. Such differences are likely to influence individual susceptibility to infections and autoimmunity and have implications for the future development of TCR-based diagnostics and therapies.

Author contributions: N.N.L., D.M.C., A.S.A., and O.V.B. designed research; N.N.L., V.V.K., E.S.E., A.P., and E.M.M. performed research; N.G.B., M.S., D.S.S., M.V.P., and J.M. contributed new reagents/analytic tools; V.V.K., P.V.S., and V.N.Z. analyzed data; and N.N.L., V.V.K., N.G.B., D.S.S., D.M.C., A.S.A., and O.V.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: The sequence data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession no. PRJNA600142).

[1]N.N.L. and V.V.K. contributed equally to this work.

[2]A.S.A. and O.V.B. contributed equally to this work.

[3]To whom correspondence may be addressed. Email: alexapt0151@gmail.com or olbritan@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2003170117/-/DCSupplemental.

First published June 1, 2020.

(19–21). However, there is still a substantial gap in our understanding of how allelic variability in the MHC Class II locus shapes the intrinsic properties of naïve TCRα and TCRβ CDR3 repertoires. It is also unclear whether these effects differ substantially for conventional $CD4^+$ T (T_{conv}) and regulatory $CD4^+$ T (T_{reg}) cells, for which the thymic and peripheral selection process is thought to differ profoundly (22, 23). Nonsynonymous amino acid substitutions within the binding groove of an MHC-II molecule would be predicted to profoundly affect CDR3 repertoires. Such substitutions may alter the surface of an MHC-II molecule involved in interaction with the TCR, the conformation of antigenic peptides, and the whole repertoire of presented peptides, thus affecting TCR binding and T cell activation (24, 25).

Previously, we demonstrated that the rare MHC-II allelic variant *H2-A*^{j} is associated with increased susceptibility to tuberculosis infection (26). Both chains of H2-A^j have a number of potentially important substitutions located in α-helices that form the peptide binding groove. To investigate this association, we established a panel of *H2*-congenic recombinant mouse strains bearing different segments of the *H2* region from tuberculosis-susceptible I/St (*H2*^{j}) mice on the genetic background of the tuberculosis-resistant C57BL/6 (B6, *H2*^{b}) strain. One of the strains from this B6.I-*H2* panel, B6.I-9.3, differs from the B6 parent by the *H2-A* allele of the classical *H2* gene complex, which bears genetic material of I/St origin within the 30.90- to 34.34-Mb interval of chromosome 17. Both B6 and B6.I-9.3 are H2-E–negative strains; thus, the H2-A molecule is the only classical MHC-II product potentially influencing $CD4^+$ T cell repertoires in B6 and B6.I-9.3 mice. B6.I-9.3 and B6 mice also differ by an unusual allelic polymorphism in the nonclassical MHC-II *H2-Oa* and *H2-Ob* genes of the *H2*^{j} haplotype (26, 27), which are involved in peptide loading onto MHC-II (28, 29). These features make the B6 and B6.I-9.3 pair an attractive tool for studying the influence of MHC-II allelic variations on T cell differentiation and maturation in general and TCR repertoires in particular.

Here, we report that mice of the B6.I-9.3 (*H2-A*^{j}) strain differ from their congenic B6 (*H2-A*^{b}) counterparts by a profound deficiency of both T_{conv} and T_{reg} cells in the thymus. The T_{reg} population in H2-A^j animals was restored in the periphery, but T_{conv} cells remained deficient. Using deep TCR repertoire profiling, we evaluated the impact of the MHC-II polymorphism on the formation of TCRα and TCRβ CDR3 repertoires in naïve conventional $CD4^+$ T (nT_{conv}) and naïve regulatory $CD4^+$ T (nT_{reg}) cells. We found that the *H2-A*^{j} MHC context produces a restricted TCR repertoire with a higher level of sequence convergence and increased MHC-restricted publicity. A relatively higher number of homologous CDR3 variants formed a denser network of larger TCR clusters. Furthermore, we show that this MHC-II locus polymorphism dramatically altered key characteristics of both CDR3α and CDR3β repertoires. The potential strength of both CDR3α and CDR3β interactions with p-MHC-II complexes increases in the *H2-A*^{j} context—presumably compensating for the hampered $CD4^+$ T cell selection—and this feature was much more prominent for the nT_{reg} cells. Taken together, our results demonstrate that the structure and composition of the TCR CDR3α and CDR3β repertoires may be prominently shaped by allelic polymorphisms in the MHC-II gene complex, with notable differences between T_{conv} and T_{reg} cells that should essentially influence individual susceptibility to infectious and autoimmune diseases and cancer.

Results

Selection of $CD4^+$ T Lymphocytes Is Altered in *H2-A*^{j} Context. The I/St-derived *H2* segment present in B6.I-9.3 mice contains not just the two genes encoding the H2-A dimer but also, a few additional proximal genes—including the nonclassical MHC-II genes *H2-Ma*, *H2-Mb1*, *H2-Mb2*, *H2-Oa*, and *H2-Ob* (26). We therefore added the parental I/St strain to our analysis as a control that allows separation of bona fide *H2-A*^{j}–controlled phenotypes from possible effects of allelic variants in other genes located in the proximal part of the relevant chromosomal region. Furthermore, unlike B6 and B6.I-9.3 congenic mice, I/St mice express the second MHC-II molecule, H2-E, which also participates in $CD4^+$ cell selection.

Flow cytometry analysis of single-positive (SP) $CD4^+$ and $CD8^+$ cell populations recovered from the thymi of B6.I-9.3 (*H2-A*^{j}) mice demonstrated a profound reduction in the proportion of $CD4^+$ cells compared with parental mouse strains B6 (*H2-A*^{b}) and I/St (*H2-A*^{j} and *H2-E*^{j}), producing an unusually inverted CD4/CD8 ratio (Fig. 1*A*). The total cellularity of B6 and B6.I-9.3 thymi was similar ($207 \pm 5.4 \times 10^6$ and $219 \pm 33 \times 10^6$, respectively; $P > 0.05$) but was significantly higher compared with parental I/St mice ($144 \pm 12 \times 10^6$; $P < 0.005$). Both positive and negative selection may contribute to an unusual CD4/CD8 ratio in the thymi of B6.I-9.3 mice. We first assessed possible alterations in positive selection by measuring the expression of activation markers CD69 and CD5 in double-positive (DP) thymocytes, both of which are known to be up-regulated in response to p-MHC–TCR interactions during positive selection. As shown in Fig. 1*B*, the proportion of $CD5^+CD69^+$ DP thymocytes was lower in B6.I-9.3 mice compared with both parental strains. The recovery of the $CD4^+$ subset in B6 × B6.I-9.3 F1 hybrids confirmed that impaired thymic selection of $CD4^+$ T cells was governed by the *H2-A*^{j} allele; codominant expression of the *H2-A*^{b} allele restored the $CD4^+$ subset and normalized the population of $CD5^+CD69^+$ DP cells in thymi (Fig. 1 *A* and *B*). Together, these results clearly indicate that positive thymic selection of $CD4^+$ T lymphocytes is hampered in the context of *H2-A*^{j} compared with selection on H2-A^b or H2-E^j MHC-II molecules.

In the periphery, $CD4^+$ deficiency was also observed in the spleens and lymph nodes of B6.I-9.3 mice, leading to a dramatic decrease in the CD4/CD8 ratio compared with the parental strains (Fig. 1 *C* and *D*). F1 hybrids showed a distribution closer to that of B6 mice (spleen $0.85 \pm 0.15/1.35 \pm 0.3$). Expression of H2-A^j on the surface of B cells in B6.I-9.3 mice was not impaired (*SI Appendix*, Fig. S1) and therefore, could not explain this defect in $CD4^+$ T cell selection. Variability in the CD4/CD8 ratio in peripheral lymphoid tissues has been described for various mouse strains, and a predominance of $CD4^+$ over $CD8^+$ T cells is common for most inbred mice (*SI Appendix*, Table S1). Accordingly, both experimental inbred mouse strains and *H2*-congenic recombinant mice, which have different haplotypes and only a single MHC-II molecule (H2-A), have a CD4/CD8 ratio ≥1. B6.I-9.3 represents an example of an inverted CD4/CD8 ratio that is <1. The fact that a typical CD4/CD8 ratio >1 persists in parental I/St mice that also bear *H2*^{j} alleles argues against the influence of genes proximal to *H2-A* on this CD4/CD8 inversion and instead, suggests the compensatory contribution of an H2-E–selected $CD4^+$ pool to the total $CD4^+$ population, as has been previously reported in a genetic study (30). Thus, total homeostatic proliferation in the periphery was unable to compensate for impaired thymic selection of naïve $CD4^+$ T cells in B6.I-9.3 mice. Profound changes in CD4/CD8 ratios indicate intrinsic differences in TCR selection in the context of distinct *H2-A* allelic variants.

T_{reg} but Not T_{conv} Populations Are Reconstituted in the Spleens of B6.I-9.3 Mice. TCR–p-MHC-II interactions determine activation thresholds during selection of T_{conv} and T_{reg} cells in the thymus. Importantly, the allowed affinities and the process of selection itself differ for these two populations (22, 31, 32). To assess how these differences might influence the TCR repertoires for nT_{conv} and nT_{reg} cells developing in B6 and B6.I-9.3 mice, we generated the B6.I-9.3.Foxp3^{GFP} mouse strain (hereafter referred to as H2-A^j

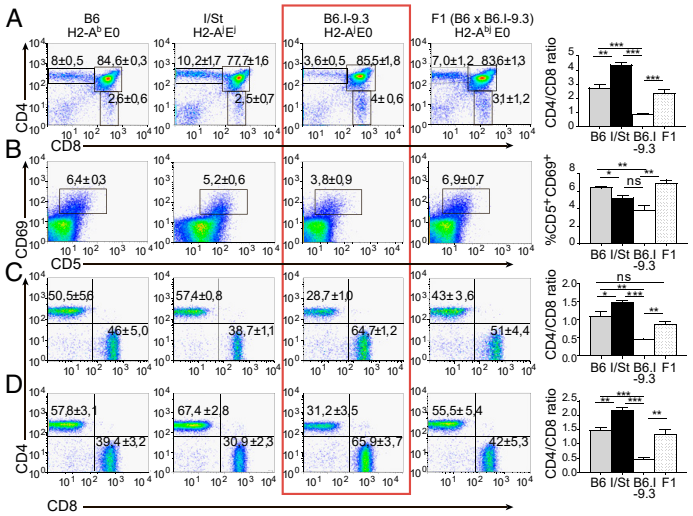

Fig. 1. B6.I-9.3 mice display an inverted CD4/CD8 ratio in the thymus and periphery. The percentage of $CD4^+$ and $CD8^+$ T cells in the (*A*) thymus, (*C*) spleen, and (*D*) peripheral lymph nodes as assessed by flow cytometry. Ratios are summarized as bar graphs in *Right*. (*B*) Thymocytes undergoing selection processes are defined as quadruple-positive $CD4^+CD8^+CD5^+CD69^+$ cells. (*C* and *D*) Results for gated $CD3^+$ T cells are shown. Representative data from one of three to five similar independent experiments are displayed, with total statistics for 6 to 12 mice per group provided in *Right* as mean ± SD (unpaired *t* test). ns, not significant; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

mice) and compared it with the B6.Foxp3^{GFP} strain (hereafter referred to as H2-A^b mice) described previously (33). Both strains produce fluorescently labeled T_{reg} cell populations. Evaluation of CD4/CD8 ratios in lymphoid organs of H2-A^j and H2-A^b Foxp3^{GFP} reporter mice confirmed defective $CD4^+$ selection in the *H2-A^j* genetic context (Fig. 2 *A* and *B*).

We next assessed whether the generation and/or maintenance of T_{reg} and T_{conv} $CD4^+$ subsets differs in H2-A^b and H2-A^j mice by counting the content and proportion of $CD4^+$ T cells in the spleen and thymus. The proportion of $Foxp3^+$ T_{reg} cells among SP $CD4^+$ thymocytes in H2-A^j and H2-A^b mice appeared to be similar (Fig. 2*C*). However, due to reduced total $CD4^+$ content, the size of the T_{reg} population in thymi of H2-A^j mice was significantly smaller (Fig. 2*D*). At the same time, in the spleens of H2-A^j mice, the total T_{reg} (Fig. 2*E*) and nT_{reg} (Fig. 2*F*) populations recovered to an extent that they matched the counts

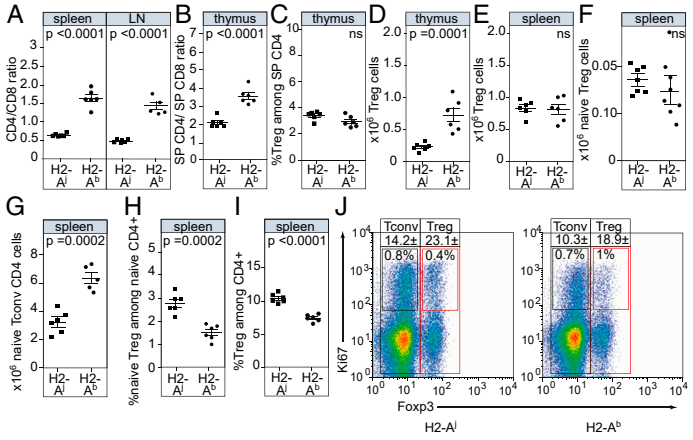

Fig. 2. $CD4^+$ T cell populations in lymphoid organs of H2-A^j and H2-A^b mice. CD4/CD8 T cell ratios in (*A*) spleens and lymph nodes (LN) and (*B*) thymi for H2-A^b ($n = 6$) and H2-A^j ($n = 6$) mice. (*C*) The percentage of nT_{reg} cells in the population of thymic SP $CD4^+$ T cells is equal in H2-A^j and H2-A^b mice. The total number of nT_{reg} cells in H2-A^b ($n = 6$) and H2-A^j ($n = 6$) mice differs for the SP $CD4^+$ thymic population (*D*) but not for the spleen (*E*). (*F*) The splenic content of nT_{reg} cells is equal in H2-A^j and H2-A^b mice, whereas (*G*) the size of the nT_{conv} cell population is significantly smaller in H2-A^j. The percentage of (*H*) nT_{reg} and (*I*) total T_{reg} cells is higher in H2-A^j compared with H2-A^b spleens. (*J*) Proliferation of splenic $CD4^+$ T cells from sorted T_{reg} cells gated for Foxp3 expression. All data are representative of at least two experiments ($n = 6$ females per group, unpaired *t* test, $P = 0.006$ for T_{conv} and $P = 0.003$ for T_{reg}). ns, not significant.

detected in H2-A^b mice, while the nT$_{conv}$ population remained significantly reduced (Fig. 2*G*). Thus, in the spleens of H2-A^j mice, the proportion of nT$_{reg}$ (Fig. 2*H*) and primed (Fig. 2*I*) T$_{reg}$ cells was increased compared with H2-A^b mice.

Since the reduction in the T$_{reg}$ population that we observed in the thymi of H2-A^j mice was restored in the spleen, we questioned whether this was due to increased proliferation of T$_{reg}$ cells in order to achieve homeostatic replenishment. We estimated the proliferation rate of splenic CD4$^+$ T cells using a Ki67 expression staining. This analysis revealed an increased proliferation rate for both T$_{conv}$ and T$_{reg}$ splenic subsets in H2-A^j compared with H2-A^b mice (Fig. 2*J*). However, T$_{reg}$ cells restored their peripheral niche more efficiently, which is consistent with other observation (34).

TCR Repertoires of H2-A^j Mice Are Distinct from H2-A^b and Characterized by Decreased Diversity and High Publicity and Convergence. The initial diversity of preimmune repertoires of naïve T cells is generated by the recombination machinery as well as thymic and peripheral selection (10). To explore the influence of MHC-II allelic variation on the TCR landscape, we analyzed repertoires of nT$_{conv}$ and nT$_{reg}$ cells isolated from splenic cell populations of homozygous H2-A^j and H2-A^b mice. We sorted 5,600 to 65,000 nT$_{reg}$ and 100,000 nT$_{conv}$ cells per mouse, with the controlled purity of cell fractions >98%. TCRα and -β complementary DNA (cDNA) libraries were obtained using rapid amplification of 5'-cDNA ends (5'-RACE) with unique molecular identifiers (UMIs) and TCR repertoires extracted using MIGEC (35) and MiXCR (36) software. Out of each sample, we extracted CDR3 repertoires for at least 2,000 nT$_{reg}$ TCRα, 8,000 nT$_{conv}$ TCRα, 5,000 nT$_{reg}$ TCRβ, and 14,000 nT$_{conv}$ TCRβ unique UMI-labeled cDNA molecules (*SI Appendix*, Table S2*A*).

For comparison of general repertoire similarity, we performed clustering analysis based on the number of amino acid CDR3 clonotypes shared between the 800 most frequent clonotypes of each pair of samples. This analysis clearly clustered the functional (in-frame, without stop codons) TCRα and TCRβ repertoires of nT$_{conv}$ CD4$^+$ subsets separately for H2-A^j and H2-A^b mice (Fig. 3*A*), suggesting that the distinct p-MHC-II context essentially shapes the structure of the TCR repertoires at the stage of thymic selection. This was not the case for nonfunctional TCR repertoires resulting from unsuccessful recombination events. Such TCRs have not been expressed and thus, preserve

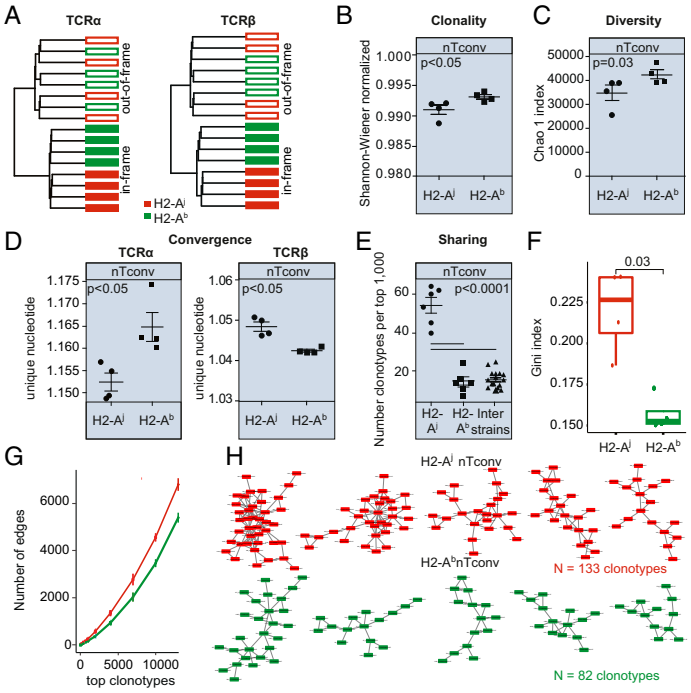

Fig. 3. TCR repertoire profiling. (*A*) Cluster analysis of the TCRα and TCRβ repertoires for nT$_{conv}$ isolated from spleens of H2-A^j (red; $n = 4$) and H2-A^b (green; $n = 4$) mice. Datasets were normalized based on the observed diversity of the smallest sample. Using a pairwise similarity metric (based on the number of identical amino acid sequences of clonotypes), we performed cluster analysis of CDR3β in-frame and nonfunctional variants for the top 800 clonotypes using VDJtools. (*B–E*) Properties of the TCRβ repertoire of nT$_{conv}$ from spleens. *P* values were computed using unpaired *t* test. (*B*) Assessment of clonality, where clonotype evenness (expressed as Shannon–Wiener index) was normalized by clonotype number. (*C*) Analysis of repertoire diversity using the Chao1 index for nT$_{conv}$ cells. We collected 14,000 UMIs per sample for CDR3β. (*D*) Convergence measurement as a ratio of unique nucleotide number encoding the same amino acid. (*E*) Sharing of TCR repertoires between normalized samples, based on the number of identical CDR3β amino acid sequences in the top 1,000 clonotypes. (*F*) The Gini index for the top 2,500 CDR3β amino acid clonotypes in the nT$_{conv}$ repertoires of H2-A^b and H2-A^j mice reflects inequality in clonotype size. *P* values were computed using Mann–Whitney *U* test. (*G*) Connectivity of repertoires in H2-A^j and H2-A^b mice. The average number of edges per node of homologous TCR clusters grows along with repertoire depth. (*H*) The top five clonotype clusters from the repertoires of a single H2-A^b and H2-A^j mouse. TCR networks were built from amino acid CDR3β sequences of the 2,500 most abundant clonotypes.

the repertoire features produced by the recombination machinery before thymic selection (37, 38). Clustering analysis did not separate the nonfunctional CDR3 repertoires of H2-A^j and H2-A^b mice, indicating that segregation of the functional repertoires was a result of distinct selection.

We assessed the normalized Shannon–Wiener and Chao1 diversity indices at the amino acid level (39) after normalization of samples by down-sampling to equal counts of unique UMI-labeled cDNA molecules. These two statistical metrics assess different aspects of the TCR repertoire; normalized Shannon–Weiner reflects the evenness of the clonotype frequency distribution, whereas Chao1 provides the lower boundary of total clonotype diversity as estimated based on the relative counts of low-frequency clonotypes. As could be expected, the drastic reduction in the count of naïve CD4$^+$ cells concomitant with elevated proliferation was reflected in the decreased diversity and increased clonality of naïve TCR repertoires. In comparison with H2-A^b mice, the amino acid TCRβ repertoires of nT$_{conv}$ cells in H2-A^j showed significantly lower Chao1 and higher Gini index, which are related to unequal distribution of clonotype size (Fig. 3 *B*, *C*, and *G*). We observed a similar tendency in the TCRβ repertoires of nT$_{reg}$, but the number of clonotypes that we obtained was too small for these data to be statistically representative (*SI Appendix*, Fig. S2 *A*–*C*).

Next, we assessed the relative convergence of TCRβ repertoires, as measured by the normalized abundance of CDR3 nucleotide sequences encoding the same amino acid sequence, which results in the appearance of identical frequently produced TCR clonotypes within and between animals (40). Remarkably, the CDR3β repertoire of nT$_{conv}$ cells in H2-A^j mice was characterized by a prominently higher degree of convergent recombination events (*t* test, $P < 0.005$) (Fig. 3*D*) in comparison with H2-A^b, while the CDR3α repertoire in H2-A^j contained less converged and rather diverse TCR variants (Fig. 3*D*). We then estimated the extent of TCR clonotype sharing (i.e., publicity) at the amino acid level within and between H2-A^j and H2-A^b mouse groups. The cohort of H2-A^j mice shared a significantly higher proportion of CDR3β clonotypes than the H2-A^b mice: on average, 54 vs. 14 shared amino acid clonotypes from the 1,000 most frequent clonotypes of each repertoire (Fig. 3*E*). Finally, network analysis of TCRβ homology clusters based on amino acid similarity showed denser network connectivity of TCRβ repertoires in H2-A^j mice (Fig. 3 *F*–*H*). Together, these observations clearly indicate that the *H2-A^j* genetic context favors selection of a narrower naïve TCRβ repertoire, characterized by high publicity, constrained diversity, and increased convergence.

Prominent and Distinct Impact of *H2-A^j* on CDR3α and CDR3β Characteristics. As described in ref. 41, we assessed average characteristics of the CDR3 repertoire, such as length, number of added nucleotides, and the physicochemical properties of amino acids located in the middle of CDR3 (*SI Appendix*, Fig. S3), which are usually in tight contact with p-MHC-II. Interestingly, the TCRβ repertoire of H2-A^j mice was skewed toward longer CDR3 loops in both nT$_{conv}$ and nT$_{reg}$ CD4$^+$ subsets (at average, weighted per-clonotype frequencies) (Fig. 4*A*). This increase in length was associated with a higher number of nongermline added nucleotides (*SI Appendix*, Fig. S3*A*). Remarkably, we observed an opposite tendency in the TCRα repertoire, with decreased CDR3 length (Fig. 4*A*). Additionally, nT$_{reg}$ cells in H2-A^j and H2-A^b mice possessed significantly shorter CDR3s in both α- and β-chains compared with nT$_{conv}$ cells. This is consistent with previous findings from other studies (42) as well as our current findings for human nT$_{reg}$ cells. The H2-A^j repertoire was also characterized by a lower average number of bulky amino acid residues based on the "volume" parameter of VDJtools (including W, R, K, Y, and F) in the middle of CDR3β but a higher number of such amino acids in CDR3α (Fig. 4*B*). We hypothesize that such reciprocal shifting of CDR3α and CDR3β characteristics might reflect distinct TCR fitting with p-MHC complex in H2-A^j mice.

Based on statistical mechanical computations of pairwise interactions between amino acids (43, 44), a set of generally strongly interacting amino acid residues has been selected. In essence, those include hydrophobic and aromatic residues that can mediate diverse types of interaction (45) and contribute greatly to TCR–p-MHC binding affinity (46). We estimated relative strength of CDR3 interaction with diverse (i.e., not necessarily bearing cognate antigen) p-MHC complexes based on the average number of such strongly interacting amino acid residues (F, I, L, M, V, W, and Y) within the central part of the CDR3 using the "strength" parameter of VDJtools. Previously, this parameter has been shown to distinguish TCR repertoires of T$_{reg}$ cells and CD4$^+$ T$_{conv}$ cells in mice (39, 47), in keeping with the higher TCR affinity allowed during thymic selection of T$_{reg}$ population (31, 32). It has been hypothesized that T$_{reg}$ cells that possess high-affinity TCRs can compete with conventional CD4$^+$ cells in the periphery for targets on antigen-presenting cells (23), preventing spontaneous proliferation of T$_{conv}$ cells. High CDR3 strength also distinguishes human nT$_{reg}$ and effector T$_{reg}$ cells vs. other helper subsets, CD8$^+$CXCR3$^+$ vs. CD8$^+$CXCR3$^-$ naive T cells (48). Furthermore, CDR3 strength and hydrophobicity are low in experienced memory B cells (49). To some extent, high CDR3 strength and hydrophobicity can be interpreted as T or B cell potential to broad, cross-reactive, and self-reactive interactions (44, 50).

In agreement with previous studies and our current data for human nT$_{reg}$ cells, we found that the average number of strongly interacting amino acids among the five residues in the middle of CDR3 was higher in nT$_{reg}$ compared with nT$_{conv}$ repertoires in both mouse strains, and for both CDR3α and CDR3β (Fig. 4*C*). An increased average number of strongly interacting amino acids in the nT$_{reg}$ CDR3s suggests the preferential selection of TCRs with high affinity to self-antigens for generation of thymic T$_{reg}$ population (22, 31, 32, 51). We also noted that the relative strength was significantly higher for H2-A^j mice compared with H2-A^b mice in the CDR3α and -β repertoires of both nT$_{reg}$ and nT$_{conv}$ cells. Furthermore, this effect was significantly more prominent in the nT$_{reg}$ subset. As a result, the difference between nT$_{reg}$ and nT$_{conv}$ cells was substantially higher in H2-A^j mice (Fig. 4*D*), which might potentially result in generally increased capacity of T$_{reg}$ cells to suppress activated T cells in H2-A^j mice.

Using cluster connectivity analysis, we searched for the specific characteristics of CDR3β amino acid sequence variants enriched in the repertoire of nT$_{reg}$ cells selected on different MHC-II molecules. We separately pooled all clonotypes for H2-A^b and H2-A^j mice and searched for similar amino acid TCR clonotypes within mouse strains. These similar TCRs may be represented as a network of CDRβ sequences having one amino acid mismatch between nodes. Considering only sequences that have more neighbors in one strain rather than in another, we found clonotypes that might carry an imprint of selection on a certain MHC-II allele. Notably, we detected distinct dominant CDR3β motifs (Fig. 4*E*). A CDR3β motif identical to that found in H2-A^j context was previously reported among public TCR clonotypes responding to respiratory syncytial virus infection (52). It is also interesting that the same CDR3β motif was previously found among cross-reactive TCRs recognizing both MHC-I and -II (53).

TCR repertoire analysis of all samples using high-dimensional features enabled us to clearly separate CDR3α and CDR3β repertoires by both functionality and genetic background (Fig. 4*F* and *SI Appendix*, Fig. S4). This demonstrates the prominent differences in the TCR CDR3 repertoire landscapes of both nT$_{reg}$ and nT$_{conv}$ subsets between H2-A^j and H2-A^b mice.

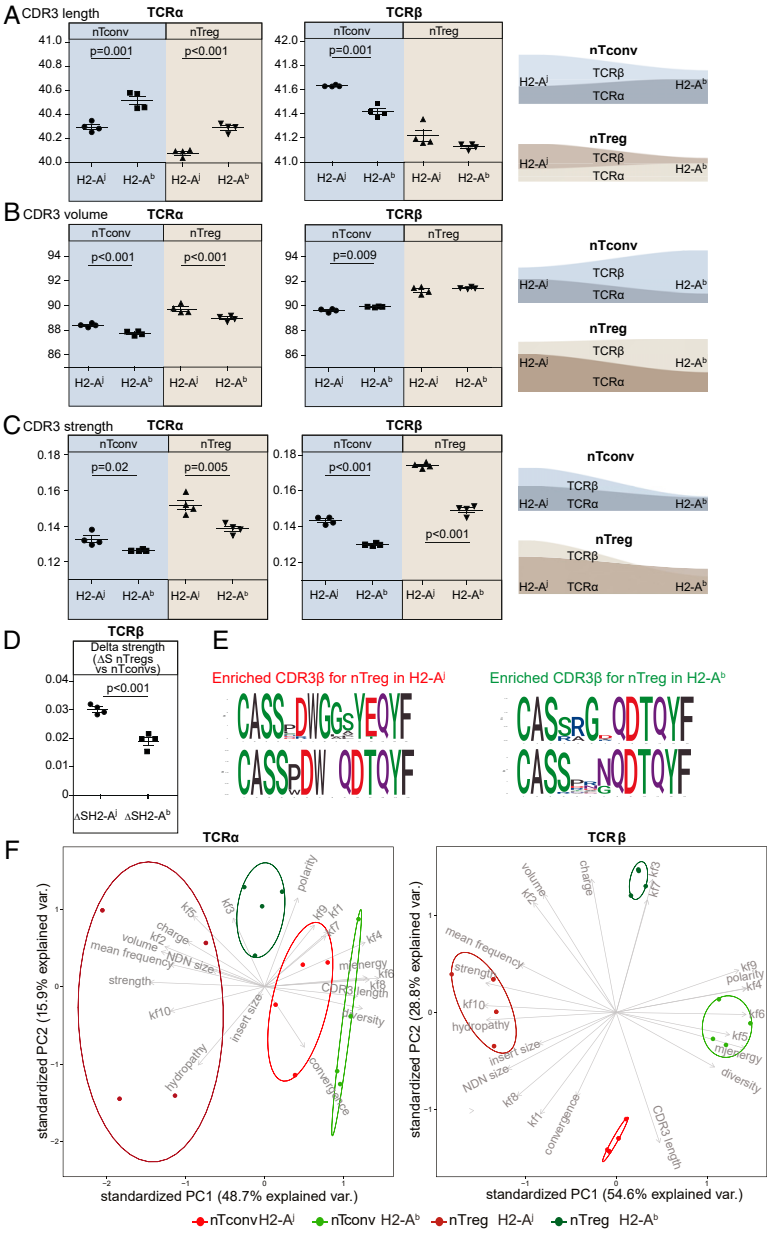

Fig. 4. Characteristics of CDR3α and CDR3β loops in naïve TCR repertoires. Average count of (*A*) CDR3 length and (*B*) volume based on the presence of bulky amino acid side chains for CDR3α and CDR3β in nT_{conv} and nT_{reg} cells from H2-A^b or H2-A^i. *Right* shows schematic illustrations of the differences in CDR3α and CDR3β length and volume parameters in mouse strains. (*C*) Predicted strength of interaction for bulk CDR3α and CDR3β repertoires of nT_{reg} and nT_{conv} cells in H2-A^i and H2-A^b mice. $n = 4$ mice in each group. *Right* shows schematic illustrations of the differences in the CDR3α and CDR3β interaction strength in the two strains. (*D*) The difference in strength values (ΔS) of CDR3β for nT_{reg} and nT_{conv} subsets in H2-A^b vs. H2-A^i. *P* values were computed using Benjamin Hochberg-corrected *t* test, based on false discovery rate <0.05. (*E*) Representative logo plots visualizing amino acid motifs of TCR CDR3β that were significantly enriched as a result of similar recombination events in nT_{reg} cells. (*F*) Cumulative contribution of 22 parameters (e.g., physicochemical properties, Kidera Factors, diversity, NDN size (NDN, added nucleotides flanking the D gene segment), and CDR3 length) in principal component analysis for the TCR repertoires of nT_{reg} and nT_{conv}. Shown here are two-dimensional principal component analysis projections of 2,000 CDR3α and 5,000 CDR3β UMIs that were randomly selected for each sample. PC1, principal component 1; PC2, principal component 2. All parameters were scaled to a similar range using *Z*-score normalization.

Skewed Usage of TCR V Segments and Impact of CDR1 and CDR2 on p-MHC-II–TCR Interaction. The germline-encoded CDR1/CDR2 regions of the TCRα and TCRβ variable domains are believed to contribute mainly to the basic interaction of TCR with MHC-II, modulating the binding angle of the hypervariable region relative to p-MHC (15, 54). Previously, it has been reported that TRBV and TRAV segment usage is strongly associated with the expressed allelic variant of MHC-II (19, 55, 56). The observation that point substitutions in conserved stretches of amino acids in MHC-II alter TRBV and TRAV segment usage in CD4 cells but do not affect the $CD4^+$ numerical count (25) provides further, direct evidence for preselection of TCR repertoires in MHC recognition.

In our experiments, H2-A^j and H2-A^b mice could be separated into clusters of V segment usage depending on MHC context (*SI Appendix*, Fig. S2*D*). Specifically, the TRAV12.2, TRAV4.3, and TRAV12.1 gene segments were used significantly more often by nT_{conv} $CD4^+$ subsets in H2-A^j mice compared with H2-A^b (*SI Appendix*, Fig. S2*D*), whereas the usage of several TRAV segments (TRAV14D.1, TRAV14.3, TRAV7.4) dropped drastically in H2-A^j mice. The observed changes in V segment usage should reflect distinct preferences in selection of those gene segments in H2-A^j mice. In other words, distinct V gene segments would be expected to provide optimal affinity for the *H2-A^j* haplotype of MHC-II, increasing the probability of passing both positive and negative selection.

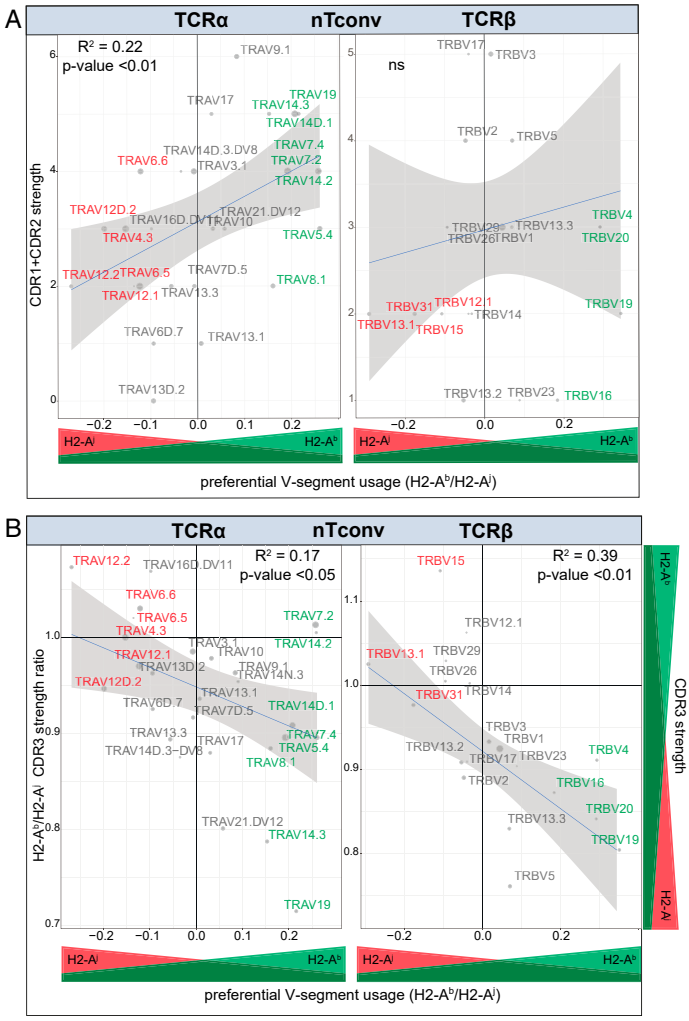

Fig. 5. V segment usage in TCR repertoires of nT_{conv} cell subsets in H2-A^b and H2-A^j mice. (*A*) Correlation of unweighted cumulative (CDR1 + CDR2) strength to ratio of average usage frequency of pooled TRBV or TRAV segments from the nT_{conv} repertoire of H2-A^b vs. H2-A^j mice. (*B*) Correlation between the ratio of weighted CDR3 strength (H2-A^b vs. H2-A^j) and ratio of average usage frequency for TRAV or TRBV segments between the nT_{conv} repertoires of H2-A^b vs. H2-A^j mice. Green and orange colors denote TRBV segments preferentially used in H2-A^b and H2-A^j repertoire, respectively.

Taking into account that TCR CDR1 and CDR2 sequences are identical in both mouse strains, we calculated the cumulative interaction strength based on the average number of strongly interacting amino acids constituting these regions. The cumulative interaction strength of CDR1 and CDR2 for each TRAV segment correlated with the fold change of its usage frequency in H2-A^b nT$_{conv}$ cells relative to H2-A^j (Fig. 5*A*). Preferential selection of defined TRAV segments (e.g., TRAV12.2, TRAV12.1, TRAV4.3, and TRAV6.5) on H2-A^j molecule required low numbers of strongly interacting amino acid residues in the middle of CDR1 and CDR2, while H2-A^b favored the selection of TRAV segments bearing CDR1 and CDR2 with high cumulative interaction strength (e.g., TRAV19, TRAV14.3, TRAV7.4) (Fig. 5*A*).

Certain TRBV segments were also enriched in H2-A^j and H2-A^b contexts but were not clearly related to the strength of corresponding CDR1 and CDR2 regions (Fig. 5*A*), suggesting preferable TCR–MHC interactions mediated by particular complementary structures. Interestingly, the TRBV segments enriched in either *H2-A^j* or *H2-A^b* context carried CDR3β variants with lower interaction strength, on average (Fig. 5*B*). This may indicate that some optimal modes of preencoded CDR1 and CDR2 binding to corresponding MHC are sufficient for these TRBV segment variants, such that strongly interacting CDR3β motifs are not required for successful thymic selection. In the nT$_{reg}$ population, we observed similar shifts in TRBV usage (*SI Appendix*, Fig. S5). However, in contrast to nT$_{conv}$ cells, only TRAV6.6 was specifically enriched in H2-A^j nT$_{reg}$ repertoire, which could reflect the influence of distinct thresholds in the course of nT$_{reg}$ and nT$_{conv}$ cells selection. A similar trend was observed for TRAV segments, where segments with weakly interacting CDR3α loops were generally preferentially selected for any allelic variants of MHC-II (Fig. 5*B*).

We can therefore conclude the following. First, distinct structural features of H2-A influence selection preferences for particular complementary TRAV and TRBV segments. Second, for TRAV segments, these differences are associated with the relative abundance of strongly interacting amino acid residues in CDR1 and CDR2 loops. Finally, in the context of highly complementary interactions between V segments and MHC, the average number of strongly interacting amino acids within the CDR3 region reciprocally decreases, potentially reflecting cumulative fitting to the optimal selection window of interaction strength required to pass thymic selection.

Nonclassical MHC-II Molecule H2-O Does Not Have Essential Impact on Selecting TCR Repertoire. We have previously observed that H2-O^j has several nonconservative substitutions in the *H2-Oa* and *H2-Ob* genes in the *H2^j* haplotype compared with H2-O^b (26) that could make the H2-O^j chaperone dysfunctional (27). To elucidate whether this nonclassical MHC-II molecule contributes to CD4 selection and TCR repertoire shaping, we analyzed repertoires of naïve CD4 subsets in H2-A$^{b\ Ob-/-}$ mice, which lack the H2-O molecule (27) (*SI Appendix*, Table S2*B*). H2-A$^{b\ Ob-/-}$ thymus and spleen had a CD4/CD8 ratio similar to H2-A^b (*SI Appendix*, Fig. S6 *A–C*). This further supports the notion that the inverted CD4/CD8 ratio in H2-A^j mice results from impaired CD4$^+$ T cell selection on the H2-A^j molecule. We assessed physicochemical properties of TCR CDR3 regions of naive CD4$^+$ T cell subsets in H2-A$^{b\ Ob-/-}$, H2-A^b, and H2-A^j. In contrast to H2-A^j mice, both H2-A$^{b\ Ob-/-}$ and H2-A^b nT$_{conv}$ cells had similar numbers of strongly interacting amino acid residues within the central part of CDR3 in both TCRα and TCRβ chains (*SI Appendix*, Fig. S6*D*). In keeping with the results described above, H2-A^j but not H2-A$^{b\ Ob-/-}$ mice tended to have more bulky amino acids in the center of CDR3α and smaller ones in CDR3β. H2-A$^{b\ Ob-/-}$ nT$_{conv}$ cells clearly have CDR3α amino acid side-chain size equal to the congenic H2-A^b strain (*SI Appendix*, Fig. S6*E*). Taken together, these results indicate that differences in the repertoire landscape of nT$_{conv}$ in H2-A^j mice arise from TCR selection on structurally different classical MHC-II molecules but not from allelic variants of H2-O.

Discussion

The preimmune, naïve TCR repertoire landscape formed by thymic and peripheral selection (12) essentially determines an organism's capacity to protect against diverse pathogenic (55, 57) and cancerous antigens (58, 59) and to achieve a balanced network of regulatory interactions between immune cells (23). Convincing data show that the interplay of the TCR repertoire with certain MHC-II allelic variants can predispose individuals to autoimmune diseases (7). The individual MHC context determines the selection of CD4$^+$ T cells and shapes the TCR repertoire, influencing both V and J segment usage (19, 25) and frequencies of particular TCR CDR3 variants (60).

Here, we have focused on identifying how selection in particular MHC-II contexts determines the structure of the TCR repertoire of nT$_{conv}$ and nT$_{reg}$ cells. We used mouse strains expressing different allelic variants of a single classical MHC-II while preserving the natural complexity of the TCR repertoires and obtained several intriguing findings. Confirming previous observations (19, 20, 25), we found prominent differences in usage of V gene segments correlating with the role of CDR1 and CDR2 in p-MHC interaction. Our data on the preferable selection of TCRs bearing defined TRAV and TRBV gene segments by each MHC allele support the germline-encoded model of TCR engagement with p-MHC (18, 61). Furthermore, the CDR3 repertoires of nT$_{conv}$ and nT$_{reg}$ subsets in H2-A^j mice (B6.I-9.3.Foxp3^{GFP}) strikingly differed from H2-A^b mice (B6.Foxp3^{GFP}) in terms of CDR3α and CDR3β amino acid composition and length.

Polymorphisms near or in the peptide binding groove in MHC-II alter the peptide repertoire and the overall TCR–p-MHC interaction interface (55, 62, 63) and consequently, cause perturbations of the TCR repertoire that successfully passes the first thymic selection. At present, no crystallographic data have been obtained for the p-H2-A^j complex, but molecular modeling of H2-A^j with CLIP peptide has suggested changes in the structure and size of binding pockets P1, P4, P6, P7, and P9 (26). Furthermore, the volume reduction of the H2-A^j P1 pocket, which is implicated in binding to H2-M (DM in humans) (64, 65), might contribute to exchange efficiency of invariant peptide (e.g., CLIP) for self-peptides in MHC-II. The root-mean-square deviation (RMSD) analysis of the molecular dynamics trajectories (*SI Appendix*, Fig. S7) supports the hypothesis of lower binding capacity of H2-A^j for CLIP than H2-A^b. The ability of H2-A^j to form a stable complex with peptides can be affected by the lack of one of the highly conserved H bonds between the peptide backbone and A74, which replaces the highly conserved E74 in the β-chain. These modeling results suggest that diversity of endogenous peptides loaded and stably presented on H2-A^j may be limited, thus hampering the positive selection of diverse CDR3 repertoire. Further comparative studies of endogenous peptide repertoires presented by H2-A^j vs. H2-A^b are required to verify this hypothesis. The most noticeable structural changes in the α-helixes of H2-A^j are the unique substitution Q61K, located in the α-chain, and the very rare substitution R70Q in the β-chain, both of which alter the polarity of interactions between chains. The position αQ61 has been reported to be a contact point between the H2-Aα chain and TCRβ CDR1, CDR2, and CDR3 (66), with targeted substitution significantly reducing the TCR activation threshold (67). Overall, all of these alterations can lead to formation of p-H2-A^j complexes that could affect interaction with TCR and influence the selection of CD4$^+$ T cells.

Based on our results, we speculate that mutual adjustment of the length of the TCRα and TCRβ CDR3 loops is mediated by the distinct docking mode of TCR–p-MHC interaction during thymic selection, such that shortening of the first CDR3 loop is

compensated for by elongation of the second one. One of the notable physicochemical characteristics of CDR3 is the number of strongly interacting (i.e., mainly hydrophobic and aromatic) amino acid residues (43) in the middle of TCRα and TCRβ CDR3. These parts of the TCR interact most closely with presented peptides (41). As shown previously, high strength is characteristic of T_{reg} CDR3 repertoires, reflecting their permissible higher affinity to self p-MHC during thymic selection (22, 31, 32). We observed a remarkable increase in the average representation of strongly interacting residues in both TCRα and TCRβ CDR3 repertoires in H2-A^j mice. The effect was more prominent in T_{reg} cells (Fig. 4*D*), which perhaps, reflects the competitive nature of thymic T_{reg} generation (51). This feature of the TCRs in the nT_{reg} population of H2-A^j mice and its numerical advantage over T_{conv} in the periphery might potentially contribute to susceptibility to tuberculosis via suppression of an efficient and timely Th1 response (68, 69)

H2-A^j genetic context was associated with profound deficiency of $CD4^+$ cells in the thymus and spleen (Fig. 2). At the same time, we found that the TCR repertoire of H2-A^j mice was characterized by narrowed diversity, increased publicity across animals of the same strain, and higher convergence—that is, increased average number of nucleotide CDR3 sequence variants per amino acid sequence variant (Fig. 3 and *SI Appendix*, Fig. S2). Cumulatively, these data led us to conclude that the *H2-A^j* context, due either to the features of the H2-A^j surface or to the restricted pool of H2-A^j–presented peptides (see below), imposes more rigorous selection of an essentially narrowed pool of TCR variants compared with H2-A^b, which also results in numerical deficiency of naive $CD4^+$ cells produced.

The number and diversity of naïve T cells are considered to be crucial for efficient and specific response to a broad range of pathogens (70). Therefore, the increased "gaps" in the narrowed repertoire of $CD4^+$ naive T cells in the *H2-A^j* context could contribute to susceptibility to particular infections, including tuberculosis. However, the diversity of antigenic molecules covered by a TCR repertoire can also be increased through higher cross-reactivity. In other words, an efficient immune response against a wide range of pathogens can be provided either by high diversity of the TCR repertoire and/or by broad TCR cross-reactivity (71–75). Accounting for specific or degenerate TCR recognition is a complex process that depends on the interaction strength of TCR with p-MHC (76). The specific feature of the H2-A^j repertoire is an increased number of strongly interacting amino acid residues in the middle of both CDR3α and CDR3β in nT_{conv} and nT_{reg} repertoires (Fig. 4*C*). Most strongly interacting amino acid residues belong to hydrophobic residues, which form high-affinity binding contacts within protein–protein interfaces (77). TCR cross-reactivity correlates with the frequency of strongly interacting residues in CDR3β (50), and we therefore cannot exclude the possibility that restricted diversity of the TCR repertoire in H2-A^j mice might be offset by potentially enhanced cross-reactivity. This could allow adaptive immunity to recognize comparable diversity of antigens with a narrower repertoire of TCRs, at the cost of potential nonspecific immune response and inflammation. More generally, such a feature of a narrowed repertoire may reflect the natural balance of diversity and cross-reactivity in terms of MHC context selection. It comes in line with the study shown that skewing properties of TCR CDR3 toward increased hydrophobicity was detected in the cases of altered T cell selection (50, 78).

The H2-A^j genome also contains nonclassical MHC-II genes of the *H2^j* haplotype, encoding H2-M and H2-O. These genes are highly conserved across organisms and participate in peptide loading onto the classical MHC-II molecule. Together, H2-O and H2-M modulate the exchange of CLIP peptide with self-peptides, thus influencing peptide repertoire editing (79, 80). Nevertheless, the exact functional role of H2-O in TCR shaping is still elusive (28, 29). H2-O expression is mainly associated with increased CLIP–MHC-II expression on the cell surface (28), and according to recent mass spectrometry data, lack of H2-O expression may alter the diversity of the presented peptide repertoire (81). Evidence suggests that the activity of the H2-O^j allelic variant is reduced, leading to enhanced H2-M function, as is observed in H2-Ob knockout mice (27). However, it has been demonstrated here and in another study (82) that the number of $CD4^+$ T cells in H2-A^b mice lacking H2-O was not diminished, and according to our data, H2-O does not have any prominent effect on generation and features of TCR repertoires of nT_{conv} cells. Thus, we can conclude that polymorphism of H2-A^j, but not H2-O^j, determines the striking differences in the selection of $CD4^+$ T cells.

The association of the *H2-A^j* allele with significantly reduced $CD4^+$ cell numbers in the thymus and spleen might be linked to certain differences in the binding groove of H2-A^j that result in less efficient positive selection of naïve $CD4^+$ T cells. However, the low count of $CD4^+$ cells may not necessarily lead to poor infection control in organisms. The composition and diversity of MHC-bound peptides have a strong effect on positive selection. Previous studies of genetically modified mice, including H2-M- or Ii (CLIP)-deficient mice, have demonstrated that limited multiplicity of low-abundance self-peptides and loose binding of peptides to MHC-II (83–85) result in reduced $CD4^+$ T cell counts. We speculate that the spectrum of peptides that are efficiently presented in H2-A^j might be more limited compared with H2-A^b. The requirement for immunodominant epitope presentation was shown in the elegant experiments by Nepal et al. (86), which also demonstrated that low $CD4^+$ cell counts in the periphery are not critical for controlling *Mycobacterium tuberculosis* (Mtb) dissemination, but proper H2-M function determines the efficiency of the immune response against Mtb infection. Thus, the inability of H2-A^j to bind and present some range of peptides might be related to Mtb susceptibility.

To summarize, in this work we show that distinct MHC-II contexts may prominently shape the features of the selected landscape for both nT_{conv} and nT_{reg} CDR3 repertoires. This influence is measurable not only as a distinct frequency of TRAV and TRBV gene segment usage but also at the level of CDR3α and CDR3β repertoire features. The reciprocal changes in CDR3α and CDR3β repertoire characteristics could reflect the distinct preferences of TCR–p-MHC accommodation. At the same time, the increased number of hydrophobic and strongly interacting amino acid residues in the middle of CDR3α and CDR3β may indicate the necessity of increased average strength of TCR–p-MHC interaction for successful positive selection. This is supported by the narrowed diversity of the more focused naive repertoire produced in the *H2-A^j* context and the decreased counts of naive $CD4^+$ T cells. Thus, we demonstrate that TCR repertoire profiling can provide a "digital" view of T cell selection. Further progress in our understanding of these processes—and ultimately, in our ability to predict the antigen-specific properties of individual TCR repertoire landscapes—should have direct implications for the development of vaccines and T cell-based immunotherapies for cancer and autoimmune disorders.

Materials and Methods

Mice and Genotyping. Mice of inbred strains I/StSnEgYCit (abbreviation I/St, H2^j), C57BL/6JCit (B6, H2^b), H2-recombinant congenic strain B6.I-9.3.19.8 (abbreviation B6.I-9.3) (26), and (B6.I-9.3xB6) F1 hybrid were bred and maintained under conventional, nonspecific pathogen-free (non-SPF) conditions at the Animal Facilities of the Central Institute for Tuberculosis in accordance with the guidelines from Russian Ministry of Health #755 and under the NIH Office of Laboratory Animal Welfare Assurance #A5502-11. Animals were killed using carbon dioxide. Transgenic mouse strain B6.Foxp3^{GFP} (abbreviated here as H2-A^b) was provided by A. Rudensky, Memorial Sloan Kettering Cancer Center, New York, NY (33). The protocol of genotyping is in *SI Appendix*. C57BL/6J *H2-Ob$^{-/-}$* mice were provided

by T. Golovkina, University of Chicago, Chicago, IL. Crossing B6.Foxp3^{GFP} and C57BL/6J *H2-Ob*$^{-/-}$, we generated F1 hybrids that were intercrossed to obtain C57BL/6 Foxp3^{GFP} *H2-Ob*$^{-/-}$ (H2-A$^{b\ Ob-/-}$) mouse strain.

Cell Preparations and Flow Cytometry. As described in detail in *SI Appendix*, spleen, thymus, and lymph nodes of 2 to 3 month-old mice were removed. Suspensions of thymus, spleen, and lymph node cells were obtained using routine procedures. The following monoclonal antibodies for fluorescence-activated cell sorting (FACS) analysis were purchased from eBiosciences, Biolegend, BD Biosciences, or Miltenyi Biotec: CD3-FITC (clone 145-2C11), CD3-PE-Cy5 (clone 17A2), CD4-PerCP or CD4-APC-cy7 (clone GK1.5), CD8-APC (or PE or PE-Cy5; clone 53-6.7), B220-APC (clone RA3-6B2), CD62L-PE (clone Mel-14), CD44-eFluor 450 (clone IM7), MHC class II I-E (clone 14-4-4S), and Foxp3-PE (clone FJK-16s). MHC class II I-A-Alexa Fluor 488 (clone 10.2-16) cross-reactive to the β-chain of H2-A^{j} was provided by T. Golovkina. Intracellular staining of Foxp3 and Ki67 (clone Sp6; Thermo Scientific) was done according to the manufacturer's protocol for Foxp3/Transcription Factor Staining Buffer Set (eBioscience). Data were acquired using a BD Biosciences FACSCalibur flow cytometer and BD Biosciences FACSCARIA III and analyzed using FlowJo software (Tree Star).

Cell Sorting. For cell sorting, CD4^{+} T cells isolated from spleen of individual H2-A^{b} and H2-A^{j} mice were enriched prior to FACS sorting by microbeads technology using negative selection with CD4^{+} T cell Isolation Kit II (Miltenyi Biotec). Purified CD4^{+} T cells were stained with CD4, CD62L, and CD44 antibodies and sorted into the two following populations: nT$_{reg}$ (CD4^{+}Foxp3 GFP^{+}CD62L^{+}CD44^{-}), nT$_{conv}$ (CD4^{+}Foxp3 GFP^{-}CD62L^{+}CD44^{-}). Each cell population was sorted at >98% purity. All cell populations were collected directly into RLT buffer (Qiagen) and stored at −70 °C.

RNA Isolation and cDNA Library Preparation. Total RNA was isolated using RNeasy mini kit (Qiagen) according to the manufacturer's instructions. In all experiments described here, cDNA libraries were obtained using 5′-RACE with UMIs (35, 41) using the Mouse TCR profiling kit (MiLaboratory LLC). Details of the procedure can be found in *SI Appendix*.

Raw Sequencing Data Analysis. We performed two independent sequencing runs. In the first run, we sequenced in parallel cDNA libraries obtained from four H2-A^{b} and four H2-A^{j} mice. In the second run, we performed sequencing of samples obtained from three H2-A^{b}, three H2-A^{j}, and five H2-A$^{b\ Ob-/-}$ mice (87). Sequencing data were analyzed using MiGEC (35), MiXCR, and VDJtools software (88). Details of the analysis can be found in *SI Appendix*.

Statistical Analysis. Statistical analysis was performed using Prism 5.0 (GraphPad) and programming language R. One-sided *t* test with unequal variances (Welch's test) was used for comparison of T cell population composition, and Mann–Whitney *U* test was used for comparison of repertoire properties. Details of the analysis can be found in *SI Appendix*.

Physicochemical CDR3 Properties and Correlation Analysis. CDR1 and CDR2 strength was calculated for mouse sequences according to http://imgt.org/ and the built-in VDJtools correspondence table (88). CDR3 strength was calculated as clonotype-weighted average, for the five amino acids in the middle of CDR3. All nonfunctional TCR sequences were excluded from analysis. Details of the analysis can be found in *SI Appendix*. The resulting *P* values were adjusted using Benjamini–Hochberg (BH) method with a false discovery rate not greater than 0.05.

Clustering Analysis. For graph construction, we take the top 2,500 most frequent clonotypes clustered by CDR3 nucleotide sequence identity out of each sample. Each node on the graph represents a single translated clonotype (clonotypes with same amino acid sequence but different nucleotide sequences do not collapse). Nodes were connected by edges if the Levenshtein distance between them was equal to zero or one and V segments of these clonotypes were identical. Further, we use only the top five biggest connected components out of the graph. Connectivity of repertoires was compared using Gini index and Mann–Whitney *U* test. Visualization of graph was made in Cytoscape.

For TCR motif detection, we separately pooled TCR clonotypes of all four H2-A^{b} and four H2-A^{j} mice and performed TCR neighborhood enrichment test, as described in ref. 89. Details of the analysis can be found in *SI Appendix*.

Principal Component Analysis. Data on the CDR3α and CDR3β repertoires were analyzed separately but using similar pipeline. For the principal component analysis we used physicochemical properties for the central five amino acids within CDR3 and other CDR3 parameters. Details of the analysis can be found in *SI Appendix*. All parameters were scaled to similar range using *Z*-score normalization; principal component analysis was performed and visualized using programming language R and package ggbiplot.

ACKNOWLEDGMENTS. We thank Prof. A. Rudensky (Sloan Kettering Institute) for providing the Foxp3^{GFP} reporter mice and T. Golovkina (The University of Chicago) for providing C57BL/6J *H2-Ob*$^{-/-}$ mice. We also thank M. Eisenstein for the English editing. Cell sorting experiments were carried out using the equipment provided by the Core facility in Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry. Sequencing and data analysis part of the work was supported by Russian Science Foundation [Grant 16-15-00149] while mouse breeding and flow cytometry data analysis was supported by Russian Foundation for Basic Research [Grant 19-015-00082 (to N.N.L.)].

1. K. A. Hogquist, S. C. Jameson, The self-obsession of T cells: How TCR signaling thresholds affect fate 'decisions' and effector function. *Nat. Immunol.* **15**, 815–823 (2014).
2. L. Racioppi, F. Ronchese, L. A. Matis, R. N. Germain, Peptide-major histocompatibility complex class II complexes with mixed agonist/antagonist properties provide evidence for ligand-related differences in T cell receptor-dependent intracellular signaling. *J. Exp. Med.* **177**, 1047–1060 (1993).
3. A. Sette, J. Alexander, K. Snoke, H. M. Grey, Antigen analogs as tools to study T-cell activation function and activation. *Semin. Immunol.* **8**, 103–108 (1996).
4. L. Yin, J. Scott-Browne, J. W. Kappler, L. Gapin, P. Marrack, T cells and their eons-old obsession with MHC. *Immunol. Rev.* **250**, 49–60 (2012).
5. J. A. Todd, J. I. Bell, H. O. McDevitt, HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* **329**, 599–604 (1987).
6. Z. Hovhannisyan *et al*., The role of HLA-DQ8 beta57 polymorphism in the anti-gluten T-cell response in coeliac disease. *Nature* **456**, 534–538 (2008).
7. M. Relle, A. Schwarting, Role of MHC-linked susceptibility genes in the pathogenesis of human and murine lupus. *Clin. Dev. Immunol.* **2012**, 584374 (2012).
8. G. Lythe, R. E. Callard, R. L. Hoare, C. Molina-París, How many TCR clonotypes does a body maintain? *J. Theor. Biol.* **389**, 214–224 (2016).
9. T. Dupic, Q. Marcou, A. M. Walczak, T. Mora, Genesis of the αβ T-cell receptor. *PLOS Comput. Biol.* **15**, e1006874 (2019).
10. A. J. Yates, Theories and quantification of thymic selection. *Front. Immunol.* **5**, 13 (2014).
11. R. D. Kilpatrick *et al*., Homeostasis of the naive CD4+ T cell compartment during aging. *J. Immunol.* **180**, 1499–1507 (2008).
12. E. G. Houston Jr., P. J. Fink, MHC drives TCR repertoire shaping, but not maturation, in recent thymic emigrants. *J. Immunol.* **183**, 7244–7249 (2009).
13. I. Stefanová, J. R. Dorfman, R. N. Germain, Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes. *Nature* **420**, 429–434 (2002).
14. J. S. Bridgeman, A. K. Sewell, J. J. Miles, D. A. Price, D. K. Cole, Structural and biophysical determinants of αβ T-cell antigen recognition. *Immunology* **135**, 9–18 (2012).
15. P. Marrack, J. P. Scott-Browne, S. Dai, L. Gapin, J. W. Kappler, Evolutionarily conserved amino acids that control TCR-MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).
16. A. Casrouge *et al*., Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *J. Immunol.* **164**, 5782–5787 (2000).
17. Q. Qi *et al*., Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13139–13144 (2014).
18. N. L. La Gruta, S. Gras, S. R. Daley, P. G. Thomas, J. Rossjohn, Understanding the drivers of MHC restriction of T cell receptors. *Nat. Rev. Immunol.* **18**, 467–478 (2018).
19. E. Sharon *et al*., Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* **48**, 995–1002 (2016).
20. K. Gao *et al*., Germline-encoded TCR-MHC contacts promote TCR V gene bias in umbilical cord blood T cell repertoire. *Front. Immunol.* **10**, 2064 (2019).
21. W. S. DeWitt 3rd *et al*., Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife* **7**, e38358 (2018).
22. C. S. Hsieh, Y. Zheng, Y. Liang, J. D. Fontenot, A. Y. Rudensky, An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.* **7**, 401–410 (2006).
23. M. Ono, R. J. Tanaka, Controversies concerning thymus-derived regulatory T cells: Fundamental issues and a new perspective. *Immunol. Cell Biol.* **94**, 3–10 (2016).
24. D. K. Cole *et al*., Human TCR-binding affinity is governed by MHC class restriction. *J. Immunol.* **178**, 5727–5734 (2007).
25. D. Silberman *et al*., Class II major histocompatibility complex mutant mice to study the germ-line bias of T-cell antigen receptors. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5608–E5617 (2016).
26. N. Logunova, M. Korotetskaya, V. Polshakov, A. Apt, The QTL within the H2 complex involved in the control of tuberculosis infection in mice is the classical class II H2-Ab1 gene. *PLoS Genet.* **11**, e1005672 (2015).

27. L. K. Denzin *et al*., Neutralizing antibody responses to viral infections are linked to the non-classical MHC class II gene H2-Ob. *Immunity* **47**, 310–322.e7 (2017).
28. Y. Gu, P. E. Jensen, X. Chen, Immunodeficiency and autoimmunity in H2-O-deficient mice. *J. Immunol.* **190**, 126–137 (2013).
29. E. D. Mellins, L. J. Stern, HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. *Curr. Opin. Immunol.* **26**, 115–122 (2014).
30. B. Yalcin *et al*., Commercially available outbred mice for genome-wide association studies. *PLoS Genet.* **6**, e1001085 (2010).
31. M. O. Li, A. Y. Rudensky, T cell receptor signalling in the control of regulatory T cell differentiation and function. *Nat. Rev. Immunol.* **16**, 220–233 (2016).
32. M. S. Jordan *et al*., Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat. Immunol.* **2**, 301–306 (2001).
33. J. D. Fontenot *et al*., Regulatory T cell lineage specification by the forkhead transcription factor foxp3. *Immunity* **22**, 329–341 (2005).
34. K. Attridge, L. S. Walker, Homeostasis and function of regulatory T cells (Tregs) in vivo: Lessons from TCR-transgenic Tregs. *Immunol. Rev.* **259**, 23–39 (2014).
35. M. Shugay *et al*., Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
36. D. A. Bolotin *et al*., MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
37. B. J. Manfras, D. Terjung, B. O. Boehm, Non-productive human TCR beta chain genes represent V-D-J diversity before selection upon function: Insight into biased usage of TCRBD and TCRBJ genes and diversity of CDR3 region length. *Hum. Immunol.* **60**, 1090–1100 (1999).
38. E. V. Putintseva *et al*., Mother and child T cell receptor repertoires: Deep profiling study. *Front. Immunol.* **4**, 463 (2013).
39. M. Izraelson *et al*., Comparative analysis of murine T-cell receptor repertoires. *Immunology* **153**, 133–144 (2018).
40. M. F. Quigley *et al*., Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 19414–19419 (2010).
41. E. S. Egorov *et al*., The changing landscape of naive T cell receptor repertoire with human aging. *Front. Immunol.* **9**, 1618 (2018).
42. A. S. Bergot *et al*., TCR sequences and tissue distribution discriminate the subsets of naïve and activated/memory Treg cells in mice. *Eur. J. Immunol.* **45**, 1524–1534 (2015).
43. S. Miyazawa, R. L. Jernigan, Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256**, 623–644 (1996).
44. A. Kosmrlj, A. K. Jha, E. S. Huseby, M. Kardar, A. K. Chakraborty, How the thymus designs antigen-specific and self-tolerant T cell receptor sequences. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16671–16676 (2008).
45. S. Chakrabarti, S. H. Bryant, A. R. Panchenko, Functional specificity lies within the properties and evolutionary changes of amino acids. *J. Mol. Biol.* **373**, 801–810 (2007).
46. N. K. Singh *et al*., Emerging concepts in TCR specificity: Rationalizing and (maybe) predicting outcomes. *J. Immunol.* **199**, 2203–2213 (2017).
47. Y. Feng *et al*., A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance. *Nature* **528**, 132–136 (2015).
48. G. De Simone *et al*., CXCR3 identifies human naive $CD8^+$ T cells with enhanced effector differentiation potential. *J. Immunol.* **203**, 3179–3189 (2019).
49. O. Grimsholm *et al*., The interplay between CD27(dull) and CD27(bright) B cells ensures the flexibility, stability, and resilience of human B cell memory. *Cell Rep.* **30**, 2963–2977.e6 (2020).
50. B. D. Stadinski *et al*., Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).
51. Z. Dembic, On integrity in immunity during ontogeny or how thymic regulatory T cells work. *Scand. J. Immunol.* **90**, e12806 (2019).
52. P. Billam *et al*., T Cell receptor clonotype influences epitope hierarchy in the CD8+ T cell response to respiratory syncytial virus infection. *J. Biol. Chem.* **286**, 4829–4841 (2011).
53. N. N. Logunova *et al*., Restricted MHC-peptide repertoire predisposes to autoimmunity. *J. Exp. Med.* **202**, 73–84 (2005).
54. L. Deng, R. J. Langley, Q. Wang, S. L. Topalian, R. A. Mariuzza, Structural insights into the editing of germ-line-encoded interactions between T-cell receptor and MHC class II by Vα CDR3. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14960–14965 (2012).
55. X. Chen, L. Poncette, T. Blankenstein, Human TCR-MHC coevolution after divergence from mice includes increased nontemplate-encoded CDR3 diversity. *J. Exp. Med.* **214**, 3417–3433 (2017).
56. P. Marrack *et al*., The somatically generated portion of T cell receptor CDR3α contributes to the MHC allele specificity of the T cell receptor. *eLife* **6**, e30918 (2017).
57. W. Luo *et al*., Limited T cell receptor repertoire diversity in tuberculosis patients correlates with clinical severity. *PLoS One* **7**, e48117 (2012).
58. E. Strønen *et al*., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
59. J. H. Cui *et al*., TCR repertoire as a novel indicator for immune monitoring and prognosis assessment of patients with cervical cancer. *Front. Immunol.* **9**, 2729 (2018).
60. A. K. Bentzen, S. R. Hadrup, Evolution of MHC-based technologies used for detection of antigen-responsive T cells. *Cancer Immunol. Immunother.* **66**, 657–666 (2017).
61. J. P. Scott-Browne, J. White, J. W. Kappler, L. Gapin, P. Marrack, Germline-encoded amino acids in the alphabeta T-cell receptor control thymic selection. *Nature* **458**, 1043–1046 (2009).
62. K. Falk, O. Rötzschke, S. Stevanović, G. Jung, H. G. Rammensee, Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351**, 290–296 (1991).
63. Z. Zavala-Ruiz, I. Strug, M. W. Anderson, J. Gorski, L. J. Stern, A polymorphic pocket at the P10 position contributes to peptide binding specificity in class II MHC proteins. *Chem. Biol.* **11**, 1395–1402 (2004).
64. C. A. Painter, L. J. Stern, Structural insights into HLA-DM mediated MHC II peptide exchange. *Curr. Top. Biochem. Res.* **13**, 39–55 (2011).
65. L. Yin, L. J. Stern, HLA-DM focuses on conformational flexibility around P1 pocket to catalyze peptide exchange. *Front. Immunol.* **4**, 336 (2013).
66. B. D. Stadinski *et al*., A role for differential variable gene pairing in creating T cell receptors specific for unique major histocompatibility ligands. *Immunity* **35**, 694–704 (2011).
67. B. D. Stadinski *et al*., Effect of CDR3 sequences and distal V gene residues in regulating TCR-MHC contacts and ligand specificity. *J. Immunol.* **192**, 6071–6082 (2014).
68. J. P. Scott-Browne *et al*., Expansion and function of Foxp3-expressing T regulatory cells during tuberculosis. *J. Exp. Med.* **204**, 2159–2169 (2007).
69. M. Kursar *et al*., Cutting edge: Regulatory T cells prevent efficient clearance of Mycobacterium tuberculosis. *J. Immunol.* **178**, 2661–2665 (2007).
70. M. P. Davenport, D. A. Price, A. J. McMichael, The T cell repertoire in infection and vaccination: Implications for control of persistent viruses. *Curr. Opin. Immunol.* **19**, 294–300 (2007).
71. H. A. Van Den Berg, C. Molina-París, A. K. Sewell, Specific T-cell activation in an unspecific T-cell repertoire. *Sci. Prog.* **94**, 245–264 (2011).
72. L. F. Su, B. A. Kidd, A. Han, J. J. Kotzin, M. M. Davis, Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity* **38**, 373–383 (2013).
73. S. C. Clute *et al*., Broad cross-reactive TCR repertoires recognizing dissimilar Epstein-Barr and influenza A virus epitopes. *J. Immunol.* **185**, 6753–6764 (2010).
74. M. E. Birnbaum *et al*., Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157**, 1073–1087 (2014).
75. D. Mason, A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395–404 (1998).
76. Y. Sun *et al*., Specificity, privacy, and degeneracy in the CD4 T cell receptor repertoire following immunization. *Front. Immunol.* **8**, 430 (2017).
77. A. A. Bogan, K. S. Thorn, Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
78. J. Henwood *et al*., Human T cell repertoire generation in the absence of MHC class II expression results in a circulating CD4+CD8- population with altered physicochemical properties of complementarity-determining region 3. *J. Immunol.* **156**, 895–906 (1996).
79. L. K. Denzin, N. F. Robbins, C. Carboy-Newcomb, P. Cresswell, Assembly and intracellular transport of HLA-DM and correction of the class II antigen-processing defect in T2 cells. *Immunity* **1**, 595–606 (1994).
80. A. B. Vogt, H. Kropshofer, G. Moldenhauer, G. J. Hämmerling, Kinetic analysis of peptide loading onto HLA-DR molecules mediated by HLA-DM. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9724–9729 (1996).
81. P. P. Nanaware, M. M. Jurewicz, J. D. Leszyk, S. A. Shaffer, L. J. Stern, HLA-DO modulates the diversity of the MHC-II self-peptidome. *Mol. Cell. Proteomics* **18**, 490–503 (2019).
82. M. Liljedahl *et al*., Altered antigen presentation in mice lacking H2-O. *Immunity* **8**, 233–243 (1998).
83. G. Kenty, W. D. Martin, L. Van Kaer, E. K. Bikoff, MHC class II expression in double mutant mice lacking invariant chain and DM functions. *J. Immunol.* **160**, 606–614 (1998).
84. S. Viville *et al*., Mice lacking the MHC class II-associated invariant chain. *Cell* **72**, 635–648 (1993).
85. C. E. Grubin, S. Kovats, P. deRoos, A. Y. Rudensky, Deficient positive selection of CD4 T cells in mice displaying altered repertoires of MHC class II-bound self-peptides. *Immunity* **7**, 197–208 (1997).
86. R. M. Nepal, B. Vesosky, J. Turner, P. Bryant, DM, but not cathepsin L, is required to control an aerosol infection with Mycobacterium tuberculosis. *J. Leukoc. Biol.* **84**, 1011–1018 (2008).
87. N. N. Logunova *et al*., Data from Bioproject: Sequencing of TCR repertoires in mice with MHC-II polymorphism. National Center for Biotechnology Information Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA600142. Deposited 9 January 2020.
88. M. Shugay *et al*., VDJtools: Unifying post-analysis of T cell receptor repertoires. *PLOS Comput. Biol.* **11**, e1004503 (2015).
89. M. V. Pogorelyy, M. Shugay, A framework for annotation of antigen specificities in high-throughput T-cell repertoire sequencing studies. *Front. Immunol.* **10**, 2159 (2019).

2.3. Heterogeneity of Th phenotypes and their involvement in response to acute viral infections

2.3.1. Manuscript2: Sort-Seq: immune repertoire-based scRNA-Seq systematization

Sort-Seq: immune repertoire-based scRNA-Seq systematization

Kriukova V.V.[1,2,*], Lukyanov D.K.[2,*], Shagina I.A.[2,3,*], Sharonov G.V.[2,3,4], Lupyr K.R.[3], Blagodatskikh K.A.[3], Staroverov D.B.[2,3], Ladell K.[5], Miners K.L. [5], Salnikova M.A.[2], Nikolaev R.V.[6], Shelikhova L.[6], Maschan M.A.[6], Britanova O.V.[2,3], Franke A.[1], Price D.A.[5,7], Chudakov D.M.[2,3,8,9]

[1] Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany
[2] Genomics of Adaptive Immunity Department, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia
[3] Institute of Translational Medicine, Pirogov Russian National Research Medical University, Moscow, Russia
[4] Institute of Experimental Oncology and Biomedical Technologies, Privolzhsky Research Medical University, Nizhny Novgorod, Russia
[5] Division of Infection and Immunity, Cardiff University School of Medicine, University Hospital of Wales, Cardiff, United Kingdom
[6] Dmitry Rogachev National Medical Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia.
[7] Systems Immunity Research Institute, Cardiff University School of Medicine, University Hospital of Wales, Cardiff, United Kingdom
[8] Central European Institute of Technology, Brno, Czech Republic
[9] Abu Dhabi Stem Cell Center, Al Muntazah, United Arab Emirates

Abstract

The functional programs chosen by B and T cell clones fundamentally determine the architecture of immune response to distinct challenges. Advances in scRNA-Seq have improved our understanding of the diversity and stability of these programs, but it has proven difficult to link this information with known lymphocyte subsets. Here, we introduce Sort-Seq, an immune repertoire-based method that allows exact positioning of phenotypically defined lymphocyte subsets within scRNA-Seq data. Sort-Seq outperformed CITE-Seq for accurate mapping of the classical CD4$^+$ T helper (Th) cell subsets (Th1, Th1-17, Th17, Th22, Th2a, Th2, and Treg), offering a more powerful approach to the surface phenotype-based scRNA-Seq classification of adaptive lymphocyte subpopulations. Using integrated scRNA-Seq, Sort-Seq, and CITE-Seq data from 122 donors, we provide a comprehensive Th cell scRNA-Seq reference map. Exploration of this dataset revealed the low plasticity and extreme sustainability of the Th17, Th22, Th2, and Th2a cell programs over years. We also develop Cultivation-based Antigen-specific T cell identificatoR in Replicates (CultivAToRR), which identified >80 SARS-CoV-2-specific CD4$^+$ TCRβ clonotypes in a single donor across a wide frequency range. We complemented these results with frequency-based capturing of COVID-19-responsive clonotypes and screening against known SARS-CoV-2-specific TCRs. Positioning within the annotated scRNA-Seq map revealed functional subtypes of Th cell clones involved in primary and secondary responses against SARS-CoV-2. The ability to capture low-frequency antigen-specific T cell clones in combination with Sort-Seq-based scRNA-Seq annotation creates an integral pipeline that links challenge-responsive clones with their exact functional subtypes, providing a solid foundation for investigating T cell roles in healthy and pathological immune responses and vaccine development.

Introduction

Adaptive immunity controls pathogen clearance, tissue homeostasis, and vaccine efficacy. These processes depend largely on antigen immunogenicity, the functionality of cognate naive and/or memory B and T cells, and the type of acquired immune response.

Clonal populations of $CD4^+$ T helper (Th) cells essentially orchestrate the course of an immune response via specific interactions with peptide epitopes presented in complex with major histocompatibility class II (MHCII) molecules. Their functional and antigen-specific diversity allows them to guide both classical (B cells, dendritic cells, macrophages) and non-classical (endothelial cells, epithelial cells, granulocytes) antigen-presenting cells to optimize effector functionality[1, 2, 3]. As a consequence, inappropriate Th responses to certain antigens have been associated with impaired pathogen clearance[4, 5, 6], inefficient responses to vaccination[7], acute and chronic hypersensitivity and inflammation[8, 9, 10], inflammaging[11], autoimmunity[12, 13, 14, 15], and cancer[16]. Accordingly, when investigating T cell responses, it is critical not only to quantify the magnitude of the antigen-specific T cell clonal expansion but also to understand functional programs, and related phenotypes of the responding and memory helper T cells.

Single-cell RNA-Seq (scRNA-Seq) techniques are shedding light on the true diversity of Th cell programs[17, 18, 19], among which classical subsets described on the basis of cytokine release profiles and patterns of surface marker expression have to find their place. However, the expression of transcripts encoding characteristic surface markers is often low in scRNA-Seq datasets, with indirect correlations between mRNA abundance and protein density[20, 21]. This problem can be overcome to some extent by the incorporation of protein-level expression data into single-cell experiments, a task that is essentially implemented in CITE-Seq technology, which makes use of barcoded antibodies directed against markers of interest expressed on the cell surface[22, 23, 24].

Here, we report an alternative approach to the surface protein-based mapping of lymphocyte subsets to scRNA-Seq data, termed Sort-Seq. This approach makes use of immune repertoires from sorted lymphocyte subsets, which are then mapped to scRNA-Seq+scBCR/TCR-Seq datasets obtained from the same donors, employing natural barcodes in the form of sequence-defined BCRs/TCRs. We demonstrate that this method outcompetes CITE-Seq for the accurate mapping of phenotypically defined Th subsets. We next integrate scRNA-Seq, Sort-Seq, and CITE-Seq outputs of 122 donors to provide a Th scRNA-Seq reference dataset and employ it to analyze long-term Th subsets stability and plasticity at the scRNA-Seq level.

In addition, dataset and improved method for the identification of low-frequency antigen-specific T cell clones among PBMCs, termed Cultivation-based Antigen-specific T cell identificatoR in Replicates (CultivAToRR).

These approaches in combination allowed the identification of functional subsets of the SARS-CoV-2-responsive Th cell clonotypes via scRNA-Seq. Collectively, our results validate a seamless integration of wet laboratory data and bioinformatic tools, enabling unique insights into the true nature of the adaptive immune system.

Results

Sort-Seq

Functional subtypes of human helper T cells, such as Th1, Th2, Th17, Th1-17, Th22, Tfh, and Treg, are classically distinguished based on the surface markers, such as CD127, CD25, CCR10, CXCR5, CXCR3, CCR6, CCR4, and CRTh2 (e.g. **Fig. 1**, middle panel).

Here we conceptualized the precise annotation of Th cells in scRNA-Seq data powered by overlapping scTCR-Seq repertoires and TCR repertoires of FACS-sorted Th cell subsets (**Fig. 1**). We hypothesized that, due to relatively high phenotypic stability and low plasticity between the helper T cell subsets[25], this approach would accurately map phenotypically defined subsets within scRNA-Seq data.

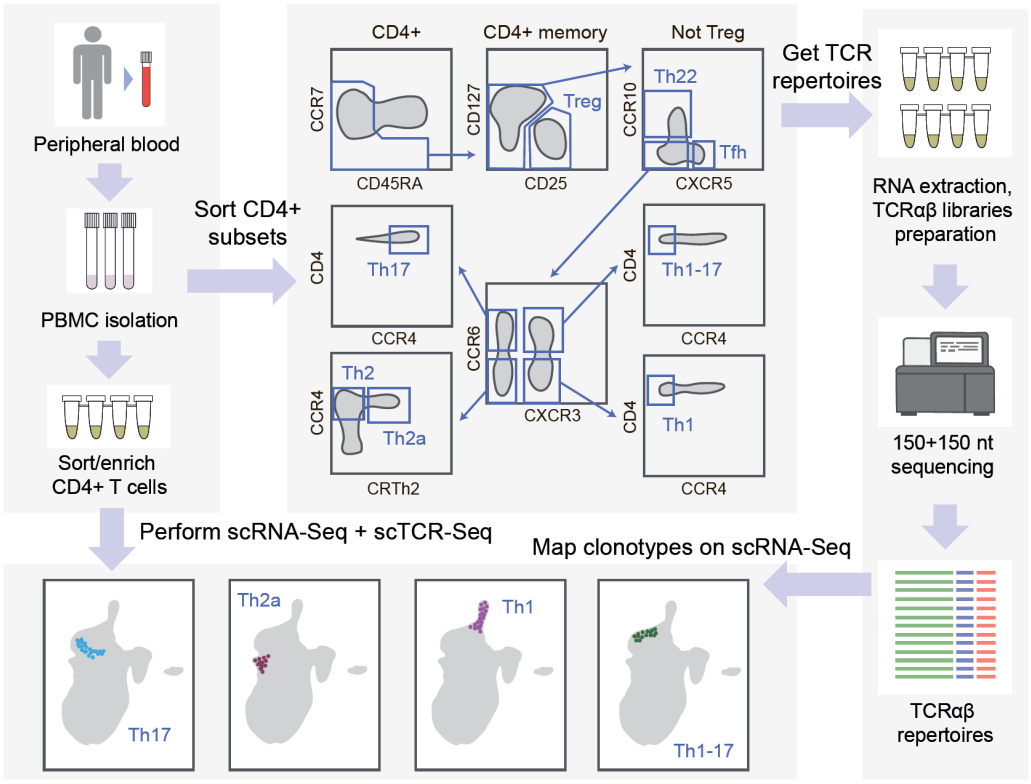

Figure 1. Sort-Seq concept. T cell or B cell subsets of interest are stained and FACS-sorted using conventional surface markers from a sample of interest, such as PBMCs, lymph nodes, or tumor tissue. Immune receptor repertoires are obtained for each sorted subset and mapped to the scRNA-Seq data obtained for the same patient or animal. Complementary scTCR-Seq or scBCR-Seq data is used to position clones of FACS-sorted subsets within scRNA-Seq. The figure is built to illustrate the content of the current work. The same concept is applicable for mapping various T cell and B cell subsets in human and model animal scRNA-Seq data.

To verify this concept, named Sort-Seq, we exploited previously obtained TCRα and TCRβ repertoires of Th1, Th17, Th1-17, Th22, Th2, Th2a, Treg, and Tfh CD4+ T cell subsets thoroughly sorted from peripheral blood of healthy donors[25]. Three participants of the latter study were available for repeated blood donation at the moment of the current study. We performed scRNA-Seq and scTCR-Seq profiling from their sorted bulk effector/memory CD4+ T cells (gated as CD4+, NOT CCR7+CD45RA+ cells to deplete naive T cells). The obtained scRNA-Seq data has then been integrated with the CD4+ scRNA-Seq reported in Ref. 26 in order to: 1) increase the power of the downstream analysis such as clustering and UMAP visualization on a larger number of donors and 2) enable comparison of our approach with the CITE-Seq method employed in the latter work. This resulted in a reference dataset composed of 147,677 cells (**Fig. 2a**), without notable donor-specific or study-specific batch effects (**Supplementary Fig**. **1a**).

In order to link scRNA-Seq clusters to the classic Th phenotypes, we mapped TCR clonotypes of the sorted Th subsets using scTCRs as natural barcodes. Remarkably, T cell clones from each of the sorted Th subpopulations formed clearly defined spots on the scRNA-Seq UMAP (**Fig. 2b,c,d, Supplementary Fig**. **1b,c**).

Of note, most of the scRNA-Seq clusters showed low clonality in the scTCR-Seq data, thus excluding clonal biases in the Sort-Seq-based annotation (**Supplementary Fig**. **1e**). High clonality was exclusively observed for the cluster annotated as Temra cytotoxic Th1 subset, which is in line with the previous reports[27].

We next attempted to annotate the same surface-marker defined Th subsets within scRNA-Seq using CITE-Seq. For this, we used CITE-Seq data from Ref. 26, and performed sequential *in silico* sorting-like gating as shown in **Supplementary Fig**. **2**.

Altogether, Sort-Seq resulted in much more accurate cluster annotations compared to CITE-Seq (**Fig**. **2b-d**, **Supplementary Figs**. **1**,**3**). Both approaches yielded similar accuracy of cluster annotations for certain subsets, such as Tregs, defined by $CD25^{high}CD127^{low}$ gating. CITE-Seq based annotation also readily distinguished Th17 and Th22 subsets. At the same time, CITE-Seq failed to annotate Th1 (gated as NOT $CD25^{high}CD127^{low}$, CCR10-CCR6-CXCR3+CCR4-) and poorly annotated Th2 and Th2a subsets (gated as NOT $CD25^{high}CD127^{low}$, CCR10-CCR6-CXCR3-CCR4+ and CRTh2- or CRTh2+, respectively), presumably due to poor surface expression of the corresponding markers. We concluded that, due to the high accuracy of FACS cell sorting, Sort-Seq approach is preferable for scRNA-Seq cluster annotation to find accurate correspondence with the studied T cell subpopulations of interest.

Based on the integrated data of scRNA-Seq, Sort-Seq, and CITE-Seq outputs of 122 donors and taking into account findings and considerations of Sakaguchi and colleagues[18], we propose a comprehensive Th scRNA-Seq reference map with 16 major Th clusters (**Fig. 2a, Supplementary Table 1**). Owing to Sort-Seq-based classification, it for the first time accurately positions Th2 and Th2a subsets. Classic Th1, Th17, Th22, as well as Temra cytotoxic[27, 28, 29], but not Th1-17 (see below) subsets also find their exact locations within the reference scRNA-Seq dataset. “Eff-mem IFN response” cluster is nearly absent in healthy donors’ blood but is well-detectable in moderate and severe COVID patients (**Supplementary Fig. 4**), probably representing the typical behavior of Th cells in acute viral infection[30]. Tfh subset is found in two dissimilar clusters differing in the expression of CXCR5 and CXCR3. This points at functional heterogeneity of Tfh subset[31], which warrants a deeper focused investigation.

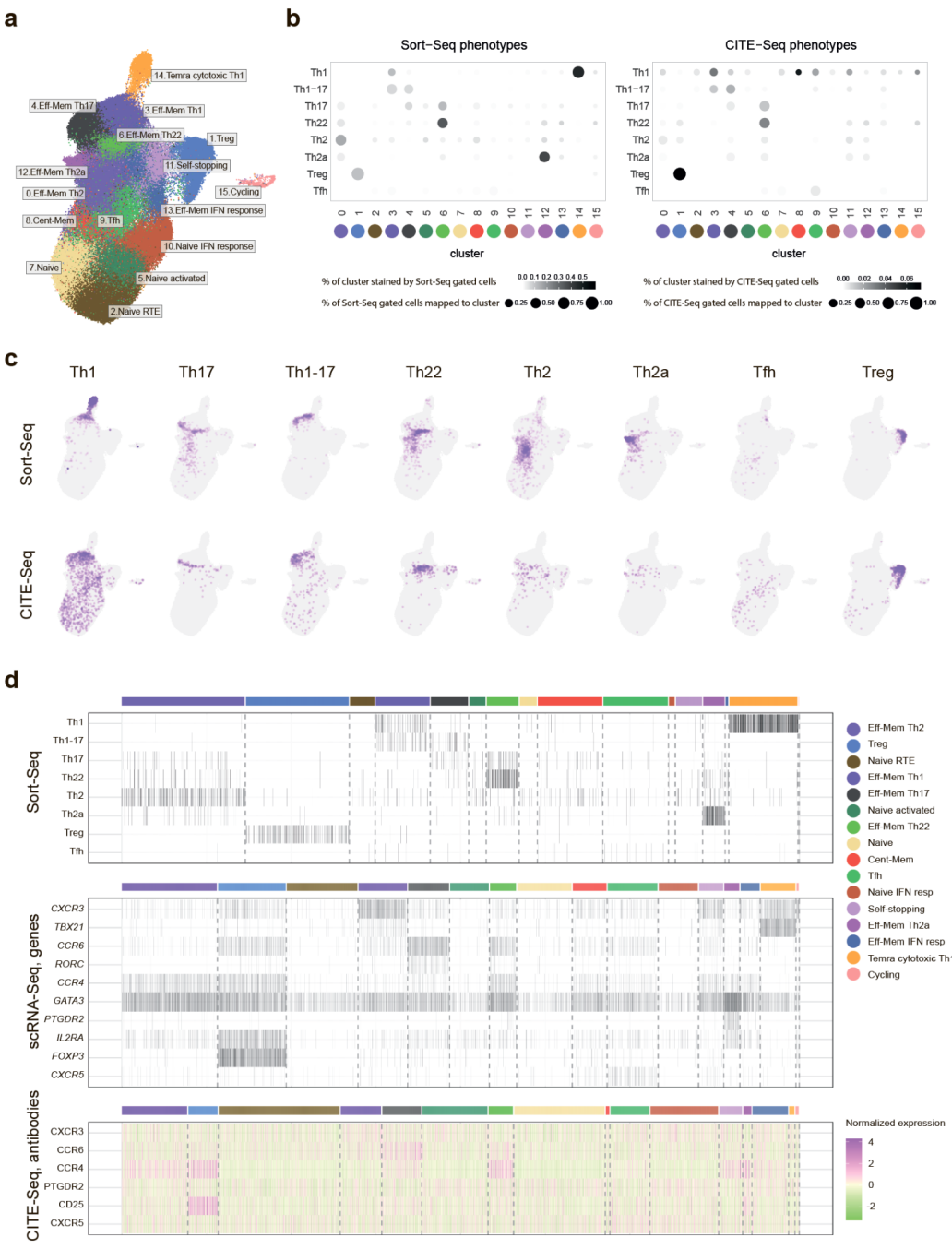
a
14.Temra cytotoxic Th1
4.Eff-Mem Th17
3.Eff-Mem Th1
6.Eff-Mem Th22
1.Treg
12.Eff-Mem Th2a
11.Self-stopping
0.Eff-Mem Th2
15.Cycling
13.Eff-Mem IFN response
8.Cent-Mem
9.Tfh
10.Naive IFN response
7.Naive
5.Naive activated
2.Naive RTE
b
Sort−Seq phenotypes
CITE−Seq phenotypes
Th1
Th1−17
Th17
Th22
Th2
Th2a
Treg
Tfh
cluster
% of cluster stained by Sort-Seq gated cells
% of Sort-Seq gated cells mapped to cluster
% of cluster stained by CITE-Seq gated cells
% of CITE-Seq gated cells mapped to cluster
c
Th1
Th17
Th1-17
Th22
Th2
Th2a
Tfh
Treg
Sort-Seq
CITE-Seq
d
Sort-Seq
scRNA-Seq, genes
CITE-Seq, antibodies
CXCR3
TBX21
CCR6
RORC
CCR4
GATA3
PTGDR2
IL2RA
FOXP3
CXCR5
CD25
Eff-Mem Th2
Treg
Naive RTE
Eff-Mem Th1
Eff-Mem Th17
Naive activated
Eff-Mem Th22
Naive
Cent-Mem
Tfh
Naive IFN resp
Self-stopping
Eff-Mem Th2a
Eff-Mem IFN resp
Temra cytotoxic Th1
Cycling
Normalized expression

Figure 2. Mapping the classic Th subsets with scRNA-Seq. a. UMAP visualization of the reference scRNA-Seq dataset of peripheral blood Th cells. Dataset built via Seurat integration of publicly available and our scRNA-Seq data. The proposed classification is based on previous knowledge and findings of the current work. **b**. Dot plots summarizing the positioning of sorted (Sort-Seq) and *in silico* gated (CITE-Seq) Th subsets within the scRNA-Seq clusters shown on the panel (c). For normalization, we used 20,000 randomly selected scRNA-Seq cells with matched CITE-Seq or Sort-Seq data for each plot. Dot intensity shows the stained proportion of the scRNA-Seq cluster. Dot size shows the proportion of Sort-Seq-identified or proportion of *in silico* CITE-Seq-based gated scRNA-Seq cells mapped to the scRNA-Seq cluster. **c**. UMAP plots showing the localization of Sort-Seq and CITE-Seq defined subsets. TCRβ clonotypes were used to define Th subsets in Sort-Seq method. Expression of the surface markers was used to gate Th subsets in the CITE-Seq-based annotation. The color intensity in Sort-Seq is proportional to the clonal frequencies in the original sorted Th bulk TCRβ repertoires. **d**. Heatmap summarizing the positioning of Sort-Seq subsets, scRNA-Seq expression of characteristic genes, and CITE-Seq signal for the corresponding surface proteins. For normalization, equal numbers of 10,000 scRNA-Seq cells were randomly selected from Sort-Seq and CITE-Seq experimental datasets. Each tile of the heatmap represents one cell. The color intensity in the scRNA-Seq plot visualizes the gene expression in a cell (white color depicts zero expression). Protein expression measured by CITE-Seq was Z-scored, and all values exceeding 99.5 percentile were trimmed.

Phenotypic versus intrinsic program plasticity of Th cells

To evaluate the relationships and cross-subset plasticity of T cell clones, we analyzed cluster stability at different clustering resolution levels (**Supplementary Fig. 5a**) and clonal intersections between the clusters (**Supplementary Fig**. **5b**). We previously suggested notable plasticity between Th22/Th17, Th17/Th2, and Th2/Th2a subsets based on the corresponding intersections of sorted Th cell subset repertoires[25]. However, the new data on clonal overlaps and cluster stability at various resolutions in the current study prompts us to reconsider these interpretations.

Indeed, T cell clones sorted as classic Th17 subset (gated as NOT $CD25^{high}CD127^{low}$, CCR10-CCR6+CXCR3-CCR4+) are found in both Th17 and Th22 scRNA-Seq clusters. However, clonal overlap between these clusters is low while the stability of both clusters is high (**Supplementary Figs**. **5a**,**b**). Furthermore, clones sorted as classic Th22 (gated as NOT $CD25^{high}CD127^{low}$, CCR10+) are almost exclusively found in Th22 scRNA-Seq cluster. This indicates that while certain Th22 clones may exhibit similar surface markers to Th17 clones due to the phenotypic plasticity, their intrinsic programs remain stable. In other words, our data suggests that Th17 clones mainly remain Th17, and Th22 mainly remain Th22. The independent origin of human Th22 clones was initially suggested[32] and reported in mouse models[33, 34]. Distinct TCR repertoire features also suggested the existence of clonally discrete Th22 subset[25]. According to this logic, the population classically sorted by phenotypic markers and described as Th17, represents the mixture of *bona fide* Th17 and *bona fide* Th22 cells. In contrast, classically sorted CCR10+ Th22 subsets almost fully coincide with the corresponding scRNA-Seq cluster, representing an almost pure population of uniformly programmed T cells (**Supplementary Table 2**).

Similarly, based on this data we suggest the self-standing nature of Th17 and Th1 subsets. T cell clones sorted as Th1-17 subset were localized in both of these scRNA-Seq clusters, but clonal overlap between Th17 and Th1 clusters is low, and stability of both clusters is high (**Supplementary Fig**. **5a**,**b**). This data suggests the existence of stable clonal T cell populations that carry either Th17 or Th1 programs, and that only their phenotypic features but not the true underlying program could change. The Th1-17 subset, in this paradigm, does not exist as a stable combo of the two major programs of Th cells (**Supplementary Table 2**).

Th2a cluster appears to be self-standing: it is stable at different clustering resolutions, and is clonally well separated from Th2 cluster (**Fig. 2**, **Supplementary Figs**. **5a**,**b**). In contrast, Th1 and Temra cytotoxic Th1 clusters demonstrate notable clonal exchange (**Supplementary Fig**. **5b**), indicating that the former could convert into the latter, as some of the recent works suggest[35, 36].

In the proposed scRNA-Seq classification shown in **Fig**. **2a**, we give priority to the cells partitioning into stable scRNA-Seq clusters (**Supplementary Fig**. **5a**), while Sort-Seq data are used to match classical surface phenotype-based sorted T cell subsets with those clusters.

Of note, T cell repertoires of the sorted subsets were obtained in 2018, while scRNA-Seq experiment was performed in 2022. Despite the 4-year distance, Sort-Seq mapped sorted clones to clearly defined positions, generally limited to one or two neighboring scRNA-Seq clusters (**Fig**. **2**, **Supplementary Fig**. **1**), indicating long-term stability of a program once chosen by an activated T cell clone.

COVID-specific TCRs discovery

To test Sort-Seq capability to determine the phenotypes of antigen-responsive T cells, we focused on searching SARS-CoV2-specific CD4+ T cell clones in donor D11, who was vaccinated with adenoviral SARS-CoV2 vaccine in Dec2020/Jan2021 and got two registered, PCR-confirmed SARS-CoV2 infections in April 2021 and January 2022 (**Supplementary Note 1, Supplementary Fig**. **6, Supplementary Table 3**).

To identify TCR variants responding to SARS-CoV2 antigens for this donor, we implemented several approaches:

A. Cultivation-based Antigen-specific T cell identificatoR in Replicates method (CultivAToRR).

In this assay, we relied on the concept of antigen-specific T cell expansion, followed by TCR repertoire profiling[37, 38]. This approach, either used in bulk[38] or focused on dividing T cells that progressively dilute cytoplasmic dyes such as carboxyfluorescein succinimidyl ester (CFSE)[37], allows the identification of T cell clones that selectively proliferate in presence of particular antigens and consequently become enriched within the TCR repertoire.

Notably, the random nature of cell sampling and further interactions between T cells and potent antigen-presenting cells in a culture, as well as the varying proliferation potential of individual memory T cells, substantially limits applicability of such approaches to relatively large clonal T cell expansions[38]. To account for the sampling issue, we exploited our previous experience with time-lapse T cell clonal tracking[39, 40] and identification of T cell clones that are locally expanded in tumor sections[41]. In both scenarios, we obtained independent biological (at the level of cells) replicates for the TCR profiling, followed by identification of reproducibly expanded clones, where we also accounted for the sampling noise defined as variability between the replicates.

Here, we cultured CFSE-labeled D11 PBMCs in the presence of three distinct mixtures of antigenic peptides, each in three independent replicates. After 7 days, divided ($CFSE^{low}$) CD4+ and CD8+ subsets were sorted from each replica, yielding 224-7,322 T cells, followed by the sensitive TCR library preparation technique. Obtained TCRβ CDR3 repertoires have been then analyzed with the pipeline based on EdgeR, a widely used method to identify differentially expressed genes in RNA-Seq[42]. Here, the same EdgeR algorithm allowed us to identify TCR CDR3 variants that were reproducibly enriched within repertoires of

replicates cultured in the presence of the corresponding mixture of antigenic peptides, compared to the other mixtures of antigens (following the same logic as in Ref. 39). Those antigenic peptide mixtures included Miltenyi peptivators comprising the overlapping 15mers of 1) SARS-CoV-2 spike protein (S) and 2) pooled membrane glycoprotein (M) + nucleoprotein (N), and 3) a mixture of irrelevant control peptides. See **Fig**. **3a** for the overview of the CultivAToRR approach.

EdgeR analysis of obtained repertoires revealed 93 TCRβ clonotypes that specifically and reproducibly proliferated *in vitro* in the presence of S (28 CD4 and 4 CD8 clonotypes) or M+N (59 CD4 and 2 CD8 clonotypes) peptivators. This bias towards CD4+ T cells could be explained by the nature of 15mers peptivators, since proper processing towards shorter 9-10mers that are presented by MHCI to CD8 T cells may not always proceed efficiently, the question that requires further investigation.

The frequencies of these clonotypes were tracked within the deep peripheral TCRβ repertoires of the donor, which were obtained at multiple time points. The analysis revealed that these clonotypes expanded in the donor's blood exclusively on two time points, which coincided with one or both SARS-CoV-2 infections (**Fig**. **3b**), confirming the accuracy of the CultivAToRR approach. A group of CD4+ Spike-specific TCRβ CDR3s (CASQEGVSNQPQHF, CASSEGASNQPQHF, CASSEGSSSQPQHF, TRBV6-1/TRBJ1-5) was identical or highly homologous to recently described TCRs specific to Spike epitope NLLLQYGSFCTQLNRAL in DRB1*15:01 context[43], also carried by D11. One of the CD4+ Spike-specific clonotypes that first appeared after vaccination also responded to the 1st SARS-CoV-2 infection and then dominated in the 2nd SARS-CoV-2 infection, suggesting vaccine-mediated protection (CASSQDSQNLGGSYEQYF, TRBV4-2/TRBJ2-7, **Fig**. **3b**).

Notably, the frequencies of each of these clonotypes were relatively low, mostly below 0.1% of the whole peripheral repertoire at the peak of response to the 2nd infection. This frequency is generally below the sensitivity of MHC tetramer staining or antigen-specific activation assays[44]. This result demonstrates the potency of our method, which integrates antigen-specific T cell culturing, replicates, and comparative data post-analysis, in identifying low-frequency T cell clones specific to the antigens of interest.

B. Clonal expansion in corresponding time points.

Although CultivAToRR approach allowed us to identify more than 90 SARS-CoV-2-specific TCRβ variants for D11, its results were limited to the three tested antigens, and biased towards CD4+ TCRs, thus essentially missing the CD8+ TCRs as well as response to non-M/N/S epitopes. To widen our fishing net, we applied a frequency-based approach to identify the TCRβ variants that specifically expanded at the time points of vaccination or SARS-CoV-2 infection in the bulk peripheral TCRβ repertoire of D11, a technique applied previously to reveal vaccine-responsive[39] or pathogen-specific[40, 45] T cell clones.

Again, edgeR analysis was applied to the bulk peripheral TCRβ repertoires, obtained in biological replicates and normalized to the 45,000 randomly sampled UMI (unique TCRβ cDNA molecules) per replicate per time point. This analysis revealed 8, 14 and 70 TCRβ clonotypes expanding at the time points of vaccination, 1st or 2nd SARS-CoV-2 infections, respectively, compared to the three reference time points (**Fig**. **3c**,**d**). A number of clonotypes that were identified as expanded at the time point of the 1st SARS-CoV-2 infection were also expanding at the 2nd infection, and *vice versa*. At the same time, similar analysis revealed only from 1 to 2 clonal expansions in the control time points compared to the challenge time points (**Supplementary Fig**. **7a**). This suggests that most of the edgeR-identified clonotypes that were expanded in peripheral blood in the challenge time points are truly associated with the corresponding challenges.

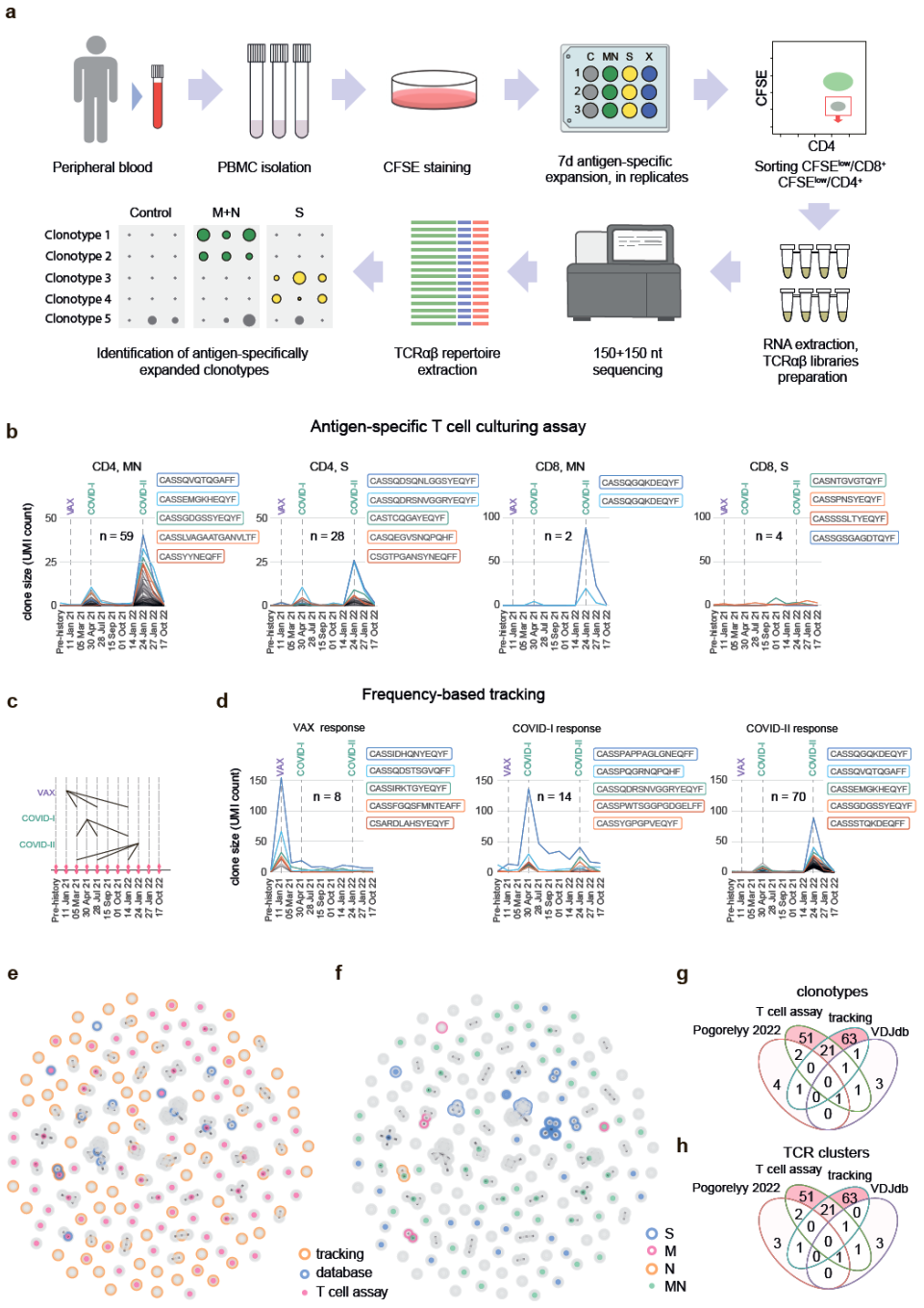
a
Peripheral blood
PBMC isolation
CFSE staining
C MN S X
7d antigen-specific expansion, in replicates
CFSE
CD4
Sorting CFSE^low/CD8+ CFSE^low/CD4+
RNA extraction, TCRαβ libraries preparation
150+150 nt sequencing
TCRαβ repertoire extraction
Control M+N S
Clonotype 1
Clonotype 2
Clonotype 3
Clonotype 4
Clonotype 5
Identification of antigen-specifically expanded clonotypes
b
Antigen-specific T cell culturing assay
CD4, MN
CD4, S
CD8, MN
CD8, S
clone size (UMI count)
n = 59
n = 28
n = 2
n = 4
c
d
Frequency-based tracking
VAX response
COVID-I response
COVID-II response
n = 8
n = 14
n = 70
e
f
g
clonotypes
T cell assay
tracking
Pogorelyy 2022
VDJdb
h
TCR clusters
tracking
database
T cell assay
S
M
N
MN

Figure 3. Three methods of capturing COVID-specifc TCR clonotypes. a. Scheme of the antigen-specific T cell culturing approach, CultivAToRR. **b**,**d**. Longitudinal tracking of TCRβ clonotypes (CDR3nt + V + J) identified with CultivAToRR (b) and based on expansion of frequency in target time points (d), using PBMC TCRβ profiling in replicates in sequential time points. Each PBMC TCR-Seq library is downsampled to 45,000 TCRβ encoding cDNA molecules (UMI). TCRβ CDR3 amino acid sequences are shown for the five largest clonotypes on each plot. **c**. Schematic representation of target and reference time points for frequency-based discovery of expanding TCRβ clonotypes. **e**,**f**. TCRβ CDR3 homology clusters identified by the three methods. Each node represents a single CD3aa + V + J clonotype, nodes having one amino acid mismatch in CDR3β are connected with edges. Method (e) and target antigen, if known (f) are shown. **g**,**h**. Venn diagrams representing clonal (g) and CDR3 homology cluster (h) overlaps between the methods.

C. Screening against known SARS-CoV-2-specific TCRs.

To exploit the accumulated knowledge of SARS-CoV-2-specific TCRs with known target peptides and/or restricting human leukocyte antigens (HLAs), we also overlapped peripheral time-tracked TCRβ repertoires of our donor with the VDJdb database[46] and the dataset of CD4+ TCRs enriched in SARS-CoV-2 patients with computationally implied HLA allele specificities[47]. We narrowed our search to the HLA class I and II alleles matching the donor's HLA (**Supplementary Note 1**). This analysis revealed 7 and 8 clonotypes matching those listed in VDJdb and Pogorelyy *et al.*, respectively. The largest expansions of identified clonotypes were observed at time points corresponding to infection with SARS-CoV-2 (**Supplementary Fig**. **7b**,**c**).

Finally, we integrated the data on SARS-CoV-2-specific TCRβ clonotypes obtained using the three approaches: 1) antigen-specific culturing; 2) frequency-based capturing; 3) database search (**Supplementary Table 4**). To account for the presence of convergent CDR3 groups[48] we built single aa-mismatch TCRβ CDR3 homology clusters on the pooled 1)+2)+3) data, visualized on **Fig 3e**,**f**. Notably, a number of CDR3 homology clusters and clonotypes were independently supported by 2 or all 3 of the approaches, cross-confirming their specificity (**Fig 3e-h**).

Mapping functionality of SARS-CoV-2-responsive T cell clones

To functionally characterize SARS-CoV-2-responsive Th cell clones, we performed scRNA-Seq of effector/memory Th cells at the time points right after COVID-I and COVID-II for the same donor D11. We next mapped D11 scRNA-Seq data on the bigger landscape of scRNA-Seq data integrated from 122 donors and assigned each D11 T cell to the corresponding scRNA-Seq cluster (**Supplementary Fig**. **8a**). Next, we highlighted SARS-CoV-2-responsive TCRβ clonotypes identified in the previous section within the scTCR-Seq of D11 (**Fig**. **4**, **Supplementary Fig**. **8b**).

17 of identified SARS-CoV-2-responsive clonotypes of D11 have been successfully mapped on Th cells localized in several scRNA-Seq clusters. Most of the cells represented "Self-stopping" (*PD-1+, TIGIT+*), Th2, Th2a, and Th17 clusters (**Fig**. **4b, Supplementary Fig**. **8c**). Most cells recognized as responding to COVID-I based on frequency expansion were also assigned to "Self-stopping", Th2, Th2a, and Th17 clusters. Most cells responding to the COVID-II as well as identified by antigen-specific culturing assay were assigned to the "Self-stopping" cluster (**Fig**. **4a**,**b, Supplementary Fig**. **8c**).

Notably, none of the SARS-CoV-2-responsive clonotypes have been identified among Th1, Temra-Th1, or "Eff-Mem IFN response" clusters. We can suggest that the Th response to COVID-I (moderate illness)

was suboptimal in the context of viral infection. The presence of “Self-stopping” SARS-CoV-2-specific T cells and their higher clonal heterogeneity could reflect the generally more prominent T cell response to COVID-II (mild illness).

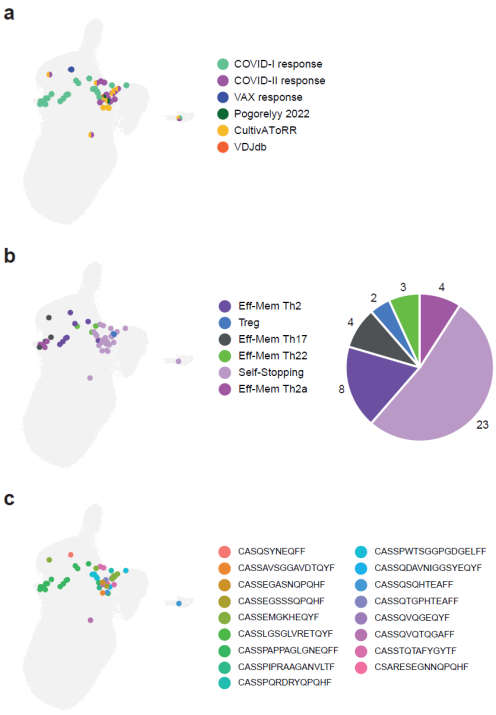

Figure 4. Mapping to the reference dataset shows functional clusters of SARS-CoV-2-specific clones. Positioning of D11 SARS-CoV-2-specific TCRβ CDR3 clonotypes. **a**. SARS-CoV-2-specific cells are colored based on identification method, clonotypes verified by more than one method are shown as small pie charts. **b**. Corresponding scRNA-Seq clusters. Large pie chart shows the proportions of scRNA-Seq clusters in the anti-SARS-CoV-2 response. **c**. TCRβ CDR3 clonotypes.

Discussion

The architecture of T cell memory essentially determines the entire pattern of our interaction with the antigens of the surrounding world, our microbiota, and our self-antigens[1, 2]. This architecture starts formation in the prenatal period[49, 50, 51, 52, 53], is actively formed in the first years of life in the contact with pathogens, airborne and food antigens, and microbiota maturation[54, 55] and then continues to be actively shaped by vaccinations and further contacts with infectious and non-infections challenges and antigens.

Clonal populations of T cells that are instructively primed by professional antigen-presenting cells[1, 2, 3] make decisions about which reaction programs to choose to respond to each specific antigen. They remember these programs as memory clones, forming stable response patterns to familiar challenges, patterns of regulation and cross-regulation of immune responses to friends and foes.

Mistakes made in such decisions cost us dearly: they lead to autoimmune diseases, inefficient elimination of pathogens, chronic inflammation, cancer, and may essentially underlie the entire phenomenon of inflammaging. Presumably for this reason, some mammalian species are likely to avoid forming such a long-term clonal memory[11].

In humans, however, the clonal memory of both CD8+ and CD4+ T lymphocytes can persist for years and decades[56, 57]. In the present work, this allowed us to map Th subset repertoires on the very same T cell clones of the scRNA-Seq data obtained for the donor 4 years later, which emphasizes the stability of the huge number of clones accumulated over previous years[58]. We also show that these clones are mapped predominantly or even exclusively within their corresponding and independent scRNA-Seq clusters, highlighting the persistence of program decisions once made by each T cell clone.

Repertoire-based Sort-Seq annotation of scRNA-Seq data turned out to be a powerful aid in matching classical subsets of lymphocytes sorted by surface markers with stable scRNAseq clusters that more comprehensively describe the diversity of functional lymphocyte programs. This effort structures our understanding of functional diversity of Th cells, critical for further progress in cancer and autoimmunity immunotherapy and vaccine development.

CITE-seq can also be used to phenotype cell populations on a single-cell level based on surface markers, having an advantage of simultaneously measuring hundreds of those. However, our work shows that Sort-Seq significantly outperforms CITE-Seq in terms of resolution capacity. Additionally, some of the CITE-Seq staining antibodies may affect cell signaling and transcriptomic profiles, introducing bias in the scRNA-Seq landscape. In contrast, in the Sort-Seq pipeline, cells stained with surface antibodies for sorting and bulk TCR-Seq are obtained and analyzed independently of the scRNA-Seq/scTCR-Seq experiment.

At the same time, Sort-Seq approach has certain limitations that must be taken into account. Firstly, similar to CITE-Seq, the method relies solely on surface markers to define target cell subpopulations. Secondly, the method is applicable only to T- and B-cells with specific clonal receptors acting as living barcodes. Thirdly, Sort-Seq is based on lymphocyte clonality, where we need to capture the same clone in sorting and scRNA-seq experiments. Therefore, Sort-Seq based annotation of e.g. central memory T cell subsets or peripheral blood Tfh cells, which are almost as diverse as naive T cells[25, 31], may require substantially deeper scRNA-Seq and bulk TCR profiling. For the naive T cell subsets, Sort-Seq application would be probably limited to innate-like, relatively clonal naive T cell subpopulations of fetal origin[51]. However, once the laborious work is done for the first time for each tissue and subsets of interest, scientific community may use the newly generated datasets to build well-annotated by Sort-Seq scRNA-Seq references of lymphocytes - as we did in the current study for peripheral blood CD4+ helper T cell subsets.

In the future works, Sort-Seq can be used to match and classify: 1) CD4+ T cell populations in the peripheral blood, lymph nodes, tertiary lymphoid structures at the sites of chronic inflammation and tumor environment, other tissues in health and disease; 2) diverse follicular helper T cells[31]; 3) known and unknown types of invariant and semi-invariant T cells, such as iNKT, NKT, MAIT, and CAIT[59, 60, 61]; 4) gamma delta T cells[62]; 5) central memory and stem cell memory T cells that would probably require deeper scRNA-Seq coverage due to their relatively lower clonality; 6) CD8+ T cell subsets[63]; 7) B cell functional subsets[64].

In the second part of the work, we report CultivAToRR, a cost-efficient method for identifying antigen-specific T cell clones in the patient's blood. This method, representing the modification of previously

described approaches[37, 38], was here empowered by biological replicates and appropriate statistical analysis of TCR repertoires. It should be widely applicable in studies of cancer, autoimmunity, infections, and vaccinations as a convenient easy-to-use tool for the rapid identification of effector/memory T cell clones specific to particular antigenic peptides or peptide mixes. Here, we employed CultivAToRR to identify more than 80 CD4+ SARS-CoV2-specific T cell clones in a single donor. Their expansion in longitudinal TCR repertoire-based tracking ideally coincided with two SARS-CoV2 infections of the donor, confirming accuracy of the method.

Finally, linking the two branches of the current work, we mapped SARS-CoV2-specific T cell clones with the scRNA-Seq data of the donor, integrated and annotated within the global CD4+ T cell reference map. This allowed us to determine their functional scRNA-Seq clusters, and to accordingly make certain assumptions about the nature of patients' immune response to the first and second infections.

Altogether, we hope that our work:

1) Clarifies correspondence between the well-studied Th subsets and scRNA-Seq landscape.
2) Proposes more accurate classification of CD4+ T cell scRNA-Seq, delineating positioning of Th2 and Th2a clusters, and refining a number of other details.
3) Shows long-term program stability and low intrinsic plasticity of Th memory clones.
4) Offers Sort-Seq, a new powerful approach for classification of T- and B-lymphocyte subsets and their exact positioning within scRNA-Seq landscape.
5) Provides an integrated easy-to-use scRNA-Seq reference dataset of peripheral Th lymphocytes.
6) Offers CultivAToRR, a powerful, cost-efficient, and easy-to-use method to identify antigen-specific TCRs in a 10-days assay that starts directly from PBMCs and does not require many laborious operations.

More generally, we hope that this work is helping to move the study of the role of programmed populations of memory T cells to a new level, making it possible to clearly distinguish the functional nature of each immune response, as well as to investigate the true plasticity between the known T cell subsets. This level of understanding serves as a necessary stepping stone to rational development of better immunotherapeutic approaches in oncology and autoimmunity, and vaccine development, where the chosen T cell programs fundamentally determine the type of immune response and decide the matter.

Data availability

Raw sequencing data is available in NCBI Sequence Read Archive (BioProject: PRJNA995237). Processed Seurat objects with Th reference and scRNA-Seq datasets from D11 are available for download at https://figshare.com/projects/T_helper_subsets_Kriukova_et_al_/173466. PBMC ds45k dataset (D11, TCRβ repertoires from PBMCs in replicates, downsampled to 45,000 UMI in each sample) is deposited in our GitHub repository (Th_kriukova/outs/ds45k) at https://github.com/kriukovav/Th_kriukova.

Code availability

Code for analysis (TCRβ repertoires from PBMCs) and all figures (scRNA-Seq and TCRβ repertoires) is available at https://github.com/kriukovav/Th_kriukova. R package wrapping Seurat[65] functions to perform reference mapping to our Th scRNA-Seq dataset is available at https://github.com/kriukovav/CD4map.

Acknowledgements

We thank Vadim Karnaukhov, Denis Syrko, Viktor Kotlyar, and Dmitry Bolotin for assistance with data analysis and Maria Vakhitova, Tatiana Grigorova, Ilgar Mamedov, and the NGS Laboratory (IKMB, Kiel University) for assistance with sequencing, which was also supported by the DFG Research Infrastructure NGS_CC (project 407495230) as part of the Next Generation Sequencing Competence Network (project 423957469).

ONLINE METHODS

IFN-γ ELISpot

PBMCs were isolated via standard density gradient centrifugation, quantified using a LUNA-II Automated Cell Counter (Logos Biosystems), and resuspended at 3×10^6 cells/mL in warm (37°C) CTL-Test™ Medium (Cellular Technology Ltd.) containing 1% L-glutamine (Thermo Fisher Scientific). Cells were then plated in 96-well anti-IFN-γ-coated ELISpot plates (Human IFN-γ Single-Color ELISPOT Kit, Cellular Technology Ltd.) at 3×10^5 cells/well in the presence of recombinant SARS-CoV-2 nucleocapsid phosphoprotein (N) or recombinant SARS-CoV-2 spike glycoprotein (S) in duplicate (each 1 µg/mL, Miltenyi Biotec). Negative control wells lacked viral proteins, and positive control wells contained phytohemagglutinin (2.5 ng/mL, Sigma-Aldrich). Plates were incubated for 18 h at 37°C in the presence of 5% CO_2. Assays were developed according to the manufacturer's instructions, and spots were counted using an ImmunoSpot S6 Universal Analyzer (Cellular Technology Ltd.).

CultivAToRR assay

Freshly isolated PBMCs were stained with CFSE (450 nM) and seeded in 48-well cell culture plates at 0.5×10^6 cells/well in complete RPMI (RPMI 1640 + 10% heat-inactivated fetal bovine serum) supplemented with IL-21 (10 ng/mL). The experiment included three conditions, each with three replicates: stimulation with PepTivator SARS-CoV-2 Prot_S (Miltenyi Biotec), stimulation with a mixture of PepTivator SARS-CoV-2 Prot_M and PepTivator SARS-CoV-2 Prot_N (Miltenyi Biotec), and control stimulation with a pool of irrelevant peptides (all at a final concentration of 0.6 nmol/peptide/mL). Cells were stained with anti-CD4–Alexa Fluor 647 (clone RPA-T4, BioLegend) and anti-CD8–BV421 (clone SK1, BioLegend) on day 6. $CFSE^{low}CD4^+$ and $CFSE^{low}CD8^+$ T cells were sorted directly into RLT buffer (200 µL, QIAGEN) using a FACSAria III (BD Biosciences). RNA was isolated using RNeasy Micro Kit (>500 cells/sample, QIAGEN) or with TRIzol Reagent (<500 cells/sample, Thermo Fisher Scientific). TCRβ libraries were constructed using all extracted RNA. Sequencing and raw data analysis were performed as described below. Significantly expanded clonotypes were identified using EdgeR[42, 66, 67]. UMI counts were filtered using the `filterByExpr` function with the following parameters: `min.count = 1, min.total.count = 5, large.n = 1, and min.prop = 0.5`. Count normalization and estimated dispersion by the default method were performed using the trimmed mean of M values (TMM). Each group was tested against all others using a quasi-likelihood (QL) negative binomial generalized log-linear model followed by the F-test.

Final clonotypes had an FDR <0.05 and a logFC ≥4. The code is available as part of the TCRgrapher library at https://github.com/KseniaMIPT/tcrgrapher (edgeR_pipeline function).

Bulk TCR-Seq library preparation and sequencing

PBMC replicates were lysed either fresh or after a freeze/thaw cycle in Buffer RLT (QIAGEN) (**Supplementary Table 3**). Total RNA was isolated using RNeasy Mini Kit (QIAGEN). UMI-based TCRβ repertoire libraries were prepared using a Human TCR RNA Multiplex Kit (MiLaboratories). Pooled samples were sequenced across 150+150 bp with a coverage of 50–100 reads per input cell on a NextSeq550 (Illumina).

Bulk TCR-Seq data analysis and clonotype table preprocessing

Raw fastq data were analyzed using MiXCR[68] (v4.1.0) with the appropriate built-in preset (either `milab-human-tcr-rna-multiplex-cdr3` or `milab-human-tcr-rna-race-cdr3`, the latter used only for pre-vaccination bulk PBMC TCR repertoire extraction, time point "pre-history"). Only cDNA molecules sequenced at least twice (based on UMI coverage) were considered in the generation of TCR clonotype tables (parameters `Massemble.consensusAssemblerParameters.assembler.minRecordsPerConsensus=2` and `MrefineTagsAndSort.parameters.postFilter=null` in MiXCR). TCR clonotype tables were additionally filtered to include only productive CDR3β amino acid sequences starting with the conserved cysteine and ending with the conserved phenylalanine. The resulting TCRβ repertoires from bulk PBMCs were all normalized to the same depth by randomly sampling 45,000 unique cDNA molecules from each sample (ds45k dataset). To complement the missing time point before the first vaccination, an artificial "pre-history" time point was generated with a pre-vaccination TCRβ repertoire composed of exactly 45,000 cDNA molecules, based on Rep-Seq data from sorted $CD4^+$ and $CD8^+$ T cells from the same donor (several time points in replicates, all collected in 2017). These samples were divided into two sets, each incorporating $CD4^+$ and $CD8^+$ TCR repertoires, and 30,000 $CD4^+$ TCR cDNA molecules and 15,000 $CD8^+$ TCR cDNA molecules were randomly sampled from each set. This procedure resulted in two replicates of TCRβ repertoires from $CD4^+$ and $CD8^+$ T cells pooled at a ratio of 2:1 (pseudo-bulk PBMCs). All table data manipulations were performed using R (v4.1.2) and tidyverse (v1.3.2).

Clonal expansions at corresponding time points

TCR clonotypes were defined as unique nucleotide TCRβ sequences in the ds45k dataset. Differentially abundant TCRs between a considered time point and three relevant reference time points, where no immune challenge was documented, were identified using edgeR (v3.36.0). Biological replicates were used to fuel edgeR statistics. The exact combinations of experimental time points together with the relevant reference time points are shown on **Fig. 3c**. Clonotypes were defined as expanded at an FDR <0.05 and a $\log2(count_{exp}/count_{ref})$ >1. Artificially generated pre-vaccination TCR repertoires (pre-history) were not used in the differential abundance analysis and were only adopted for the purpose of visualization.

Screening against known SARS-CoV-2-specific TCRs

Overlaps were sought between clonotypes in ds45k dataset and the VDJdb-2020-03-30 database[46] or the Pogorelyy *et al.* dataset[47], with exact TRBV and TRBJ segment matches and a maximum of one amino acid mismatch within the CDR3β (cdr3aa-v-j-1mm intersect type). Analysis and data visualization were confined to clonotypes specific for SARS-CoV-2 restricted by a donor-matched HLA. Clonotypes represented by at least three unique cDNA molecules at the VAX, COVID1, or COVID2 time points in the ds45k dataset were also selected for reliability.

TCRβ CDR3 homology clusters identified by the three methods

Clonotypes (defined by amino acid CDR3β sequence and TRBV-TRBJ segments) identified via CultivAToRR, clonal expansion at corresponding time points, and mapping against known SARS-CoV-2-specific TCRs were pulled together before clustering, linked by TRBV-TRBJ segment and CDR3β sequence identity. This set was complemented by including clonotypes from the ds45k dataset with one amino acid mismatch. Only clonotypes with the same TRBV-TRBJ combination were allowed in any one cluster. For cluster visualization, each node represents one TCRβ clonotype, and each edge links nodes with one amino acid mismatch.

scRNA-Seq and scTCR-Seq library preparation and sequencing

Frozen PBMCs from D11 (COVID1 and COVID2 samples) were thawed according to 10x Genomics recommendations and rested for 1 h in complete RPMI (RPMI 1640 + 10% autologous serum) at 37°C in the presence of 5% CO_2. Cells were then stained with anti-CD4–Alexa Fluor 647 (clone RPA-T4, BioLegend), anti-CD8–PerCP (clone SK1, BioLegend), anti-CD19–FITC (clone J3-119, Beckman Coulter), anti-CD27–eFluor 780 (clone LG.7F9, Thermo Fisher Scientific), anti-CD45RA–eFluor 450 (clone HI100, Thermo Fisher Scientific), and 7-AAD (Thermo Fisher Scientific). Effector/memory $CD4^+$ T cells gated as $CD4^+$ $CD8^-$ $CD19^-$ after excluding $CD27^+$ $CD45RA^+$ events were sorted into complete RPMI using a FACSAria III (BD Biosciences). Dead cells were excluded based on morphology and staining with 7-AAD. Sorted effector/memory $CD4^+$ T cells were washed twice in phosphate-buffered saline containing 0.04% bovine serum albumin and loaded onto a Next GEM Chip (10x Genomics) in one replicate with a target count of 10,000 cells for the COVID1 sample and in two replicates with a target count of 10,000 cells for the COVID2 sample. The emulsion from the COVID1 sample was divided into two parts, resulting in the generation of two technical replicates, each originating from approximately 5,000 cells. Samples were prepared using a Chromium Next GEM Single Cell V(D)J Reagent Kit v1.1 (10x Genomics). Pooled samples were sequenced with a coverage of 16,000 reads per input cell for scRNA-Seq and 5,000 reads per input cell for scTCR-Seq on a NextSeq 550 System (Illumina).

Fresh PBMCs from D01, D04, D05 were stained with anti-CCR7–PE-Cy7 (clone 3D12, BD Biosciences), anti-CD3–APC-Fire750 (clone SK7, BioLegend), anti-CD4–PE-Cy5.5 (clone S3.5, Thermo Fisher Scientific), anti-CD14–V500 (clone M5E2, BD Biosciences), anti-CD19–V500 (clone HIB19, BD Biosciences), anti-CD45RA–PE-Cy5 (clone HI100, BioLegend), and LIVE/DEAD Fixable Aqua (Thermo Fisher Scientific). Viable effector/memory $CD4^+$ T cells gated as $CD3^+$ $CD4^+$ $CD14^-$ $CD19^-$ after excluding $CCR7^+$ $CD45RA^+$ events were sorted in two replicates for D01 and without replicates for D04 and D05 using a custom-modified FACSAria II (BD Biosciences) and loaded onto a Chromium Controller (10x

Genomics). Samples were prepared using a Chromium Next GEM Single Cell 5' Reagent Kit v2 (10x Genomics). Pooled samples were sequenced with a coverage of 100,000 reads per input cell for scRNA-Seq and 25,000 reads per input cell for scTCR-Seq on a NovaSeq 6000 System with an S4 Flow Cell (Illumina).

scRNA-Seq and scTCR-Seq data analysis

Raw scRNA-Seq and scTCR-seq fastq files were processed using the `count` and `vdj` pipelines in cellranger (v6.1.2). Filtered gene expression matrices were uploaded into the Seurat R package (v4.2.0)[65]. TCR data were processed using the CellaRepertorium R package (v1.4.0). The most abundant TCR chain was used for cells with two relevant transcripts (TRA or TRB). Publicly available data were downloaded via E-MTAB-10026 and converted to Seurat object. $CD4^+$ T cell clusters were chosen based on the average cluster expression of *CD4* and *CD3E*. The public dataset was then split by sample origin. CITE-Seq ADT counts were normalized using the CLR method. Additional filtering steps were performed on individual datasets, including the elimination of outlier clusters with low UMI counts and/or high percentages of mitochondrial genes and cells expressing *CD8A* and *CD8B*. The stress score was calculated for each cell using `AddModuleScore()` function in Seurat and included genes upregulated as a consequence of the dissociation procedure[69]. Integration was carried out using the Seurat reference-based reciprocal PCA protocol, with the largest 3' and 5' datasets chosen as references to account for differences in methodology[65]. The percentage of mitochondrial genes and the stress gene signature were regressed out of individual and integrated datasets. All TCR and IG genes were removed from variable features used in the PCA and from the anchor features used for integration. The number of dimensions used for running the UMAP algorithm was 25.

Reference mapping

D11_COVID1, D11_COVID2_rep1, and D11_COVID2_rep2 samples were mapped to the integrated reference dataset Seurat Reference Mapping procedure with default parameters[65]. Plots were designed to highlight amino acid-defined TCRβ clonotypes with specificity for SARS-CoV-2.

Ethics statement

Ethical approval was granted by the Cardiff University School of Medicine Research Ethics Committee (16/55) and the Institutional Review Board of the Pirogov Russian National Research Medical University.

References

1. Kunzli, M. & Masopust, D. CD4(+) T cell memory. *Nat Immunol* **24**, 903-914 (2023).
2. Sallusto, F. Heterogeneity of Human CD4(+) T Cells Against Microbes. *Annual review of immunology* **34**, 317-334 (2016).
3. Borst, J., Ahrends, T., Babala, N., Melief, C.J.M. & Kastenmuller, W. CD4(+) T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol* **18**, 635-647 (2018).

4. McDonald, D.R. TH17 deficiency in human disease. *The Journal of allergy and clinical immunology* **129**, 1429-1435; quiz 1436-1427 (2012).

5. Cook, M.C. & Tangye, S.G. Primary immune deficiencies affecting lymphocyte differentiation: lessons from the spectrum of resulting infections. *Int Immunol* **21**, 1003-1011 (2009).

6. Hernandez-Santos, N. *et al.* Th17 cells confer long-term adaptive immunity to oral mucosal Candida albicans infections. *Mucosal immunology* **6**, 900-910 (2013).

7. Misiak, A. *et al.* Addition of a TLR7 agonist to an acellular pertussis vaccine enhances Th1 and Th17 responses and protective immunity in a mouse model. *Vaccine* **35**, 5256-5263 (2017).

8. Bacher, P. & Scheffold, A. Antigen-specific regulatory T-cell responses against aeroantigens and their role in allergy. *Mucosal immunology* **11**, 1537-1550 (2018).

9. McGee, H.S. & Agrawal, D.K. TH2 cells in the pathogenesis of airway remodeling: regulatory T cells a plausible panacea for asthma. *Immunologic research* **35**, 219-232 (2006).

10. Finotto, S. T-cell regulation in asthmatic diseases. *Chem Immunol Allergy* **94**, 83-92 (2008).

11. Izraelson, M. *et al.* Distinct organization of adaptive immunity in the long-lived rodent Spalax galili. *Nat Aging* **1**, 179-189 (2021).

12. Osnes, L.T., Nakken, B., Bodolay, E. & Szodoray, P. Assessment of intracellular cytokines and regulatory cells in patients with autoimmune diseases and primary immunodeficiencies - novel tool for diagnostics and patient follow-up. *Autoimmunity reviews* **12**, 967-971 (2013).

13. Costa, N. *et al.* Two separate effects contribute to regulatory T cell defect in systemic lupus erythematosus patients and their unaffected relatives. *Clin Exp Immunol* **189**, 318-330 (2017).

14. Bonelli, M. *et al.* Quantitative and qualitative deficiencies of regulatory T cells in patients with systemic lupus erythematosus (SLE). *Int Immunol* **20**, 861-868 (2008).

15. Miyara, M. *et al.* Global natural regulatory T cell depletion in active systemic lupus erythematosus. *J Immunol* **175**, 8392-8400 (2005).

16. Protti, M.P., De Monte, L. & Di Lullo, G. Tumor antigen-specific CD4+ T cells in cancer immunity: from antigen identification to tumor prognosis and development of therapeutic strategies. *Tissue Antigens* **83**, 237-246 (2014).

17. Andreatta, M. *et al.* A CD4(+) T cell reference map delineates subtype-specific adaptation during acute and chronic viral infections. *eLife* **11** (2022).

18. Yasumizu, Y. *et al.* Single-cell transcriptome landscape of circulating CD4⁺ T cell populations in human autoimmune diseases. *bioRxiv*, 2023.2005.2009.540089 (2023).

19. Radtke, D. *et al.* Th2 single-cell heterogeneity and clonal distribution at distant sites in helminth-infected mice. *eLife* **11** (2022).

20. Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382-387 (2014).

21. Reimegard, J. *et al.* A combined approach for single-cell mRNA and intracellular protein expression analysis. *Commun Biol* **4**, 624 (2021).

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

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

24. Shahi, P., Kim, S.C., Haliburton, J.R., Gartner, Z.J. & Abate, A.R. Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Scientific reports* **7**, 44447 (2017).

25. Kasatskaya, S.A. *et al.* Functionally specialized human CD4(+) T-cell subsets express physicochemically distinct TCRs. *eLife* **9** (2020).

26. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med* **27**, 904-916 (2021).

27. Patil, V.S. *et al.* Precursors of human CD4(+) cytotoxic T lymphocytes identified by single-cell transcriptome analysis. *Science immunology* **3** (2018).

28. Cheroutre, H. & Husain, M.M. CD4 CTL: living up to the challenge. *Semin Immunol* **25**, 273-281 (2013).

29. Hashimoto, K. *et al.* Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc Natl Acad Sci U S A* **116**, 24242-24251 (2019).

30. Scharf, L. *et al.* Longitudinal single-cell analysis of SARS-CoV-2-reactive B cells uncovers persistence of early-formed, antigen-specific clones. *JCI insight* **8** (2023).

31. Brenna, E. *et al.* CD4(+) T Follicular Helper Cells in Human Tonsils and Blood Are Clonally Convergent but Divergent from Non-Tfh CD4(+) Cells. *Cell reports* **30**, 137-152 e135 (2020).

32. Duhen, T., Geiger, R., Jarrossay, D., Lanzavecchia, A. & Sallusto, F. Production of interleukin 22 but not interleukin 17 by a subset of human skin-homing memory T cells. *Nat Immunol* **10**, 857-863 (2009).

33. Barnes, J.L. *et al.* T-helper 22 cells develop as a distinct lineage from Th17 cells during bacterial infection and phenotypic stability is regulated by T-bet. *Mucosal immunology* **14**, 1077-1087 (2021).

34. Perez, L.G. *et al.* TGF-beta signaling in Th17 cells promotes IL-22 production and colitis-associated colon cancer. *Nature communications* **11**, 2608 (2020).

35. Knudson, C.J. *et al.* Mechanisms of Antiviral Cytotoxic CD4 T Cell Differentiation. *J Virol* **95**, e0056621 (2021).

36. Hoeks, C., Duran, G., Hellings, N. & Broux, B. When Helpers Go Above and Beyond: Development and Characterization of Cytotoxic CD4(+) T Cells. *Frontiers in immunology* **13**, 951900 (2022).

37. Klinger, M. *et al.* Combining next-generation sequencing and immune assays: a novel method for identification of antigen-specific T cells. *PLoS One* **8**, e74231 (2013).

38. Danilova, L. *et al.* The Mutation-Associated Neoantigen Functional Expansion of Specific T Cells (MANAFEST) Assay: A Sensitive Platform for Monitoring Antitumor Immunity. *Cancer immunology research* **6**, 888-899 (2018).

39. Pogorelyy, M.V. *et al.* Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. *Proc Natl Acad Sci U S A* **115**, 12704-12709 (2018).

40. Minervina, A.A. *et al.* Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. *eLife* **9** (2020).

41. Yuzhakova, D.V. *et al.* Measuring Intratumoral Heterogeneity of Immune Repertoires. *Frontiers in oncology* **10**, 512 (2020).

42. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

43. Wragg, K.M. *et al.* Establishment and recall of SARS-CoV-2 spike epitope-specific CD4(+) T cell memory. *Nat Immunol* **23**, 768-780 (2022).

44. Dolton, G. *et al.* Optimized Peptide-MHC Multimer Protocols for Detection and Isolation of Autoimmune T-Cells. *Frontiers in immunology* **9**, 1378 (2018).

45. Minervina, A.A. *et al.* Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *eLife* **10** (2021).

46. Goncharov, M. *et al.* VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat Methods* **19**, 1017-1019 (2022).

47. Pogorelyy, M.V. *et al.* Resolving SARS-CoV-2 CD4(+) T cell specificity via reverse epitope discovery. *Cell Rep Med* **3**, 100697 (2022).

48. Venturi, V. *et al.* Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci U S A* **103**, 18691-18696 (2006).

49. King, C.L. *et al.* Acquired immune responses to Plasmodium falciparum merozoite surface protein-1 in the human fetus. *J Immunol* **168**, 356-364 (2002).

50. Li, N. *et al.* Memory CD4(+) T cells are generated in the human fetal intestine. *Nat Immunol* **20**, 301-312 (2019).

51. Pogorelyy, M.V. *et al.* Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLoS computational biology* **13**, e1005572 (2017).

52. Prescott, S.L. *et al.* Development of allergen-specific T-cell memory in atopic and normal children. *Lancet* **353**, 196-200 (1999).

53. Huygens, A., Dauby, N., Vermijlen, D. & Marchant, A. Immunity to cytomegalovirus in early life. *Frontiers in immunology* **5**, 552 (2014).

54. Renz, H. & Skevaki, C. Early life microbial exposures and allergy risks: opportunities for prevention. *Nat Rev Immunol* **21**, 177-191 (2021).

55. Pieren, D.K.J., Boer, M.C. & de Wit, J. The adaptive immune system in early life: The shift makes it count. *Frontiers in immunology* **13**, 1031924 (2022).

56. Naniche, D. *et al.* Decrease in measles virus-specific CD4 T cell memory in vaccinated subjects. *J Infect Dis* **190**, 1387-1395 (2004).

57. Jokinen, S., Osterlund, P., Julkunen, I. & Davidkin, I. Cellular immunity to mumps virus in young adults 21 years after measles-mumps-rubella vaccination. *J Infect Dis* **196**, 861-867 (2007).

58. Yoshida, K. *et al.* Aging-related changes in human T-cell repertoire over 20years delineated by deep sequencing of peripheral T-cell receptors. *Experimental gerontology* **96**, 29-37 (2017).

59. LeBlanc, G., Kreissl, F.K., Melamed, J., Sobel, A.L. & Constantinides, M.G. The role of unconventional T cells in maintaining tissue homeostasis. *Semin Immunol* **61-64**, 101656 (2022).

60. Rosati, E. *et al.* A novel unconventional T cell population enriched in Crohn's disease. *Gut* (2022).

61. Crowther, M.D. *et al.* Genome-wide CRISPR-Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1. *Nat Immunol* **21**, 178-185 (2020).

62. Davey, M.S. *et al.* The human Vdelta2(+) T-cell compartment comprises distinct innate-like Vgamma9(+) and adaptive Vgamma9(-) subsets. *Nature communications* **9**, 1760 (2018).

63. Galletti, G. *et al.* Two subsets of stem-like CD8(+) memory T cell progenitors with distinct fate commitments in humans. *Nat Immunol* **21**, 1552-1562 (2020).

64. Stewart, A. *et al.* Single-Cell Transcriptomic Analyses Define Distinct Peripheral B Cell Subsets and Discrete Development Pathways. *Frontiers in immunology* **12**, 602539 (2021).

65. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

66. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297 (2012).

67. Chen, Y., Lun, A.T. & Smyth, G.K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438 (2016).

68. Bolotin, D.A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* **12**, 380-381 (2015).

69. Denisenko, E. *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol* **21**, 130 (2020).

2.4. Involvement of Th cells in response to chronic viral infections

2.4.1. Manuscript 3: Inhibitory IL-10-producing CD4+ T cells are T-bet-dependent and facilitate cytomegalovirus persistence via coexpression of arginase-1

eLife

RESEARCH ARTICLE

Inhibitory IL-10-producing CD4$^+$ T cells are T-bet-dependent and facilitate cytomegalovirus persistence via coexpression of arginase-1

Mathew Clement[1,2]*, Kristin Ladell[1], Kelly L Miners[1], Morgan Marsden[1], Lucy Chapman[1], Anna Cardus Figueras[1], Jake Scott[1], Robert Andrews[1,2], Simon Clare[3], Valeriia V Kriukova[4,5,6], Ksenia R Lupyr[4,5,7], Olga V Britanova[5,6,7], David R Withers[8], Simon A Jones[1,2], Dmitriy M Chudakov[4,5,7,9], David A Price[1,2], Ian R Humphreys[1,2]

[1]Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, United Kingdom; [2]Systems Immunity Research Institute, School of Medicine, Cardiff University, Cardiff, United Kingdom; [3]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom; [4]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russian Federation; [5]Genomics of Adaptive Immunity Department, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation; [6]Institute of Clinical Molecular Biology, Christian-Albrecht-University of Kiel, Kiel, Germany; [7]Institute of Translational Medicine, Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow, Russian Federation; [8]Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, United Kingdom; [9]Abu Dhabi Stem Cell Center, Al Muntazah, United Arab Emirates

***For correspondence:** clementm@cardiff.ac.uk

Competing interest: The authors declare that no competing interests exist.

Funding: See page 17

Preprinted: 27 January 2022
Received: 01 April 2022
Accepted: 11 May 2023
Published: 13 July 2023

Reviewing Editor: Stipan Jonjic, University of Rijeka, Croatia

Abstract Inhibitory CD4$^+$ T cells have been linked with suboptimal immune responses against cancer and pathogen chronicity. However, the mechanisms that underpin the development of these regulatory cells, especially in the context of ongoing antigen exposure, have remained obscure. To address this knowledge gap, we undertook a comprehensive functional, phenotypic, and transcriptomic analysis of interleukin (IL)-10-producing CD4$^+$ T cells induced by chronic infection with murine cytomegalovirus (MCMV). We identified these cells as clonally expanded and highly differentiated T_H1-like cells that developed in a T-bet-dependent manner and coexpressed arginase-1 (Arg1), which promotes the catalytic breakdown of L-arginine. Mice lacking Arg1-expressing CD4$^+$ T cells exhibited more robust antiviral immunity and were better able to control MCMV. Conditional deletion of T-bet in the CD4$^+$ lineage suppressed the development of these inhibitory cells and also enhanced immune control of MCMV. Collectively, these data elucidated the ontogeny of IL-10-producing CD4$^+$ T cells and revealed a previously unappreciated mechanism of immune regulation, whereby viral persistence was facilitated by the site-specific delivery of Arg1.

Editor's evaluation

This study analyzes the development and functional relevance of IL-10-producing regulatory T cells in a mouse model of cytomegalovirus infection. The results indicate that IL-10-producing CD4$^+$ T cells express genes associated with chronically activated T_H1-like cells, undergo clonal expansion,

and inhibit antiviral T cell responses via the secretion of arginase, an enzyme that breaks down an amino acid required for T cell activation and proliferation. These findings reveal a novel and important immunoregulatory mechanism that facilitates viral persistence.

Introduction

Immune dysregulation occurs during many persistent viral infections. High levels of ongoing viral replication, which characterize human immunodeficiency virus (HIV), hepatitis B virus (HBV), and, in mice, lymphocytic choriomeningitis virus (LCMV), typically lead to T cell exhaustion, defined by impaired effector functions, the expression of inhibitory cytokines and receptors (***Wherry, 2011***), and substantial alterations in cellular gene expression (***Doering et al., 2012***). Moreover, inducible and naturally occurring FoxP3$^+$ regulatory T cells accumulate during many chronic viral infections, presumably to limit excessive immune activation (***Veiga-Parga et al., 2013***), and T helper (T_H)1-like cells that express the immunosuppressive cytokine interleukin (IL)-10 can be induced by LCMV (***Parish et al., 2014***), HIV (***Graziosi et al., 1994***), and human/murine cytomegalovirus (HCMV/MCMV) (***Clement et al., 2016***; ***Jones et al., 2010***; ***Mason et al., 2013***). Evidence from parasitic infections suggests that IL-10-producing T_H1-like cells protect against immune pathology, akin to classical FoxP3$^+$ regulatory T cells (***Anderson et al., 2007***; ***Jankovic et al., 2007***). However, experimental deletion of IL-10 production in T cells has been shown to promote the clearance of LCMV without any obvious collateral effects (***Clement et al., 2016***; ***Parish et al., 2014***; ***Richter et al., 2013***), suggesting a potential therapeutic role for similar manipulations in humans, albeit with the possibility of an attendant risk to the development of CD8$^+$ T cell memory (***Laidlaw et al., 2015***).

The mechanisms that induce IL-10 expression in T cells require further clarification, despite proposed roles for costimulatory receptors, cytokines, transcription factors, and signals delivered via the T cell receptor (TCR) (***Saraiva and O'Garra, 2010***). For example, chronic antigen exposure and the transcription factor Blimp-1 appear to be important for the development of IL-10-producing CD4$^+$ T cells in mice infected with LCMV (***Parish et al., 2014***), and the inhibitory receptor TIGIT is known to act upstream of IL-10 (***Schorer et al., 2020***). However, it is clear that viral persistence can be facilitated by IL-10, exemplified in the context of MCMV infection by ongoing replication in the salivary glands (SGs) (***Humphreys et al., 2007***; ***Mandaric et al., 2012***).

Interferon (IFN)γ-expressing CD4$^+$ T cells have been shown to limit viral replication in the SGs of mice infected with MCMV (***Jonjić et al., 1989***; ***Lucin et al., 1992***; ***Walton et al., 2011***). Nonetheless, CD4$^+$ T cells also represent an important source of IL-10 (***Clement et al., 2016***; ***Humphreys et al., 2007***), the production of which is promoted by IL-27 during acute infection with MCMV. In contrast, less is known about the mucosal IL-10-producing CD4$^+$ T cells that appear during chronic infection with MCMV, which are phenotypically distinct from type 1 regulatory T (Tr1) cells, specifically lacking concurrent expression of CD49d and LAG-3, and develop independently of IL-27 (***Clement et al., 2016***). These cells express high levels of various transcription factors, such as c-Maf and T-bet (***Clement et al., 2016***), and often coexpress other molecules with putative inhibitory functions, such as PD-1, TIM-3, and IL-21 (***Apetoh et al., 2010***; ***Awasthi et al., 2007***; ***Chihara et al., 2018***; ***Pot et al., 2009***; ***Zhu et al., 2015***). However, the functional relevance of these characteristics has remained obscure, along with the ontogeny of IL-10-producing CD4$^+$ T cells during chronic infection with MCMV.

To address these issues, we performed a comprehensive functional, phenotypic, and transcriptomic analysis of IL-10-producing CD4$^+$ T cells isolated from the SGs of mice infected with MCMV. Our data revealed that these cells were clonally expanded and highly differentiated T_H1-like cells with gene expression signatures that indicated a key developmental role for T-bet. In addition, we identified an inhibitory effect attributable to arginase-1 (Arg1), which was upregulated among IL-10-producing CD4$^+$ T cells during viral chronicity and facilitated the site-specific persistence of MCMV.

Results

IL-10-producing CD4$^+$ T cells display a T_H1-like profile

To better understand the development and functionality of inhibitory CD4$^+$ T cells that develop during viral chronicity, we infected MCMV IL-10 reporter (10BiT) mice with MCMV. These mice express Thy1.1 under the *Il10* promoter (***Maynard et al., 2007***). Unlike mucosal sites in the respiratory tract

(**Zhang et al., 2019**), ongoing viral replication in this model occurs primarily in the SGs, facilitated by the induction of $CD4^+$ T cells that produce IL-10 (**Humphreys et al., 2007**), which peak on day 14 post-infection (p.i.) (**Clement et al., 2016**). At this time point, we found that approximately 10–30% of $CD4^+$ T cells in the SGs were $Thy1.1^+$, of which ~95% displayed an effector memory phenotype ($CD44^{hi}$ $CD62L^{lo}$) (**Figure 1—figure supplement 1A**). IL-10-producing $CD4^+$ T cells were also induced by polyclonal stimulation and universally expressed Thy1.1 (**Figure 1—figure supplement 1B**).

We then compared the transcriptional profiles of endogenously generated $IL\text{-}10^+$ and $IL\text{-}10^-$ $CD4^+$ T cells, isolated via fluorescence-activated cell sorting (FACS) as $Thy1.1^+$ ($IL\text{-}10^+$) and $Thy1.1^-$ ($IL\text{-}10^-$) $CD44^{hi}$ $CD62L^{lo}$ $CD4^+$ T cells (**Figure 1—figure supplement 1A**). Principal component analysis (PCA) of the RNA-seq data revealed that $Thy1.1^+$ $CD4^+$ T cells were transcriptionally distinct from $Thy1.1^-$ $CD4^+$ T cells (**Figure 1—figure supplement 1C**). As expected, *Il10* was highly upregulated in $Thy1.1^+$ $CD4^+$ T cells (**Figure 1A**), and chromatin was more open in the *Il10* promoter region compared with $Thy1.1^-$ $CD4^+$ T cells (**Figure 1B**). Genes associated with localization and cell migration (*Ccl7*, *Cxcl2*, *Cxcl12*, *Ccl5*, *Cxcl14*, *Ccl28*, *Ccl12*, *Ccr1*, and *Ccr5*), cell signaling (*Ceacam1*, *Havcr2*, *Tigit*, *Lag3*, *Cd40*, *Cd36*, and *Itgb4*), regulation of cellular processes (*Prdm1*, *Gata2*, *Yes1*, *Card10*, and *Il33*), and metabolism (*Elovl7*, *Galnt3*, *Car13*, *Aldh1l1*, and *Ildrl*), including glycolysis and the tricarboxylic acid cycle (*Fbp2* and *Sdhc*), oxidative phosphorylation (*Osgin1*), and the mitochondrial respiratory chain (*Mt-Nd1*, *Ndufs6*, *Ndufb8*, *Uqcrfs1*, and *Uqcr11*), were also upregulated in $Thy1.1^+$ $CD4^+$ T cells, alongside genes associated with activation (*Fgl2*, *Cxcr2*, and *Nfil3*) and antiviral effector functions (*Gzmb*, *Prf1*, *Gzmk*, and *Lyz2*) (**Figure 1A, C, D** and **Figure 1—figure supplement 1D**). These latter gene profiles suggested the potential for cytolytic activity, but we found no evidence of a concomitant increase in the expression levels of granzyme B protein among $Thy1.1^+$ $CD4^+$ T cells (data not shown), which also lacked gene signatures classically associated with cytotoxic $CD4^+$ T cells, such as the upregulation of *Klrc1* and *Crtam* (https://doi.org/10.5281/zenodo.7243956).

$Thy1.1^+$ $CD4^+$ T cells are known to express the T_H1-associated chemokine receptors CXCR3 and CCR5 (**Clement et al., 2016**). We found that MCMV-induced $Thy1.1^+$ $CD4^+$ T cells shared many transcripts with $CD4^+$ T_H1 cells generated in vitro (**Stubbington et al., 2015**; **Figure 1E**), encompassing genes associated with numerous cellular and immunological processes (https://doi.org/10.5281/zenodo.7447477), and further expressed IFNγ in response to polyclonal stimulation at a population frequency of ~25% (**Figure 1—figure supplement 1E**). These data suggested that $Thy1.1^+$ $CD4^+$ T cells were commonly derived from antigen-specific T_H1 cells, especially given that IFNγ detection via flow cytometry likely underestimates the composite frequency of $CD4^+$ T cells that specifically recognize MCMV (**Jeitziner et al., 2013**). However, genes associated with the induction of IFNγ, including *Il18r1* and *Il18rap*, were actually downregulated in $Thy1.1^+$ $CD4^+$ T cells (**Figure 1A, C, D**), and in two of three replicates, a similar pattern was observed for *Ifng* (https://doi.org/10.5281/zenodo.7243956). Comparable findings were reported previously in functional studies of IFNγ expression at the protein level among IL-10-producing $CD4^+$ T cells specific for HCMV or MCMV (**Clement et al., 2016**; **Mason et al., 2013**).

Other genes that were downregulated in $Thy1.1^+$ $CD4^+$ T cells included *Il7* and *Tcf1/7* (**Figure 1A, D**), which extended to the protein level (**Figure 1F**). The relative underexpression of these cell survival-associated factors coincided temporally with the rapid contraction of virus-specific IL-10-producing $CD4^+$ T cells that typically occurs during the early stages of viral chronicity (**Clement et al., 2016**). In contrast, $Thy1.1^+$ and $Thy1.1^-$ $CD4^+$ T cells expressed similar levels of transcripts encoding DR5 (https://doi.org/10.5281/zenodo.7243956), which engages natural killer (NK) cell-expressed TRAIL and induces $CD4^+$ T cell death in the SGs (**Schuster et al., 2014**).

IL-10 production among $CD4^+$ T cells has been associated with the expression of inhibitory molecules and markers of exhaustion (**Saraiva and O'Garra, 2010**). Counterintuitively, we found that MCMV-induced $Thy1.1^+$ $CD4^+$ T cells downregulated the exhaustion-associated transcription factor *Tox1* but nonetheless expressed a module of inhibitory genes, including *Lag3*, *Fgl2*, *Havcr2*, and *Entpd1* (**Figure 1D**). These inhibitory molecules were also expressed at the protein level, alongside PD-1 (**Figure 1G** and **Figure 1—figure supplement 1F**). In addition, differential bystander activation seemed unlikely, because $Thy1.1^+$ $CD4^+$ T cells expressed LAG-3 and PD-1 more commonly than $Thy1.1^-$ $CD4^+$ T cells after preselection based on the induction of IFNγ (**Figure 1—figure supplement 1E**).

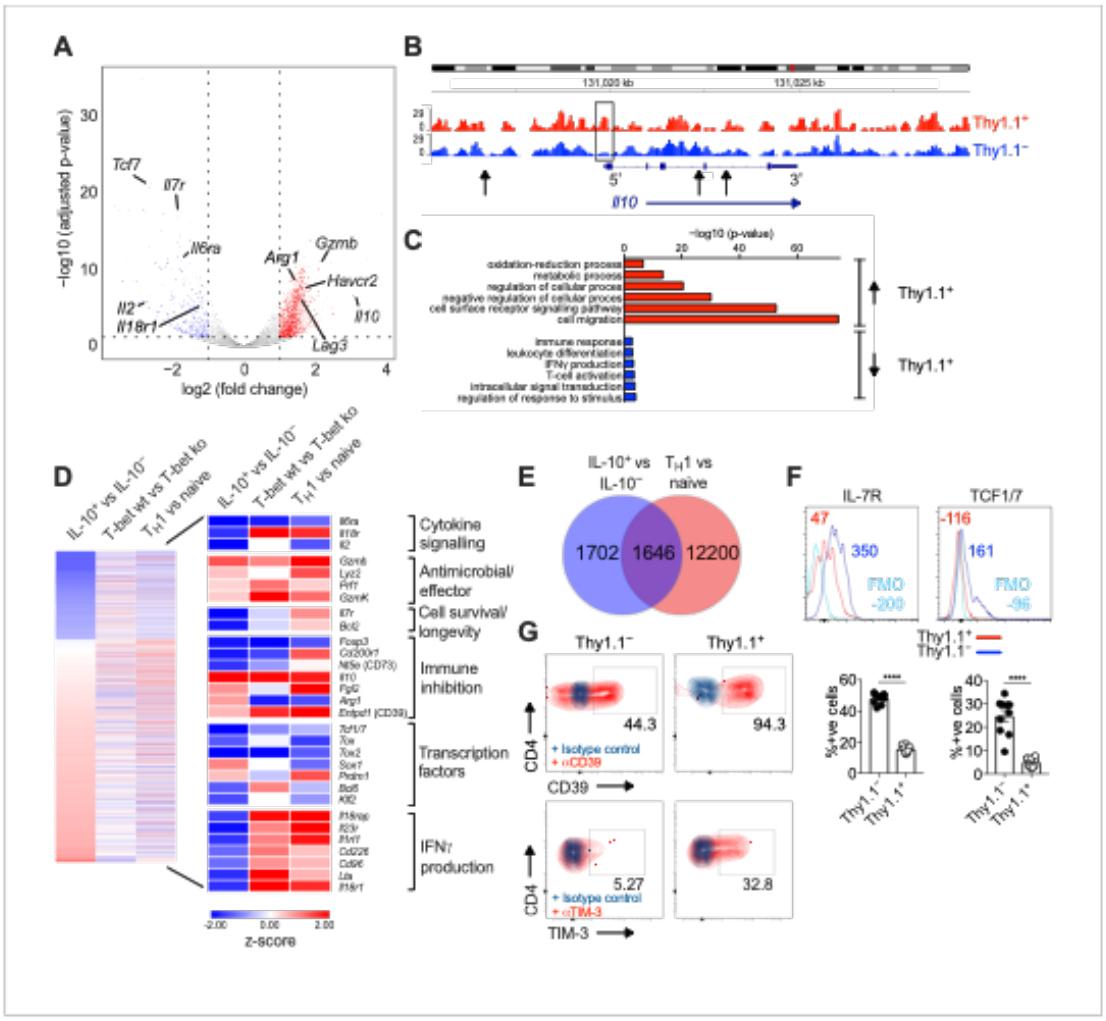

Figure 1. Interleukin (IL)-10-producing CD4$^+$ T cells display a T_H1-like profile. 10BiT mice were infected with 3 × 10^6 pfu of murine cytomegalovirus (MCMV). Leukocytes were isolated from the salivary glands (SGs) on day 14 p.i. and sorted as CD4$^+$ CD44$^+$ CD62L$^-$ CD90/90.1$^+$ (Thy1.1$^+$) or CD90/90.1$^-$ (Thy1.1$^-$) populations via fluorescence-activated cell sorting (FACS). (**A**) Volcano plot highlighting differentially upregulated genes in Thy1.1$^+$ CD4$^+$ T cells (red) versus Thy1.1$^-$ CD4$^+$ T cells (blue). (**B**) ATAC-seq profiles showing accessible chromatin regions in the *Il10* gene for Thy1.1$^+$ CD4$^+$ T cells (red) and Thy1.1$^-$ CD4$^+$ T cells (blue). Data are shown as normalized values accounting for the total number of reads per lane. The black box indicates a major difference in chromatin accessibility. Black arrows indicate binding motifs for Tbx21. (**C**) Gene ontology analysis of data from (**A**) indicating the top six modules that were upregulated (red) or downregulated (blue) in Thy1.1$^+$ CD4$^+$ T cells. (**D**) Heatmap comparing data from (**A**) (left column) with published data from T-bet$^+$ versus T-bet-knockout CD4$^+$ T cells (middle column, GSE38808) and T_H1 versus naive CD4$^+$ T cells (right column, E-MTAB-2582). Displayed genes were selected according to relevant pathways identified via gene ontology analysis and tabulated against respective functions (all $p < 0.05$). (**E**) Venn diagram showing the overlap between genes enriched in Thy1.1$^+$ CD4$^+$ T cells (**A, D**) and genes enriched in T_H1-like CD4$^+$ T cells (E-MTAB-2582). Data in (**A–E**) are shown as pooled analyses from a minimum of $n = 5$ mice per group representing three independent experiments. (**F**) Representative histograms (top) and summary bar graphs (bottom) showing the expression of IL-7R and TCF1/7 among Thy1.1$^+$ CD4$^+$ T cells (red) and Thy1.1$^-$ CD4$^+$ T cells (blue). The fluorescence-minus-one control is shown in sky blue (top). Bottom: data are shown as mean ± standard error of the mean (SEM; $n = 10$ mice per group representing two independent experiments). ****$p < 0.0001$ (Mann–Whitney *U* test). (**G**) Representative flow

Figure 1 continued on next page

Figure 1 continued

cytometry plots showing the expression of CD39 and TIM-3 among Thy1.1$^+$ CD4$^+$ T cells (red) and Thy1.1$^-$ CD4$^+$ T cells (blue). Data are shown as pooled analyses from a minimum of *n* = 10 mice per group representing two independent experiments.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Interleukin (IL)-10-producing CD4$^+$ T cells display a T_H1-like profile.

Figure supplement 1. Characterization of CD4$^+$ T cells isolated from the salivary glands (SGs) after infection with murine cytomegalovirus (MCMV).

Figure supplement 1—source data 1. Characterization of CD4$^+$ T cells isolated from the salivary glands (SGs) after infection with murine cytomegalovirus (MCMV).

Collectively, these data showed that IL-10-producing CD4$^+$ T cells exhibited a highly differentiated T_H1-like profile, characterized by the upregulation of various inhibitory molecules and the downmodulation of IFNγ expression lacking concordance with known signatures of exhaustion, during chronic infection with MCMV.

IL-10-producing CD4$^+$ T cells exhibit prominent clonal structures

IL-10-producing CD4$^+$ T cells recognize a broad range of viral antigens during chronic infection with MCMV (***Clement et al., 2016***). To characterize these interactions in more detail and evaluate the clonal relationship between IL-10$^+$ (Thy1.1$^+$) and IL-10$^-$ (Thy1.1$^-$) CD4$^+$ T cells, we used a next-generation approach to sequence the corresponding TCRs.

The repertoires of Thy1.1$^+$ CD4$^+$ T cells were less diverse and incorporated more prominent clonal expansions compared with the repertoires of Thy1.1$^-$ CD4$^+$ T cells (***Figure 2A, B*** and ***Figure 2—figure supplement 1A***). Several features also indicated that these expansions represented antigen-focused responses confined largely to Thy1.1$^+$ CD4$^+$ T cells (***Figure 2C–G***). First, the number of nucleotide variants that encoded each complementarity-determining region (CDR)3α and CDR3β amino acid sequence, an indicator of antigen-specific convergence (***Logunova et al., 2020***), was higher overall among Thy1.1$^+$ CD4$^+$ T cells versus Thy1.1$^-$ CD4$^+$ T cells (***Figure 2C***). Second, there were some differences in *Trbv* gene use that distinguished Thy1.1$^+$ CD4$^+$ T cells from Thy1.1$^-$ CD4$^+$ T cells, albeit with a general preference for *Trbv3*, *Trbv5*, and *Trbv31* (***Figure 2D***). Third, clusters of homologous TCRβ variants, identified using the statistical model ALICE (***Pogorelyy et al., 2019***), were detected predominantly among Thy1.1$^+$ CD4$^+$ T cells (***Figure 2E–G***). Importantly, this latter model accounts for generation probabilities, reliably separating immunologically relevant and irrelevant public TCRs.

It should be noted that none of these differences were absolute. For example, the clusters of TCRβ variants identified among Thy1.1$^+$ CD4$^+$ T cells also occurred at lower cumulative frequencies among Thy1.1$^-$ CD4$^+$ T cells (***Figure 2E–G***), and the TCRα and TCRβ repertoires overlapped considerably between Thy1.1$^+$ CD4$^+$ T cells and Thy1.1$^-$ CD4$^+$ T cells (***Figure 2—figure supplement 1B***). Moreover, there were no prominent differences in the physicochemical properties of amino acids in the central parts of the CDR3α and CDR3β loops, which generally differ among functionally discrete subsets of CD4$^+$ T cells (***Kasatskaya et al., 2020***), to indicate an ontogenetic divergence between Thy1.1$^+$ CD4$^+$ T cells and Thy1.1$^-$ CD4$^+$ T cells (***Figure 2—figure supplement 1C***).

Collectively, these data revealed the presence of common molecular signatures that predominated among Thy1.1$^+$ CD4$^+$ T cells, consistent with the notion of an antigen-driven process of differentiation leading to the production of IL-10.

IL-10-producing CD4$^+$ T cells are enriched for expression of Arg1

Our analysis of inhibitory gene expression revealed one particularly intriguing feature, namely that Thy1.1$^+$ CD4$^+$ T cells significantly upregulated *Arg1* (***Figure 1D***). Arg1 promotes the catalytic breakdown of L-arginine (***Munder, 2009***) and has been shown to inhibit the proliferation of T cells (***Czystowska-Kuzmicz et al., 2019***; ***Rodriguez et al., 2004***; ***Rodriguez et al., 2002***). A previous study also reported that T cells could express *Arg1* (***Washburn et al., 2019***), although the functional relevance of this observation has remained obscure.

To address this knowledge gap, we first confirmed expression at the protein level via Western blotting (***Figure 3A***) in experiments incorporating control mice lacking the ability to express Arg1 in the CD4$^+$ lineage (*Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$). We then revealed the open chromatin structure around *Arg1* in Thy1.1$^+$ CD4$^+$ T cells (***Figure 3B***) and further probed the expression of Arg1 versus Thy1.1 among

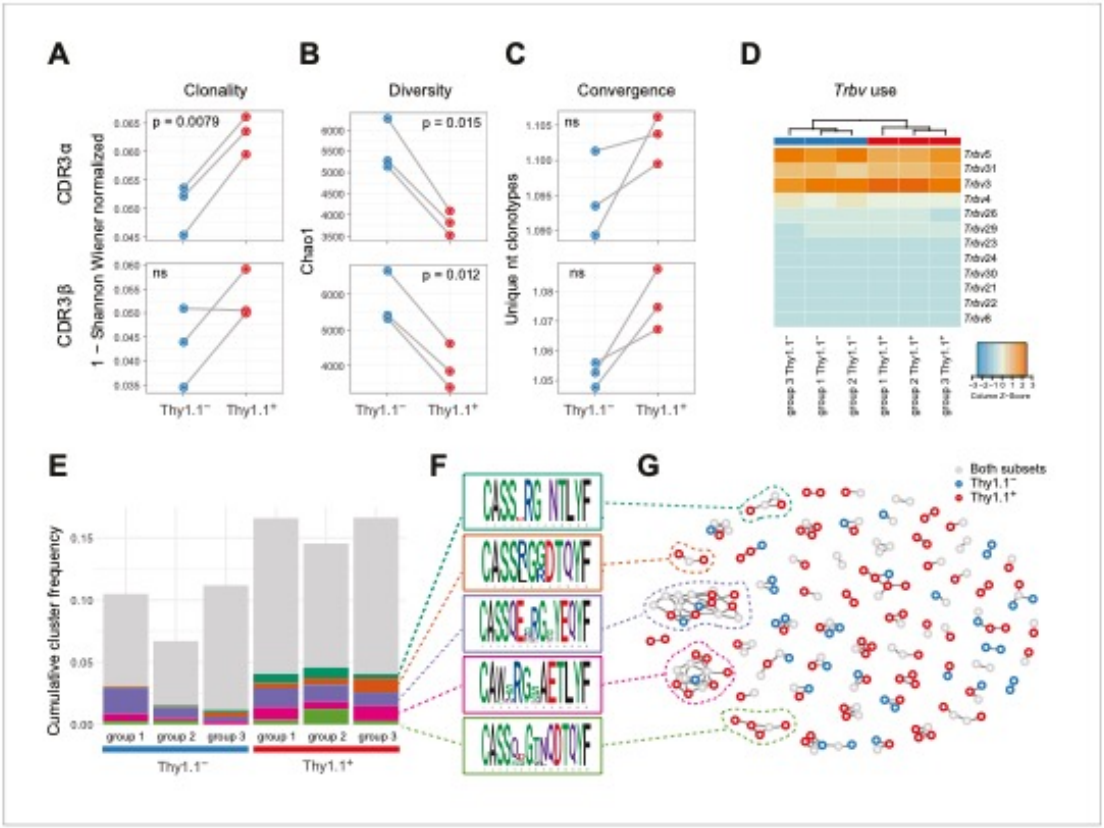

Figure 2. Interleukin (IL)-10-producing CD4⁺ T cells exhibit prominent clonal structures. 10BiT mice were infected with 3 × 10^4 pfu of murine cytomegalovirus (MCMV). Leukocytes were isolated from the salivary glands (SGs) on day 14 p.i. and sorted as CD4⁺ CD44⁺ CD62L⁻ CD90/90.1⁺ (Thy1.1⁺) or CD90/90.1⁻ (Thy1.1⁻) populations via fluorescence-activated cell sorting (FACS). (**A**) Clonality and (**B**) diversity metrics calculated for the T cell receptor (TCR)α (top) and TCRβ repertoires (bottom) derived from Thy1.1⁺ CD4⁺ T cells and Thy1.1⁻ CD4⁺ T cells. (**C**) TCR convergence measured as the average number of nucleotide sequences encoding amino acid-identical complementarity-determining region (CDR)3α (top) and CDR3β loops (bottom) across the 2000 most prevalent clonotypes. (**A–C**) p values were calculated using a paired t-test with Benjamini–Hochberg correction. ns, not significant. (**D**) Hierarchical clustering of *Trbv* gene use weighted by clonotype frequency. (**E–G**) Cluster analysis of the 500 most prevalent TCRβ clonotypes using the tcrgrapher pipeline. (**E**) The cumulative frequency of tcrgrapher hits per sample is shown in gray. The frequency of each cluster comprising at least two tcrgrapher hits was calculated for each sample and averaged across all six repertoires. The five most prevalent clusters are shown in color. (**F**) Amino acid logos for each of the five most prevalent clusters. (**G**) Visual representation of clusters comprising at least two tcrgrapher hits. Nodes represent unique amino acid sequences. Edges connect sequences with a single amino acid mismatch. Amino acid sequences present only in Thy1.1⁺ CD4⁺ T cells are shown in red, amino acid sequences present only in Thy1.1⁻ CD4⁺ T cells are shown in blue, and amino acid sequences present in both Thy1.1⁺ CD4⁺ T cells and Thy1.1⁻ CD4⁺ T cells are shown in gray. Data are shown as pooled analyses from n = 4 mice per group representing three independent experiments (groups 1–3).

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Analysis of repertoire diversity and overlap and the physicochemical properties of T cell receptors (TCRs).

CD4⁺ T cells via flow cytometry (***Figure 3C***). Our data showed that Arg1 was expressed by CD4⁺ T cells almost exclusively in the SGs (***Figure 3A, C***). Of note, Arg1 expression was also detected among CD8⁺ T cells but not among NK T cells via flow cytometry, although intracellular discrimination was subtle (***Figure 3—figure supplement 1A***). Depletion experiments nonetheless revealed that only CD4⁺ T

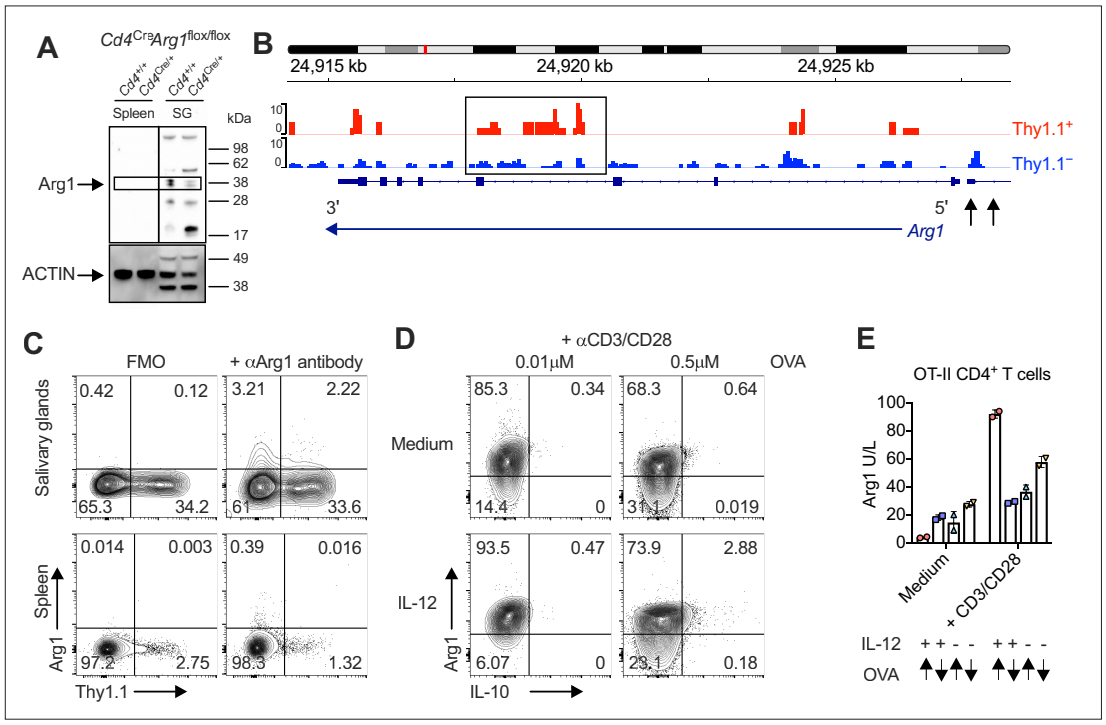

Figure 3. Interleukin (IL)-10-producing CD4$^+$ T cells are enriched for expression of arginase-1 (Arg1). (**A**) Expression of Arg1 among leukocytes isolated via magnetic separation from the salivary glands (SGs) and spleens of *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ or *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice on day 14 p.i. detected by Western blot. (**B, C**) 10BiT mice were infected with 3 × 10^4 pfu of murine cytomegalovirus (MCMV). (**B**) Leukocytes were isolated from the SGs on day 14 p.i. and sorted as CD4$^+$ CD44$^+$ CD62L$^-$ CD90/90.1$^+$ (Thy1.1$^+$) or CD90/90.1$^-$ (Thy1.1$^-$) populations via fluorescence-activated cell sorting (FACS). ATAC-seq profiles show accessible chromatin regions in the *Arg1* gene for Thy1.1$^+$ CD4$^+$ T cells (red) and Thy1.1$^-$ CD4$^+$ T cells (blue). Data are shown as normalized values accounting for the total number of reads per lane. The black box indicates a major difference in chromatin accessibility. Black arrows indicate binding motifs for Tbx21. Data are shown as pooled analyses from a minimum of *n* = 5 mice per group representing three independent experiments. (**C**) Representative flow cytometry plots showing the expression of Arg1 versus Thy1.1 among CD4$^+$ T cells isolated from the SGs and spleens on day 14 p.i. (**A, C**) Data are shown as pooled analyses from a minimum of *n* = 7 mice per group representing three independent experiments. (**D**) Representative flow cytometry plots showing the expression of Arg1 versus IL-10 among OT-II-specific CD4$^+$ T cells generated in vitro in the absence or presence of IL-12 (3 ng/ml) ± $OVA_{323–339}$ for 7 days and then stimulated with anti-CD3/CD28 for 4 hr. (**E**) Summary bar graph showing Arg1 protein concentrations in culture supernatants from (**D**) after stimulation with anti-CD3/CD28 for 48 hr. Up arrows indicate the higher concentration of $OVA_{323–339}$ (0.5 μM), and down arrows indicate the lower concentration of $OVA_{323–339}$ (0.01 μM). Data are shown as mean ± standard error of the mean (SEM).

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Interleukin (IL)-10-producing CD4$^+$ T cells are enriched for expression of arginase-1 (Arg1).

Source data 2. Interleukin (IL)-10-producing CD4$^+$ T cells are enriched for expression of arginase-1 (Arg1).

Figure supplement 1. T cell-specific deletion of arginase-1 (Arg1) does not impact the function or phenotype of T cells in naive mice.

Figure supplement 1—source data 1. T cell-specific deletion of arginase-1 (Arg1) does not impact the function or phenotype of T cells in naive mice.

cells contributed significantly to Arg1 protein concentrations in SG homogenates during chronic infection with MCMV (***Figure 3—figure supplement 1B***).

IL-10-producing T_H1 cells can be induced experimentally via high-dose antigen stimulation in the presence of IL-12 (***Saraiva et al., 2009***). Accordingly, we hypothesized that Arg1$^+$ IL-10-producing CD4$^+$ T cells might develop under similar conditions in vitro, given the corresponding T_H1-like profile observed during chronic infection with MCMV. To test this notion, we stimulated ovalbumin

(OVA)-specific transgenic CD4$^+$ T cells from OT-II mice with high or low doses of the cognate peptide ($OVA_{323–339}$) in the absence or presence of IL-12. Despite constitutively high expression levels of Arg1, only high-dose $OVA_{323–339}$ in combination with IL-12 induced the development of CD4$^+$ T cells that expressed Arg1 and IL-10, and importantly, all IL-10-producing CD4$^+$ T cells coexpressed Arg1 (***Figure 3D***). Arg1 protein concentrations in culture supernatants also increased substantially after stimulation with high-dose $OVA_{323–339}$ in the presence of IL-12 (***Figure 3E***).

Collectively, these data suggested that Arg1 expression was a hallmark of IL-10-producing CD4$^+$ T cells, which developed almost exclusively in the SGs during chronic infection with MCMV.

CD4$^+$ T cells promote viral persistence via expression of Arg1

To explore the biological relevance of our findings, we infected *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ and *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice with MCMV. Lineage-specific deletion of *Arg1* did not impact the function or phenotype of CD4$^+$ T cells in naive mice (***Figure 3—figure supplement 1C, D***). Higher numbers of IFNγ-expressing CD4$^+$ T cells (***Figure 4A***) and higher frequencies of proliferating (Ki-67$^+$) CD4$^+$ T cells (***Figure 4B***) were nonetheless observed after viral antigen stimulation in the SGs of *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ versus *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ mice during the chronic phase of infection with MCMV. These data suggested that Arg1 inhibited the proliferation of CD4$^+$ T cells in vivo, consistent with a previous in vitro study (***Munder et al., 2006***). Similarly, higher numbers of virus-specific CD8$^+$ T cells, quantified using tetrameric antigen probes, were detected in the spleens of *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ versus *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ mice on day 30 p.i. (***Figure 4C***). In contrast, *Arg1* deletion had no significant impact on the development of virus-specific IL-10-producing CD4$^+$ T cells in the SGs (***Figure 4D***).

IFNγ-expressing CD4$^+$ T cells are known to restrict MCMV replication in the SGs (***Walton et al., 2011***). Accordingly, we found that viral shedding in the saliva (***Figure 4E***) and viral replication in the SGs (***Figure 4F***) were reduced in *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ versus *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ mice during the chronic phase of infection with MCMV. Importantly, no differences in viral replication were observed between *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ and control mice at an earlier time point (day 14 p.i.) (***Figure 4—figure supplement 1A***), and deletion of *Arg1* in myeloid cells, achieved using *Lyz2*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice, had no impact on viral shedding in the saliva or viral replication in the SGs (***Figure 4—figure supplement 1B, C***). Of note, there was also no evidence that *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice were more susceptible to virus-induced autoimmunity, as indicated by anti-Sjögrens syndrome antigen (SSA) titers comparable to those observed in *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ mice (***Figure 4—figure supplement 1D***).

Collectively, these data indicated that Arg1 expression by CD4$^+$ T cells selectively inhibited the accumulation of virus-specific CD4$^+$ and CD8$^+$ T cells in the SGs, leading to suboptimal immune control of viral replication during chronic infection with MCMV.

IL-10-producing CD4$^+$ T cells develop in a T-bet-dependent manner

In line with our previous work (***Clement et al., 2016***), we noted that Thy1.1$^+$ CD4$^+$ T cells expressed higher amounts of T-bet at the protein level than Thy1.1$^-$ CD4$^+$ T cells (***Figure 5A***), and concordantly, we found that open chromatin was enriched in the *Tbx21* region of Thy1.1$^+$ CD4$^+$ T cells versus Thy1.1$^-$ CD4$^+$ T cells (***Figure 5B***). No such differences in expression intensity were observed for the related T-box transcription factor Eomesodermin (Eomes) (***Figure 5—figure supplement 1A, B***), which promotes the development of Tr1 cells (***Roessner et al., 2021***; ***Zhang et al., 2017***). We also detected considerable overlap between the gene expression profiles of Thy1.1$^+$ CD4$^+$ T cells and T-bet-orchestrated CD4$^+$ T_H1 cells generated in vitro (***Zhu et al., 2012***; ***Figure 5C***). Moreover, T-bet suppresses the expression of TCF1/7 and IL-7R (***Dominguez et al., 2015***; ***Oestreich et al., 2011***), mimicking key phenotypic features of Thy1.1$^+$ CD4$^+$ T cells (***Figure 1D, F***). On the basis of these observations, we hypothesized that T-bet promoted the development of Arg1$^+$ IL-10-producing CD4$^+$ T cells during chronic infection with MCMV.

To test this notion, we crossed tamoxifen-inducible *Cd4*$^{CreERT2/+}$ mice with *Tbx21*$^{flox/flox}$ mice (***Intlekofer et al., 2008***), allowing us to suppress T-bet expression at the onset of viral chronicity (day 7 p.i.). These mice were further crossed to incorporate a *Rosa26*^{flSTOPtdRFP} allele (***Luche et al., 2007***), enabling the identification of cells in which *Cre* was expressed via the detection of tandem-dimer red fluorescent protein (tdRFP). Mice were then infected with MCMV. Tamoxifen was administered daily for 5 days from day 7 p.i. to deplete T-bet after initial antigen exposure in a CD4-dependent manner, an

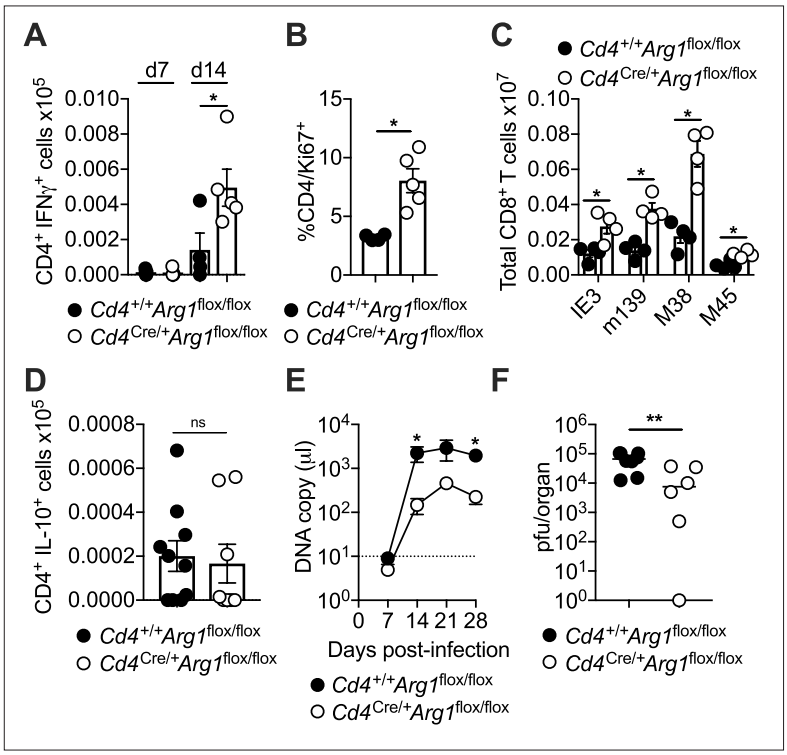

Figure 4. CD4⁺ T cells promote viral persistence via expression of arginase-1 (Arg1). *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ and *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice were infected with 3 × 10⁴ pfu of murine cytomegalovirus (MCMV). (**A**) MCMV-specific CD4⁺ T cell responses in the salivary glands (SGs) on days 7 and 14 p.i. measured using flow cytometry to detect interferon (IFN)γ. Immunodominant peptides were pooled for stimulation. Data are shown as mean ± standard error of the mean (SEM; *n* = 4–6 mice per group representing three independent experiments). (**B**) Expression of Ki-67 among CD4⁺ T cells isolated from the SGs on day 14 p.i. measured via flow cytometry. Data are shown as mean ± SEM (*n* = 4–5 mice per group representing two independent experiments). (**C**) MCMV tetramer⁺ CD8⁺ T cells quantified in spleens on day 30 p.i. via flow cytometry. Data are shown as mean ± SEM (*n* = 4 mice per group representing two independent experiments). (**D**) MCMV-specific CD4⁺ T cell responses in the SGs on days 7 and 14 p.i. measured using flow cytometry to detect interleukin (IL)-10. Immunodominant peptides were pooled for stimulation. Data are shown as mean ± SEM (*n* = 8–10 mice per group representing two independent experiments). (**A–D**) *p < 0.05 (Mann–Whitney *U* test). (**E**) Viral genomes in saliva on days 7, 14, 21, and 28 p.i. measured via qPCR. Data are shown as mean ± SEM (*n* = 8 mice per group representing two independent experiments). *p < 0.05 (Mann–Whitney *U* test). (**F**) MCMV replication in SG homogenates on day 30 p.i. measured via plaque assay. Data are shown as individual points with median values (*n* = 6–7 mice per group representing two or three independent experiments). **p < 0.01 (Mann–Whitney *U* test).

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. CD4⁺ T cells promote viral persistence via expression of arginase-1 (Arg1).

Figure supplement 1. Arginase-1 (Arg1) expression in myeloid cells does not impact the replication of murine cytomegalovirus (MCMV).

Figure supplement 1—source data 1. Arginase-1 (Arg1) expression in myeloid cells does not impact the replication of murine cytomegalovirus (MCMV).

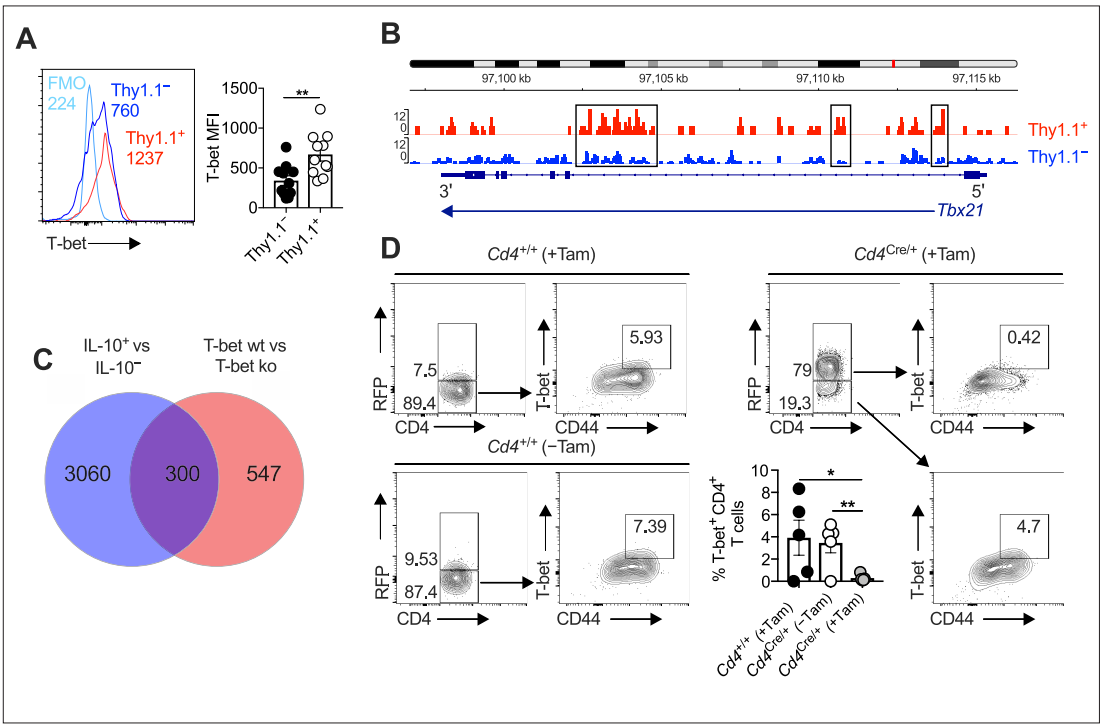

Figure 5. Interleukin (IL)-10-producing CD4$^+$ T cells express T-bet and T-bet-inducible genes. (**A, B**) 10BiT mice were infected with 3 × 10^4 pfu of murine cytomegalovirus (MCMV). (**A**) Leukocytes were isolated from the salivary glands (SGs) on day 14 p.i. Representative histograms (left) and summary bar graph (right) show the expression of T-bet among Thy1.1$^+$ CD4$^+$ T cells (red) and Thy1.1$^-$ CD4$^+$ T cells (blue). Data are shown as mean ± standard error of the mean (SEM; *n* = 5–6 mice per group pooled from two independent experiments). MFI, median fluorescence intensity. **p < 0.01 (Mann–Whitney *U* test). (**B**) Leukocytes were isolated from the SGs on day 14 p.i. and sorted as CD4$^+$ CD44$^+$ CD62L$^-$ CD90/90.1$^+$ (Thy1.1$^+$) or CD90/90.1$^-$ (Thy1.1$^-$) populations via fluorescence-activated cell sorting (FACS). ATAC-seq profiles show accessible chromatin regions in the *Tbx21* gene for Thy1.1$^+$ CD4$^+$ T cells (red) and Thy1.1$^-$ CD4$^+$ T cells (blue). Data are shown as normalized values accounting for the total number of reads per lane. The black boxes indicate major differences in chromatin accessibility. Data are shown as pooled analyses from a minimum of *n* = 5 mice per group representing three independent experiments. (**C**) Venn diagram showing the overlap between genes enriched in Thy1.1$^+$ CD4$^+$ T cells (**A, D**) and genes enriched in T-bet$^+$ CD4$^+$ T cells (GSE38808). (**D**) *Cd4*$^{+/+}$*Tbx21*$^{flox/flox}$ (*Cd4*$^{+/+}$) and *Cd4*$^{CreERT2/+}$*Tbx21*$^{flox/flox}$ (*Cd4*$^{Cre/+}$) mice were infected with 3 × 10^4 pfu of MCMV. Tamoxifen was administered (+Tam) or withheld (−Tam) from days 7 to 12 p.i. Leukocytes were isolated from the SGs on day 14 p.i. Representative flow cytometry plots show the expression of CD4 versus RFP (left) and CD44 versus T-bet for the gated populations (right). Data in the bar graph are shown as mean ± SEM (*n* = 4–5 mice per group representing four or five independent experiments). *p < 0.05, **p < 0.01 (Mann–Whitney *U* test).

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. Interleukin (IL)-10-producing CD4$^+$ T cells express T-bet and T-bet-inducible genes.

Figure supplement 1. Interleukin (IL)-10-producing CD4$^+$ T cells are not selectively regulated by Eomes.

Figure supplement 1—source data 1. Interleukin (IL)-10-producing CD4$^+$ T cells are not selectively regulated by Eomes.

effect that clearly associated with the coincident expression of RFP (***Figure 5D*** and ***Figure 5—figure supplement 1C***).

CD4-specific T-bet depletion reduced the accumulation of virus-specific IL-10-producing CD4$^+$ T cells (***Figure 6A***) and Arg1$^+$ CD4$^+$ T cells by day 14 p.i. (***Figure 6B, C***). However, it seemed unlikely that T-bet directly stimulated the expression of IL-10 and Arg1, because the corresponding binding motifs were not preferentially accessible in the *Il10* and *Arg1* regions (***Figure 1B*** and ***Figure 3B***). Instead, we observed a shift toward a less differentiated phenotype in *Cd4*$^{CreERT2/+}$*Tbx21*$^{flox/flox}$ mice (***Figure 6D–G***),

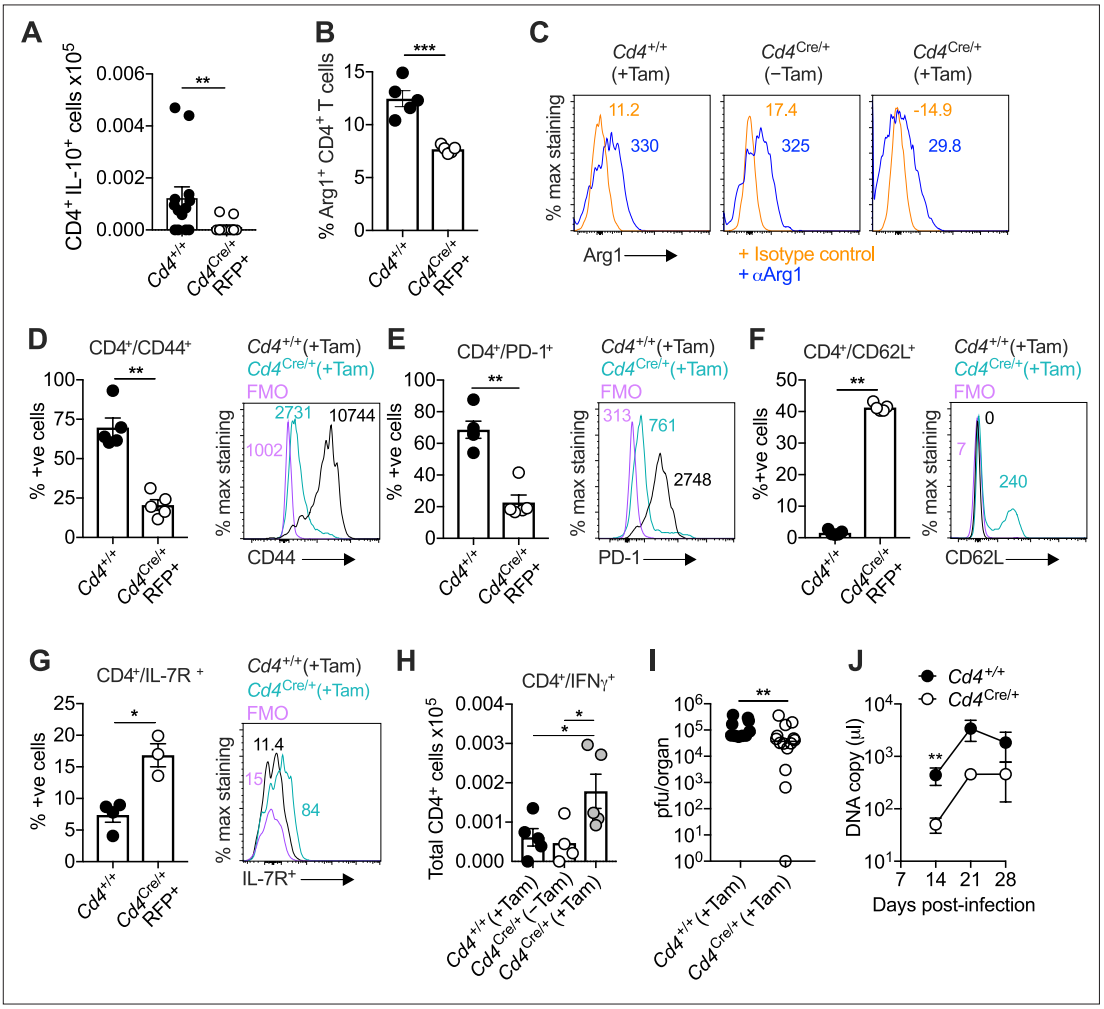

Figure 6. Interleukin (IL)-10-producing CD4⁺ T cells develop in a T-bet-dependent manner. (**A–J**) *Cd4*$^{+/+}$*Tbx21*$^{flox/flox}$ (*Cd4*$^{+/+}$) and *Cd4*$^{CreERT2/+}$*Tbx21*$^{flox/flox}$ (*Cd4*$^{Cre/+}$) mice were infected with 3 × 10⁴ pfu of murine cytomegalovirus (MCMV). Tamoxifen was administered (+Tam) or withheld (−Tam) from days 7 to 12 p.i. Leukocytes were isolated from the salivary glands (SGs) on day 14 p.i. (**A**) MCMV-specific CD4⁺ T cell responses in the SGs measured using flow cytometry to detect IL-10. Immunodominant peptides were pooled for stimulation. Data are shown as mean ± standard error of the mean (SEM; *n* = 4–5 mice per group representing two independent experiments). **p < 0.01 (Mann–Whitney *U* test). Summary bar graph (**B**) and representative histograms (**C**) showing the expression of arginase-1 (Arg1) among CD4⁺ T cells measured via flow cytometry. Data are shown as mean ± SEM (*n* = 5 mice per group representing two independent experiments). ***p < 0.001 (Mann–Whitney *U* test). Summary bar graphs (left) and representative histograms (right) showing the expression of CD44 (**D**), PD-1 (**E**), CD62L (**F**), and IL-7R (**G**) among CD4⁺ T cells measured via flow cytometry. Data in (**D–F**) are shown as mean ± SEM (*n* = 5 mice per group representing two independent experiments). Data in (**G**) are shown as mean ± SEM (*n* = 3–4 mice per group representing two independent experiments). *p < 0.05, **p < 0.01 (Mann–Whitney *U* test). (**H**) MCMV-specific CD4⁺ T cell responses in the SGs measured using flow cytometry to detect interferon (IFN)γ. Immunodominant peptides were pooled for stimulation. Data are shown as mean ± SEM (*n* = 4–5 mice per group representing two independent experiments). *p < 0.05 (Mann–Whitney *U* test). (**I**) MCMV replication in SG homogenates on day 14 p.i. measured via plaque assay. Data are shown as individual points with median values (*n* = 10–14 mice per group pooled from three independent experiments). **p < 0.01 (Mann–Whitney *U* test). (**J**) Viral genomes in saliva on days 7, 14, 21, and 28 p.i. measured via qPCR. Data are shown as mean

Figure 6 continued on next page

Figure 6 continued

± SEM (*n* = 5 mice per group representing two independent experiments). **p < 0.01 (Mann–Whitney *U* test). Data in (**A–I**) show all groups after the administration of tamoxifen.

The online version of this article includes the following source data for figure 6:

Source data 1. Interleukin (IL)-10-producing CD4$^+$ T cells develop in a T-bet-dependent manner.

with decreased expression of CD44 (***Figure 6D***) and PD-1 (***Figure 6E***) and increased expression of CD62L (***Figure 6F***) and IL-7R (***Figure 6G***) in the absence of T-bet. In addition, T-bet depletion was associated with increased numbers of virus-specific IFNγ-expressing CD4$^+$ T cells (***Figure 6H***) and enhanced control of viral replication in the SGs (***Figure 6I***), as well as reduced salivary shedding of MCMV (***Figure 6J***). In accordance with the observation that IL-10-producing CD4$^+$ T cells occur transiently during the early stages of viral chronicity (***Clement et al., 2016***), we also found that CD4-specific T-bet depletion was no longer protective by day 28 p.i. (***Figure 6J*** and ***Figure 5—figure supplement 1D***).

Collectively, these results suggested that T-bet exhibited dual functionality under conditions of recurrent antigen stimulation, concurrently driving the accumulation of virus-specific T_H1 responses and promoting the development of Arg1$^+$ IL-10-producing CD4$^+$ T cells, which together shaped immune control of viral replication during the early stages of chronic infection with MCMV.

Discussion

In this study, we identified IL-10-producing CD4$^+$ T cells as highly differentiated T_H1-like cells, which occurred transiently at mucosal sites of viral persistence under the strict governance of T-bet. These cells also expressed the metabolic enzyme Arg1. Further investigation revealed that Arg1 expression was a critical immune regulatory function of CD4$^+$ T cells, suppressing antiviral immunity and facilitating viral chronicity in the context of infection with MCMV.

Our finding that *Arg1* was one of the putative immune regulatory genes expressed by IL-10-producing CD4$^+$ T cells was initially counterintuitive, given the reported lack of Arg1 protein expression by human CD4$^+$ T cells in vitro (***Geiger et al., 2016***). However, *Arg1* expression by CD4$^+$ and CD8$^+$ T cells has been described in patients with sepsis (***Washburn et al., 2019***), indicating a potential role for generic infectious stimuli and/or an antigen-driven process mediated via the TCR. The latter possibility is consistent with our repertoire data, which revealed that prominent clonal expansions were associated with the production of IL-10. Importantly, we demonstrated that Arg1 was expressed at the protein level, both in secreted form and intracellularly. It should be noted that relatively weak intracellular signals were detected in flow cytometric analyses of C57BL/6 mice, akin to an earlier study of patients infected with HBV (***Pallett et al., 2015***). Western blotting experiments nonetheless clearly showed that CD4$^+$ T cells expressed Arg1 in the SGs of mice chronically infected with MCMV. Our flow cytometry experiments also suggested that CD8$^+$ T cells could express Arg1, albeit with the caveat of suboptimal visualization, but importantly, CD4$^+$ T cells secreted far greater amounts of soluble Arg1, at least in the milieu of the SGs.

L-arginine deficiency results in cell cycle arrest via the downregulation of CD3 ζ , a key component of the TCR (***Rodriguez et al., 2002***), and T cell proliferation is impaired in the absence of L-arginine (***Rodriguez et al., 2007***). In line with these observations, we found that improved control of viral replication in mice lacking Arg1$^+$ CD4$^+$ T cells was associated with increased numbers of MCMV-specific effector CD4$^+$ T cells, which are known to limit viral replication in the SGs (***Lucin et al., 1992***; ***Walton et al., 2011***). Accordingly, the acquisition of Arg1 expression by highly differentiated CD4$^+$ T cells impinges on antiviral immunity, representing an inhibitory mechanism that operates alongside the production of IL-10 (***Clement et al., 2016***; ***Humphreys et al., 2007***). It is also notable here that regulatory T cells may act similarly via the expression of Arg2 (***Lowe et al., 2019***).

IL-10 is regulated by a number of transcription factors, including those required for T cell differentiation, and is induced downstream of the TCR (***Saraiva et al., 2020***). Our discovery that T-bet promotes the development of IL-10-producing CD4$^+$ T cells aligns with the concept that chronic antigen stimulation can trigger a negative feedback loop via these mechanisms, which concurrently drive activation and differentiation. It nonetheless remains less clear to what extent T-bet directly orchestrates the expression of other inhibitory molecules by CD4$^+$ T cells. There was no notable

increase in chromatin-accessible binding motifs for T-bet within either the *Il10* or *Arg1* genes in IL-10-producing CD4$^+$ T cells. However, we did find that deletion of T-bet after acute infection limited the accumulation of highly differentiated CD4$^+$ T cells, including those expressing Arg1 and IL-10. In line with this observation, which suggested a differentiation-linked process of functional remodeling driven by chronic antigen exposure, T-bet is known to promote the differentiation of CD8$^+$ T cells (***Dominguez et al., 2015***; ***Omilusik et al., 2015***) and inhibit the expression of TCF1/7 (***Omilusik et al., 2015***) and IL-7R (***Intlekofer et al., 2007***), both of which were downregulated among Arg1$^+$ IL-10-producing CD4$^+$ T cells induced by MCMV.

The relationship between mucosal IL-10-producing CD4$^+$ T cells and classical Tr1 cells requires further clarification. In contrast to Tr1 cells (***Gagliani et al., 2013***), IL-10-producing CD4$^+$ T cells in the SGs do not coexpress CD49b and LAG-3 (***Clement et al., 2016***), highlighting the diverse nature of regulatory CD4$^+$ T cells defined by the secretion of IL-10 (***Brockmann et al., 2018***). However, these cells do express multiple inhibitory receptors, including LAG-3, alongside other markers that typically characterize Tr1 cells, such as CCR5 and granzyme K (***Brockmann et al., 2018***; ***Roessner et al., 2021***; ***Thelen et al., 2023***). In our model, Thy1.1$^+$ CD4$^+$ T cells also expressed Eomes, akin to clonally expanded T-bet-dependent Tr1 cells (***Thelen et al., 2023***). Accordingly, mucosal IL-10-producing CD4$^+$ T cells may develop along similar lines to conventional Tr1 cells, which differentiate under the control of Eomes in conjunction with Blimp-1 (***Roessner et al., 2021***; ***Zhang et al., 2017***).

The mechanisms that directly induce the expression of *Il10* and *Arg1* in CD4$^+$ T cells remain obscure. Although IL-10-producing CD4$^+$ T cells were characterized by prominent clonal expansions, repertoire analysis revealed no obvious determinative role for the TCR. However, the transcription factor c-Maf, which promotes the expression of IL-10 in multiple T_H cell subsets (***Gabryšová et al., 2018***), is known to be upregulated in IL-10-producing CD4$^+$ T cells during chronic infection with MCMV (***Clement et al., 2016***). We also demonstrated previously that ICOS signaling promotes the accumulation of IL-10-producing CD4$^+$ T cells under the same conditions (***Clement et al., 2016***), and c-Maf acts downstream of ICOS (***Bauquet et al., 2009***; ***Nurieva et al., 2003***). It therefore seems likely that an ICOS–c-Maf axis participates in the direct induction of *Il10*, although it is less clear how this applies to *Arg1*. Moreover, we found that Arg1$^+$ IL-10-producing CD4$^+$ T cells developed optimally in vitro in response to high-dose antigen stimulation in the presence of IL-12, suggesting determinative roles for ERK MAP kinase and the transcription factor STAT4 (***Saraiva et al., 2009***).

The evolutionary advantage of a mechanism that leads to the production of immune regulatory molecules and consequently facilitates viral persistence without impacting autoimmunity is difficult to explain. One possibility is that some of these molecules, potentially including Arg1, serve to maintain tissue health in the presence of ongoing inflammation. It would be informative in this context to evaluate how Arg1 delivery via CD4$^+$ T cells impacts mucosal pathology over time at sites beyond the SGs. Alternatively, local immune suppression may somehow limit the emergence of more pathogenic viral strains under conditions of persistent replication, thereby protecting the host and the wider population. Irrespective of the precise explanation, it seems clear that viral persistence can be facilitated by this intrinsic regulatory mechanism, a phenomenon that could feasibly extend to pathogens other than MCMV.

In summary, we have demonstrated that T-bet activity during a chronic viral infection can impede antiviral immune control by driving the development of highly differentiated T_H1-like cells that express genes encoding inhibitory molecules, including IL-10 and Arg1. Importantly, these Arg1$^+$ IL-10-producing CD4$^+$ T cells developed in vitro under conditions of extreme antigen stimulation, especially in the presence of IL-12, and the expression of Arg1, which was also secreted in vivo as a soluble immune regulatory protein, facilitated viral replication during the chronic phase of infection with MCMV. These observations are conceivably important not only from a biological perspective but also from a translational perspective, revealing a previously unappreciated mechanism through which CD4$^+$ T cells can suppress potentially harmful immune responses via the regulation of L-arginine.

Materials and methods

Mice

10BiT reporter mice were originally derived by Padraic Fallon (Trinity College Dublin, Republic of Ireland). These mice express CD90/CD90.1 (Thy1.1) under the control of the *Il10* promoter and retain

endogenous expression of IL-10 (**Maynard et al., 2007**). *Cd4*$^{+/+}$*Tbx21*$^{flox/flox}$ mice and *Cd4*$^{CreERT2/+}$*Tbx21*$^{flox/flox}$ mice were generated by crossing *Rosa26*^{flSTOPtdRFP} mice (**Luche et al., 2007**) with *Tbx21*$^{flox/flox}$ mice (**Intlekofer et al., 2008**). *Arg1*$^{flox/flox}$ mice (JAX stock #008817) and *Lyz2*$^{Cre/+}$ mice (JAX stock #004871) were purchased from The Jackson Laboratory. These mice were bred to generate *Lyz2*$^{+/+}$*Arg1*$^{flox/flox}$ mice and *Lyz2*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice. *Arg1*$^{flox/flox}$ mice were further bred with *Cd4*$^{Cre/+}$ mice to generate *Cd4*$^{+/+}$*Arg1*$^{flox/flox}$ mice and *Cd4*$^{Cre/+}$*Arg1*$^{flox/flox}$ mice. C57BL/6 WT mice (JAX stock #000664) were purchased from Charles River or Envigo. OT-II mice (JAX stock #004194) were purchased from The Jackson Laboratory. *Cd4*$^{Cre/+}$ mice (JAX stock #022071) were a kind gift from Sarah Dimeloe (University of Birmingham, UK).

Infections and treatments

MCMV was prepared via sorbital gradient purification as described previously (**Stacey et al., 2014**). Mice were infected with 3×10^4 pfu of MCMV intraperitoneally. *Cd4*$^{+/+}$ and *Cd4*$^{Cre/+}$ mice were injected intraperitoneally with tamoxifen (Sigma-Aldrich) as indicated for 5 days from day 7 p.i. at a daily dose of 75 mg/kg (20 mg/ml). In some experiments, mice were injected intraperitoneally with anti-CD4 (100 µg of clone GK1.5 and 100 µg of clone YTS191) and/or anti-CD8 (100 µg of clone YTS169.4 and 100 µg of clone YTS156.7.7) or an isotype control, rat IgG2b anti-keyhole limpet hemocyanin (400 µg of clone LTF-2), on days 6 and 8 p.i. (all from BioXell).

Next-generation sequencing

Leukocytes were isolated directly ex vivo from the SGs of 10BiT mice on day 14 p.i. (minimum $n = 5$ mice per group with three replicates). Pooled cells were labeled using a Zombie Aqua Fixable Viability Kit (BioLegend) and stained with anti-CD16/CD32 (Fc block, BioLegend) followed by anti-CD4–BV605 (clone RM4-5, BioLegend), anti-CD44–FITC (clone IM7, BioLegend), anti-CD62L–PE-Cy7 (clone MEL-14, BioLegend), and anti-CD90/90.1–PE (clone OX-7, BioLegend). Cells were sorted as CD4$^+$ CD44$^+$ CD62L$^-$ CD90/90.1$^+$ (Thy1.1$^+$) or CD90/90.1$^-$ (Thy1.1$^-$) populations directly into Buffer RLT or Buffer RLT Plus (QIAGEN) using a modified FACS Aria II (BD Biosciences). Total RNA was extracted using an RNeasy Micro Kit or an RNeasy Mini Kit (QIAGEN), and RNA integrity scores were determined using an RNA 6000 Pico Kit (Agilent).

RNA-seq

Multiplexed mRNA libraries were obtained using a TruSeq RNA Library Prep Kit v2 (Illumina) and analyzed using a Bioanalyzer High Sensitivity DNA Chip (Agilent). Libraries were sequenced using a HiSeq 2500 System (Illumina). Paired-end reads (100 bp) were trimmed using Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and assessed for quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the mouse GRCm38 reference genome using STAR (**Dobin et al., 2013**). Counts were assigned to transcripts using featureCounts (**Liao et al., 2014**) with the GRCm38.84 Ensembl gene build GTF (http://www.ensembl.org/info/data/ftp/index.html/). Differential gene expression analyses were performed using DESeq2 (**Love et al., 2014**). Genes were discarded from the analysis if differential expression failed to reach significance (adjusted $p < 0.05$ with Benjamini–Hochberg correction).

ATAC-seq

ATAC-seq was performed as described previously (**Buenrostro et al., 2013**) using a Nextera DNA Sample Preparation Kit (Illumina). DNA was isolated after library preparation using a MiniElute PCR Purification Kit (QIAGEN). Size selection was performed using a BluePippin System (Sage Science) with 2% Agarose Gel Cassettes (Sage Science). Libraries were sequenced using a HiSeq 4000 System (Illumina). Paired-end reads (100 bp) were trimmed using Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and assessed for quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were mapped to the mouse GRCm38 reference genome using BWA (**Li and Durbin, 2009**). Duplicate reads were removed using MarkDuplicates (http://broadinstitute.github.io/picard/). Peaks were called using MACS2 (**Zhang et al., 2008**) in BAMPE mode (adjusted $p < 0.05$ with Benjamini–Hochberg correction). Differential analysis of open regions was performed using DiffBind (http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf).

TCR-seq

Unique molecular identifier (UMI)-labeled 5′-RACE TCR cDNA libraries were synthesized using a Mouse TCR Profiling Kit (MiLaboratories). Indexed samples were pooled and sequenced using a MiSeq System (Illumina). Paired-end reads (150 bp) were extracted and clustered by UMI using MiGEC (**Shugay et al., 2014**). Sequences were discarded from the analysis if the read count was <4 per cDNA. Error correction was performed using MiGEC. Repertoires were extracted using MiXCR (**Bolotin et al., 2015**). The weighted average use of bulky, charged, and strongly interacting (aromatic and hydrophobic) amino acids positioned centrally in the CDR3α/β sequences and *Trbv* gene use (weighted by frequency) were calculated using VDJtools (**Shugay et al., 2015**). Diversity was calculated by downsampling the repertoires to an equal number of UMIs (n = 4300 for TCRα and n = 3900 for TCRβ) three separate times and plotting the mean Chao1 index, reflecting lower bound total diversity, and (1 − normalized Shannon–Wiener index), reflecting clonality. Visualization was achieved using PlotQuantileStats in VDJtools. The mean number of unique nucleotide sequences encoding the same amino acid sequence (convergence) was calculated for the 2000 most prevalent clonotypes in each sample using VDJtools. Overlap was calculated for the 2000 most prevalent clonotypes using F2 metrics to estimate sharing at the amino acid level among *V* gene-matched sequences in each sample. Cluster analysis was performed using the 500 most prevalent clonotypes in pooled samples (Thy1.1$^+$ versus Thy1.1$^-$). These datasets were analyzed using tcrgrapher (https://github.com/KseniaMIPT/tcrgrapher), an R library based on ALICE (**Pogorelyy et al., 2019**). The real and expected numbers of neighbors were calculated for each clonotype with a maximum of n = 1 amino acid mismatch. Clonotypes with a significantly higher number of neighbors than expected on statistical grounds (adjusted $p < 0.0001$ with Benjamini–Hochberg correction) were identified as tcrgrapher hits. The expected number of neighbors was estimated via generation probabilities calculated for each clonotype using OLGA (**Sethna et al., 2019**), with the selection factor set at Q = 6.27 (**Elhanati et al., 2018**).

Bioinformatics

RNA-seq analysis was performed using DESeq2/1.32.0, dplyr/1.0.7, and GenomicRanges/1.44.0. Volcano plots were drawn using ggplot2/3.3.5 and ggrepel/0.9.1. TCR-seq analysis was performed using MiGEC/1.2.9, MiXCR/3.0.13, VDJtools/1.2.1, tidyverse/1.3.1, igraph/1.2.6, ggnetwork/0.5.10, msa/1.22.0, tcrgrapher/0.0.09000, stringdist/0.9.6.3, and ggseqlogo/0.1. Other software packages used across next-generation sequencing platforms included Nextflow (https://www.nextflow.io/) (**Di Tommaso et al., 2017**), trimgalore/0.6.4 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), FastQC/0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), multiqc/1.7 (https://multiqc.info/) (**Ewels et al., 2016**), STAR/2.7.3 (https://github.com/alexdobin/STAR) (**Dobin et al., 2013**), BWA/0.7.10 (http://bio-bwa.sourceforge.net/) (**Li and Durbin, 2009**), picard/2.20.2 (http://broadinstitute.github.io/picard/), samtools/1.9 (http://www.htslib.org/) (**Danecek et al., 2021**), bamtools/2.5.1 (https://github.com/pezmaster31/bamtools) (**Barnett et al., 2011**), featurecounts/2.0.0 (http://subread.sourceforge.net/) (**Liao et al., 2014**), and MACS2/2.1.2 (https://pypi.org/project/MACS2/) (**Zhang et al., 2008**). Venn diagrams were drawn using InteractiVenn (http://www.interactivenn.net) (**Heberle et al., 2015**). Heatmaps were drawn using Morpheus (https://software.broadinstitute.org/morpheus). ATACseq data were visualized using Integrative Genomics Viewer (**Robinson et al., 2011**). Gene ontology analysis was performed using GOTermFinder (https://go.princeton.edu/cgi-bin/GOTermFinder). ATAC-seq motif analysis was performed using The MEME Suite (https://meme-suite.org/meme/doc/cite.html?man_type=web) (**Bailey et al., 2015**). *Tbx21* binding motifs were obtained using JASPAR^{2020} (**Fornes et al., 2020**).

Quantification of MCMV

Infectious virus was quantified in organs via plaque assay as described previously (**Stack et al., 2015**). 3T3 cells were purchased directly from the American Type Culture Collection. *Mycoplasma* infection was excluded prior to assay. Viral loads in oral lavage were quantified via qPCR (**Clement et al., 2016**; **Kamimura and Lanier, 2014**).

Western blotting

Leukocytes were isolated from SGs and spleens as described previously (**Stacey et al., 2011**). CD4$^+$ T cells were purified via magnetic separation using a MagniSort Mouse CD4 Positive Selection Kit

(Thermo Fisher Scientific). Cell lysates were generated from equal numbers of cells using NuPAGE LDS Sample Buffer (Thermo Fisher Scientific) supplemented with 100 mM dithiothreitol (Sigma-Aldrich). Samples were loaded onto 4–12% NuPAGE Bis-Tris Gels (Thermo Fisher Scientific) after boiling and transferred to a PVDF membrane using an XCell II Blot Module (Thermo Fisher Scientific). Blots were probed with anti-arginase-1 (rabbit polyclonal, Thermo Fisher Scientific) and developed using anti-rabbit IgG–HRP (Bio-Rad). Band intensity was determined using a G:BOX Gel Imaging System (Syngene). Blots were then stripped using Restore PLUS Western Blot Stripping Buffer (Thermo Fisher Scientific) and probed again with anti-actin (rabbit polyclonal, Abcam).

Flow cytometry

Leukocytes were isolated from SGs and spleens as described previously (***Stacey et al., 2011***). Cells were labeled using a Zombie Aqua Fixable Viability Kit (BioLegend) and stained with anti-CD16/CD32 (Fc block, BioLegend) followed by combinations of anti-CD4–APC-Cy7, anti-CD4–BV605, or anti-CD4–Pacific Blue (clone RM4-5, BioLegend), anti-CD8–BV605 (clone 53.6.7, BioLegend), anti-CD39–Alexa Fluor 647 (clone Duha59, BioLegend), anti-CD44–FITC (clone IM7, BioLegend), anti-CD62L–BV711 or anti-CD62L–PE-Cy7 (clone MEL-14, BioLegend), anti-CD90/90.1–APC, anti-CD90/90.1–BV605, or anti-CD90/90.1–PE (clone OX-7, BioLegend), anti-CD127–BV711 (clone A7R34, BioLegend), anti-CD223–PE (LAG-3) (clone C9B7W, BioLegend), anti-NK1.1–PE (clone PK136, BioLegend), anti-PD-1–BV421 (clone 29F.1A12, BioLegend), anti-TIGIT–APC (clone 1G9, BioLegend), and anti-TIM-3–BV711 or anti-TIM-3–PerCP-Cy5.5 (clone RMT3-23, BioLegend). Tetramer staining was performed as described previously (***Clement et al., 2016***). The following specificities were used in this study: H-2D^b M45 residues 985–993 (HGIRNASFI), H-2K^b IE3 residues 416–423 (RALEYKNL), H-2K^b M38 residues 316–323 (SSPPMFRV), and H-2K^b m139 residues 419–426 (TVYGFCLL) (National Institutes of Health Tetramer Core Facility). Internal antigen expression was determined after fixation/permeabilization in BD Cytofix/Cytoperm Solution (BD Biosciences) or FoxP3/Transcription Factor Staining Buffer (eBioscience). Cells were stained with combinations of anti-arginase-1–APC (clone Met1-Lys322, Bio-Techne) or anti-arginase-1–PE-Cy7 (clone A1ex5, eBioscience), anti-Ki-67–Alexa Fluor 488 (clone 11F6, BioLegend), anti-T-bet–BV605 or anti-T-bet–Pacific Blue (clone 4B10, BioLegend), and anti-TCF1/7–Alexa Fluor 647 (clone C63D9, Cell Signaling Technology). Cells were acquired using an Attune NxT (Thermo Fisher Scientific) or an LSR Fortessa (BD Biosciences). All flow cytometry data were analyzed using FlowJo v10.5.3 or v10.8.1 (FlowJo LLC).

Quantification of anti-SSA

Plasma samples were obtained from cardiac punctures and assessed for anti-SSA/Ro 60 IgG levels via ELISA (Alpha Diagnostics).

In vitro induction of Arg1

Spleens were harvested from OT-II mice (n = 13) and digested for 30 min via the injection of collagenase-D (Thermo Fisher Scientific). Cells were isolated and purified via magnetic separation using a Pan Dendritic Cell Isolation Kit (Miltenyi Biotec) and a MagniSort Mouse CD4 Positive Selection Kit (Thermo Fisher Scientific). Cultures were established using 2 × 10^4 purified CD11c$^+$ dendritic cells/ml and 1 × 10^5 purified CD4$^+$ T cells/ml as described previously (***Saraiva et al., 2009***). Cells were cultured in the absence or presence of IL-12 (3 ng/ml, PeproTech) without or with the OVA peptide ISQAVHAAHAEINEAGR (residues 323–339, GenScript) at a concentration of 0.01 or 0.5 µM in RPMI medium supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin, 2 mM L-glutamine, and 10% heat-inactivated fetal calf serum (all from Thermo Fisher Scientific). On day 7, cells were unexposed (medium control) or exposed to plate-bound anti-CD3 (2 µg/ml, clone 145-2C11, BioLegend) and anti-CD28 (2 µg/ml, clone 37.51, BioLegend). After 2 hr, the cultures were supplemented without or with brefeldin A (2 µg/ml, Sigma-Aldrich), and after a further 2 hr, cells were labeled using a Zombie Aqua Fixable Viability Kit (BioLegend) and stained with anti-CD16/CD32 (Fc block, BioLegend) followed by anti-CD4–Pacific Blue (clone RM4-5, BioLegend) and anti-CD8–BV605 (clone 53.6.7, BioLegend). Internal antigen expression was determined after fixation/permeabilization in BD Cytofix/Cytoperm Solution (BD Biosciences). Cells were stained with anti-arginase-1–PE-Cy7 (clone A1ex5, eBioscience), anti-IFNγ–FITC (clone XMG1.2, BioLegend), and anti-IL-10–APC (clone JES5-16E3, eBioscience). Cells were acquired using an Attune NxT (Thermo Fisher Scientific), and data were analyzed using FlowJo

v10.5.3 (FlowJo LLC). Supernatants were harvested from cultures that lacked brefeldin A after a further 48 hr and analyzed for Arg1 protein expression via ELISA (Aviva Systems Biology).

Functional assays

Leukocytes from SGs and spleens were stimulated with peptides at a final concentration of 3 µg/ml for 2 hr at 37°C. The following peptides were used to stimulate $CD4^+$ T cells: m09 residues 133–147 (GYLYIYPSAGNSFDL), M25 residues 409–423 (NHLYETPISATAMVI), m139 residues 560–574 (TRPYRYPRVCDASLS), and m142 residues 24–38 (RSRYLTAAAVTAVLQ). The cultures were then supplemented with brefeldin A (2 µg/ml, Sigma-Aldrich) and incubated for a further 4 hr at 37°C. After stimulation, cells were labeled using a Zombie Aqua Fixable Viability Kit (BioLegend) and stained with anti-CD16/CD32 (Fc block, BioLegend) followed by anti-CD4–APC-Cy7 or anti-CD4–BV605 (clone RM4-5, BioLegend). Internal antigen expression was determined after fixation/permeabilization in BD Cytofix/Cytoperm Solution (BD Biosciences). Cells were stained with combinations of anti-IFNγ–APC-Cy7, anti-IFNγ–FITC, or anti-IFNγ–Pacific Blue (clone XMG1.2, BioLegend), anti-IL-10–APC or anti-IL-10–FITC (clone JES5-16E3, eBioscience), and anti-T-bet–BV605 (clone 4B10, BioLegend). Cells were acquired using an Attune NxT (Thermo Fisher Scientific), and data were analyzed using FlowJo v10.5.3 (FlowJo LLC).

Statistics

Sample sizes for next-generation sequencing experiments were calculated using G*Power (https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html), where a minimum of $n = 5$ pooled mice per group was used to detect a difference in means with 90% power and an α value set at 0.05 across a minimum of three replicates. All outliers were included in the final datasets. Comparisons between two groups were performed using the Mann–Whitney *U* test. Significance across all tests is reported as *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.

Acknowledgements

We thank Steven Reiner (Columbia University, New York, USA) for providing *Tbx21*$^{flox/flox}$ mice, Joerg Fehling (Ulm University, Germany) for providing *Rosa26*^{flSTOPtdRFP} mice, and Sarah Dimeloe (University of Birmingham, UK) for providing *Cd4*$^{Cre/+}$ mice. Biotinylated monomers were obtained from the NIH Tetramer Core Facility. MC was supported by a Systems Immunity Research Institute Fellowship (Cardiff University). VVK received infrastructure support from the Deutsche Forschungsgemeinschaft (DFG) Cluster of Excellence "Precision Medicine in Chronic Inflammation" (PMI). OVB, DMC, and DAP were supported by a Ministry of Science and Higher Education of the Russian Federation Subsidy Grant (075-15-2019-1789). DRW was supported by a Wellcome Trust Senior Research Fellowship (110199/Z/15/Z). DAP was supported by a Wellcome Trust Senior Investigator Award (100326/Z/12/Z). IRH was supported by a Medical Research Council Confidence in Concept Award, a Wellcome Trust Collaborator Award (209213/Z/17/Z), and a Wellcome Trust Senior Research Fellowship (207503/Z/17/Z).

Additional information

Funding

Funder	Grant reference number	Author
Wellcome Trust	207503/Z/17/Z	Ian R Humphreys
Wellcome Trust	100326/Z/12/Z	David A Price
Wellcome Trust	110199/Z/15/Z	David R Withers
Deutsche Forschungsgemeinschaft	Cluster of Excellence "Precision Medicine in Chronic Inflammation"	Valeriia V Kriukova
Ministry of Science and Higher Education of the Russian Federation	075-15-2019-1789	Dmitriy M Chudakov

Funder	Grant reference number	Author
Wellcome Trust	209213/Z/17/Z	Ian R Humphreys

The funders had no role in study design, data collection, and interpretation, or the decision to submit the work for publication. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Author contributions

Mathew Clement, Conceptualization, Data curation, Formal analysis, Validation, Investigation, Methodology, Writing – original draft, Project administration, Writing – review and editing; Kristin Ladell, Morgan Marsden, Anna Cardus Figueras, Valeriia V Kriukova, Ksenia R Lupyr, Data curation, Formal analysis; Kelly L Miners, Lucy Chapman, Jake Scott, Simon Clare, Data curation; Robert Andrews, Simon A Jones, Resources, Formal analysis; Olga V Britanova, Dmitriy M Chudakov, Data curation, Formal analysis, Funding acquisition, Writing – review and editing; David R Withers, Resources, Formal analysis, Funding acquisition; David A Price, Resources, Formal analysis, Supervision, Funding acquisition, Methodology, Writing – original draft, Writing – review and editing; Ian R Humphreys, Conceptualization, Resources, Formal analysis, Supervision, Funding acquisition, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Mathew Clement http://orcid.org/0000-0002-9280-5281
David A Price http://orcid.org/0000-0001-9416-2737
Ian R Humphreys http://orcid.org/0000-0002-9512-5337

Ethics

All mouse experiments were approved by the Biological Services Facility (Cardiff University) and performed under UK Home Office Project License P7867DADD.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.79165.sa1
Author response https://doi.org/10.7554/eLife.79165.sa2

Additional files

Supplementary files

- MDAR checklist

Data availability

All data needed to evaluate the conclusions of this study are presented in the paper and/or the Supplementary Materials. RNA-seq data generated as part of this study have been deposited in ArrayExpress (E-ERAD-445). A fully annotated version of the RNA-seq dataset is freely available via Zenodo (https://doi.org/10.5281/zenodo.7243956). An annotated list of genes shared between the dataset presented here (E-ERAD-445) and the in vitro T_H1 dataset (E-MTAB-2582) is also freely available via Zenodo (https://doi.org/10.5281/zenodo.7447477). TCR-seq data generated as part of this study have been deposited in the NCBI Sequence Read Archive (PRJNA860054). Processed repertoire datasets are freely available via Figshare (https://figshare.com/projects/Clement_InhibitoryCD4/143541).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Clement M, Humphreys IR	2021	Characterization of IL-10-producing $CD4^+$ T cells in a mucosal herpesvirus infection	https://www.ebi.ac.uk/arrayexpress/experiments/E-ERAD-445/	ArrayExpress, E-ERAD-445

Continued on next page

Continued

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Clement M	2023	TCR-seq data generated as part of this study	https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA860054	NCBI Sequence Read Archive, PRJNA860054
Clement M	2022	RNA-seq data comparing salivary gland IL-10-producing $CD4^+$ T cells versus IL-10-nonproducing $CD4^+$ T cells on day 14 after infection with MCMV	https://doi.org/10.5281/zenodo.7243956	Zenodo, 10.5281/zenodo.7243956
Clement M	2022	Shared gene use between salivary gland IL-10-producing $CD4^+$ T cells and T_H1-like $CD4^+$ T cells generated in vivo	https://doi.org/10.5281/zenodo.7447477	Zenodo, 10.5281/zenodo.7447477
Kriukova V	2022	metadata_Clement_TCR	https://doi.org/10.6084/m9.figshare.20310990.v1	figshare, 10.6084/m9.figshare.20310990.v1

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Stubbington M, Mahata B, Svensson V, Deonarine A, Nissen JK	2015	An mRNA-sequencing atlas of mouse $CD4^+$ T cell transcriptomes	https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2582/	ArrayExpress, E-MTAB-2582
Zhu J, Jankovic D, Oler AJ, Wei G, Sharma S, Hu G, Guo L, Yagi R, Yamane H, Punkosdy G, Feigenbaum L, Zhao K, Paul WE	2012	The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell program during Th1 cell responses	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38808	NCBI Gene Expression Omnibus, GSE38808

References

Anderson CF, Oukka M, Kuchroo VJ, Sacks D. 2007. $CD4^+CD25^-Foxp3^-$ Th1 cells are the source of IL-10-mediated immune suppression in chronic cutaneous leishmaniasis. *Journal of Experimental Medicine* **204**:285–297. DOI: https://doi.org/10.1084/jem.20061886, PMID: 17283207

Apetoh L, Quintana FJ, Pot C, Joller N, Xiao S, Kumar D, Burns EJ, Sherr DH, Weiner HL, Kuchroo VK. 2010. The aryl hydrocarbon receptor interacts with c-Maf to promote the differentiation of type 1 regulatory T cells induced by IL-27. *Nature Immunology* **11**:854–861. DOI: https://doi.org/10.1038/ni.1912, PMID: 20676095

Awasthi A, Carrier Y, Peron JPS, Bettelli E, Kamanaka M, Flavell RA, Kuchroo VK, Oukka M, Weiner HL. 2007. A dominant function for interleukin 27 in generating interleukin 10-producing anti-inflammatory T cells. *Nature Immunology* **8**:1380–1389. DOI: https://doi.org/10.1038/ni1541, PMID: 17994022

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Research* **43**:W39–W49. DOI: https://doi.org/10.1093/nar/gkv416, PMID: 25953851

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. Bamtools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**:1691–1692. DOI: https://doi.org/10.1093/bioinformatics/btr174, PMID: 21493652

Bauquet AT, Jin H, Paterson AM, Mitsdoerffer M, Ho IC, Sharpe AH, Kuchroo VK. 2009. The costimulatory molecule ICOS regulates the expression of c-Maf and IL-21 in the development of follicular T helper cells and T_H-17 cells. *Nature Immunology* **10**:167–175. DOI: https://doi.org/10.1038/ni.1690, PMID: 19098919

Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM. 2015. MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods* **12**:380–381. DOI: https://doi.org/10.1038/nmeth.3364, PMID: 25924071

Brockmann L, Soukou S, Steglich B, Czarnewski P, Zhao L, Wende S, Bedke T, Ergen C, Manthey C, Agalioti T, Geffken M, Seiz O, Parigi SM, Sorini C, Geginat J, Fujio K, Jacobs T, Roesch T, Izbicki JR, Lohse AW, et al. 2018. Molecular and functional heterogeneity of IL-10-producing $CD4^+$ T cells. *Nature Communications* **9**:5457. DOI: https://doi.org/10.1038/s41467-018-07581-4, PMID: 30575716

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**:1213–1218. DOI: https://doi.org/10.1038/nmeth.2688, PMID: 24097267

Chihara N, Madi A, Kondo T, Zhang H, Acharya N, Singer M, Nyman J, Marjanovic ND, Kowalczyk MS, Wang C, Kurtulus S, Law T, Etminan Y, Nevin J, Buckley CD, Burkett PR, Buenrostro JD, Rozenblatt-Rosen O, Anderson AC, Regev A, et al. 2018. Induction and transcriptional regulation of the co-inhibitory gene module in T cells. *Nature* **558**:454–459. DOI: https://doi.org/10.1038/s41586-018-0206-z, PMID: 29899446

Clement M, Marsden M, Stacey MA, Abdul-Karim J, Gimeno Brias S, Costa Bento D, Scurr MJ, Ghazal P, Weaver CT, Carlesso G, Clare S, Jones SA, Godkin A, Jones GW, Humphreys IR. 2016. Cytomegalovirus-specific IL-10-producing CD4$^+$ T cells are governed by type-I IFN-induced IL-27 and promote virus persistence. *PLOS Pathogens* **12**:e1006050. DOI: https://doi.org/10.1371/journal.ppat.1006050, PMID: 27926930

Czystowska-Kuzmicz M, Sosnowska A, Nowis D, Ramji K, Szajnik M, Chlebowska-Tuz J, Wolinska E, Gaj P, Grazul M, Pilch Z, Zerrouqi A, Graczyk-Jarzynka A, Soroczynska K, Cierniak S, Koktysz R, Elishaev E, Gruca S, Stefanowicz A, Blaszczyk R, Borek B, et al. 2019. Small extracellular vesicles containing arginase-1 suppress T-cell responses and promote tumor growth in ovarian carcinoma. *Nature Communications* **10**:3000. DOI: https://doi.org/10.1038/s41467-019-10979-3, PMID: 31278254

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**:giab008. DOI: https://doi.org/10.1093/gigascience/giab008, PMID: 33590861

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nature Biotechnology* **35**:316–319. DOI: https://doi.org/10.1038/nbt.3820, PMID: 28398311

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Doering TA, Crawford A, Angelosanto JM, Paley MA, Ziegler CG, Wherry EJ. 2012. Network analysis reveals centrally connected genes and pathways involved in CD8$^+$ T cell exhaustion versus memory. *Immunity* **37**:1130–1144. DOI: https://doi.org/10.1016/j.immuni.2012.08.021, PMID: 23159438

Dominguez CX, Amezquita RA, Guan T, Marshall HD, Joshi NS, Kleinstein SH, Kaech SM. 2015. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *Journal of Experimental Medicine* **212**:2041–2056. DOI: https://doi.org/10.1084/jem.20150186, PMID: 26503446

Elhanati Y, Sethna Z, Callan CG, Mora T, Walczak AM. 2018. Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. *Immunological Reviews* **284**:167–179. DOI: https://doi.org/10.1111/imr.12665, PMID: 29944757

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**:3047–3048. DOI: https://doi.org/10.1093/bioinformatics/btw354, PMID: 27312411

Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **48**:D87–D92. DOI: https://doi.org/10.1093/nar/gkz1001, PMID: 31701148

Gabryšová L, Alvarez-Martinez M, Luisier R, Cox LS, Sodenkamp J, Hosking C, Pérez-Mazliah D, Whicher C, Kannan Y, Potempa K, Wu X, Bhaw L, Wende H, Sieweke MH, Elgar G, Wilson M, Briscoe J, Metzis V, Langhorne J, Luscombe NM, et al. 2018. c-Maf controls immune responses by regulating disease-specific gene networks and repressing IL-2 in CD4$^+$ T cells. *Nature Immunology* **19**:497–507. DOI: https://doi.org/10.1038/s41590-018-0083-5, PMID: 29662170

Gagliani N, Magnani CF, Huber S, Gianolini ME, Pala M, Licona-Limon P, Guo B, Herbert DR, Bulfone A, Trentini F, Di Serio C, Bacchetta R, Andreani M, Brockmann L, Gregori S, Flavell RA, Roncarolo M-G. 2013. Coexpression of CD49b and LAG-3 identifies human and mouse T regulatory type 1 cells. *Nature Medicine* **19**:739–746. DOI: https://doi.org/10.1038/nm.3179, PMID: 23624599

Geiger R, Rieckmann JC, Wolf T, Basso C, Feng Y, Fuhrer T, Kogadeeva M, Picotti P, Meissner F, Mann M, Zamboni N, Sallusto F, Lanzavecchia A. 2016. L-arginine modulates T cell metabolism and enhances survival and anti-tumor activity. *Cell* **167**:829–842. DOI: https://doi.org/10.1016/j.cell.2016.09.031, PMID: 27745970

Graziosi C, Pantaleo G, Gantt KR, Fortin JP, Demarest JF, Cohen OJ, Sékaly RP, Fauci AS. 1994. Lack of evidence for the dichotomy of T_H1 and T_H2 predominance in HIV-infected individuals. *Science* **265**:248–252. DOI: https://doi.org/10.1126/science.8023143, PMID: 8023143

Heberle H, Meirelles GV, da Silva FR, Telles GP, Minghim R. 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**:169. DOI: https://doi.org/10.1186/s12859-015-0611-3, PMID: 25994840

Humphreys IR, de Trez C, Kinkade A, Benedict CA, Croft M, Ware CF. 2007. Cytomegalovirus exploits IL-10-mediated immune regulation in the salivary glands. *Journal of Experimental Medicine* **204**:1217–1225. DOI: https://doi.org/10.1084/jem.20062424, PMID: 17485516

Intlekofer AM, Takemoto N, Kao C, Banerjee A, Schambach F, Northrop JK, Shen H, Wherry EJ, Reiner SL. 2007. Requirement for T-bet in the aberrant differentiation of unhelped memory CD8$^+$ T cells. *Journal of Experimental Medicine* **204**:2015–2021. DOI: https://doi.org/10.1084/jem.20070841, PMID: 17698591

Intlekofer AM, Banerjee A, Takemoto N, Gordon SM, Dejong CS, Shin H, Hunter CA, Wherry EJ, Lindsten T, Reiner SL. 2008. Anomalous type 17 response to viral infection by CD8+ T cells lacking T-bet and eomesodermin. *Science* **321**:408–411. DOI: https://doi.org/10.1126/science.1159806, PMID: 18635804

Jankovic D, Kullberg MC, Feng CG, Goldszmid RS, Collazo CM, Wilson M, Wynn TA, Kamanaka M, Flavell RA, Sher A. 2007. Conventional T-bet+Foxp3− Th1 cells are the major source of host-protective regulatory IL-10 during intracellular protozoan infection. *Journal of Experimental Medicine* **204**:273–283. DOI: https://doi.org/10.1084/jem.20062175, PMID: 17283209

Jeitziner SM, Walton SM, Torti N, Oxenius A. 2013. Adoptive transfer of cytomegalovirus-specific effector CD4+ T cells provides antiviral protection from murine CMV infection. *European Journal of Immunology* **43**:2886–2895. DOI: https://doi.org/10.1002/eji.201343690, PMID: 23921569

Jones M, Ladell K, Wynn KK, Stacey MA, Quigley MF, Gostick E, Price DA, Humphreys IR. 2010. IL-10 restricts memory T cell inflation during cytomegalovirus infection. *Journal of Immunology* **185**:3583–3592. DOI: https://doi.org/10.4049/jimmunol.1001535, PMID: 20713884

Jonjić S, Mutter W, Weiland F, Reddehase MJ, Koszinowski UH. 1989. Site-restricted persistent cytomegalovirus infection after selective long-term depletion of CD4+ T lymphocytes. *Journal of Experimental Medicine* **169**:1199–1212. DOI: https://doi.org/10.1084/jem.169.4.1199, PMID: 2564415

Kamimura Y, Lanier LL. 2014. Rapid and sequential quantitation of salivary gland-associated mouse cytomegalovirus in oral lavage. *Journal of Virological Methods* **205**:53–56. DOI: https://doi.org/10.1016/j.jviromet.2014.03.029, PMID: 24747009

Kasatskaya SA, Ladell K, Egorov ES, Miners KL, Davydov AN, Metsger M, Staroverov DB, Matveyshina EK, Shagina IA, Mamedov IZ, Izraelson M, Shelyakin PV, Britanova OV, Price DA, Chudakov DM. 2020. Functionally specialized human CD4+ T-cell subsets express physicochemically distinct TCRs. *eLife* **9**:e57063. DOI: https://doi.org/10.7554/eLife.57063, PMID: 33289628

Laidlaw BJ, Cui W, Amezquita RA, Gray SM, Guan T, Lu Y, Kobayashi Y, Flavell RA, Kleinstein SH, Craft J, Kaech SM. 2015. Production of IL-10 by CD4+ regulatory T cells during the resolution of infection promotes the maturation of memory CD8+ T cells. *Nature Immunology* **16**:871–879. DOI: https://doi.org/10.1038/ni.3224, PMID: 26147684

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: https://doi.org/10.1093/bioinformatics/btp324, PMID: 19451168

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**:923–930. DOI: https://doi.org/10.1093/bioinformatics/btt656, PMID: 24227677

Logunova NN, Kriukova VV, Shelyakin PV, Egorov ES, Pereverzeva A, Bozhanova NG, Shugay M, Shcherbinin DS, Pogorelyy MV, Merzlyak EM, Zubov VN, Meiler J, Chudakov DM, Apt AS, Britanova OV. 2020. MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naive CD4+ T cells. *PNAS* **117**:13659–13669. DOI: https://doi.org/10.1073/pnas.2003170117, PMID: 32482872

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. DOI: https://doi.org/10.1186/s13059-014-0550-8, PMID: 25516281

Lowe MM, Boothby I, Clancy S, Ahn RS, Liao W, Nguyen DN, Schumann K, Marson A, Mahuron KM, Kingsbury GA, Liu Z, Munoz Sandoval P, Rodriguez RS, Pauli ML, Taravati K, Arron ST, Neuhaus IM, Harris HW, Kim EA, Shin US, et al. 2019. Regulatory T cells use arginase 2 to enhance their metabolic fitness in tissues. *JCI Insight* **4**:e129756. DOI: https://doi.org/10.1172/jci.insight.129756, PMID: 31852848

Luche H, Weber O, Nageswara Rao T, Blum C, Fehling HJ. 2007. Faithful activation of an extra-bright red fluorescent protein in "knock-in" Cre-reporter mice ideally suited for lineage tracing studies. *European Journal of Immunology* **37**:43–53. DOI: https://doi.org/10.1002/eji.200636745, PMID: 17171761

Lucin P, Pavić I, Polić B, Jonjić S, Koszinowski UH. 1992. Gamma interferon-dependent clearance of cytomegalovirus infection in salivary glands. *Journal of Virology* **66**:1977–1984. DOI: https://doi.org/10.1128/JVI.66.4.1977-1984.1992, PMID: 1312614

Mandaric S, Walton SM, Rülicke T, Richter K, Girard-Madoux MJH, Clausen BE, Zurunic A, Kamanaka M, Flavell RA, Jonjic S, Oxenius A. 2012. IL-10 suppression of NK/DC crosstalk leads to poor priming of MCMV-specific CD4 T cells and prolonged MCMV persistence. *PLOS Pathogens* **8**:e1002846. DOI: https://doi.org/10.1371/journal.ppat.1002846, PMID: 22876184

Mason GM, Jackson S, Okecha G, Poole E, Sissons JGP, Sinclair J, Wills MR. 2013. Human cytomegalovirus latency-associated proteins elicit immune-suppressive IL-10 producing CD4+ T cells. *PLOS Pathogens* **9**:e1003635. DOI: https://doi.org/10.1371/journal.ppat.1003635, PMID: 24130479

Maynard CL, Harrington LE, Janowski KM, Oliver JR, Zindl CL, Rudensky AY, Weaver CT. 2007. Regulatory T cells expressing interleukin 10 develop from Foxp3+ and Foxp3− precursor cells in the absence of interleukin 10. *Nature Immunology* **8**:931–941. DOI: https://doi.org/10.1038/ni1504, PMID: 17694059

Munder M, Schneider H, Luckner C, Giese T, Langhans CD, Fuentes JM, Kropf P, Mueller I, Kolb A, Modolell M, Ho AD. 2006. Suppression of T-cell functions by human granulocyte arginase. *Blood* **108**:1627–1634. DOI: https://doi.org/10.1182/blood-2006-11-010389, PMID: 16709924

Munder M. 2009. Arginase: an emerging key player in the mammalian immune system. *British Journal of Pharmacology* **158**:638–651. DOI: https://doi.org/10.1111/j.1476-5381.2009.00291.x, PMID: 19764983

Nurieva RI, Duong J, Kishikawa H, Dianzani U, Rojo JM, Ho I, Flavell RA, Dong C. 2003. Transcriptional regulation of Th2 differentiation by inducible costimulator. *Immunity* **18**:801–811. DOI: https://doi.org/10.1016/s1074-7613(03)00144-4, PMID: 12818161

Oestreich KJ, Huang AC, Weinmann AS. 2011. The lineage-defining factors T-bet and Bcl-6 collaborate to regulate Th1 gene expression patterns. *Journal of Experimental Medicine* **208**:1001–1013. DOI: https://doi.org/10.1084/jem.20102144, PMID: 21518797

Omilusik KD, Best JA, Yu B, Goossens S, Weidemann A, Nguyen JV, Seuntjens E, Stryjewska A, Zweier C, Roychoudhuri R, Gattinoni L, Bird LM, Higashi Y, Kondoh H, Huylebroeck D, Haigh J, Goldrath AW. 2015. Transcriptional repressor ZEB2 promotes terminal differentiation of CD8$^+$ effector and memory T cell populations during infection. *Journal of Experimental Medicine* **212**:2027–2039. DOI: https://doi.org/10.1084/jem.20150194, PMID: 26503445

Pallett LJ, Gill US, Quaglia A, Sinclair LV, Jover-Cobos M, Schurich A, Singh KP, Thomas N, Das A, Chen A, Fusai G, Bertoletti A, Cantrell DA, Kennedy PT, Davies NA, Haniffa M, Maini MK. 2015. Metabolic regulation of hepatitis B immunopathology by myeloid-derived suppressor cells. *Nature Medicine* **21**:591–600. DOI: https://doi.org/10.1038/nm.3856, PMID: 25962123

Parish IA, Marshall HD, Staron MM, Lang PA, Brüstle A, Chen JH, Cui W, Tsui Y-C, Perry C, Laidlaw BJ, Ohashi PS, Weaver CT, Kaech SM. 2014. Chronic viral infection promotes sustained Th1-derived immunoregulatory IL-10 via BLIMP-1. *Journal of Clinical Investigation* **124**:3455–3468. DOI: https://doi.org/10.1172/JCI66108, PMID: 25003188

Pogorelyy MV, Minervina AA, Shugay M, Chudakov DM, Lebedev YB, Mora T, Walczak AM. 2019. Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLOS Biology* **17**:e3000314. DOI: https://doi.org/10.1371/journal.pbio.3000314, PMID: 31194732

Pot C, Jin H, Awasthi A, Liu SM, Lai CY, Madan R, Sharpe AH, Karp CL, Miaw SC, Ho IC, Kuchroo VK. 2009. Cutting edge: IL-27 induces the transcription factor c-Maf, cytokine IL-21, and the costimulatory receptor ICOS that coordinately act together to promote differentiation of IL-10-producing Tr1 cells. *Journal of Immunology* **183**:797–801. DOI: https://doi.org/10.4049/jimmunol.0901233, PMID: 19570826

Richter K, Perriard G, Behrendt R, Schwendener RA, Sexl V, Dunn R, Kamanaka M, Flavell RA, Roers A, Oxenius A. 2013. Macrophage and T cell produced IL-10 promotes viral chronicity. *PLOS Pathogens* **9**:e1003735. DOI: https://doi.org/10.1371/journal.ppat.1003735, PMID: 24244162

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**:24–26. DOI: https://doi.org/10.1038/nbt.1754, PMID: 21221095

Rodriguez PC, Zea AH, Culotta KS, Zabaleta J, Ochoa JB, Ochoa AC. 2002. Regulation of T cell receptor CD3 ζ chain expression by L-arginine. *Journal of Biological Chemistry* **277**:21123–21129. DOI: https://doi.org/10.1074/jbc.M110675200, PMID: 11950832

Rodriguez PC, Quiceno DG, Zabaleta J, Ortiz B, Zea AH, Piazuelo MB, Delgado A, Correa P, Brayer J, Sotomayor EM, Antonia S, Ochoa JB, Ochoa AC. 2004. Arginase I production in the tumor microenvironment by mature myeloid cells inhibits T-cell receptor expression and antigen-specific T-cell responses. *Cancer Research* **64**:5839–5849. DOI: https://doi.org/10.1158/0008-5472.CAN-04-0465, PMID: 15313928

Rodriguez PC, Quiceno DG, Ochoa AC. 2007. L-arginine availability regulates T-lymphocyte cell-cycle progression. *Blood* **109**:1568–1573. DOI: https://doi.org/10.1182/blood-2006-06-031856, PMID: 17023580

Roessner PM, Llaó Cid L, Lupar E, Roider T, Bordas M, Schifflers C, Arseni L, Gaupel A-C, Kilpert F, Krötschel M, Arnold SJ, Sellner L, Colomer D, Stilgenbauer S, Dietrich S, Lichter P, Izcue A, Seiffert M. 2021. EOMES and IL-10 regulate antitumor activity of T regulatory type 1 CD4$^+$ T cells in chronic lymphocytic leukemia. *Leukemia* **35**:2311–2324. DOI: https://doi.org/10.1038/s41375-021-01136-1, PMID: 33526861

Saraiva M, Christensen JR, Veldhoen M, Murphy TL, Murphy KM, O'Garra A. 2009. Interleukin-10 production by Th1 cells requires interleukin-12-induced STAT4 transcription factor and ERK MAP kinase activation by high antigen dose. *Immunity* **31**:209–219. DOI: https://doi.org/10.1016/j.immuni.2009.05.012, PMID: 19646904

Saraiva M, O'Garra A. 2010. The regulation of IL-10 production by immune cells. *Nature Reviews Immunology* **10**:170–181. DOI: https://doi.org/10.1038/nri2711, PMID: 20154735

Saraiva M, Vieira P, O'Garra A. 2020. Biology and therapeutic potential of interleukin-10. *Journal of Experimental Medicine* **217**:e20190418. DOI: https://doi.org/10.1084/jem.20190418, PMID: 31611251

Schorer M, Rakebrandt N, Lambert K, Hunziker A, Pallmer K, Oxenius A, Kipar A, Stertz S, Joller N. 2020. TIGIT limits immune pathology during viral infections. *Nature Communications* **11**:1288. DOI: https://doi.org/10.1038/s41467-020-15025-1, PMID: 32152316

Schuster IS, Wikstrom ME, Brizard G, Coudert JD, Estcourt MJ, Manzur M, O'Reilly LA, Smyth MJ, Trapani JA, Hill GR, Andoniou CE, Degli-Esposti MA. 2014. TRAIL$^+$ NK cells control CD4$^+$ T cell responses during chronic viral infection to limit autoimmunity. *Immunity* **41**:646–656. DOI: https://doi.org/10.1016/j.immuni.2014.09.013, PMID: 25367576

Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T. 2019. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* **35**:2974–2981. DOI: https://doi.org/10.1093/bioinformatics/btz035, PMID: 30657870

Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, Bolotin DA, Staroverov DB, Putintseva EV, Plevova K, Linnemann C, Shagin D, Pospisilova S, Lukyanov S, Schumacher TN, Chudakov DM. 2014. Towards error-free profiling of immune repertoires. *Nature Methods* **11**:653–655. DOI: https://doi.org/10.1038/nmeth.2960, PMID: 24793455

Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI, Kirgizov KI, Skorobogatova EV, Chudakov DM. 2015. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLOS Computational Biology* **11**:e1004503. DOI: https://doi.org/10.1371/journal.pcbi.1004503, PMID: 26606115

Stacey MA, Marsden M, Wang ECY, Wilkinson GWG, Humphreys IR. 2011. IL-10 restricts activation-induced death of NK cells during acute murine cytomegalovirus infection. *Journal of Immunology* **187**:2944–2952. DOI: https://doi.org/10.4049/jimmunol.1101021, PMID: 21849677

Stacey MA, Marsden M, Pham N TA, Clare S, Dolton G, Stack G, Jones E, Klenerman P, Gallimore AM, Taylor PR, Snelgrove RJ, Lawley TD, Dougan G, Benedict CA, Jones SA, Wilkinson GWG, Humphreys IR. 2014. Neutrophils recruited by IL-22 in peripheral tissues function as TRAIL-dependent antiviral effectors against MCMV. *Cell Host & Microbe* **15**:471–483. DOI: https://doi.org/10.1016/j.chom.2014.03.003, PMID: 24721575

Stack G, Jones E, Marsden M, Stacey MA, Snelgrove RJ, Lacaze P, Jacques LC, Cuff SM, Stanton RJ, Gallimore AM, Hussell T, Wilkinson GWG, Ghazal P, Taylor PR, Humphreys IR. 2015. CD200 receptor restriction of myeloid cell responses antagonizes antiviral immunity and facilitates cytomegalovirus persistence within mucosal tissue. *PLOS Pathogens* **11**:e1004641. DOI: https://doi.org/10.1371/journal.ppat.1004641, PMID: 25654642

Stubbington MJ, Mahata B, Svensson V, Deonarine A, Nissen JK, Betz AG, Teichmann SA. 2015. An atlas of mouse $CD4^+$ T cell transcriptomes. *Biology Direct* **10**:14. DOI: https://doi.org/10.1186/s13062-015-0045-x, PMID: 25886751

Thelen B, Schipperges V, Knörlein P, Hummel JF, Arnold F, Kupferschmid L, Klose CSN, Arnold SJ, Boerries M, Tanriver Y. 2023. Eomes is sufficient to regulate IL-10 expression and cytotoxic effector molecules in murine $CD4^+$ T cells. *Frontiers in Immunology* **14**:1058267. DOI: https://doi.org/10.3389/fimmu.2023.1058267, PMID: 36756120

Veiga-Parga T, Sehrawat S, Rouse BT. 2013. Role of regulatory T cells during virus infection. *Immunological Reviews* **255**:182–196. DOI: https://doi.org/10.1111/imr.12085, PMID: 23947355

Walton SM, Mandaric S, Torti N, Zimmermann A, Hengel H, Oxenius A. 2011. Absence of cross-presenting cells in the salivary gland and viral immune evasion confine cytomegalovirus immune control to effector CD4 T cells. *PLOS Pathogens* **7**:e1002214. DOI: https://doi.org/10.1371/journal.ppat.1002214, PMID: 21901102

Washburn ML, Wang Z, Walton AH, Goedegebuure SP, Figueroa DJ, Van Horn S, Grossman J, Remlinger K, Madsen H, Brown J, Srinivasan R, Wolf AI, Berger SB, Yi VN, Hawkins WG, Fields RC, Hotchkiss RS. 2019. T cell- and monocyte-specific RNA-sequencing analysis in septic and nonseptic critically ill patients and in patients with cancer. *Journal of Immunology* **203**:1897–1908. DOI: https://doi.org/10.4049/jimmunol.1900560, PMID: 31484735

Wherry EJ. 2011. T cell exhaustion. *Nature Immunology* **12**:492–499. DOI: https://doi.org/10.1038/ni.2035, PMID: 21739672

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**:R137. DOI: https://doi.org/10.1186/gb-2008-9-9-r137, PMID: 18798982

Zhang P, Lee JS, Gartlan KH, Schuster IS, Comerford I, Varelias A, Ullah MA, Vuckovic S, Koyama M, Kuns RD, Locke KR, Beckett KJ, Olver SD, Samson LD, Montes de Oca M, de Labastida Rivera F, Clouston AD, Belz GT, Blazar BR, MacDonald KP, et al. 2017. Eomesodermin promotes the development of type 1 regulatory T (T_R1) cells. *Science Immunology* **2**:eaah7152. DOI: https://doi.org/10.1126/sciimmunol.aah7152, PMID: 28738016

Zhang S, Caldeira-Dantas S, Smith CJ, Snyder CM. 2019. Persistent viral replication and the development of T-cell responses after intranasal infection by MCMV. *Medical Microbiology and Immunology* **208**:457–468. DOI: https://doi.org/10.1007/s00430-019-00589-7, PMID: 30848361

Zhu J, Jankovic D, Oler AJ, Wei G, Sharma S, Hu G, Guo L, Yagi R, Yamane H, Punkosdy G, Feigenbaum L, Zhao K, Paul WE. 2012. The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell program during Th1 cell responses. *Immunity* **37**:660–673. DOI: https://doi.org/10.1016/j.immuni.2012.09.007, PMID: 23041064

Zhu C, Sakuishi K, Xiao S, Sun Z, Zaghouani S, Gu G, Wang C, Tan DJ, Wu C, Rangachari M, Pertel T, Jin H-T, Ahmed R, Anderson AC, Kuchroo VK. 2015. An IL-27/NFIL3 signalling axis drives Tim-3 and IL-10 expression and T-cell dysfunction. *Nature Communications* **6**:6072. DOI: https://doi.org/10.1038/ncomms7072, PMID: 25614966

3. Discussion

3.1. Single-cell Research in Immunology

3.1.1. Achievements

The development and spreading of single-cell technologies have revolutionized modern experimental biology. However, most probably none of the fields has profited as much as immunology. Indeed, single-cell technologies are especially applicable to immunological studies. Immune cells show enormous heterogeneity in underlying programs, while the small size and relatively simple morphology of immune cells make them one of the most easy-to-handle cell types for droplet-based scRNA-Seq protocols, e.g., 10x genomics. Moreover, when coupled with other techniques (scTCR-Seq, scBCR-Seq, and CITE-Seq), single-cell transcriptomics provides an invaluable advantage to researchers, streamlining TCRα and TCRβ (heavy and light chain) pairing, cell phenotyping, and even profiling immune repertoire specificities[80]. Another significant advantage, quite commonly overlooked, is that single-cell transcriptomics has opened the door to advanced immunological studies in animals other than mice or humans. Before, studies on non-model animals were extremely challenging due to the lack of commercially available recombinant proteins and fluorescently labeled antibodies to various cell markers. With the rise of scRNA-Seq, we are no longer limited by the availability of species-specific reagents.

3.1.2. Challenges

Even though we produce more and more scRNA-Seq data daily, the problem of unified data storage still needs to be solved. Leaders in the field still use classic transcriptomics databases like NCBI GEO or ArrayExpress. During submission, the user is required to complete only sample-related metadata fields, which makes perfect sense for bulk RNA-Seq data. However, the user is not obliged to fill in the cell-related metadata fields, making the whole entry to the database unstandardized. There is a need to switch to a separate public database that would allow the deposition of matched raw fastq data, unfiltered expression matrices, filtered (high-quality cells only) and annotated expression matrices, and the processed Seurat, scanpy, or scExperiment objects. One of the promising examples is the Single Cell Portal (https://singlecell.zendesk.com/hc/en-us), and I hope such transparent data-sharing initiatives will be more widely used by our research community soon.

Standardized data deposition is also crucial for easier data reuse. The input of dataset-producing centers is extremely valuable; however, nowadays, one may do research and test their hypotheses without generating any new data and using publicly available data only. I genuinely believe that our

research benefited greatly from the direct use of data produced by Stephenson et al. [81]. The authors of the original research performed scRNA-Seq, scTCR-Seq, and scBCR-Seq in PBMC from 130 donors with well-documented metadata of COVID-19 disease history. They reported many significant findings related to the innate and adaptive immune response to SARS-CoV-2, including the increased proliferative response in unconventional monocytes, tuning of hematopoiesis towards platelet production, and the kinetics of T and B cell subsets' expansions across the disease severity range. Our analysis only focused on CD4+ T cells out of this extensive PBMC collection, allowing us to dissect them in greater detail. With the aid of Sort-Seq-powered helper T-cell annotation, we revised the initial findings made by the authors. According to our annotation, the most prominent changes arose in naive IFN-response and effector-memory IFN-response cell clusters. Such cells are almost absent in healthy individuals and accumulate in patients with moderate COVID-19 severity, probably mirroring the typical IFN Type I response to the virus in immunocompetent patients. Our results support findings reported by another research group based on an independent scRNA-Seq dataset[82]. Alternatively, severe and critical patients were reported to have impaired IFN Type I response[83], which also aligns well with our findings. To conclude, even this tiny example shows the importance of scRNA-Seq data reuse and reanalysis.

3.2. Heterogeneity of Th cells

Our immune system keeps the information about initial decisions made in response to various antigens in the form of memory Th clones. We traditionally classify helpers with different cytokine and surface marker expression profiles into different Th subsets. However, Th cells may transiently change the expression of their marker genes *in vivo* or *in vitro* under specific conditions[84,85]. In this light, surface marker expression may not always serve as a tool for ultimate Th annotation. Some attempts to classify naive and effector-memory Th subsets based on their TCR repertoire features were also made by other researchers[7,33,37] and partially by us in this study. However, as we may see now, none of the TCR features may definitely predict the functionality and subset of a given Th cell, and these assumptions are only applicable in a statistical manner.

We consider that truly existing core self-standing Th effector programs should represent distinct and conserved global transcriptional states. Within these programs, two types of Th "plasticity" might occur. There might be some phenotypic and often transient changes in gene expression under specific conditions or real program plasticity when some part of the effector clone redirects the global transcriptional program entirely. Using standard flow cytometry and a small set of markers, one may not easily distinguish small transient versus global changes. This may result in gating cells having

various real Th programs into the same flow cytometry gate. Single-cell technologies allow us to look simultaneously at marker gene expression and all other transcripts. We strongly believe that such global gene expression essentially defines various Th programs. We introduced a new approach, Sort-Seq, to draw direct connections between classically defined Th subsets and the corresponding scRNA-Seq clusters, which are defined based on global transcriptional diversity. We compared Sort-Seq to the routinely used surface marker profiling method in single cells, known as CITE-Seq, and achieved greater precision in annotation of Th2, Th2a, and Th1 subsets. However, I believe these technologies should not compete but rather mutually complement each other, aiming to achieve the most advanced Th annotation in scRNA-Seq experiments. Thus, we combined our findings based on Sort-Seq, CITE-Seq, and scRNA-Seq and published the CD4map - a reference dataset of peripheral blood CD4+ T cells.

One of the possible next steps towards the definitive systematization of Th programs might be a parallel scRNA-Seq + scATAC-Seq + scTCR-Seq profiling of effector and memory CD4+ T cells, to ultimately link gene expression with chromatin accessibility in single cells within T cell clones. This would give us a snapshot of epigenetic potential for Th plasticity[86]. Ideally, the number of cells profiled from one individual for such a study should be increased far beyond 10,000. We aim to profile many cells belonging to the same clones to study the clonal plasticity. Unfortunately, with the current scRNA-Seq standard sample size limit, we capture clonal expansions almost exclusively in terminally differentiated cells.

The CD4map reference dataset includes original samples from our study and the data published by Stephenson et al. [81]. The inclusion of samples from COVID-19 patients with moderate disease severity yielded the presence of two disease-associated IFN-response clusters in the dataset. This observation highlights the importance of profiling CD4+ T cells in various disease conditions to catch transcriptional programs inactive in health. Yasumizu et al. performed a similar integrative analysis of peripheral blood CD4+ lymphocytes involving patients with autoimmune diseases (myasthenia gravis, multiple sclerosis, systemic lupus erythematosus) to complement the picture observed in healthy donors[87]. Andreatta et al., in contrast, built their reference atlas of antigen-specific T cells using mouse models of chronic and acute viral infections to cover as many T cell transcriptional programs as possible[61]. Extending research to other conditions like atopy could enrich such reference datasets and make them even more widely applicable.

3.3. Studying antiviral immunity

3.3.1. Studying antiviral immunity in humans

Usually, studies on human T-cell responses to viruses are pretty challenging. Researchers mainly used the vaccination against yellow fever model to track the kinetics of antiviral response. Yellow fever vaccine is a safe live-attenuated virus provoking a solid antibody and T cell response[88]. The game changed dramatically during the COVID-19 pandemic. Sadly, the pandemic resulted in a global emergency, many human deaths, and extreme challenges to healthcare systems, while the "Long COVID" is still affecting the quality of life of many patients[89]. Nevertheless, the pandemics allowed immunologists to collect and access enormous data documenting all aspects of the human immune response to viral infection. For example, we and other researchers[90] tracked the involvement of T cell clones in primary and secondary antiviral immune responses, which is usually challenging with other human viruses due to an often unexpected onset of infection.

3.3.2. Studying antiviral immunity in mouse models

Mouse models give immunologists much more freedom in the experiment design, and the availability of transgenic animals allows to disentangle the role of different immune cells and molecules in a highly controlled setting. In this study, we made use of such IL-10 reporter mice to study the immune response to MCMV. IL-10 producing Th cell in salivary glands of mice infected with MCMV attenuate viral clearance and promote MCMV chronicity[65,91]. We particularly questioned the clonal relation of IL-10-producing cells to their IL-10-negative counterparts in salivary glands. We observed evidence of an antigen-driven ongoing immune response based on the overrepresented clusters of similar TCRβ in the TCR-Seq data. Remarkably, IL-10-positive and IL-10-negative cells showed clonal overlaps, and we assumed that the generation of IL-10-producing cells results from Th clonal plasticity in the context of viral infection.

In this study, our collaborators identified the arginase-1 expression as a feature of pathological IL-10-positive phenotype in MCMV infection. Strikingly, another study[92] led by other researchers and published almost immediately after **Manuscript 3** also reported the role of T cell-intrinsic arginase-1 in CD4+ T cell-mediated antiviral immunity. They observed the effects arginase-1 in a mouse model of acute influenza infection. *Arg1* deletion in CD4+ T cells resulted in an altered lung Th antiviral response kinetics with an accelerated IL-10-mediated contraction stage. In such conditions, viral clearance was achieved with diminished collateral lung tissue injury. Overall, the revelation of the arginase-1 role in Th biology by two independent groups provides an additional valuable cross-validation to these studies. However, the discoveries were made in different mouse models, namely

in the model of chronic MCMV persistence in salivary glands versus acute influenza infection in the lungs. These critical differences in the models probably resulted in observing the partially discordant effects of arginase-1 on Th phenotypes.

3.4. Conclusions

In summary, within the scope of this thesis, we

a) described the unique TCR repertoire features of nTreg and nTconv lymphocytes selected on different MHC-II alleles;

b) profiled the scRNA-Seq phenotypes of peripheral blood Th lymphocytes and their involvement in an immune response to acute SARS-CoV-2 infection;

c) characterized the TCR repertoire features of a phenotypically distinct IL-10-producing Th subset, promoting chronic MCMV infection.

The three manuscripts presenting the result of this thesis collectively give us valuable insights into the functional composition of the naive and effector-memory Th pools, the structure of their TCR repertoires, and their involvement in antiviral immune responses.

Bibliography

1. Logunova, N. N. *et al.* MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naïve CD4 $^{+}$ T cells. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13659–13669 (2020).

2. Clement, M. *et al.* Inhibitory IL-10-producing CD4+ T cells are T-bet-dependent and facilitate cytomegalovirus persistence via coexpression of arginase-1. *eLife* **12**, e79165 (2023).

3. Lefranc, M.-P. IMGT, the International ImMunoGeneTics Information System. *Cold Spring Harb Protoc* **2011**, pdb.top115 (2011).

4. 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).

5. Silberman, D. *et al.* Class II major histocompatibility complex mutant mice to study the germ-line bias of T-cell antigen receptors. *Proc. Natl. Acad. Sci. U.S.A.* **113**, (2016).

6. Deng, L., Langley, R. J., Wang, Q., Topalian, S. L. & Mariuzza, R. A. Structural insights into the editing of germ-line–encoded interactions between T-cell receptor and MHC class II by Vα CDR3. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14960–14965 (2012).

7. Camaglia, F. *et al.* Quantifying changes in the T cell receptor repertoire during thymic development. *eLife* **12**, e81622 (2023).

8. Yamano, T. *et al.* Thymic B Cells Are Licensed to Present Self Antigens for Central T Cell Tolerance Induction. *Immunity* **42**, 1048–1061 (2015).

9. Stritesky, G. L. *et al.* Murine thymic selection quantified using a unique method to capture deleted T cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4679–4684 (2013).

10. Friesen, T. J., Ji, Q. & Fink, P. J. Recent thymic emigrants are tolerized in the absence of inflammation. *Journal of Experimental Medicine* **213**, 913–920 (2016).

11. Sprent, J. & Surh, C. D. Normal T cell homeostasis: the conversion of naive cells into memory-phenotype cells. *Nat Immunol* **12**, 478–484 (2011).

12. Guichard, V. *et al.* Calcium-mediated shaping of naive CD4 T-cell phenotype and function. *eLife* **6**, e27215 (2017).

13. Ciucci, T. *et al.* The Emergence and Functional Fitness of Memory CD4+ T Cells Require the Transcription Factor Thpok. *Immunity* **50**, 91-105.e4 (2019).

14. Bevington, S. L. *et al.* Stable Epigenetic Programming of Effector and Central Memory CD4 T Cells Occurs Within 7 Days of Antigen Exposure In Vivo. *Front. Immunol.* **12**, 642807 (2021).

15. Akondy, R. S. *et al.* Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).

16. Miron, M. *et al.* Maintenance of the human memory T cell repertoire by subset and tissue site.

Genome Med **13**, 100 (2021).

17. Cano-Gamez, E. *et al.* Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nat Commun* **11**, 1801 (2020).

18. Bacher, P. *et al.* Human Anti-fungal Th17 Immunity and Pathology Rely on Cross-Reactivity against Candida albicans. *Cell* **176**, 1340-1355.e15 (2019).

19. Schiffer, R., Baron, J., Dagtekin, G., Jahnen-Dechent, W. & Zwadlo-Klarwasser, G. Differential regulation of the expression of transporters associated with antigen processing, TAP1 and TAP2, by cytokines and lipopolysaccharide in primary human macrophages. *Inflamm. res.* **51**, 403–408 (2002).

20. Myles, A., Gearhart, P. J. & Cancro, M. P. Signals that drive T-bet expression in B cells. *Cellular Immunology* **321**, 3–7 (2017).

21. Wambre, E. R. G., DeLong, J. H., James, E. A., Robinson, D. & Kwok, W. W. TH2A Cells As a Unique TH2 Cell Subset in Allergic Individuals: Steps Toward a T Cell Biomarker For Allergy. *Journal of Allergy and Clinical Immunology* **129**, AB129 (2012).

22. Durham, S. R. & Shamji, M. H. Allergen immunotherapy: past, present and future. *Nat Rev Immunol* **23**, 317–328 (2023).

23. Miller, L. S. & Cho, J. S. Immunity against Staphylococcus aureus cutaneous infections. *Nat Rev Immunol* **11**, 505–518 (2011).

24. Jukic, A., Bakiri, L., Wagner, E. F., Tilg, H. & Adolph, T. E. Calprotectin: from biomarker to biological function. *Gut* **70**, 1978–1988 (2021).

25. Al-Saadany, H. M., Hussein, M. S., Gaber, R. A. & Zaytoun, H. A. Th-17 cells and serum IL-17 in rheumatoid arthritis patients: Correlation with disease activity and severity. *The Egyptian Rheumatologist* **38**, 1–7 (2016).

26. Krebs, C. F. *et al.* Pathogen-induced tissue-resident memory T_H 17 (T_{RM} 17) cells amplify autoimmune kidney disease. *Sci. Immunol.* **5**, eaba4163 (2020).

27. Zepp, J., Wu, L. & Li, X. IL-17 receptor signaling and T helper 17-mediated autoimmune demyelinating disease. *Trends in Immunology* **32**, 232–239 (2011).

28. Moniruzzaman, M. *et al.* Interleukin-22 suppresses major histocompatibility complex II in mucosal epithelial cells. *Journal of Experimental Medicine* **220**, e20230106 (2023).

29. Duhen, T., Geiger, R., Jarrossay, D., Lanzavecchia, A. & Sallusto, F. Production of interleukin 22 but not interleukin 17 by a subset of human skin-homing memory T cells. *Nat Immunol* **10**, 857–863 (2009).

30. Perez, L. G. *et al.* TGF-β signaling in Th17 cells promotes IL-22 production and colitis-associated colon cancer. *Nat Commun* **11**, 2608 (2020).

31. Barnes, J. L. *et al.* T-helper 22 cells develop as a distinct lineage from Th17 cells during bacterial

infection and phenotypic stability is regulated by T-bet. *Mucosal Immunology* **14**, 1077–1087 (2021).

32. Belpaire, A., Van Geel, N. & Speeckaert, R. From IL-17 to IFN-γ in inflammatory skin disorders: Is transdifferentiation a potential treatment target? *Front. Immunol.* **13**, 932265 (2022).

33. Kasatskaya, S. A. *et al.* Functionally specialized human CD4+ T-cell subsets express physicochemically distinct TCRs. *eLife* **9**, e57063 (2020).

34. Kurup, S. P. *et al.* Regulatory T cells impede acute and long-term immunity to blood-stage malaria through CTLA-4. *Nat Med* **23**, 1220–1225 (2017).

35. Dolina, J. S. *et al.* Developmentally distinct $CD4^+$ T_{reg} lineages shape the $CD8^+$ T cell response to acute *Listeria* infection. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2113329119 (2022).

36. Zelinskyy, G. *et al.* The regulatory T-cell response during acute retroviral infection is locally defined and controls the magnitude and duration of the virus-specific cytotoxic T-cell response. *Blood* **114**, 3199–3207 (2009).

37. Lagattuta, K. A. *et al.* Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat Immunol* **23**, 446–457 (2022).

38. Izraelson, M. *et al.* Comparative analysis of murine T-cell receptor repertoires. *Immunology* **153**, 133–144 (2018).

39. Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat Immunol* **17**, 946–955 (2016).

40. Petersone, L. *et al.* IL-21 shapes germinal center polarization via light zone B cell selection and cyclin D3 upregulation. *Journal of Experimental Medicine* **220**, e20221653 (2023).

41. Ettinger, R. *et al.* IL-21 Induces Differentiation of Human Naive and Memory B Cells into Antibody-Secreting Plasma Cells. *The Journal of Immunology* **175**, 7867–7879 (2005).

42. Olatunde, A. C., Hale, J. S. & Lamb, T. J. Cytokine-skewed Tfh cells: functional consequences for B cell help. *Trends in Immunology* **42**, 536–550 (2021).

43. Maceiras, A. R. *et al.* T follicular helper and T follicular regulatory cells have different TCR specificity. *Nat Commun* **8**, 15067 (2017).

44. Ke, F. *et al.* Germinal center B cells that acquire nuclear proteins are specifically suppressed by follicular regulatory T cells. *eLife* **12**, e83908 (2023).

45. Hashimoto, K. *et al.* Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24242–24251 (2019).

46. Elyahu, Y. *et al.* Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Sci. Adv.* **5**, eaaw8330 (2019).

47. Hasegawa, T. *et al.* Cytotoxic CD4+ T cells eliminate senescent cells by targeting

cytomegalovirus antigen. *Cell* **186**, 1417-1431.e20 (2023).

48. Decout, A., Katz, J. D., Venkatraman, S. & Ablasser, A. The cGAS–STING pathway as a therapeutic target in inflammatory diseases. *Nat Rev Immunol* **21**, 548–569 (2021).

49. Grossman, W. J. *et al.* Differential expression of granzymes A and B in human cytotoxic lymphocyte subsets and T regulatory cells. *Blood* **104**, 2840–2848 (2004).

50. Yang, E. & Li, M. M. H. All About the RNA: Interferon-Stimulated Genes That Interfere With Viral RNA Processes. *Front. Immunol.* **11**, 605024 (2020).

51. Kim, A. S. *et al.* Pan-protective anti-alphavirus human antibodies target a conserved E1 protein epitope. *Cell* **184**, 4414-4429.e19 (2021).

52. Khoury, D. S. *et al.* Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. *Nat Med* **27**, 1205–1211 (2021).

53. Lucas, C. *et al.* Delayed production of neutralizing antibodies correlates with fatal COVID-19. *Nat Med* **27**, 1178–1186 (2021).

54. Hor, J. L. *et al.* Spatiotemporally Distinct Interactions with Dendritic Cell Subsets Facilitates CD4+ and CD8+ T Cell Activation to Localized Viral Infection. *Immunity* **43**, 554–565 (2015).

55. Cullen, J. G. *et al.* $CD4^+$ T help promotes influenza virus-specific $CD8^+$ T cell memory by limiting metabolic dysfunction. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4481–4488 (2019).

56. Laidlaw, B. J. & Ellebedy, A. H. The germinal centre B cell response to SARS-CoV-2. *Nat Rev Immunol* **22**, 7–18 (2022).

57. Zuniga, E. I., Macal, M., Lewis, G. M. & Harker, J. A. Innate and Adaptive Immune Regulation During Chronic Viral Infections. *Annu. Rev. Virol.* **2**, 573–597 (2015).

58. Lanfermeijer, J. *et al.* Age and CMV-Infection Jointly Affect the EBV-Specific CD8+ T-Cell Repertoire. *Front. Aging* **2**, 665637 (2021).

59. Paley, M. A. *et al.* Progenitor and Terminal Subsets of $CD8^+$ T Cells Cooperate to Contain Chronic Viral Infection. *Science* **338**, 1220–1225 (2012).

60. Van Leeuwen, E. M. M., Remmerswaal, E. B. M., Heemskerk, M. H. M., Ten Berge, I. J. M. & Van Lier, R. A. W. Strong selection of virus-specific cytotoxic CD4+ T-cell clones during primary human cytomegalovirus infection. *Blood* **108**, 3121–3127 (2006).

61. Andreatta, M. *et al.* A CD4+ T cell reference map delineates subtype-specific adaptation during acute and chronic viral infections. *eLife* **11**, e76339 (2022).

62. Zander, R. *et al.* Tfh-cell-derived interleukin 21 sustains effector CD8+ T cell responses during chronic viral infection. *Immunity* **55**, 475-493.e5 (2022).

63. Bonsignori, M. *et al.* Antibody-virus co-evolution in HIV infection: paths for HIV vaccine

development. *Immunological Reviews* **275**, 145–160 (2017).

64. Fallet, B. *et al.* Chronic Viral Infection Promotes Efficient Germinal Center B Cell Responses. *Cell Reports* **30**, 1013-1026.e7 (2020).

65. Clement, M. *et al.* Cytomegalovirus-Specific IL-10-Producing CD4+ T Cells Are Governed by Type-I IFN-Induced IL-27 and Promote Virus Persistence. *PLoS Pathog* **12**, e1006050 (2016).

66. Corcoran, M. *et al.* Archaic humans have contributed to large-scale variation in modern human T cell receptor genes. *Immunity* **56**, 635-652.e6 (2023).

67. Gras, S. *et al.* Allelic polymorphism in the T cell receptor and its impact on immune responses. *Journal of Experimental Medicine* **207**, 1555–1567 (2010).

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

69. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).

70. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**, 825–835 (2021).

71. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).

72. 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).

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

74. Hudson, W. H. & Sudmeier, L. J. Localization of T cell clonotypes using the Visium spatial transcriptomics platform. *STAR Protocols* **3**, 101391 (2022).

75. Izraelson, M. *et al.* Distinct organization of adaptive immunity in the long-lived rodent Spalax galili. *Nat Aging* **1**, 179–189 (2021).

76. Wickham, H. *et al.* Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).

77. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

78. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

79. Pogorelyy, M. V. *et al.* Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol* **17**, e3000314 (2019).

80. Minervina, A. A. *et al.* SARS-CoV-2 antigen exposure history shapes phenotypes and specificity

of memory CD8+ T cells. *Nat Immunol* **23**, 781–790 (2022).

81. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med* **27**, 904–916 (2021).

82. Scharf, L. *et al.* Longitudinal single-cell analysis of SARS-CoV-2–reactive B cells uncovers persistence of early-formed, antigen-specific clones. *JCI Insight* **8**, e165299 (2023).

83. Hadjadj, J. *et al.* Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).

84. Bittner-Eddy, P. D., Fischer, L. A. & Costalonga, M. Transient Expression of IL-17A in Foxp3 Fate-Tracked Cells in Porphyromonas gingivalis-Mediated Oral Dysbiosis. *Front. Immunol.* **11**, 677 (2020).

85. Wang, J., Ioan-Facsinay, A., van der Voort, E. I. H., Huizinga, T. W. J. & Toes, R. E. M. Transient expression of FOXP3 in human activated nonregulatory CD4 $^{+}$ T cells. *Eur J Immunol* **37**, 129–138 (2007).

86. Renaude, E. *et al.* Epigenetic Reprogramming of CD4+ Helper T Cells as a Strategy to Improve Anticancer Immunotherapy. *Front. Immunol.* **12**, 669992 (2021).

87. Yasumizu, Y. *et al. Single-cell transcriptome landscape of circulating CD4 $^{+}$ T cell populations in human autoimmune diseases*. http://biorxiv.org/lookup/doi/10.1101/2023.05.09.540089 (2023) doi:10.1101/2023.05.09.540089.

88. Minervina, A. A. *et al.* Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. *eLife* **9**, e53704 (2020).

89. Klein, J. *et al.* Distinguishing features of Long COVID identified through immune profiling. *Nature* (2023) doi:10.1038/s41586-023-06651-y.

90. Minervina, A. A. *et al.* Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *eLife* **10**, e63502 (2021).

91. Humphreys, I. R. *et al.* Cytomegalovirus exploits IL-10–mediated immune regulation in the salivary glands. *The Journal of Experimental Medicine* **204**, 1217–1225 (2007).

92. West, E. E. *et al.* Loss of CD4+ T cell-intrinsic arginase 1 accelerates Th1 response kinetics and reduces lung pathology during influenza infection. *Immunity* **56**, 2036-2053.e12 (2023).

Acknowledgments

Hereby, I would like to thank all great people who helped me during the work on my PhD Thesis.

First of all, I would like to thank my supervisor, Prof. Dr. Andre Franke, who hosted me in Kiel and provided me with support, guidance, and new opportunities in science. Your support during the hardest times was crucial and, at the latest stages of my research, generated in me enthusiasm to complete this thesis.

I am also grateful to Dr. Dmitriy Chudakov for being my mentor during all stages of my research. Your enthusiasm for immunology always gives me the drive in my own work. Thank you for your guidance; this work would never be completed without your support.

I would also like to thank my co-supervisor, Prof. Dr. Petra Bacher, and Dr. Olga Britanova. I value a lot all your comments and advice. Thank you for expanding my knowledge in immunology and molecular biology.

I am grateful to all members of the groups led by Andre Franke and Dmitriy Chudakov. I would like especially thank Aya Mahdy, Tim Steiert, Hesham El Abd, and Alba Troci from our group in Kiel. You did not only teach me a lot in science but also helped me with various challenges that I faced as an international student. I am immensely grateful to Nadezhda Logunova, Irina Shagina, Daniil Lukyanov, Ekaterina Merzlyak, Dmitriy Staroverov, Kseniia Lupyr, and Pavel Shelyakin. You taught me almost everything I know about lab work and TCR data analysis, and I value your contribution to my work a lot.

I thank all donors who voluntarily donated their blood for this research. And I am also grateful to the members of the DNA lab, NGS lab, and the Cell Biology lab in Kiel, Yewgenia Dolshanskaya and Sabine Kock, for their contribution.

Finally, I would like to thank my husband Ivan, my friends and family. Your support is truly invaluable.

Declaration

I hereby declare,

I. that apart from my supervisor's guidance, the content and design of this thesis is completely my own work. Contributions of other authors are listed in the following section.

II. this thesis has not been submitted either partially or completely as part of a doctoral degree to another examining institution. No other materials are published or submitted for publication than indicated in this thesis.

III. this thesis was prepared in compliance with the "Rules of Good Scientific Practice" of the German Research Foundation (DFG)

IV. that I did not have an academic degree revoked.

_____ _____

Valeriia Kriukova

Authors contributions

Manuscript 1: MHC-II alleles shape the CDR3 repertoires of conventional and regulatory naive CD4+ T cells (Logunova *et al. PNAS* 2020)

N.N.L., D.M.C., A.S.A., and O.V.B. designed the research; N.N.L., **V. Kriukova**, E.S.E., A.P., and E.M.M. performed research; N.G.B., M.S., D.S.S., M.V.P., and J.M. contributed new reagents/analytic tools; **V. Kriukova**, P.V.S., and V.N.Z. analyzed data; N.N.L., **V. Kriukova**, N.G.B., D.S.S., D.M.C., A.S.A., and O.V.B. wrote the paper.

N.N.L. and **V. Kriukova** contributed equally to this work.

In detail, the contribution of **V. Kriukova** was

a) the isolation of splenocytes from H2-Ab $^{0b-/-}$ mice and sample preparation for cell sorting;
b) the TCR-Seq for H2-Ab $^{0b-/-}$ mice;
c) TCR repertoire data analysis together with P.V.S. and V.N.Z.;
d) Figures preparation, together with P.V.S., N.N.L., and O.V.B.;
e) Preparation of the manuscript draft and editing of the final version together with N.N.L., D.S.S., D.M.C., A.S.A., and O.V.B.

Manuscript 2: Sort-Seq: immune repertoire-based scRNA-Seq systematization (Kriukova et al., under review in *Nature Communications*)

D.M.C. designed the research; **V. Kriukova**, I.A.S., K.M., and K.L. performed the scRNA-Seq experiments; I.A.S. performed the TCR-Seq; I.A.S., D.B.S., and G.V.S. performed T cell functional assays; N.R.V., L.S., and M.M.A. performed the ELISPOT assays; M.A.S. performed the HLA typing; K.R.L. and D.M.C. analyzed the TCR-Seq data following the functional T cell assays; **V. Kriukova** and D.M.C. analyzed the TCR-Seq data following the longitudinal PBMC tracking; D.K.L**., V. Kriukova**, K.A.B., and D.M.C. analyzed the scRNA-Seq data; **V. Kriukova** curated samples and data; O.V.B., A. Franke, D.P., D.M.C. supervised the study; **V. Kriukova**, D.K.L, and D.M.C. prepared figures and the draft of the manuscript.

Manuscript 3: Inhibitory IL-10-producing CD4+ T cells are T-bet-dependent and facilitate cytomegalovirus persistence via coexpression of arginase-1 (Clement et al. *eLife* 2023)

M.C. and I.R.H. conceptualized the study and administrated the project; M.C., K.L., K.L.M., M.M., L.C., A.C.F., J.S., S.C., **V. Kriukova**, and K.R.L. curated the data; M.C., K.L., M.M., A.C.F., R.A., **V. Kriukova**, K.R.L., O.V.B., D.R.W., S.A.J., D.M.C., D.A.P., and I.R.H. performed the formal analysis; R.A., O.V.B., D.R.W., D.M.C., D.A.P., and I.R.H. acquired resources; M.C., O.V.B., D.M.C., D.A.P., and I.R.H. wrote and edited the manuscript.

In detail, the contribution of **V. Kriukova** was

a) TCR-Seq experiments;
b) TCR repertoire data analysis together with K.R.L.;
c) TCR repertoire-related figures preparation together with O.V.B. and D.M.C.