Modelling the T-cell repertoires of circulating T-cells and its application in cardiovascular diseases

Kenz Le Gouge

Sorbonne Université

Ecole doctorale

UMRS959 / Immunologie, immunopathologie, immunothérapeutique

Modelling the T-cell repertoires of circulating T-cells and its application in cardiovascular diseases

By Kenz Le Gouge

Thèse de doctorat d'Immunologie

Directed by Encarnita Mariotti-Ferrandiz

Presented and publicly defended on December, 14th of 2023

Evaluated by a jury composed of :

Pr Ziad Mallat, MD, PhD, thesis reviewer

Pr José Borghans, PhD, thesis reviewer

Dr Anne Eugster, PhD, thesis reviewer

Pr Pierre Boudinot, PhD, thesis examiner

Pr Yves Allenbach, MD, PhD, thesis examiner

Dr Encarnita Mariotti-Ferrandiz, thesis supervisor

Acknowledgements

I would like to thank my supervisor Mrs. Mariotti-Ferrandiz, the reviewers and examiners of my thesis jury to accept to evaluate my work.

Gustavo Ramos and Peter Rainer for their help throughout my 4 years of PhD. It was a pleasure to work with you on such a fantastic topic that is immuno-cardiology.

David Klatzmann, merci de m'avoir accueilli dans votre laboratoire et de m'avoir fourni un environnement idéal pour laisser part à ma créativité sans bornes sur PowerPoint.

Paul Stys, un amant, un frère, un père, beaucoup de personnes en une seule. Tu es ma plus belle *bromance*, je n'imagine pas ma vie au labo sans toi, et le reste de l'étage rêve de notre départ. Le feu et la glace, le yin et le yang, l'ordre et le « ah je pensais qu'on était hier ». Pour notre futur professionnel à tous les deux, il vaut mieux se séparer, considère ces remericements comme une déclaration de rupture. Prends ton envol, petit ange, pour éviter le prochain « ils ont fait QUOI ? ».

A ma famille, mes parents, Merwan, Yasmine, qui ont toujours cru en moi et m'ont soutenu jusqu'au bout dans cette aventure. « Quand est-ce que tu finis tes études », « et après tu fais quoi comme vrai métier », « ça avance ton rapport de stage ? » sont des phrases qui ont été curieusement motivantes et désormais gravées dans ma mémoire.

Nicolas C., mon collègue de musique, compagnon de crêpes et binôme de blind-test. Au-delà du travail et de ton éternel esprit positif, tu m'auras ouvert les yeux sur la prolifique année 1996, et notamment *Freed from Desire* de Gala. Au plaisir d'échanger avec toi au bord des platines !

Lucas (dit le L) et Pierre (dit le P), de nos jeunes années à l'UPMC jusqu'au COVID des Apero Electroniques, il ne nous manque qu'un brunch sur une certaine péniche pour pouvoir dire « on a tout fait ».

Julie et Matthias, les deux électrons libres de Clermont-Ferrand. Merci Julie, pour ton amour de la Nièvre et ta volonté de la protéger. Pour le gratton, ce n'est pas nécessaire. Grâce à vous, je sais placer Michelin City : dans le 13ème.

Helena, petit ange partie trop tôt. Tu auras été pour moi un modèle de ce que doit être un docteur pendant, et après sa thèse ! Ta bonne humeur permanente et contagieuse aura beaucoup manqué, et j'espère de tout cœur que nos chemins se rencontreront à nouveau autour d'un dîner.

A l’équipe médicale: Paul, Marie, Chloé, Aurore, Solène, Cathy et tous les autres, merci pour tous ces moments, escapades tant physiques que sensorielles !

Keren, pour ton aide précieuse dans les moments qui ont compté, ta gentillesse et ton empathie à toute épreuve.

Merci à toute l’équipe TCR, présente et passée, qui m’a ~~supporté~~ accompagné lors de toutes ces années et permis aujourd’hui de présenter un travail acceptable sur le plan scientifique, ainsi que tous les autres membres du laboratoire. Je n’ai rien réalisé seul, et chacun de vous a pris part à mes travaux.

Encarnita, à laquelle je laisse le dernier paragraphe. Pour ton encadrement, ton humanité et ta bienveillance. Tu m’as vu arriver en petit étudiant perdu qui voulait apprendre la bioinformatique, et tu me vois partir en vieux docteur, toujours perdu. Entre mes blagues, ma lecture des emails et ma légère tendance à voir le verre à moitié vide, je suis admiratif que tu n’aies pas décidé de m’abandonner sur la route. Je n’ai pas été un étudiant facile mais tu as été d’une patience infinie (même si tu quittes souvent la pièce quand j’ouvre la bouche). **Un immense merci pour ces 4 ans**.

Acronyms table

Abbreviation	Full form
AIR	*Adaptive Immune Repertoire*
AIRE	*Autoimmune Regulator*
AIRR	*Adaptive Immune Recepetor Repertoire*
AIRR-seq	*Adaptive Immune Recepetor Repertoire sequencing*
AMI	*Acute myocardial infarction*
APC	*Antigen presentating cell*
cDNA	*Complementary DNA*
CDR	*complementary determining region*
CMV	*Cytomegalovirus*
CTL	*Cytotoxic T lymphocyte*
CD	*Cluster of differenciation*
CTLA-4	*cytotoxic T-lymphocyte-associated protein 4*
CVD	*Cardiovascular disease*
DNA	*Desoxyribonucleic acid*
DP	*Double positive*
FFPE	*Formalin-fixed and paraffin-embedded*
EFS	*Etablissement Français du Sang*
ExAmp	*Exclusion amplification*
Foxp3	*Forkhead box protein 3*
GWAS	*Gene wide association study*
HLA	*Human leukocyte antigen*
HTS	*High throughput sequencing*
Ig	*Immunolgobulin*
IL	*Interleukin*
IMGT	*International ImMunoGeneTics information system*
IPEX	*Immunodysregulation polyendocrinopathy enteropathy X-linked syndrome*
MHC	*Major histocompatibility complex*
MI	*Myocardial infarction*
MIS-C	*Multiple inflammatory syndrome in children*
mRNA	*messenger RNA*
Mtb	*Mycobacterium tuberculosis*
NGS	*New geenration sequencing*
PBMC	*peripheral blood cell*
Pgen	*Probability of generation*
RACE	*Rapid amplification of complementary ends*
RAG	*Recombination activating gene*
RIN	*RNA integrity number*
RNA	*Ribonucleid acid*
RSS	*Recombining signal sequence*
SA	*Staphylococus aureus*
SARS-CoV-2	*Severe acute respiratory syndrome coronavirus 2*
SBS	*Sequencing by synthesis*
SCID	*Severe combined immunodeficiency*
T1D	*Type 1 Diabetes*
TAC	*Transversal aortic constriction*
TCR	*T cell receptor*
TCR-seq	*T cell receptor sequencing*
TdT	*Terminal deoxynucleotidyl transferase*
TEC	*Thymic epithelial cell*
TGF	*Transforming growth factor*
Th	*T helper*
TNF	*Tumor necrosis factor*
TRA	*T cell receptor alpha*
TRB	*T cell receptor beta*
Treg	*Regulatory T cell*
TSS	*Toxic shock syndrom*
UCI	*Universal cellular identifier*
UDI	*Unique dual indexing*
UMI	*Unique molecular indexing*
RDI	*Redundant dual indexing*

1 CHAPTER 1: INTRODUCTION

The immune system is a vast and complex network of molecules, cells and organs, tightly orchestrated to function as safeguards against self and foreign threats. It involves multitudes of specialised cell types such as T cells, B cells, macrophages, and dendritic cells, each with distinct functions in maintaining the individual's homeostasis. The immune system also countains diverse set of mediators: antibodies, cytokines, chemokines that effectively coordinate immune responses and signals between cells. Immune receptors, whether they are dedicated to the recognition of pathogenics patterns or mediators, enable signal transduction to immune cells and regulate their functions. Dedicated tissues, called lymphoid organs, function as monitoring stations distributed across the body, to mount an effective and quick response. From this intricate and multilayer network of cooperating molecules, proteins, cells and tissues, emerges the integrative notion of system.

The immune system complexity requires an almost perfect orchestration of many different cellular actors to function and effectively maintain the host's homeostatic state. Because of its complexity and compartmentalisation, a first encounter is always long to address (Farber et al., 2016) and complete pathogen clearance can take weeks. However, subsequent infections with the same pathogen are resolved much faster, typically in a few days. This process called immunological memory highlights the ability of the immune system to "store" memories of past encounters. The recognition and rapid response is possible due to pre-existening cells that are specifically designed to identify and kill the pathogen upon repeated encounters. Although this has been known for centuries and in many cultures (Gross and Sepkowitz, 1998), immunological memory remained mysterious until science progressed. Work of Paul Erlich (1892) demonstrated that immune memory could not be genetically heritable in the traditional sense, but instead each individual shape their own immune system (Ehrlich and Hübener, 1894). Memory is a very dynamic system, which can "remember" past events and rapidly respond to threats previously encountered, while still being able to learn and adapt to new encounters. The feature is primarily possessed by cells that are part of the the adaptive immune system, although recently the innate immune response has been proposed to contain a certain degree of memory, leading to the concept of trained immunity (reviewed by Netea et al., 2020). This concept will not be covered in this introduction.

The immune system is divided in two main compartments, the innate and the adaptive compartment, distinct but deeply interconnected. The latter is called adaptive based on its

ability to constantly evolve and adapt to new threats at the scale of an individual's lifespan. The adaptive immune system consists of two main cell subsets; the B cells and the T cells. At the interface of innate and adaptive immune system, antigen-presenting cells are responsible for capture and presentation of antigens to other immune cells. Their role fulfils a dual purpose: it is crucial for initiating the adaptive immune response, while also modulating and maintaining the tolerance to self. In this manuscript, I will focus on the T cell subpopulation, as this has been the principal focus of my PhD thesis work. I will detail how it manages to not only recognize a specific antigen, but also any new antigen that a pathogen can harbour.

1.1 T CELLS IN THE IMMUNE SYSTEM

1.1.1 TCR-bearing cells, or T cells

T cells are a distinct population from the white blood cells, also called leukocytes. They are the major lymphocytes population, representing around 70% of lymphocytes in the peripheral blood of healthy adults (Reichert et al., 1991). The main feature/characteristics of T cells is their antigen-specific receptor, known as the T cell receptor (TCR), expressed on their surface. The TCR exhibits a high affinity for a multi-protein complex consisting of a peptide, originating from a pathogen or derived from a host cell, and the major histocompatibility complex (MHC). Such complex is expressed at the surface of antigen presenting cells (APCs) allowing the presentation of a given peptide, within the context of a particular MHC to T-cells. Such peptide in such context is named an antigen, which means a molecule that may have the capacity to induce an immune response through the activation of the T-cell upon its recognition by the TCR. As such, it is hypothesised that this particular complex is responsible for T-cell specificity, as T cells are activated and therefore carry out their functions through the TCR/pMHC complex binding. T cells are a diverse population, comprising different phenotypical subsets that will accomplish specific roles. T cells can be divided in two major groups, effectors and regulator. Effector T cells are experienced, having been activated by the presentation of an antigen. They have functions to destroy the threat, whether it is by recruiting and activating other cells of the immune system, or directly killing it. In contrast, regulatory T cells are present much more restricted quantity and possess this ability to control immune response. They play a major role in self-tolerance and preventing the effector compartment to wreak havoc. In this chapter, I will cover the plurality of T cells from their phenotypic

description, and their functions, to the generation of their TCR that imparts their exceptional diversity.

1.1.1.1 T for thymus

T cells originate from a unique organ: the thymus. It has long been established that the thymus was an organ where lymphocytes were produced, but remained unexplored for a long time due to being considered vestigial. It was in 1961 that Jacques Miller showed that the thymus was, indeed, critical to mount immune responses. He was the first to demonstrate how neonatal thymectomy on mice lead to their inability to fight against viruses or reject skin grafts (Miller, 1961a, 1961b). These experiments paved the way to show how the thymus was the central organ to mount a thymic-dependant response, with thymic lymphocytes: the **T cells**.

Although T cells are generated in the thymus, T cell progenitors, thymocytes, are derived from hematopoietic stem cells in the bone marrow (Kondo et al., 1997). Once generated, they migrate to the thymus where they undergo different stages of maturation. This process, called thymopoiesis, is when T cells will acquire their phenotype and express markers at their surface that will determine their role and fate. We can distinguish two main subtypes of T cells, depending on the cluster of differentiation (CD) acting as a co-receptor associated to the TCR: $CD4^+$ T cells and $CD8^+$ T cells. Each subtype further subdivides into different subsets of specialised cells.

1.1.1.2 $CD4^+$ T cells

$CD4^+$ T cells represent the majority of T cells in an adult (~60% of T cells). They play a major role in orchestrating the immune response and cover a wide range of functions. They express a large panel of receptors and ligand, which can activate B and T cells through cognate interaction. They also secrete a large range of cytokines such as IFN-γ or GM-CSF, which stimulate macrophages and other cells of the innate immune system, as well as non-immune cells (Boehm et al., 1997; Pawlak et al., 2020). More than mere activators of other cells, $CD4^+$ T cells can fuel inflammation or polarise immune responses (Tuzlak et al., 2021). $CD4^+$ T cells are restricted by the MHC-II, meaning their TCR can exclusively recognise peptides bound to MHC-II molecules at the surface of APCs. While peptides are presented by cells from the self, they can originate from intra-cellular or extra-cellular milieu through dedicated processes covered in 1.1.3.1.

Distinct $CD4^+$ T cell subtypes accomplish distinct functions, with most of each subtype are specialising based on their phenotype. First identified in 1986 by Mosmann *et al*., T helper -1 (Th1) and -2 (Th2) cells are the most common type of helper T cells. They are distinguished by their set of secreted molecules, or secretomes (Mosmann et al., 1986). They play a role in activating other immune cells, by releasing factors that will orient the immune reponse. Th1 are associated to the cellular response. By producing IL-2 and interferon-γ, the Th1 subset drives a strong, global inflammatory T cell response that mainly targets intracellular pathogens and favour cellular debris clearance by stimulating macrophages (Hsieh et al., 1993; Hwang et al., 2005; Nathan et al., 1983). Th2, on the other hand, secrete IL-4, IL-5 and IL-13 and are associated to extra-cellular pathogens response (Glatman Zaretsky et al., 2009). Mostly known for its role in allergic responses, its secretome favours B-cell antibody class switch to immunoglobulin E (IgE), specialised in anti-parasite immune responses (Mohrs et al., 2000). Interestingly, Th1 and Th2 responses are polarised, meaning one inhibits the other. This mechanism reinforces the specialisation of T cells in orchestrating the immune responses (Brinkmann et al., 1993).

In 2005, a series of published studies redefined the well dichotomized world of helper T cells with the discovery of Th17 T cells (Harrington et al., 2005; Langrish et al., 2005; Park et al., 2005), a pro-inflammatory subtype induced during bacterial infections which secretes IL-17. Although these specialised T cells were discovered in the gut, their role in inflammatory diseases pathogenesis was almost immediately investigated and demonstrated in colitis, uveitis, scleritis and rheumatoid arthritis (Amadi-Obi et al., 2007; Hue et al., 2006; Sato et al., 2006; Yen et al., 2006). Further studies on patients showed that these cells recirculate in healthy conditions and are elevated in inflammatory diseases (Acosta-Rodriguez et al., 2007; Ellul et al., 2021; Nistala et al., 2010).

1.1.1.3 CD8 T cells

$CD8^+$ T cells, also called the cytotoxic T lymphocytes (CTL), are specialised in the identification and elimination of infected cells, whether they are viruses, bacteria, or tumoral cells (Kägi et al., 1994; Walsh et al., 1994). As opposed to $CD4^+$ T cells, their TCR can exclusively recognize peptides bound to MHC-I. At first called cellular antibodies (Govaerts, 1960), $CD8^+$ T cells are coined cytotoxic as they acquire a cell-killing capacities after their

activation in secondary lymphoid organs. As poorly regulated cytotoxicity can cause uncontrolled damage, $CD8^+$ T cells leave the thymus as naïve cells. In order to unlock their full arsenal, they need to be activated by a presentation of a peptide that complement their TCR in secondary lymphoid organs. CTLs rely on three distinct but complementary mechanisms to kill cells. First they utilise release of the perforine/granzyme molecules, located in secretory granules of CTL (Kägi et al., 1994; Rosenau and Moon, 1961). Upon contact with its target, the CTL will release the granules. Perforin creates pores on target's surface, allowing granzyme to activate a caspase-induced apoptosis of the target cell. Second mechanism relies on the so-called "cell surface death receptor" Fas molecule and its ligand, FasL. Cross-linking of CTL's FasL to its target Fas leads to a similar apoptosis caspase-dependant cascade signal, effectively killing its target. The two mechanism are independent and show synergetic action (Hassin et al., 2011). A similar system with the TRAIL molecule completes the artillery of CTLs.

1.1.1.4 Regulatory T cells

T helper and CTL are often called conventional T cells, as opposed to regulatory T cells (T_{REG}). T_{REG} are a subtype of T cells specialised in the maintenance of self-tolerance and the regulation of immune response. The vast majority of T_{REG} exhibit the CD4 co-receptor, although some also express the CD8 (Mayer et al., 2011). Additionally, they express the forkhead box P3 (FoxP3), a transcription factor that will heavily influence their functionality. FoxP3 is the master regulatory factor of T cells (Fontenot et al., 2003). Deleterious mutations in *FOXP3* gene induce immunodysregulation polyendocrinopathy enteropathy X-linked syndrome (IPEX syndrome) (Powell et al., 1982), and murine models are short lived, spontaneously developing a wide range of autoimmune disorders (Hadaschik et al., 2015). T_{REG} exert an active suppression of the immune system to prevent pathological autoimmunity and prevent deleterious consequences of uncontrolled inflammation during infections. Additionally, T_{REG} play key roles in cancer development as well on wound healing. T_{REG} use a large array of suppressive mechanisms to exert their immunoregulatory functions, with dozens described, including both cognate interactions and production of soluble factors.

The primary impact of T_{REG} is directly on conventional T cells. They were initially identified by their high expression of CD25 (Sakaguchi et al., 1995). CD25 is the high affinity receptor for IL-2, and T_{REG} outcompete conventional T cells by depriving them of an exogenous source of IL-2. T_{REG} cells also express several immune checkpoints ligands on their surface, which are key molecules involved in the immunoregulation of T cells. Moreover, T cells

promote a shift towards a pro-tolerogenic state in antigen-presenting cells, such as dendritic cells (Onishi et al., 2008). More precisely, they express inhibitory molecules, such as the Cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), which will indirectly induce an anergic state in conventional T cells. They also secrete immunosuppressive cytokines, such as IL-10 or IL35. IL-10 mediates direct suppression on the T compartment, and was found to be specifically involved in the regulation of Th17 response (Chaudhry et al., 2011). Finally, T_{REG} are heavily implicated in wound healing. Skin-resident T_{REG} show active attenuation of inflammation on cutaneous wounds, and reduced pro-inflammatory macrophage accumulation (Nosbaum et al., 2016).

All T_{REG} are not created equal, and they differ depending on their origin. There are native T_{REG}, generated in the thymus (tT_{REG}) and peripheral T_{REG} (pT_{REG}), generated in the peripheral from the pool of conventional T cells by stimulation with tumour growth factor β (TGF-β) (Chen et al., 2003). It was shown that tT_{REG} are more transcriptomically stable than pT_{REG}, the latter displaying a "plasticity", losing their regulatory phenotype in absence of TGF-β (Komatsu et al., 2009).

1.1.2 Origins of the T cell diversity

T cell are defined by their TCR, which is central for their recognition of foreign and self-peptides. However, two mysteries puzzled scientists for quite a while. It became clear that T cell diversity was immense, and that the potential diversity could not be accounted for by the ~23 000 coding genes of our genome (Martin et al., 2023). Furthermore, T cells had to be able to recognise peptides from yet unknown pathogens whether an individual had yet to encounter them, or because they would emerged from the never-ending race of viral mutations. Specificity is another black cloud in the TCR sky, as estimates in mice show that it would require more than 100 times a mouse volume in T cells alone to hold 1 naïve helper T cell for each possible peptide-MHC-II complex (Mason, 1998). From this apparent contradiction, immunologists started to unravel the secrets of TCR generation, delving into the modalities of peptide presentation and antigen recognition by the TCR.

1.1.2.1 T cell structural organization

The T cell receptor was identified in early 1983 by several group concurrently (Kappler et al., 1983; Meuer et al., 1983), and lead to a 20 years race to unfold its secrets. Its primary structure was resolved in mice in 1984 (Saito et al., 1984a, 1984b) and the αβ chains were

structurally resolved at 2.5 Å in 1996 (Garcia et al., 1996) . Expressed by all T cells of the periphery, the TCR is a heterodimer made of an α and a β chain, associated to a larger 5-mer complex called the CD3 consisting of the γ, δ, ε and ζ subunits, and either the CD4 or CD8

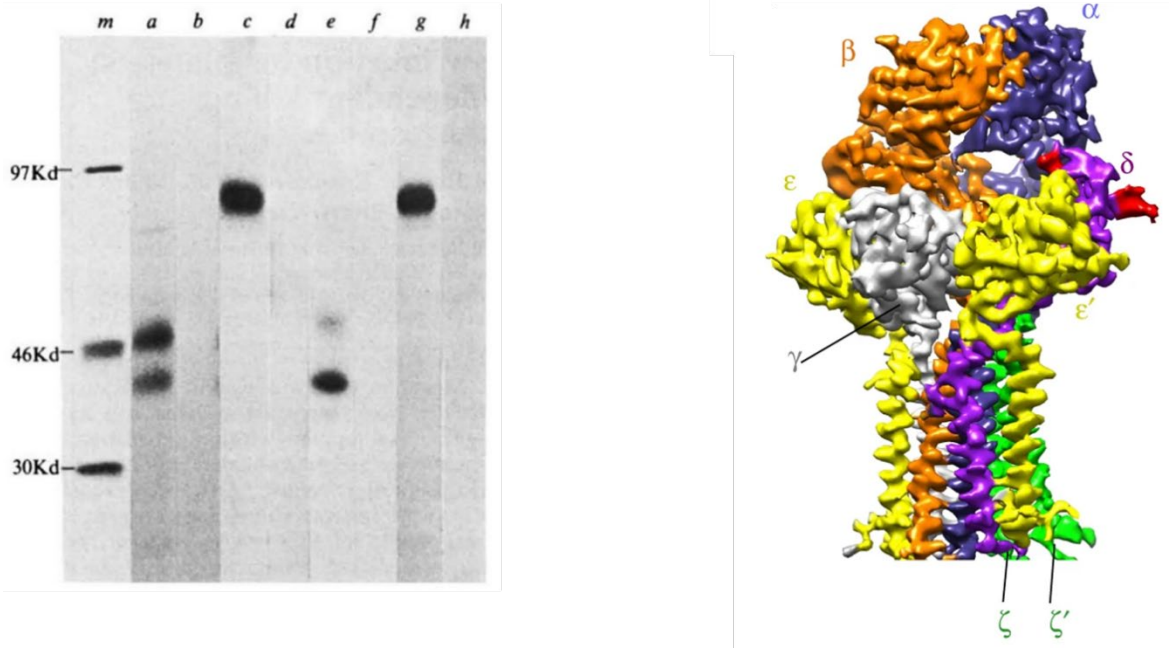


Figure 1: ***From the first observation to complete characterization of the TCR complex.*** *(**left**) Autoradiography of an SDS-polyacrylamide gel electrophoresis of a CD8 and CD4 T cell receptor. a : reduced form clone CT8, c : dimerized form, e reduced form clone CT4, : dimerized form CD4. From Meuer et al, 1983. (**right**) The crystallography resolved complete TCR complex, with α and β chain, CD3γ, CD3ε CD3δ, and homodimer CD3ζζ. From Dong et al., 2019*

accessory molecule. This octamer structure, although known for more than 40 years, was only structurally resolved in 2019 by De Dong *et al.* through cryo-electron microscopy (Dong et al., 2019). TCR final assembled form has a molecular mass of 58 kD, with a 28kD α chain and 30 kD β chain.

Extracellular domain: TCR ectodomain is an immunoglobin-like receptor, with a constant region (C) and variable region (V) derived from somatic rearrangement of several loci (see 1.1.2.2). It is composed of six loops, 3 per chain, called complementary determining regions, or CDR. The CDR1 and CDR2 sequences are mostly constant, whereas the CDR3 is highly variable and is considered to define the specificity of the TCR. The joining of the two CDR3 of the heterodimer creates the peptide-binding region with the MHC, which will be described in further details. Although CD3 extracellular domains are immunoglobin-like, their role is not related to antigen recognition, but rather involved in the signal transduction (Frank et al., 1990; Irving and Weiss, 1991).

Transmembrane domain: Both α and β chains cross the membrane only once. This transmembrane (TM), conserved region is necessary for the assembly of the full TCR-CD3

complex, as TM-truncated α chains failed to assemble to CD3 subunits (Alcover et al., 1990; Manolios et al., 1990), and lead to impaired signalling.

Cytoplasmic domain: Cytoplasmic region of the αβ TCR is too short to be able to transduce the activation signal but rather rely on the CD3. Especially, the CD3-ξξ dimers contain ITAMs motifs that will be phosphorylated by the Lck protein. These phosphorylated ITAMs will provide docking sites for further adaptor kinases and lead to the transduction of the signal (Marie-Cardine and Schraven, 1999).

1.1.2.2 T cell genetic organization

The genetic mechanisms responsible for TCR generation are very similar to those of antibodies, and literature often refer to them as antibody-like. TCR loci cover a large portion of the genome, respectively 930kb and 510kb for the TRA and TRB loci (Martin et al., 2023). TCR are the result of somatic recombination during T cell maturation. Akin to antibodies, it involves 3 family of genes: the Variable (V), Diversity (D) and Joining (J) genes (Davis and Bjorkman, 1988; Siu et al., 1984; Tonegawa, 1983). The considerable work of drawing a cartography of the immune receptor (BCR, TCR, MHC) genes has been initiated by Marie-Paule Lefranc in 1989 (Lefranc et al., 1999) through the international ImMunoGeneTics (IMGT) laboratory that hold a database. This online platform enumerates all the Ig of TCR genes of vertebrates. As of 2023, the human database indicates that a TCR α chain can be assembled from 54 TRAV, 61 TRAJ and 1 TRAC, and the β chain from 64 TRBV, 2 TRBD and 14 TRBJ and 2 TRBC genes (Lefranc and Lefranc, 2001).

The database has remained unchanged since 2003, yet it is still widely considered as the primary reference for immunoglobin analysis. Nonetheless, new genome versions keep getting refined (Grh38) and gene tables need to be updated. For instance, CellRanger 3.1.0 (10X Genomics) single-cell alignment uses a custom TCR reference, which features an adapted version the IMGT TCR gene database according to their observations (pseudo-genes, alternate splicing versions, custom mutations in mice backgrounds) (https://support.10xgenomics.com/single-cell-vdj/software/pipelines/3.1/advanced/built-in-refs).

1.1.2.3 The TCR, from finite elements to virtually infinite diversity

The V, D and J genes undergo rearrangment into a single DNA sequence prior to their expression as a single chain at the surface of T cells. However, we have discussed their distribution across several hundreds of kilobases (see 1.1.2.2). The process of somatic rearrangements each thymocyte undergoes, called the V(D)J recombination, is crucial for the T cell differenciation and for the TCR diversity generation (Figure 2). During this process, one V, one D (on the TRB locus only) and one J genes will be randomly selected among the various members of each of these three families and joined together to form separately the α and β chains of the TCR. This process is regulated through genetic and molecular mechanisms to prevent aberrant recombination, and subsequent reediting of these loci. We will detail here the mechanisms underlying the fundamental engine driving antigen receptor diversity.

VDJ recombination

Each VDJ gene is flanked either by a 12-base pair recombination signal sequence (12RSS) or by a 23RSS. An RSS is defined as a 12 or 23 bases "spacer" sequence flanked by a palindromic heptamer (CACAGTG) and a palindromic conserved nonamer (ACAAAAACC) (Ramsden et al., 1994). Genes can only join between a 12RSS and a 23RSS, a principle known as the 12/23 rule (Early et al., 1980; Kurosawa et al., 1981). V_α and V_β genes feature a 23RSS at their 3', D_β segments are flanked by 12RSS in 5', and 23RSS in 3', J_α and J_β feature 12RSS on 5' (Figure 3).

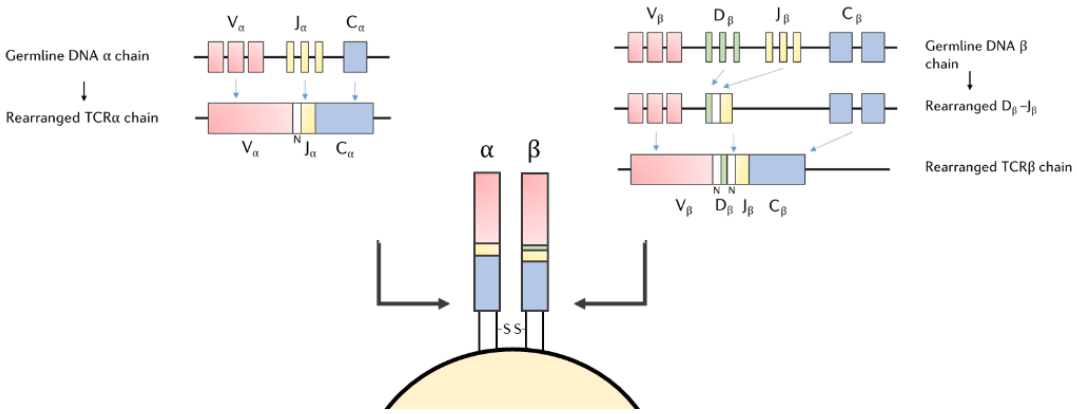


*Figure 2: **Schematic mechanisms of the TCRαβ VDJ recombination**. α chain is recombined through VJ recombination, while β chain by VDJ recombination. Both chain are assembled to form a complete αβ TCR.*

The RAG1 and RAG2 proteins, acronyms for Recombination-Activating Genes 1 and 2, respectively, are two conserved, co-expressed, lymphoid-specific enzymes that are responsible for the joining of the V(D)J and C genes (Oettinger et al., 1990). The RAG1/2 complex will bind to an RSS and create two single-stranded nicks at the start and end of the RSS (Hirokawa et al., 2020). This process leads the cleavage and hairpin formation of a sealed "coding" end, and a blunt "signal" end. Hairpins of one 12RSS and 23RSS coding ends will be reopened and sealed again. This step involves the terminal deoxynucleotidyl transferase (TdT), a protein that performs a non-templated nucleotide enzyme addition, with random deletion of bases during the joining step through exonucleolytic activity (Bodger et al., 1983). For the β chain, this process is in 2-step to avoid recombining a V segment to a J segment, with a DJ recombined first, followed by V-DJ to form a complete VDJ exon. The α chain is a 1 step process and does not involve D segment (Figure 3).

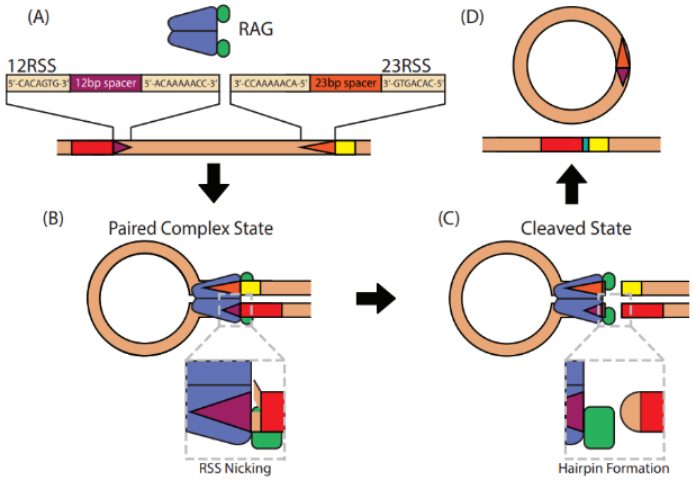


Figure 3: ***Schematic focusing on the initial steps of V(D)J recombination.*** *(**A**) The RAG protein complex binds to the 12- and 23RSSs (purple and orange triangles, respectively) neighboring gene segments (shown as red and yellow boxes on the DNA), (**B**) forming the paired complex (PC). At any point when it is bound to an RSS, RAG can introduce a nick in the DNA between the heptamer and gene segment (shown with the magnified 12RSS) and must do so to both sites before (**C**) it cleaves the DNA to expose the gene segments. As indicated by the magnified gene segment end, the exposed DNA strands of the gene segment are connected to form a DNA hairpin. (**D**) Additional proteins join these segments together. In this work, the stages subsequent to DNA cleavage are not monitored. From Hirokawa et al, 2019*

Importantly, while this mechanism is highly effective in generating the immense diversity of TCRs, it is also highly inefficient since 2/3 of rearrangements will produce non-productive chains due to the random addition and deletion of nucleotides happening at each recombination event, resulting in out-of-frames chains. Moreover, it presents a mere 1/9 chance to produce productive α and β chain. To reduce the chance of producing non-productive TCR, T cells successively rearrange their β, then their α loci. In a process called allelic exclusion, a β mono-allelic rearrangement is initiated. If non-functional, the loci is silenced and a second rearrangement occurs on the second copy of the chromosome. Subsequently, the α locus is rearranged (Petrie et al., 1993). During the α rearrangement, there is evidence that both chromosomes are rearranged simultaneously, with ~25% of cells having both α loci rearranged (Malissen et al., 1992). This is true at the genomic, transcriptional (Eltahla et al., 2016; Howie

et al., 2015) and at the phenotypic scale where T cells with two functional receptors have been observed (Heath et al., 1995; Padovan et al., 1993). Multiple rounds of α rearrangement can occur, replacing previous Vα-Jα rearrangements (Petrie et al., 1993; Wang et al., 1998). The relevance of double TCR cells in the physiopathological context is still debated (Balakrishnan and Morris, 2016; Muhowski and Rogers, 2023; Schuldt and Binstadt, 2019).

In some cases, the 2-step β chain rearrangement can become a 3-step mechanism, with evidence of TRBD1-TRBD2 prior to D-J rearrangements in mouse and humans (Hempel et al., 1998; Smirnova et al., 2023). This mechanism adds another N-addition, which contributes to increasing the diversity of the CDR3 region.

The lack of RAG1/2 enzymes leads to a complete absence of mature B and T cells, called the severe combined immunodeficiency (SCID) syndrome (Schwarz et al., 1996), fatal without treatment. RAG can recombine any gene that is flanked by this sequence making RSS sequences a powerful genetic tool. RSS-like or perfect matched sequences can be observed in other loci, and are called cryptic RSS. These "illegitimate" recombination sites were identified as linked to T cell acute lymphoid leukaemia and dysregulation of key leukaemia driver genes (Aplan et al., 1990).

The VDJ recombination as a pseudo-random process

V(D)J recombination is not a perfectly random process. The RAG complex binds to the consensus nonamer and heptamer of the 12- and 23RSS. These sequences are highly conserved, as mutations affects the probability of generating a hairpin, and ultimately form an exon with the corresponding gene (Hirokawa et al., 2020; Lee et al., 2003). Even though spacers are considered to be less important for the binding, their variability could result to several fold difference in recombination activity (Akira et al., 1987; Hirokawa et al., 2020; Larijani et al., 1999). This directly links low VDJ gene usage to RAG binding efficiency, even in cases of a single nucleotide mutation. TdT enzyme is also well known for being biased, and having a strong preference to adding G nucleotides (Alt and Baltimore, 1982; Gangi-Peterson et al., 1999). This leads to D regions to be preferentially enriched in G nucleotides.

Early observations showed skewed gene usage depending on gene distributions on the loci. A chain rearrangement typically biases towards gene segments located at the 5′ end of the Jα region, and Vα tends to favour the 3′ end of the Vα region (Pasqual et al., 2002; Thompson et al., 1990). Work from Park *et al*. has leveraged single-cell sequencing to link genetic rearrangements to genetic position on loci. They showed that VDJ gene usage was skewed

during pre-mature stages, and revealed that these biases could not be explained by RSS scoring alone. These results suggest the importance of a conformational looping structure of the TCR loci, already hinted by others (Carico et al., 2017; Hu et al., 2015; Park et al., 2020), favouring the pairing of proximal genes together, before any further selection.

Epigenetic factors also contribute to the gene usage. Enhancers positions on downstream 3' position of the TCR locus ensures the transcription of TCR genes (Bories et al., 1996; Bouvier et al., 1996), while dynamic methylation of histones ensures only specific portion of the genome are accessible at a given stage (Morshead et al., 2003). This also ensures that recombination first happens in a DJ, then V-DJ order. On B cells, it was shown that histone 4 acetylation correlated with V gene usage, while dimethylated H3/K9 correlated with poorly rearranged genes (Espinoza and Feeney, 2007).

Adopting a probabilistic approach, Murugan and colleagues modelled the probability of recombination events in non-selected T cell populations (Murugan et al., 2012). They showed how V-J combination does not correlate with linear location, and how D genes are constraint by J gene choice. However, analysis shows that V gene choice is independent of prior DJ rearrangement. Using Markov chains, they demonstrated how nucleotide insertion was not random and could be predicted by the previously inserted nucleotide on 5' strand. Similarly deletions were not stochastic, but rather influenced by the rearranged genes, as noted by others works previously (Gauss and Lieber, 1996).

Estimating the theoretical diversity

In 1988, solid estimates of the TCR theoretical diversity were made by immunologists about the three main mechanisms of diversity: i) the genetic VDJ segment diversity, ii) the random nucleotide addition and deletion at segment junctions and iii) the αβ chain pairing. Mark Davis and Pamela Bjorkman placed at around 10^{15} the number of possible TCR combinations that could be generated in humans, with similar calculations in mice placing it around 10^{20} (Lieber, 1991). These extravagant numbers, on par with astronomical distance scales, show how a small set of genes can lead to an enormous number of combinations. These numbers were not challenged for more than 30 years, until a team of physicists published an entropy-based model, placing the diversity of human TCR to be around 10^{61} possible rearrangements (Mora and Walczak, 2016), later revised to rather be around 10^{19} (Dupic et al., 2019). The last estimate emphasised on the existence of other factors influencing the seemingly random process of VDJ recombination (see above).

This chapter introduced T cell receptor generation as pseudo-random process. With such an immense possible number of TCRs, we can define the notion of repertoire; the collection of all possible TCRs that can be generated. We can further distinguish the *theoretical* repertoire, defined by Mora and Walczak, which encompasses all possibilities without considering impossible rearrangements or pseudogenes, from the *virtual* repertoire, which comprises all viable combinations, and the *available* repertoire, which is the collection of all TCRs in an individual. The theoretical repertoire is distinct from the other two, as it considers all possible combinations beforethymic selection. In the following sections, we'll see the principles of T cell antigen-specific recognition through the TCR and its importance during the thymic differentiation in driving clonal deletion.

1.1.3 Modalities of MHC recognition

Scientists in the 1950's were trying to understand the principles of tissue rejection in transplantation. The discovery of a set of proteins that genetically determined structures on the cell surface that regulate immunological reactions, and the associated genetic mapping awarded them the Nobel prize in 1980. The study of MHC recognition by the TCR in 1974 rewarded Rolf Zinkernagel and Peter Doherty the Nobel prize in 1996 (R. Zinkernagel and Doherty, 1974; R. M. Zinkernagel and Doherty, 1974). The major histocompatibility complex, or human leukocyte antigen (HLA) in humans, is a polygenic and highly polymorphic set of genes that are at the centre of self to non-self-recognition. MHC molecules can be distinguished into two major type of genes, class I (MHC-I) and class II (MHC-II) MHC, with distinct expression patterns in cells. A third group of MHC molecule will not be covered here as it is not involved in antigen presenting and recognition. Likely due to evolution stress, the MHC complex contains more than 200 genes, thousands of alleles, spans 4 megabases and more than 20 different loci (The MHC sequencing consortium, 1999).

1.1.3.1 Processing of antigens

MHC-I are expressed by all nucleated cells of the body and platelets, although some cells are notoriously known to have low expression of it. The first MHC-I crystal structures (Bjorkman et al., 1987; Saper et al., 1991) revealed a heterodimer comprised of three α domains (α_1 to α_3) and an invariant chain called the β2-microglobulin. The α_1 and α_2 form a groove, spacious enough to accommodate a small peptide - the antigen. MHC-I presents peptides from the intracellular components, such as self-molecules or altered-self (viruses, cancers). Briefly,

antigens are processed by the proteasome, a complex expressed in all cells responsible for the pre-processing of peptides before their loading onto the MHC. Proteins tagged for ubiquitination (Wei et al., 2017) are sampled by the proteasome, where various families of proteolytic molecules will prepare peptides. Peptides are then trafficked outside of the endoplasmic reticulum and loaded onto the MHC by the transporter associated complex (TAP) (Ortmann et al., 1994). If required, endoplasmic reticulum associated amino peptidase (ERAAP) can further trim the peptide to accommodate the binding to the MHC-I (Serwold et al., 2002). Final peptides are typically in the 8 to 10 amino-acid range, although the complexity of non-covalent interactions and hundreds of polymorphisms allow exceptions.

On the other hand, the MHC-II is expressed by a restricted subset of cells, coined professional antigen presenting cells, or APC. Among them, we can find dendritic cells, naïve B cells, and activated macrophages. This limited expression is linked to the source of peptides, as MHC-II antigens are not internally produced, but rather captured through various mechanisms such as endocytosis or phagocytosis. The processing pathway is also distinct from the MHC-I. Native extracellular proteins are directed towards endocytosomes, acidic and reducing compartments that facilitate their degradation. MHC-II is assembled in the endoplasmic reticulum and transported to endosomes, where the suitable peptides are loaded onto the MHC-II and then directed to the cell membrane. Final antigens presented are typically in the 13-17 amino-acid range although, far greater were observed (Chicz et al., 1992).

A dedicated process called cross-presentation allows the presentation of extracellular antigens, typically from MHC-II pathway, to be loaded and presented by MHC-I. This process happens in healthy conditions, notably through autophagy (Dengjel et al., 2005). Recently, Barbet *et al*. uncovered a novel function of how TAP blockade by viral infection was bypassed by non-canonical presentation, allowing an efficient CD8 T cell priming and clearance of the infection (Barbet et al., 2021). Thus, cross-presentation elicits balanced $CD8^+$ and $CD4^+$ responses.

1.1.3.2 TCR, peptide and MHC

Typical cells express tens of thousands MHC-I molecules simultaneously (Kowalewski et al., 2015), with constant renewal of antigen presentations. TCR-pMHC recognition works in a multiple step program, where the T cells "scan" the pMHC complex with its TCR. In the canonical model, the docking between TCR and pMHC first needs a stabilisation step, independent of the peptide, to allow the docking of the TCR and MHC. CDR1 and CDR2 loops permit the MHC scaffolding to bind to the TCR with hydrogen bounds and van der Waals interactions (Wu et al., 2002). The first MHC recognition step is then followed by the "peptide – scanning" phase. The affinity for the peptide is linked with stabilisation energy, where higher stability leads to prolonged contact. The whole process is aided by co-receptors CD4 and CD8

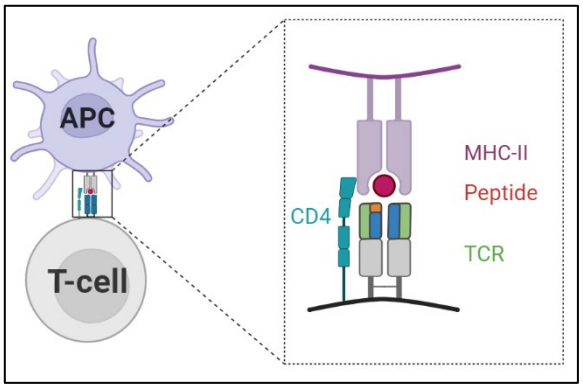


Figure 4: (left) A schematic representation of an αβ TCR expressed on the surface of a CD4⁺ cytotoxic T cell, supported by a CD4 coreceptor, recognises an antigenic peptide presented by an MHC class II molecule at the surface of an antigen presenting cell.

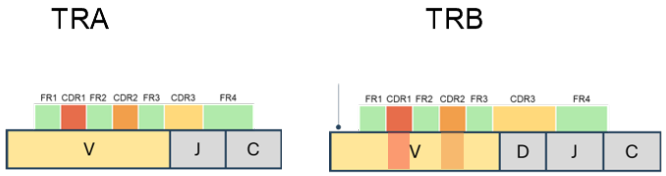
and is highly dynamic (Figure 4). While CDR1 and CDR2 are considered static, CDR3 offer the largest range of motion, and large conformational adjustment are observed between the bound and unbound states (Garcia et al., 1998). These changes are linked to recognition degeneracy, as a single TCR can accommodate multiple peptide bindings, sometimes with very different properties. The angle of recognition can also be altered during docking, with reported "tilt" of TCR angle to accommodate pMHC binding (Pierce and Weng, 2013).

MHC presentation is also dependant on the VJ genes usage.

MHC can influence the V(D)J usage. Indeed, V and J genes are the main contributors to the TCR-pMHC complex stabilisation (Garcia et al., 2009; Sim et al., 1996; Zerrahn et al., 1997). In the immunodominant public TCR to CMV peptide, isoleucine$_{54}$ located in the CDR2β region of the TRBV6-5 is implied in 7 van der Walls contacts with the HLA, the tyrosine$_{31}$of CDR1α in TRAV49 is implicated in 25 contacts with the peptide. This outnumbering of contacts by CDR1α and CDR2β exceeds the number from CDR3α and CDR3β (Gras et al., 2009). CDR1 and CDR2 are germline-encoded in the V region, while we have shown how CDR3 diversity was generated through specific mechanisms (see 1.1.2.3). The higher

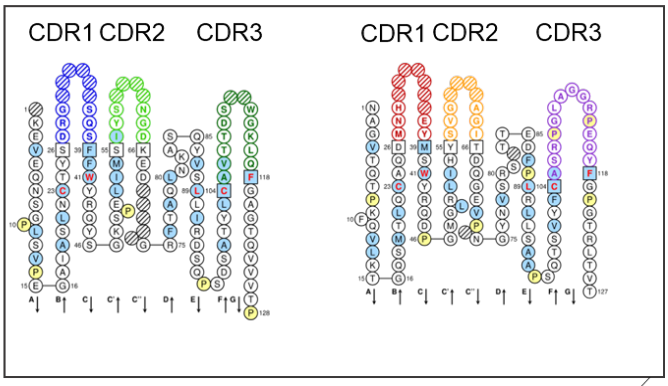
Gene structure

Gene structure of α and β chains. CDR1 and 2 are located on the V gene, while CDR3 is sitting between the V(D)J portion.



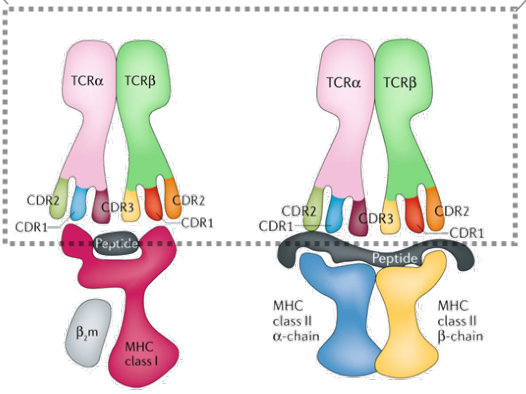
IMGT collier de perle

IMGT Collier de Perles visualisation of the complete α and β chains of AT6 clone. CDR1, CDR2 and CDR3 are located on the outside part of the secondary structure and form distinct loops, allowing contact with external structures.



Integrated schematic structure with MHC

Integrated schematic structure of the TCR-pMHC binding. TCR α and β chain are in close interaction with the pMHC complex. MHC-I (left) peptides are smaller.



3-dimensionnal structure of TCR-pMHC docking

Structural representation of the MHC-peptide complex, with the co-receptor CD4 (left) or CD8 (right). Peptide, in red, in presented by the APC on the MHC, forming a close synapse with the TCR. The co-receptor stabilizes the docking. Adapted from David Goodsell, 2005.

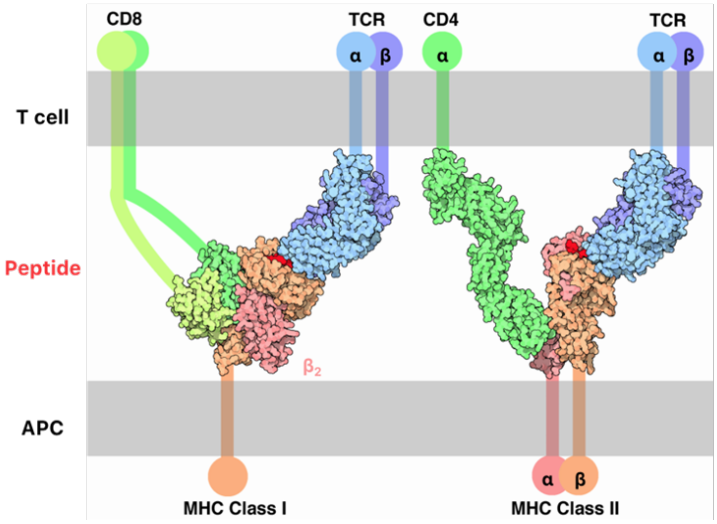


Figure 5: From primary structure to 3 dimensional structure of docking.

contribution of germline-encoded regions to stabilise the TCR-pMHC was later studied as the "TCR germline bias". (Feng et al., 2007; Garcia et al., 2009). this advocates for a co-evolution between TCR and MHC (Blackman et al., 1986; Lu et al., 2019). The complete intricacies of VDJ recognition,

from the primary structure of TCR chains to the 3 dimensionnal docking is represented in Figure 5.

1.1.3.3 Cheating recognition with superantigens

A non classical way of recognition of antigens by TCR is independent of the peptide presentation by the MHC. The term was proposed by Marrack and colleagues (White et al., 1989), after they showed that mice administered with staphylococcal enterotoxin B (SEB) would experience the activation of all T cells, both mature and immature, bearing Vβ3 and Vβ8 gene families.

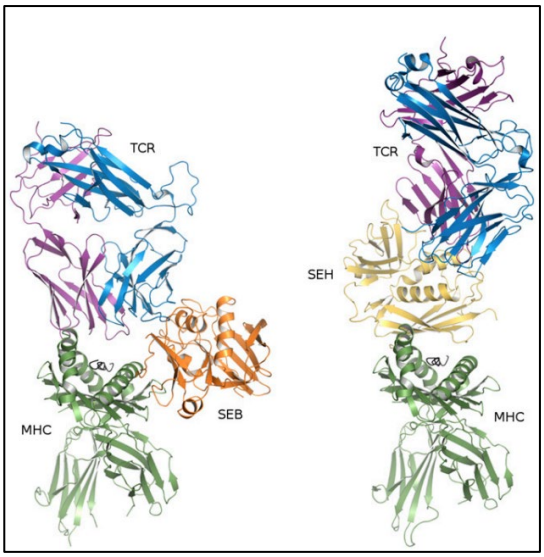


Figure 6: Example of MHC-independent binding of Staphiloccocus endotoxins B and H (SEB, SEH) binding to TCRV beta region. Comparison of the ternary TCR–superantigen–MHC complexes with SEB and SEH. All proteins are shown in ribbon representation. SEB (left panel) is shown in orange and SEH (right panel) in yellow, the TCR a-chain in purple, the TCR b-chain in blue, and MHC class II in green. From Rödström, 2014

Superantigen were demonstrated to be linked to MHC-II (Mehindate et al., 1995). For the case of Staphylococcal enterotoxin A (SEA), it would first bind to the MHC-II, then bind to Vβ regions. This would results in broad, polyclonal activation of T cells bearing those Vβs, of the order of 5 to 20% of T cell population. In humans, this non-specific activation of T cells result in a toxic shock syndrome (TSS), caused by endotoxins produced notably from *Streptococcus pyogenes,* S*taphylococcus aureus (SA)*, and others (Marrack and Kappler, 1990). When administered to an immature population, SA induces complete depletion of target T cells bearing Vβ in murine models; whereas on mature population it triggers a cytokine storm and T cells unresponsiveness (Jenkinson et al., 1990; Kappler et al., 1987; Murphy et al., 1990).

1.1.4 Autoimmunity, a necessary evil

The Nobel recipient and father of Immunology, Paul Erlich (Nobel prize, 1908) postulated the "horror autotoxicus", a theory suggesting that the human body was incapable of mounting an immune response against itself. MHC was identified early on as the molecule binding to the TCR and presenting the peptide by their hypervariable regions (Chothia et al., 1988; Davis and Bjorkman, 1988). However, given the very high number of possible T cell structures, some of the generated T would inevitably react to self-peptides. Indeed, mechanisms of tolerance exist in the immune system to avoid such a tragic fate. Kappler and colleagues demonstrated that immature, autoimmune thymocytes would be deleted in the thymus during their maturation process (Huseby et al., 2003; Kappler et al., 1987). I will detail now how the selection processes in the thymus limits the egress of autoreactive clones, and the mechanisms involved in the establishment of tolerance.

1.1.4.1 The thymus, deadly school for T cells

Every day, millions of thymocytes migrate to the thymus, but only a small fraction of them actually survive its supposedly strict selection. Both experimental (Egerton et al., 1990) and mathematical approaches (Thomas-Vaslin et al., 2008) estimate that around 95% of thymocytes die before reaching maturity. This figure is significantly higher than the theoretical prediction of 70% producing a non-productive TCR (see 1.1.2.3). Indeed, T cells undergo two highly selective rounds of selection that profoundly shape the available repertoire of TCR as they leave the thymus and enter the periphery. Predicted by Burnett in 1959 (Nobel Prize 1960) (Burnet, 1959), this process known as the thymic positive and negative selection, addresses the need to select T cells that can interact with an MHC molecule, but also do not react to self-peptides. Indeed, an effective selection would only allow cells that can recognise the self (self MHC molecules) but also distinguish foreign peptides, preventing auto-reactive disorders and therefore supporting tissue and organism integrity.

The thymus located in the anterior superior mediastinum is composed of two pyramidal lobes, each comprising of a central medulla and an outer cortex. Immature thymocytes produced in the bone marrow enter the thymus through the blood circulation at the corticomedullary junction, before joining the cortex region. Thymocytes are a heterogeneous population, where their phenotype characterises the steps of their maturation. When they enter the thymus, they do not express the lineage marker of mature T cells, CD4 and CD8, and are called double

negative (DN). DN will undergo the rearrangement of their TCRβ locus, during four stages of maturation, ranging from DN1 to DN4. Thymocytes will express both CD4 and CD8 after successfully rearranging their β chain. These $CD4^+$ $CD8^+$ double positive cells (DP) migrate to the thymic cortex and rearrange their TCRα locus. DP cells with a functional αβ TCR will be challenged by central thymic epithelial cells (cTec) with self-peptide-MHC complexes. DP unable to recognise a pMHC will undergo apoptosis, also called death by neglect (Bertolino et al., 1999). T cells that react with a high affinity to pMHC will also go through apoptosis, a process called clonal deletion. This step, called positive selection, is also the first one to initiate the lineage commitment to CD4 or CD8 single positive (SP) cells. SP cells then migrate to the medulla where they will be challenged by medullary TEC (mTEC) to recognise self-peptide-MHC complexes. T cell bearing a TCR with a high affinity for self will be clonally deleted in a similar manner as during positive selection. In the $CD4^+$ thymocytes however, T cells with an intermediate/high affinity for self-peptide will be selected and enter in a distinct, specific program and commit to the $CD4^+$ T_{REG} cell subset (Coutinho, 2005) . After 4 days of medullary challenge (McCaughtry et al., 2007), SP thymocytes will express the sphingosine-1-phosphate receptor (S1PR1) and CD62L (Carlson et al., 2006), which will allow their egress to the blood vessel and their new journey into the periphery and be locked into a quiescent, G_0 state (Zhang et al., 2018) until their reactivation.

This schematic view of the thymus highlights its role as a very strict selective school for T cells. The multiple rounds of selection are here to ensure that all T cells can interact with self-MHC, but not necessarily get activated to self-peptides, unless for a restricted subtype of regulatory T cells. From this step, emerges the concept of tolerance, or how the immune system controls the unavoidable autoimmunity imposed by the diversity of TCR, MHC, and peptides.

1.1.4.2 Central tolerance and its imperfections

During their generation in the thymus, T cells arising with reactivity to self-antigens are deleted. This process prevents autoimmunity and has been coined central tolerance. We have mentioned in 1.1.4.1 how clonal deletion ensured that TCR with high affinity would undergo apoptosis. I will detail here how TECs are the key actors of central tolerance.

Cortical TECs are responsible for positive selection, ensuring TCRs can recognise the MHC. DP thymocytes survive the positive selection by repetitive engagement of their newly rearranged TCR (Sakaguchi et al., 2003). TCR engagement restores the transcriptional activity

of those thymocytes, while those cells dying by neglect would engage in caspase-3 mediated apoptosis (Mingueneau et al., 2013). Evidence exists for clonal deletion at this stage but it accounts for a small part of the total(McCaughtry et al., 2008).

Medullary TEC are crucial for the negative selection. mTEC express the autoimmune regulator (*AIRE*) gene. The AIRE gene (Autoimmune Regulator) is critically important for the development of central immune tolerance in the thymus. Indeed, thymic selection would require thymocyte export the periphery for them to encounter MHC bound to tissue-specific antigen (TSA) and comprehensively sample the self-peptidome. This process was not retained during evolution, probably because it would allow autoreactive antigens to roam freely, instead of having them trapped and centralised in an organ. Instead, evolution selected this unique mechanism of ectopic expression of self-antigens (Saltis et al., 2008). Chromatin analysis revealed how AIRE regulates ectopic gene expression through histone modifications and activates the RNA polymerase II recruitment (Kumar et al., 2001; Org et al., 2009), and its transcriptional targets could be modified by its microenvironment (Guerau-de-Arellano et al., 2008). AIRE-expressing mTEC display a high renewal and form a sponge-like network in the medulla, optimised with their presentation functions (Gray et al., 2007). Each mTEC shows a high single-cell heterogeneity. Genes are expressed in modules in a semi-random fashion (Brennecke et al., 2015; Dhalla et al., 2020; Meredith et al., 2015) and the proteome they express can change during their lifespan (Pinto et al., 2013; Tykocinski et al., 2010). As such, AIRE-expressed genes seem to work in small clusters of co-expression. Authors have hypothesised that it could be related to organ-expression patterns, and mTEC would be specialised cell types from deep tissues, however this was not supported by confronting transcriptomes of mTEC to known cell atlases (Meredith et al., 2015).

mTEC also guide the fate of regulatory T cells, crucial actors of peripheral tolerance. AIRE-deficient mice are unable to redirect T_{REG} biased clones to a regulatory phenotype (Malchow et al., 2016, 2013), which will harbor a typical $CD4^+$ effector phenotype, leading to systemic autoimmune disorders.

Altogether, TEC are professional self-peptides presenting factories, either through i) optimal peptide generation or ii) efficient self-proteome sampling and presentation to thymocytes. They efficiently suppress autoreactive TCRs bearing cells with high affinity contact to MHCs or self-peptides antigens by triggering apoptosis on immature thymocytes.

1.1.4.3 The issue of leaky central tolerance

It was quickly observed that the thymic quality control is not perfect and autoreactive clones manage to escape to the periphery (Gammon and Sercarz, 1989). This observation was expected given the prevalence of autoimmune diseases in the population.

Eventually, negative selection cannot stop all autoreactive clones, as it may only spot the most reactive ones, but let the low avidity ones escape (Zehn and Bevan, 2006). Don Mason (Mason, 1998) estimated there are more than 10^{17} "presentable" peptides on a MHC-II. If the number of genes expressed in a tissue is around 11 000, AIRE allows the ectopic expression of more than 19 000 genes out of the >25 000 coding genes in the genome, but only a fraction at a time (Meredith et al., 2015; Ramsköld et al., 2009; Sansom et al., 2014). The difference in these numbers suggests that i) AIRE cannot express all the self-peptides, ii) each thymocyte is unlikely to encounter every peptide during its 4 days span and be actively deleted.

1.1.4.4 Peripheral tolerance

At 20 years old, there are more than 300 million mature T cells produced from the thymus every day (Bains et al., 2009). To account for the likely massive flow of autoreactive clones egressing from the thymus, a second type of tolerance has been selected during evolution and take place in the peripheral organs. This highly distributed and efficient immune surveillance system ensures that those uninvited guests do not cause uncontrolled havoc. Peripheral tolerance relies on 5 main mechanisms: quiescence, ignorance, anergy, exhaustion, and peripheral induction of T_{REG} (reviewed by ElTanbouly and Noelle, 2021) happening at the different stages of T cell life (Figure 7).

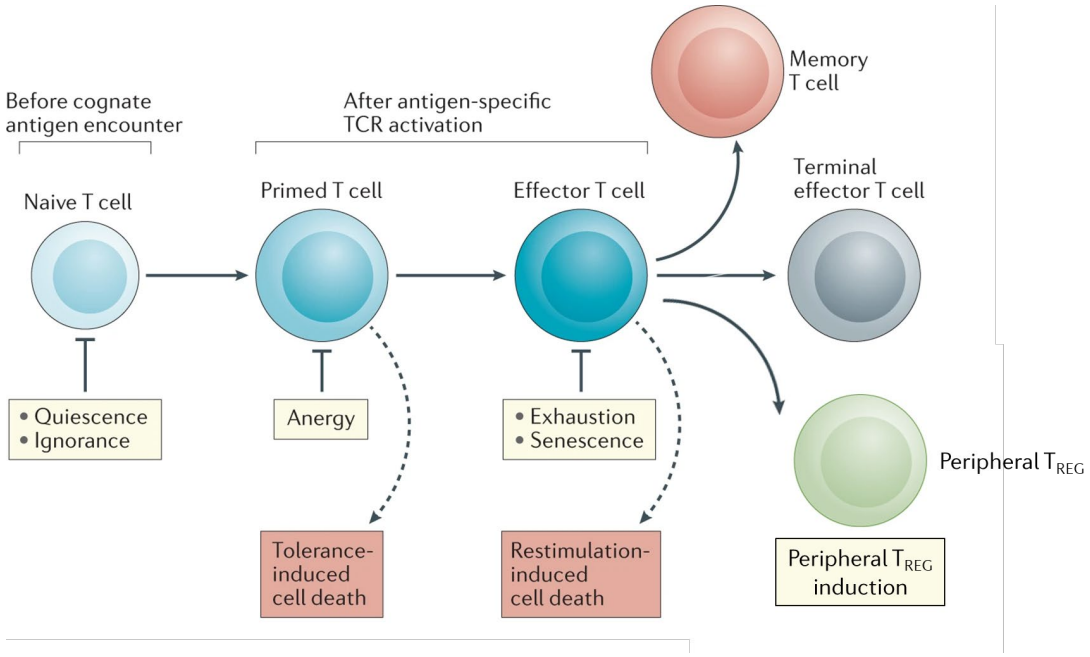


Figure 7: ***Temporal schematic integrating the tolerance checkpoints at each stage of the peripheral T cell lifespan.*** *Six tolerance checkpoints exist and integrate to regulate T cell responses at all stages. These T cell regulatory checkpoints start at the naive T cell stage, where quiescence and ignorance enforce T cell tolerance. These checkpoints occur before T cell activation by cognate antigen encounter and priming. After antigen-specific T cell activation, co-stimulation-deficient T cell receptor (TCR) signalling can trigger anergy, which enforces T cell hyporesponsiveness and limits T cell responses to inappropriate stimuli (such as self-antigens). Such tolerogenic activation can also induce peripheral T cell deletion, known as tolerance-induced cell death. In periphery, effector T cells can be converted to peripheral T_{REG} under specific signals. Adapter from ElTanbouly and Noelle (2021)*

Quiescence: Recent thymic emigrant T cells (RTE) have been shown to be quiescent, a state that is tightly regulated by overlapping genetic pathways (Zhang et al., 2018). Such programs set a threshold for activation of naïve T cells, meaning that only those that will receive a sufficient signal through their TCR will effectively leave this resting state. Therefore, quiescence prevents T-cell spontaneous activation. Disruption of the genetic pathways involved in quiescence such as transcription factors Runx1 or c-Rel leads to autoimmune diseases in mice models (Chang et al., 2011; Wong et al., 2012).

Ignorance : Transgenic murine models expressing a viral protein failed to elicit an autoimmune response (Jolicoeur et al., 1994). However, tolerance was removed if mice were then inoculated with the virus and T cells would eventually destroy β-islets (Ohashi et al., 1991; Oldstone et al., 1991). These early experiments showed how the immune system would ignore antigens not filtered by central tolerance, Ignorance is the absence of immune response of T cells against an antigen that would normally, given a specific context such as inflammation, trigger their activation. It does not affect their capacity to activate, would it be in the presence of the correct co-stimulation signals or cytokines (Cao et al., 2015; Heath et al., 1992; Mamula et al., 1992; Zehn and Bevan, 2006).

Anergy : Distinct from ignorance and quiescence, anergy is the induced unresponsiveness of a T cell after the delivery of a strong TCR signal mediated by other cells (Jenkins and Schwartz, 1987; Schwartz, 2003). This state is reversible in the presence of IL-2 (Beverly et al., 1992; Essery et al., 1988), but also integrates many other signal with other co-stimulatory molecules such as CD28, CTLA-4, or Ox40 (Krummel, 1995; Lathrop et al., 2004).

Exhaustion: Extensively covered because of its major role in tumor escape, T cell exhaustion does not play a role in pathogenesis of autoimmune diseases. Rather, exhaustion is used as a marker of T cell activation in the immune response in systemic lupus erythematous (SLE), type 1 diabetes (T1D) or multiple sclerosis (MS) (Lima et al., 2021; Pender et al., 2017; Wiedeman et al., 2020). Occurring after their continuous activation, T cell will gradually lose their effector functions and enter in an exhausted state. Often seen through the prism of cancer, and thus detrimental, exhaustion has been demonstrated beneficial for maintaining tissue integrity in case of prolonged inflammation.

Peripheral T_{REG} induction: mechanisms uncovered was the extra-thymic generation of T_{REG}. CD4 T cells display a phenomenal plasticity, and can transdifferentiate into many subsets (see 1.1.1). *De novo* generation of T_{REG} happens in the periphery, where conventional T cells will acquire a partial T_{REG} transcriptional and functional signature (Feuerer et al., 2010; Kretschmer et al., 2005). Albeit not as fully suppressive as their native, thymic counterparts, induced T_{REG} are nevertheless critical for tissue homeostasis, notably gut (Lathrop et al., 2008), as well as several allergic (Curotto de Lafaille et al., 2008; Josefowicz et al., 2012) and autoimmune context (Huter et al., 2008; Lombardi et al., 2012; Mottet et al., 2003).

Understanding T cell control of autoimmunity in the periphery provides a foundational insight into the intricate mechanisms that regulate our immune system's responses. This

knowledge not only sheds light on the prevention and management of various autoimmune disorders but also carries significant implications for the field of cardiovascular health. By exploring how T cell dysregulation can extend its influence into the cardiovascular system, we can uncover the intricate interplay between immunity and heart health.

1.2 CIRCULATING T CELLS IN CARDIOVASCULAR DISEASES

1.2.1 Current state of cardiovascular diseases

Cardiovascular diseases (CVD) is a general term that encompass a wide range of circulatory or cardiac diseases with very different aetiologies. The European Society of Cardiology (ESC) conducts an annual survey to describe the evolution of CVD landscape in all its 56 member countries. According to the 2022 survey from the European organisation, CVD is the leading cause of death in Europe (Timmis et al., 2018). With more than 4 million deaths in 2019 (2.2 million death in females and 1.9 million in males), it is far more common than cancer (2 million death reported). Among the different CVDs, ischemic heart diseases are the major cause of mortality represented, accounting for 45% of these deaths in females and 39% in males. This poses a major financial burden on economies with an estimated cost of 210 billion euros per year in 2015, with more than half of it (110 billion, 53%) due to healthcare costs. Inequalities between higher and lower income countries participate to further increase the burden of lower income countries, where higher income countries have a better access to high quality care with fewer patients.

An abundant literature review from clinical trials and epidemiologic studies have concluded that CVDs were largely preventable, especially for younger people. Indeed, despite rare genetic factors, most CVD risk factors are non-heritable. They include poor lifestyle habits, such as dietary risk factors, low physical activity, smoking, high systolic blood pressure, high cholesterol, or high body mass index (BMI) (Wilkins et al., 2017). Alcohol consumption has been reported as not impacting, or even associated to protective effect in some countries, however it is not recommended (Suliga et al., 2019).

Ischemic heart diseases are heritable. These complex diseases are not monogenetic, but rather very polygenic. The 1000 genome project (McVean et al., 2012), massive sequencing of 14 populations and 1092 complete human genomes, set the ground for genome-wide association studies (GWAS). The CARDIoGRAMplusC4D Consortium identified 44 loci across the genome linked to coronary artery disease, where mutations in these loci were found associated to vessels walls, lipid metabolism or inflammation (Nikpay et al., 2015). Similar studies were conducted for ischemic stroke (Malik et al., 2018) or peripheral artery disease (Klarin et al., 2019). They highlighted the complexity of circulatory diseases, where the inheritance of genetic factors is aggravated in presence of behavioural factors. In peripheral artery disease (PAD), tobacco consumption has been directly linked to thrombotic sequelae (Holst et al., 2010), and

carriers of factor V coagulation variant (F5 p.R506Q) had greater effects of PAD on smokers than non-smokers. This is even more conspicuous when seemingly innocent habits can have long time impacts patients' health. It was demonstrated how nuts consumption had a beneficial effect on CVDs, where walnut intake was associated to 17% reduce risk of strokes, and global intake of nuts is beneficial for individuals in reducing CVD (Guasch-Ferré et al., 2017).

1.2.2 Immunological basis of CVDs: example of myocardial infarction

The CVD landscape is very large and encompasses many aetiologies. Here, I will focus on how the immune system, and especially T cells, are involved in acute myocardial infarction as an example, before diving into the other CVDs. I will detail here what are the origins of this disease, and what are the immunological mechanisms for its onset and resolution, and how we can extrapolate these findings to other diseases.

1.2.2.1 Myocardial infarction

Acute myocardial infarction is a circulatory incident happening after one or more of the coronary arteries stop perfusing the heart. Blood flow interruption is often caused by the obstruction of the coronary arteries, followed by blood clot or rupture of atherosclerosis plaques, the fatty deposits that build up on vessel walls and can break off suddenly.

During infarction, oxygenated blood does not perfuse the heart anymore of a prolonged period of time. Affected cardiac cells die *en masse* under these hypoxic conditions, around a damaged darker zone of necrotic cells called the infarcted area. Size of initial infarct varies greatly between patients, and impacts further recovery.

The heart however, is a unique tissue. In adult mammals, myocardium displays poor regeneration functions. The cardiomyocytes, crucial for the pumping function, are non-mitotic after birth, which explains how rare primary cardiac tumours are. Estimates from ^{14}C tracking showed how only 0.5 to 1% of cardiomyocytes are regenerated a year (Bergmann et al., 2009). Due to this limited proliferation, injuries are not healed through regeneration but with scarring. The process is far less efficient than restoring the initial cells, as scarring involves fibrosis, a collection of cross-linked collagen fibres. This makeshift repair offers diminished cardiac contractility, and ultimately leads to complications in surviving patients (Bayat et al., 2022).

A massive inflammation arises from the massive, sudden death of cardiomyocytes, fibroblasts and other cells from the myocardium. Indeed, hypoxia-mediated necrosis releases

many pro inflammatory factors, triggering the Toll-like receptors of surrounding cells and switching them toward a pro-inflammatory state (Chong et al., 2004; Fu et al., 2018; Heidt et al., 2014). This early response is necessary for the recruitment of the innate immune system and the rapid clearance of debris by macrophages. Dying cells will secrete and activate matrix metalloproteinases (MMPs). These proteinases play a role in breaking down cellular and matrix components, assisting phagocytic macrophages in the removal of necrotic tissue (Spinale, 2002). This highly inflammatory environment activates monocytes that differentiates into pro-inflammatory macrophages, which further fuels inflammation for several days. Cell recruitment is at peak during days 3 to 7, where the myocardium infiltrate slowly switches towards pro-repair cells, such as tissue macrophages or T_{REG}. (Yan et al., 2013).

1.2.3 An ambiguous role for T cells in CVD, example in myocardial infarction

Despite research heavily focusing on myeloid populations such as neutrophils or monocytes, T cells are key actors in myocardial infarction. T cells are present in relatively low numbers compared to other immune cell types in the myocardium, even after myocardial damage (Yan et al., 2013). Their contribution is however crucial, as abundance does not correlate with function here, as they will rather interact and shape surrounding cells, rather than exert direct functions on the tissue (reviewed by Hofmann and Frantz, 2015). One hallmark of lymphocytes implication here is the expansion of cardiac draining lymph nodes. In multiple models, myocardial injury leads to enlargement of mediastinal lymph nodes, whether it is murine (Ramos et al., 2017), rats (Ramos et al., 2012), or humans (Rieckmann et al., 2019; Sintou et al., 2020).

T cells are observed in the infarcted area minutes after restoration of blood flow, likely brought from the circulating population (Hoffmann et al., 2012; Yang et al., 2006). Their migration toward the injured organ is helped with the upregulation of adhesion molecules, notably ICAM-1, by local myocardial cells caused by inflammation (Devaux et al., 1997). T cells will localise around vessels and extravascular space; In a murine model of transversal aorta constriction (TAC), global T cell depletion by TCRα knock-out (KO) mice did not develop heart failure compared to sham controls. Similar results were obtained independently by depleted B and T cells. In RAG2-KO mice, depletion of adaptive immune populations prevented cardiac dilation and attenuated cardiac dysfunction at 6 weeks post intervention (Laroumanie et al., 2014). T cell subsets contribution to the immune response is still debated among T subsets.

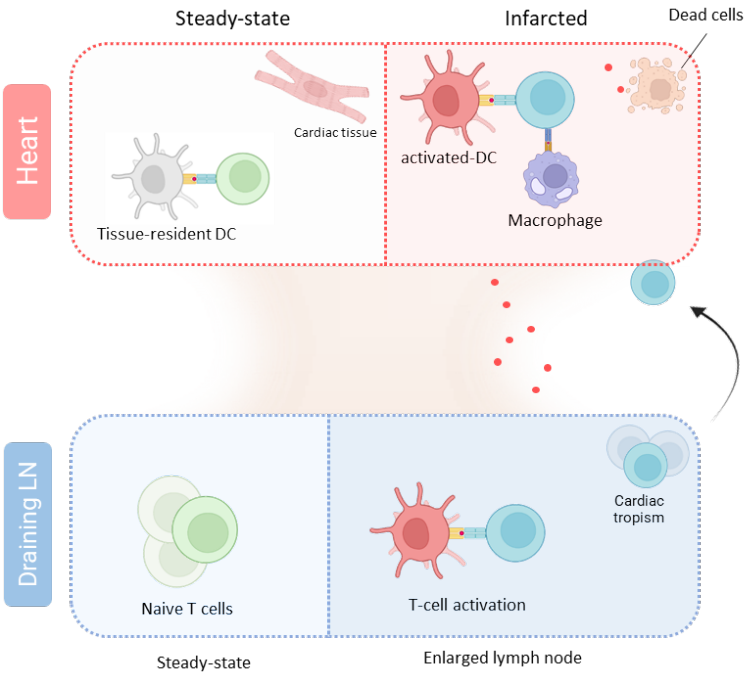


Figure 8: Interplay between the cardiac and draining lymph nodes compartments. At the steady state, cardiac tissue and infiltrating macrophages present self-antigens to naïve T cells, contributing to peripheral tolerance. During infarction (right), self-antigens and inflammatory signals are released. It activates dendritic cells and macrophages, which in turn can present cardiac antigens to infiltrating T cells. Antigen can also be transported to lymph nodes where they can be presented to T cells. Activated T cells can then move to the infarcted zone by chemotaxis.

1.2.3.1 CD4⁺ effector T cells

Investigating the role of $CD4^+$ T cells in myocardial infarction was done by several groups. Using MHC-II KO mice, Stefan Frantz's group showed that absence of $CD4^+$ T cells leads to increased mortality in mice at 56 days (Hofmann et al., 2012). The poorer cardiac repair in CD4 depleted mice was linked with an increased pro-inflammatory monocyte infiltration, suggesting a beneficial role for CD4 T cells in modulating the innate response. Two other experimental studies on the same model found a detrimental role of T cells after their depletion, showing that $CD4^+$ T cells in TAC model was associated with adverse cardiac remodelling, increased fibrosis, increased macrophage infiltration and worsening of cardiac contractility (Laroumanie et al., 2014; Nevers et al., 2015). This opposition between positive and detrimental inflammation can be attributed to the different dynamics of the chronic (TAC) or acute (MI) damage. Nevertheless, these results confirm the highly crucial role of $CD4^+$ T cells in the myocardial immune response.

Authors suggested distinct roles of T cell subsets, notably Th1, Th2 and Tregs, which may explain the contradicting results. Th1 subset has been involved in atherosclerosis, an inflammatory driven CVD, where they have been shown to secrete pro-inflammatory IFN-γ

and fuel tissue damage (Frostegård et al., 1999). In patients with inflamed hearts, IFN-γ levels are increased and Th1 imbalance is associated with adverse cardiac remodelling and poorer prognostic (Cheng et al., 2005; Fukunaga et al., 2007). IFN-γ-related deleterious effect of Th1 has been confirmed in murine TAC models, where they were shown to transition myofibroblast towards a pro-fibrotic profile (Nevers et al., 2017). IFN-γ secretion also drives the polarisation of macrophages towards a pro-inflammatory profile (Sica and Mantovani, 2012). Overabundance of inflammatory macrophages, or lack of switch towards regulatory profile, is associated with cardiac dysfunction (Shaojun Liu et al., 2020).

Th2 cells have classically been reported as beneficial for cardiac repair. While no specific mechanism has been demonstrated, the results focus on the dampening of negative outcomes induced by the polarisation towards a pro-inflammatory Th1 response. Of note, one study found that Th2 response could be considered as pro-inflammatory, and failing murine hearts induced local damage from Th2 cytokines (Bansal et al., 2017).

Pro-inflammatory Th17 cells have been described with ambivalent role in atherosclerosis (Taleb et al., 2015), or promoting heart failure in human myocarditis (Myers et al., 2016). Expanded after MI in the peripheral blood, with all other T cell subsets (Bansal et al., 2017), Th17 also infiltrate the infarcted tissue after their activation. Interestingly, IL-17A, Th17 signature cytokine, has been associated with being proapoptotic in cardiomyocytes, profibrotic in cardiac fibroblast, and proinflammatory in macrophages *in vitro* (Yan et al., 2012). Authors noted however that 90% of IL-17A producing cells are γδ $CD4^-$ T cells, suggesting a modest contribution of Th17. A rare study in neonatal mice investigated how Th17 cells prevented cardiomyocytes proliferation *in vitro* (Li et al., 2020).

1.2.3.2 Regulatory T cells

Although T effector role is subject to debate, there seem to be a scientific consensus on the contribution of TREG. Abundantly present in circulating blood and as tissue-residents, they display specific phenotype and functions depending on the target organ, whether it is in muscle (Burzyn et al., 2013), aorta (Li et al., 2022), heart (Xia et al., 2020), their unique properties confer a protective effect on tissues. TREG play a central role in myocardial repair, as they regulate inflammation, provide cytokines for phenotype switch of macrophages and dendritic cell, and induce neo-angiogenesis. Depletion of TREG has been extensively done and reported. Frangogiannis group reported a modest but positive anti-inflammatory role (Saxena et al., 2014)

on fibroblast by secreting immunomodulatory proteins IL-10 and transforming growth factor β (TGF-β). T_{REG} also exert paracrine effects on cardiomyocytes through the secretion of 6 other molecules, promoting their proliferation *in vivo* (Zacchigna et al., 2018). Non-specific activation of T_{REG} following infarction showed improved healing and survival in mice compared to sham, where T_{REG} fostered M2 polarisation of macrophages population and myofibroblast activation (Weirather et al., 2014). Confirming the beneficial role of activated T_{REG}, the peripheral induction of T_{REG} (iT_{REG}) by pro-tolerogenic dendritic cells or cardiac progenitor cells has been reported as beneficial for cardiac repair (Choo et al., 2017; Mishra et al., 2022). In line with their diverse arsenal, T_{REG} phenotype characterisation showed all cells do not contribute equally, as $CD69^+$ were positively correlated with patient outcome by dampening IL-17 pro-inflammatory role (Blanco-Domínguez et al., 2022). Investigating specificity, Delgobo and colleagues showed how the injection of monoclonal T_{REG} specific to α-myosin lead to rapidly blunted inflammation, improved fibrotic repair and re-vascularisation (Delgobo et al., 2023).

An ongoing, promising, clinical trial has been started, harnessing the positive impact of T_{REG} in myocardial repair. The Low dose interleukin-2 in patients with stable ischaemic heart disease and acute coronary syndrome (LILACS) trial is using the autologous expansion of T_{REG} to patients with stable ischaemic heart disease by providing low-dose IL-2 (Zhao et al., 2020).

1.2.3.3 $CD8^+$ T cells

$CD8^+$ T cells might be the most intriguing subset today. They were reported early on as activated after MI and able to kill cardiomyocytes *in vitro*. Even if present in large numbers after the infarct, they were initially reported as bystanders, without much importance. Indeed, studies have shown that depletion of CD8 T cells had no impact on cardiac functions or mice survival in various cardiovascular contexts (Elhage et al., 2004; Laroumanie et al., 2014; Li et al., 2020). Recently, a new interest has been found in dissecting $CD8^+$ T cells role in cardiac diseases. A thorough, multi-model study demonstrated however how CD8 T cells drive adverse post-ischemic cardiac remodelling (Santos-Zas et al., 2021). Ait-Fella and colleagues experimentally determined that activated $CD8^+$ T cells infiltrated the myocardium after MI. They demonstrated that granzyme B-mediated cytotoxicity was the main mechanism for cardiac destruction from $CD8^+$ T cells, and ablation of T cells or granzyme B deficiency was protective in mice. Similar results were obtained in pig models of cardiac ischemia. Authors further proved clinical relevance by showing that patients with elevated granzyme B had poorer prognostic.

Single-cell observations in human failing hearts suggested that most $CD8^+$ T cell rapidly become exhausted, a positive outcome given their cytotoxic functions (Rao et al., 2021). $CD8^+$ T cells have been more broadly associated with CVD. A subset of $CD8^+$ T cells, $CD95^+CD8^+$ stem cell memory T (CD8 T_{SCM}) cells, have been linked with CVD risk at large, estimated by Gensini score (Padgett et al., 2020). In this cohort of patients, authors could positively correlate the amount of T_{SCM} to higher risks of strokes, myocardial infarction or atherosclerosis.

1.2.4 T antigens in cardiovascular diseases

1.2.4.1 Cardiac antigens

CVDs are inflammatory driven diseases with antigen specific responses identified. In the heart, α-myosin has been described very early on autoimmune target (Donermeyer et al., 1995; Pummerer et al., 1996). Specifically, these studies highlighted the cardiac myosin heavy chain alpha (Myhca) residues 334-352 to elicit autoimmune myocarditis by MHC-II restricted presentation. Other positions have been described in autoimmune contexts (Axelrod et al., 2022; Krebs et al., 2007; Nindl et al., 2012). Screening for other cardiac epitopes in mice has been unsuccessful. Rickemann and colleagues identified a collection of 8 heart-enriched proteins and tested their immunogenicity (Rieckmann et al., 2019). They generated 15 mers, MHC-II restricted from the putative autoantigens and cultured them with splenocytes, however only Myhca lead to IFN-γ and IL-2 production from T cells. Interestingly, Myhca was also found to activate T cells *in vitro* in sham mice, confirming the natural immunogenicity of the protein. The same group investigated a MI cohort and identified cardiac-specific isoform adrenergic receptor β1(ADRB-1) as a target in patients but not in healthy volunteers (Hapke et al., 2022). Cardiac-specific T cells seems to target native proteins normally expressed by cells, but not neo-antigen or shock proteins. Administration of α-myosin specific $CD4^+$ T cells (TCR-M) in mice has shown to improve post-MI cardiac remodelling by *in vivo* reconversion of Th cells towards iT_{REG} profile (Nindl et al., 2012; Rieckmann et al., 2019). α-myosin specific T cells are not restricted to the injured organ and can be found in the circulation of patients with cardiac diseases (Fanti et al., 2022).

Neoantigens formed during oxidative stress have also been identified. Isolevuglandins (isoLV) are arachidonic acid derivative formed under oxidative stress that were known to activate pro-inflammatory DC and T cells, and promote hypertension (Kirabo et al., 2014). In

non-ischaemic infarction, the same group demonstrated how isoLVs acted as MHC-II neoantigens in cardiac pressure overload models (Ngwenyama et al., 2021).

1.2.4.2 Vascular antigens

Atherosclerosis is another CVD where T cell autoantigens were investigated, as T cells dominate the atherosclerotic plaque immune landscape (Fernandez et al., 2019; Jonasson et al., 1986). There is active presentation of peptide in plaque, with many APC infiltration, and evidence of recent TCR engagement in activated plaque T cells (Depuydt et al., 2023). There is document basis for of an autoimmune, T specific driven response in atherosclerosis. Atherosclerotic infiltrating T cells reactive to oxidised lipoproteins and heat shock proteins have been identified in murine Apoe-KO mice by Immunoscope (Paulsson et al., 2000). TREG specific T cells specific to ApoB-100 protein, promoter of atherosclerosis, have shown to be linked to worsening of the disease (Wolf et al., 2020). Zinc transporter protein 9 (Zip9) protein, an androgen receptor and atherosclerosis promoter, was found to be expressed in the cardiovascular endothelium (Thomas et al., 2014). Chowdbhury and colleagues showed how APC presenting Zip9 peptides APC induced $CD4^+$ T cell activation *in vitro*. Tetraspanin 17, another endothelial transmembrane protein, was also found to be an autoantigen in atherosclerotic plaques (Chowdhury et al., 2022).

1.2.4.3 No tolerance policy

The number of cardiac or endothelial autoantigens identified demonstrates how most inflammatory diseases could be branded as autoimmune. The basis for tolerance rupture do not seem to be identical across aetiologies, however.

One main driver identified has been molecular mimicry, where a pathogen and a host display similar peptides. It was described for coxsackie B3 virus induced myocarditis. In this example, molecular similarities between the M protein of coxsackie capsid protein and several cardiac myosin lead to myocarditis in patient and murine models (Manjula et al., 1985). Other viruses have been reported to induce myocarditis by similar mimicry mechanisms such as influenza viruses or myxovirus infection. In atherosclerosis, *Chlamydia pneumoniae* proteins are suspected to fuel immune response (Kuo et al., 1993). Expansion of $CD8^+$ T cells specific to common viral antigens influenza, cytomegalovirus, or severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) were reported inside coronary plaques by Chowdbhury and

colleagues, but not by Depuydt et al. (2023). The protozoan parasite *Trypanosoma cruzi* responsible of Chagas disease induces a cross-reactive T cell immune response targeting myosin-B13 (Coatnoan et al., 2009; Cunha-Neto et al., 1996), eventually leading to myocarditis.

Activation of T cells does not necessarily require the recognition of an epitope by the TCR, as they can be activated by bystander effect (Unutmaz et al., 1994). This is a known mechanism of activation in very inflammatory conditions such as hepatitis (Kim et al., 2018) or coxsackie-induced type I diabetes (Horwitz et al., 1998). Several reports have linked superantigen and autoimmune diseases. In CVDs, such mechanisms were reported as contributing but not founding events. For instance, viral infections have been reported as detrimental in atherosclerosis, by fuelling inflammation from the inflammatory cytokines released to fight the infection (reviewed by Libby et al., 1997).

1.2.4.4 Super antigens

Although superantigens not being strict antigen responses, there are documented superantigen responses implied in CVD aetiologies. Kawasaki syndrome is an acute systemic inflammation affecting primarily children under 5 (McCrindle et al., 2017). It often implies cardiovascular inflammatory complications, such as myocarditis or polyarteritis potentially leading to coronary artery aneurysms. Although the causative agents are, as of today, not completely understood, there is strong suspicion of an SA trigger. Indeed, the circulating T cells in Kawasaki syndrome represent abnormal elevation of certain Vβ families, where Vβ2 is consensual (Abe et al., 1992; Curtis et al., 1995; Pietra et al., 1994). $V\beta2^+$ $CD4^+$ and $CD8^+$ will also preferably infiltrate the inflamed organs such as the myocardium or coronary arteries, suggesting a sustained superantigen activation in those organs (Curtis et al., 1995). As typically observed in SA responses, $V\beta2^+$ cells are highly polyclonal (Abe et al., 1993).

Risk factors meta-analysis identified genetic polymorphisms associated with Kawasaki diseases, most of them being linked to immune system activation. Authors identified for instance secreted cytokines IL-6, IL-10, TNF, lymphocytes activating factors CD40, PD1 or HLA genes (Onouchi et al., 2007; Xie et al., 2018), confirming the instrumental role of T cells. Single-cell sequencing has been recently performed on circulating immune cells of Kawasaki patients before and after therapy, confirming most of the GWAS observations made above (Wang et al., 2021). Although the authors used single-cell sequencing to do TCR tracking

between timepoints, they discarded clonotypes with counts below 3, which does not capture the expected diversity of a polyclonal response.

Superantigens were also associated to endocarditis. Infective endocarditis is a transmittible disease caused by *Staphylococcus aureus.* It brought concern to the scientific community as being linked to multi drug resistance. In a rabbit model, it was shown that infective endocarditis was dependant on the SEC produced by *Staphylococcus aureus* (Salgado-Pabón et al., 2013). Investigation of the presence of SA expression in strains from sepsis shock with endocarditis patients found the presence of several expressed genes, but could not conclude on their definitive association with the pathology (Chung et al., 2014).

I have highlighted how pivotal T cells are for the progression and resolution of CVD, including myocardial infraction, as well as in the initiation of autoimmune myocarditis (Anzai et al., 2019; Won et al., 2022). The multiple T cell antigen described confirms the mounting of an effective immune response. Several studies have employed T cell repertoire sequencing to dissect the diversity, specificity and dynamics of T cells in patients with CVD. In atherosclerosis, TCR sequencing was used to demonstrate expansions of autoimmune clones. In immune checkpoint inhibitors treated patients, it identified cardiac-specific clones (Axelrod et al., 2022), while in ischaemic hearts, it found diversity differences between healthy and diseased hearts (Rieckmann et al., 2019; Tang et al., 2019).

1.3 DECIPHERING THE TCR LANDSCAPE – A METHODOLOGICAL OVERVIEW

One complex, uncertain, profoundly stochastic object would remain hardly explored: the TCR repertoire. It can be defined as the collection of all possible TCR that can be generated, and then further declined between the theoretical and effectively selected one (see 1.1.2.3). The issues were mainly technical, as its vast diversity could not be easily studied by cloning each individual T cells. From the pioneering breakthrough in molecular biology and miniaturisation emerged new technologies that will ultimately unleash the potential of TCR repertoire sequencing, making possible to sequence billions of DNA reads in less than a week. From the first experimental proofs to high throughput sequencing, I will detail the methods employed, and the consecutive improvements brought to them that lead to the current state of the art methods to dissect large-scale repertoires. In a constantly evolving field, analysis methods paradigm opened the molecular-biology-centred TCR niche to ecology, mathematics and informatics.

1.3.1 From T cells to TCR before NGS

1.3.1.1 The early days of TCR discovery

The first TCR observations were done on monoclonal murine and human "monospecific" T cell lines, exploiting hybridomas technique (Baker et al., 1979; Gillis et al., 1978). T cells were first identified by their reactivity to antibodies. The targets of antibodies derived from clones OKT3, OKT4, OKT8 (Reinherz et al., 1980b, 1980a) would later become CD3, CD4 and CD8 proteins. First experimental isolation of the TCR was made using monoclonal antibodies directed against an MHC-I specific human T-cell clone, $CT8_{III}$ (Meuer et al., 1983). Immunoprecipitation of ^{125}I radiolabelled $CT8_{III}$ TCR were then electrophoresed on reducing and non-reducing SDS-PAGE. This would reveal the α and β chains of the heterodimer TCR (see Figure 1). From a murine TCR clone, Saito and colleagues (Saito et al., 1984b) published the very first TCR complementary DNA (cDNA) complete sequences with the Maxam–Gilbert method (Maxam and Gilbert, 1980) using radioactive labelling of bases.

The process got quickly refined and by 1990, the method used was already very similar to current ones (Bragado et al., 1990). Retro-transcription (RT) of RNA of T cell clones was performed, followed by polymerase chain reaction. Bragado *et al*. methods used a multiplex approach, using the common constant β chain as their first primer, and a combination of 20

oligonucleotides for each V-β family described. Direct Sanger sequencing (Sanger et al., 1977), instead of relying on long and hazardous bacterial cloning, came as a relief as it was a faster and more reliable sequencing approach (Bhardwaj et al., 1993). Indeed, the use of multiple primers scaled poorly, and in 1993 Bhardwaj *et al*. introduced the very first challenge to multiplex approach. Rather than relying on 20 V-β primers, they designed a single *universal* Vβ-primer. Abstraction comes at the cost of specificity, with 1/3rd of the N-terminal Vβ sequence not being recovered. Authors noted that "*residual two-thirds of the Vβ gene analysed in this study would give enough sequence information to identify Vβ genes at the level of their subfamily members in most cases (62 of 71 known Vβ members), providing that the unanalysable N-terminal sequences are identical to the known Vβ sequences*". Using universal Vβ primers, it could be demonstrated how a single T cell could rearrange two TCR β chains (Obata et al., 1993). As shown in Figure 9, amplification of the region spanning the Vβ-Cβ region of a T cell clone in lane 2 is very blurry. Authors hypothesised it was likely due to multiple TCRβ being rearranged. By using two different J primers (in lane 3 and 4) instead of the Cβ, they could demonstrate how the clone in lane 2 rearranged two β chains.

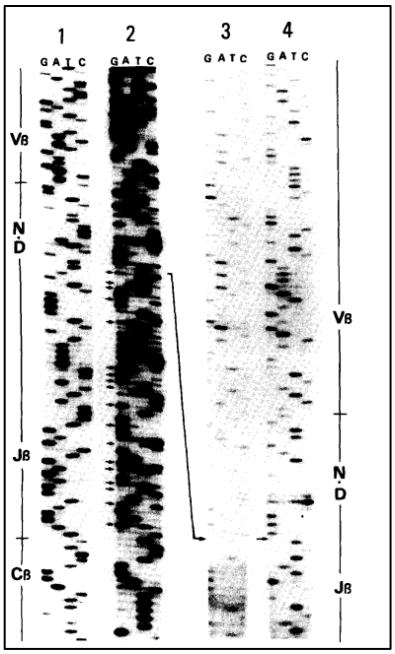


Figure 9: Direct sequencing of TCRβ cDNA amplified with the Vβ-universal primer. Lane I, T-cell clone 341-1 with Cβ sequence primer. Lane 2, T-cell clone 341-62 with Cβ sequence primer. Arrows indicate the positions where Jβ1.5 and JB2.3 segments formed the mixed ladder. Lane 3, T-cell clone 341-62 with Jβ1.5 sequence primer; and lane 4, T-cell clone 341-62 with JB2.3 sequence primer. From Obata et al, 1993)

1.3.1.2 Proteomic to molecular biology transition: the immunoscope era

These methods however, did not alleviate the need for more comprehensive data. Indeed, with millions of T cells per mL of blood, manually sequencing dozens of TCR would reveal too weak to observe large, dynamic immune responses. Fluorescent-conjugated antibodies specific to each Vβ were first used in mice to demonstrate how negative selection lead to the whole deletion of Vβ8.1 (Kappler et al., 1988), used in cytometry to quantify expansions of TCRs using these Vβ segments. Although rather limited as it confounds abundance and clonality, this approach requires minimal preparation of sample while providing solid insights on some specific diagnosis. Vβ cytometry is used in many clinical contexts : in haematology,

where T cell lymphoma can be detected by abnormally high Vβ usage in T cell (Potoczna et al., 1996; Puisieux et al., 1996); or during the first HIV experiments where major oligoclonal expansions could be observed (Pantaleo et al., 1994). Still relevant today, Vβ cytometry can detect abnormal expansion of T cells restricted to a specific TRBV as a fast and relatively cheap alternative to sequencing. Due to its limitation (limitation of coverage of VJ genes and simultaneous colors), it must be confirmed by molecular techniques (Langerak et al., 2001). Vβ cytometry has been used to investigate multiple-inflammatory syndrome in children (MIS-C) which harbour a high percentage of Vβ21.3^{+} cells (Moreews et al., 2021), X-linked immunodeficiency (Goldman et al., 1992), multiple sclerosis (Hafler et al., 1996) or rheumatoid arthritis (Struyk et al., 1995).

Nevertheless, single cytometry does not cover α chains, as very few antibodies directed against Vα families exist. Moreover, TRV cytometry does not provide information on the clonality. To tackle this issue, a French laboratory lead by the immunologist Philippe Kourilsky harnessed the properties of VDJ recombination and immune expansion. Given the varying size of CDR3, he studied the TCR repertoire changes through its CDR3 length distributions (Cochet et al., 1992; Pannetier et al., 1993). Integrating the intensity of CDR3 length fragments sizes on a gel, they were able to estimate what they called the relative index of stimulation (RIS), for which they derived a reliable indication of the emergence of a specific T cell clone. This technique initially called Immunoscope, based on the software used to integrate peaks, allowed for instance to identify oligoclonality in T lymphomas (Puisieux et al., 1996). Other groups developed similar scores based on the same approach, but maintaining the important notion of perturbation (Bomberger et al., 1998; Collette et al., 2004; Collette and Six, 2002; Gorochov et al., 1998; Mariotti-Ferrandiz et al., 2016).

1.3.2 Emergence of NGS and ready-to-use kits

Usage of fluorescent compounds rather than radiolabelled di-deoxyribonucleotides (ddNTP) (Smith et al., 1986) and capillary electrophoresis (Gocayne et al., 1987) opened new perspectives for large-scale sequencing. Although the Sanger method was considered robust (99,99% accuracy), it proved its limit for scalability. Indeed, the Human Genome Project, which goal was to sequence the whole human genome, was heavily relying on it. It offered a solid incentive for the development of better methods and is considered instrumental in the development of NGS (Gibbs, 2020).

1.3.2.1 2nd generation sequencing

NGS is a misleading term, as it refers to pretty much any technology that is done at scale. Instead, we will refer to it as n^{th}-generation sequencing. TCR studies, as of now, are all based on second-generation sequencing (SGS). SGS is based on the synthesis-based sequencing (SBS) (Nyren et al., 1993), an assay where the addition of a labelled base could be detected on an elongated DNA strand. This technique brought many industrials to race and develop SBS-based sequencers, where Illumina (Illumina Inc.) is now leader and provides most sequencing services for the TCR community.

Illumina developed a 4 channel SBS on their platforms HiSeq and MiSeq, based on academic breakthrough (Bentley et al., 2008; Shendure et al., 2005). DNA templates extremities are fixed on a solid surface and extended in a bridge-fashion (Adessi et al., 2000; Nyren et al., 1993). During sequencing, 3′-modified nucleotides are used to elongate millions of DNA strands in parallel. This prevents the elongation of multiple bases, until the addition of a reagent restores the missing -OH, allowing a new cycle of 1 base elongation step. The terminator can be removed enzymatically, hence termed reversible. This technology allows the synchronous sequencing of millions of DNA bases simultaneously. Analysing the laser-excitation imaging from single-molecule laser of the sequencing flowcells of every elongation step determines which bases are incorporated for each cluster on the flowcell (Lundquist et al., 2008).

This offers four main advantages compared to previous methods, in the sense that is -relatively- cheap, fast, reliable and highly scalable.

Illumina will further refine its method by using patterned flowcell. By having evenly spaced and defined clusters, this greatly improves yield and better resolution of elongation clusters. Illumina also introduced exclusion amplification (ExAmp) to this generation of sequencer (Shen et al., 2014). Brought to HiSeq 4000, and above, and NovaSeq series, ExAmp only needs two channels to call the four bases based on their binary detection leveraging the kinetic properties of elongation.

1.3.2.2 Multiplex, RACE, DNA

Most of the current protocols to sequence TCR start from RNA transcripts. As Illumina's technology requires DNA templates, it would be required to retro-transcribe (m)RNA to DNA first. TCR however, is made from semi-random rearrangements and although it would be

possible to obtain full length TRC extremities, the TRV would require either i) a universal primer before the TRV, ii) one primer for each known TRV, or iii) use a common V portion, at the cost of not covering the complete V gene as already done in the past. Groups from the 90's used either ii) or iii).

One more approach was developed from (Frohman et al., 1988) looking to obtain full length cDNA from variants mRNA, such as alternative splicing, they had to devise a method that could capture all mRNA based on their poly-A tail, and then used a single primer to amplify the different variants of a gene. The method, called rapid amplification of complementary ends (RACE), was quickly used to analyse and discover TCR genes (Loh et al., 1989). Another improvement came from template oligo switching (SMART technology), a reverse transcriptase adding a custom anchor at the end of its first strand (Zhu et al., 2001). The most common 5'RACE TCR sequencing kit is now acquired and developed by Takara Bio, using both 5'-RACE and templated oligo-switching. Many other commercial kits exist, all with specific constructs. An extensive list of all major protocols are listed and documented by MiLaboratories (https://mixcr.com).

The biggest competitor to 5'-RACE has been multiplex PCR. A multi-center project lead by our laboratory (Barennes et al., 2021) offered an overview of different methods used to sequence repertoires. In a thorough investigation, all partnered laboratories amplified and sequenced TCRαβ from standardised input samples. This showed the superiority of the 5'-RACE approach in both coverage and reproducibility when starting from RNA. In house multiplex methods have also been developed independently, showing the interest of the community and the accessibility of the method (Montagne et al., 2020).

A big concern of multiplex based is the amplification bias due the high multiplexing of primers (Okino et al., 2016). A similar doubt can be casted on RACE methods, which can also be subject to early termination (Scotto–Lavino et al., 2006) or poor efficiency of the template switch that adds the 5′ adapter in only 20%–60% of RNA molecules (Wulf et al., 2019).

To alleviate PCR biases, the addition of unique molecular barcodes (UMI) to DNA fragments extremities before the amplification steps has been developed (Kivioja et al., 2012). In short, UMI are small, random sequences used to track single cDNA molecules introduced at early steps, so that each cDNA is tagged with a unique UMI through the complete protocol. During preprocessing, sequences are first aligned to a reference genome and sequences with a same UMI are merged, as they were amplified from the same original cDNA. UMI length vary

between methods and manufacturers. For instance, TaKaRa uses 12 random nucleotides, with a fixed 4 bases linker, along with their 8 base long indexing (SMART-Seq® Human TCR, TaKaRa Bio), which would allow a maximum of ~10^4 unique molecules barcoding with single-end sequencing (Best et al., 2015).

In the case of the TCR, the addition of UMI is at a trade-off with the depth, as UMI-based amplifications tend to retrieve a lower amount of sample's diversity (Barennes et al., 2021), but offers a better quantification. The methods, however, showed poor reproducibility compared to some 5'RACE protocols.

The main pitfall of those high throughput methods is the loss of cell identity information. Indeed, all the aforementioned PCR-based methods are mainly applied to bulk T cells. They cannot recover the αβ pairing, as all sequences get pooled. A dedicated pipeline would require the cell-level barcoding of transcripts before their amplification to attempt the recovery. Another issue is the read length: with maximum 300bp per read, it is not possible to cover the whole TCR. Other attempts on IonTORRENT, a sequencing device tailored for long reads (>10kbp) were recorded, but authors used 200bp paired sequencing (Fang et al., 2014).

1.3.2.3 Alignment of sequences

Alignment refers to the pairing of raw nucleotides sequences to known sequences from a reference genome. As sequencing throughput increased, new methods have been developed and published. There is now an extensive number of methods optimised for different applications, whether it is bacterial unknown transcripts (Marçais et al., 2018), whole human genome (Altschul et al., 1990; Thompson et al., 1994), or RNA-seq focused data (Dobin et al., 2013). Adaptive immune receptor repertoires (AIRR) data is no different, and multiple suites have been designed to optimise TCR data alignment. Indeed, contrary to most genomic constructs, Adaptive immune repertoires (AIR) structure is very different in the sense that the tools need to correctly annotate each read to a V, D, J and C gene, with vastly different anchor points to determine the highly polymorphic CDR3. The most common tool to date is MiXCR (Bolotin et al., 2015), a constantly updated program specifically designed to align TCR datasets, with built-in quality control and standardised outputs based on IMGT conventions (Lefranc et al., 2003). It has been adapted to current constraints of kits and sequencers that cannot sequence full length with a seed-and-vote approach to map genes (Liao et al., 2013). When multiple genes can be mapped to a single-read, as it is the case with very homologous TCR genes, a collection of

subreads (seed) is used to determine the optimal mapping by consensus (vote). These approaches are preferred over k-aligners as they do not have sources of randomness. Other suites provide similar outputs, for instance Decombinator is the 4th iteration of a TCR-focused python pipeline that handles raw data demultiplexing, including UMI, standart compliant annotation of reads (Peacock et al., 2021). On the other hand, pRESTO is a more general toolkit for AIR alignment, with deep possible customisation of alignments parameters, and many sub-tools for BCR downstreams alignment (Vander Heiden et al., 2014).

The main feature of these tools is how they handle amplification errors, with built-in PCR-error correction and clonotype aggregation algorithm. Briefly, these algorithms will "rescue" reads with low quality initially discarded during alignment by aggregating them to high count, high quality clonotypes. MiXCR uses hierarchical clustering and fuzzy matching criteria (Bolotin et al., 2015).

Other open-source aligners are available, especially in the field of RNA-seq without amplification of TCR. Among them, TRUST4 is specialised in reconstructing TCR from bulk and single cell RNA-seq (Song et al., 2021) and has been found the most reliable current available tool, outperforming MiXCR in the recovery of full length TCR sequences from transcriptomes. Other alternatives such as ImRep (Mandric et al., 2020) or CATT (Chen et al., 2020) are available as well. Recent comparison of alignment tools (MiXCR, TRUST4, CATT and ImReP) to recover reads from single-cell or RNA-seq from tissues have been published, but stirred controversy on the methodology used (Davydov et al., 2023; Huang et al., 2023; Peng et al., 2023).

1.4 MEASURING THE TCR REPERTOIRE

1.4.1 What is the unit in TCR repertoire analysis

T cell clonality is a tricky aspect of their biology. A "clone" refers to a population of immune cells with identical TCR that originated from a single precursor cell, whereas a "clonotype" refers to the specific sequence of an antigen receptor on a T cell within that clone. The latter is the most commonly used definition, defined by final product of the somatic rearrangements following VDJ recombination, with the V, CDR3 and J genes used. All clones bear the same clonotypes, but not all clonotypes are derived from the same clones. For the sake of simplicity and standardisation, a nomenclature has been proposed that defines the TCR clonotype as "*a unique nucleotide sequence that arises during the gene rearrangement process*

for that receptor" (Yassai et al., 2009). When working with paired chain receptor, the final rearrangement of both chain is called a cell clonotype. In AIRR analysis, immunologists often go by the clonotype definition, although older records might use different wordings.

1.4.1.1 VntJ, VaaJ, CDR3, a definition for a question?

As we have covered in the 1st part of the introduction, TCR specificity is defined by the interaction of the TCR with a peptide loaded on an MHC and the product of the V(D)J rearrangements, fused to a constant region. Therefore, to describe a TCR, portion in contact with the MHC and its complementary chain, one should describe the genes of the rearrangement and the CDR3 region that is not hardcoded in the germline. We define a clonotype as a V-CDR3-J. Then comes the choice of the level at which we define the clonotype, whether at the nucleotide (nt) or at the amino acid (aa) level. Hence V-CDR3$_{NT}$-J, or V-CDR3$_{AA}$-J. The latter is the classic clonotype definition, as it conveys the sense of the TCR specificity. V-CDR3$_{NT}$-J can also be used in specific contexts, such as clone tracking. In fact, two clonotypes with similar CDR3s can be made from different nucleotides sequences, discriminating convergence from clonality. Indeed, the genetic code is redundant, and a single amino acid can be derived from several codons. Cysteine (C) can be encoded by the codons `UGU` and `UGC`, while leucine (L) can be coded by six different ones (`CUU`, `CUC`, `CUA`, `CUG`, `UUA`, `UUG`). Thus, if we consider for example an HIV-specific CDR3, consisting of the following 14 bases "**`CASSALASLNEQFF`**", based on the genetic-code redundancy it can theoretically be derived from $2 * 4 * 6 * 6 * 4 * 6 * 4 * 6 * 6 * 2 * 2 * 2 * 2 * 2 = 31{,}850{,}496$ distinct nucleotide combinations. This concept is widely used in B cell repertoire analysis, where B cell receptor will undergo iterative random mutations to improve their affinity to an antigen (Maul and Gearhart, 2010). From this, it is possible to map the phylogeny of each antibody (Nouri and Kleinstein, 2018; Shlomchik et al., 1987).

Aggregating TCR repertoires by their CDR3 alone is rarely used. CDR3 is the main contact point of the TCR with the pMHC complex (Dong et al., 2019), and predominantly drives antigen specificity. Studies focusing on CDR3 alone often only consider the specificity associated to the CDR3, and not at the clonotype level. This point will be further elaborated in 1.5.2.2.

1.4.2 The grand ecological robbery

The T cell receptor repertoire is the collection of all TCR that can be observed at a given time. There are many different scales of the repertoire, whether it is theoretical during generation, observable in an individual, or sampled during experimentation. If we consider TCRs as species, AIRR analysis is strikingly similar to ecology; birds' species are now TCRs, forests are organs, biomes are individuals, cytokines are nutrients. Many authors have already developed this theory, comparing the T cell repertoire as an ecological niche, in a competing environment (Bautista et al., 2009; Freitas and Rocha, 2000; Schulenburg et al., 2009).

These similarities also lead to the same question: how do I measure diversity? TCR are diverse, with sequences similarity that can be objectively measured and quantified. TCR repertoire are distinct at the scale of a cell subset, an organ, an individual, a population. Repertoires are dynamic, as they evolve in response to infections, vaccinations, much like to ecosystems. These properties have been widely conceptualised and developed by ecologists, from whom many tools have been adapted to TCR diversity studies.

1.4.2.1 Indices for diversity

Ecology distinguishes 3 types of diversity, α, β and γ (Whittaker, 1972). α-diversity is the diversity of a single sample, β-diversity is the evaluation of an assemblage of samples, and γ for comparing assemblages of several groups of samples.

α-diversity is now widely used to evaluate repertoire diversity, and uses the notion of entropy to convey the idea of quantity of information contained in a repertoire. The Rényi entropy formula uses a parameter α on the exponent to quantify the number of species (α=0, Hill index), the relative evenness (α=1, Shannon index), the probability of interspecies collision (α=2, Simpson index), or the contribution of the main species (α=inf, Berger-Parker index). Although abstract, or even simplistic as it reduces a whole repertoire to a single value (Laydon et al., 2015), the Renyi entropy is a powerful tool that finds various relevant applications. Intra-tumoral diversity has an abundant literature on the use of α-diversity to assess for the success of immunotherapies, predict diagnosis or disease outcome in patients (reviewed by Porciello et al., 2022).

β-diversity in TCR repertoire analysis is studied through the similarity of two samples composition. Three approaches exist, either presence/absence based or abundance based, or similarity based. Presence/absence-based methods see data as binary, either the species have

been seen, or not. The Jaccard index, from the botanist Paul Jaccard (Jaccard, 1901), computes the overlap between two samples, and normalise it by the union of them. This is often contrasted with the abundance-based method which take into account the distribution of samples to the computation. The Morisita-Horn index, for instance, weights the similarity of frequencies between the repertoires to define similarity. Schober et al. (2020) used the Morisita-Horn index to track immune response convergence of mice after CMV infection, where they showed that T cell repertoire stabilisation occurred in 3 months after inoculation. The final approach considers the relative distance between species. Originally used for taxonomic or phylogenetic data, it can work from any pairwise distance matrix coupled with abundance data. Several methods have been proposed (Olson et al., 2022; Yokota et al., 2017) but their poor scaling due to quadratic complexity in NGS has hindered their use and development.

Each of these β-diversity approaches has its pros and cons, Jaccard similarity is sensitive to rare clonotypes and sample size, while Morisita-Horn is less sensitive to these. Distance-based approaches rely on a pairwise distance matrix, which can be computationally intensive and biased depending on the chosen downstream representation (Chari and Pachter, 2023; Wang et al., 2023).

1.4.3 To normalise or not normalise, that is the question

Counts normalisation is a sensitive topic in the AIRR analysis. Even if implementation of technical solutions such as UMI or single-cell barcoding greatly improve count precision, it does not alleviate the issue of highly unequal sampling sizes. There is no consensus yet on whether or not to perform normalisation, and the methods used often differ between research groups, studies and topics involved. This plethora of approaches is not bad *per se*, but rather reflects the necessity to tailor normalisation to each dataset specificity. Whether they are original methods, or derived from other fields such as ecology or transcriptomics, they all carry strong assumptions on the data and its outcome.

1.4.3.1 Ecology

Ecology is the biggest contributor to normalisation in the TCR field. As a science bridging mathematics and biology, ecologists have been interested in the differences in samplings in an ecosystem for decades. The oldest, and perhaps most intuitive approach is to transform counts into frequencies, a method going with the fancy name of “Total Sum Normalisation”, which yields “normalised counts” or frequencies. Simple, yet powerful, the pitfalls of this method

when performing correlations are extensively covered by Jackson (Jackson, 1997). As an ecologist, Jackson criticises the use of compositional data (frequencies, percentages, proportions), as it changes the covariance and correlation matrices. He notes how switching from abundances to frequencies can create spurious relationships between variables. Another commonly used method is the random subsampling of an assemblage of datasets to the smallest, at the cost of discarding many potentially useful rare species.

In bulk TCR analysis, converting counts to frequencies is widely used, although sometimes with additional filters to account for technical biases (Greiff et al., 2015b). Subsampling on the other hand, has been performed for different reasons. Zhang and colleagues downsampled mice repertoires, before aggregating them by condition and computing similarity matrices (J. Zhang et al., 2021), while (Vujović et al., 2023) used it to reduce computational load during pairwise comparisons. In a sense, the concept of pooling individuals before sequencing (meta-individual) is analogous to a *pre-hoc* normalisation.

1.4.3.2 TCR-seq

The TCR-seq community has not yet validated a widely consensual approach for normalisation yet, but several attempts were recorded. They can be divided into 2 categories, experimental or computational. Experimental occurs during the library preparation, while the latter focuses on aligned data. Normalisation performed by aligner has been covered in 1.3.2.3 .

Experimental approaches are the least biased, but probably the most difficult to set up. They rely on the introduction of a known, quantifiable quantity to allow correction further on the analysis pipeline. We have described how UMI introduction in pipeline can be a robust addition (Barennes et al., 2021; Kivioja et al., 2012), but other relevant methods have been described. In a multiplexed PCR targeting TCRG genes, Carlson *et al.* (Carlson et al., 2013) were adding equimolar synthetic DNA templates of all primers to their pool. They would then normalise downstream data by removing computationally the residual differences in amplification efficiency. In TCRpower (Dahal-Koirala et al., 2022), authors amplified by 5'RACE peripheral blood mononuclear cells (PMBC) samples spiked with a collection of 45 known TCRs. Detection of the various spike-ins of known concentrations would then allow them to evaluate the replicability of experiments and set threshold for false discovery rates

(Howie et al., 2015). Such methods, although providing useful experimental standards, require a careful experimental design and access to those biological controls.

Downstream normalisation without ground truth controls performed after datasets alignment is the most common version. Many groups have adopted the removal of "singletons", clonotypes with a count of one (Greiff et al., 2015a; Mhanna et al., 2021; Peng et al., 2023). The argumentation around this filter is that these clonotypes seen only once are artefacts, due to faulty PCR or error during library production, or errors during sequencing (Greiff et al., 2015b). Two methods are possible, either completely filter out singletons, leaving the lowest count at two, or decrement one to each clonotype and filter clonotypes with null counts, which bear limited impact to downstream analysis. However, other groups are however confident in their pipeline and integrate them to their analysis (Britanova et al., 2014; Ronel et al., 2021; Sheng et al., 2021; Singleton et al., 2016). Singletons removal has been criticised by others in a comment to Peng *et al.* methods comparison (Davydov et al., 2023).

Most authors rely on converting sums to frequencies and comparing them as is. Other teams prefer to focus on top clones (Amoriello et al., 2020; Poran et al., 2020). Other dedicated methods have been described. Chaara et al., (2018) propose an entropy-based threshold for normalisation, where multiple subsamplings are done at the Shannon entropy level and a "consensus" repertoire from the multiple subsampling is then used for subsequent analysis. A robust benchmark of low-read sample normalisation has demonstrated how correcting Shannon entropy estimates yields better results when performing downsampling (Bortone et al., 2021). However, the low traction gained by these studies highlights how reluctant are research groups to perform normalisation. One argument is that subsampling is at the cost of rare TCR coverage, as sample rarefaction profits to large clonotypes.

1.4.4 Estimating the T cell diversity

There have been numerous estimations of the relationship between T cells in an individual and TCR repertoire diversity. From these first observations refined thanks to technological improvements, the TCR field slowly progressed from naïve estimations to complex modelling of the virtually infinite TCR repertoire possibilities.

First estimates did not take into account the number of clonotypes, but rather the number of T cells in an individual. The first estimates in humans date back to the 70's. A 200g rat was estimated to hold 4-5 $\times 10^9$ T cells (Trepel, 1974). From extrapolation using density ratios, a 70kg human was estimated to contain 3.3×10^{11} T cells. Main tissue holders are lymph nodes

(41.2%), spleen (15.2%), thymus and bone marrow (10.8% each). Blood accounted for 2.2% of circulating T cells. 20 years later, a similar estimate will be produced, distinguishing CD4 and CD8 compartment, but placing the total of number of T cells in an adult human at 3.5 x 10^{11}, of which 2.5 x 10^{11} CD4 T cells, and 1 x 10^{11} $CD8^+$ T cells (Clark et al., 1999). In line with these results, another group reported that a mouse contain 10^8 T cells, without publishing their method, and estimated the human body to hold 4.5 × 10^{11} T cells (Jenkins et al., 2010). Assessment in mice through extensive counting puts the estimate at 6 x 10^7 T cells in a 25g individual mouse (Boyer et al., 2019), while a recent meta-analysis re-evaluated at 7 x 10^{11} T cells in an adult human (Sender and Milo, 2021; Westermann and Pabst, 1992).

After developing Immunoscope (Pannetier et al., 1993), Philippe Kourilsky made several estimations of the link between T cell and TCR clonality. Sequencing the circulating T cell repertoire of adult donors showed that TRB diversity was about 10^6 clonotypes, and overall diversity of 10^8 TCRs (Arstila et al., 1999). In a response to Arstila's findings, Keşmir, Borghans and de Boer (Keşmir et al., 2000) argued that these results should rather place total diversity at 10^{11}. Further work on mice spleens revealed the same trend with 10^6 TRB chain in an individual (Casrouge et al., 2000). Robins *et al.* (Robins et al., 2009) also used Immunoscope to estimate the TCRβ repertoire of activated T cells to account for at least 3-4 x 10^5 TCRβ CDR3 in an adult human, which places the number of unique TRB in an individual at around 3-4 x 10^6 according to their estimations.

NGS has again revolutionised the field by making easily accessible the use of rarefaction and extrapolation models, and revise upwards previous estimates. (Britanova et al., 2014) estimated at 4 - 7 x 10^6 the number of TRB in a donor. A much bolder estimate was made using Chao2 estimation, placing at 0.1-1 x 10^8 the number of unique TCR β clonotypes in an individual (Qi et al., 2014). This method has been criticised by others (Laydon et al., 2015), showing that the Chao2 index method was not robust with varying sampling size. Complex modelling approaches were developed based on thymic output and observed diversity (Bains et al., 2009; Baltcheva et al., 2012, 2012; Thomas-Vaslin et al., 2008). This estimated the number of unique TCRs in the body to 10^{10} (Lythe et al., 2016). An even higher estimate of 10^8 to 10^9 unique TRB clones was reached by Mora and Walczak (Mora, 2019) with probabilistic Bayesian modelling, which would be consistent with the 19 years old remarks from Keşmir and colleagues (Keşmir et al., 2000).

1.5 TRANSLATIONAL ASPECTS OF MODELLING

The evolution of quantification methods of the T cell repertoire demonstrated how the TCR field has slowly evolved from a bench sided science towards an integrative one, bridging and immunology, mathematics and bioinformatics. Systems biology seeks to model, theorise, simulate how life works. It administers drugs with algorithms, dissects animals with mathematical models, and evolve with a constant flow of input data. Here, we will describe how the T cell repertoire has been modelled as a system, rather than a static object.

1.5.1 The TCR repertoire: the top model

1.5.1.1 Captivating equations

Modelling the TCR repertoire has been a fundamental step in our understanding of the immune system. Although the genetic machinery producing the immense diversity of TCR repertoire has been thoroughly studied, a generative model of the thymic generation is still needed. Naïve attempts of *in silico* TCR generation have been recorded (Venturi et al., 2008) with random drawing of nucleotides, but did not reflects the genetic complexity of the complete VDJ recombination. Alexandra Walczak and Thierry Mora, two physicists, took an interesting approach by studying the TCR generation as a probabilistic process. Their reasoning is as follows: VDJ recombination is a stochastic process involving the probability of a V, D, J genes to be rearranged with one (alpha) or 2 processes of addition or deletion of nucleotides. If it cannot be deduced from a sequence which particular event led to its generation, the probability of the final outcome to happen is the sum of each path that led to it (Murugan et al., 2012). From here, probabilistic approach is used to estimate each parameter from empirical set of sequences (V, D, J gene usage, probability of N insertions or deletions). TCR observed in peripheral T cells are the product of selection after VDJ recombination, also authors used non-productive TCR to exclude the impact of selection. The final probability of generating a TCR (Pgen) was dissociated from the probability of observing a sequence after selection (Ppost), with Q called the selection factor (Elhanati et al., 2014). This work brought i) a generation model for any sequence in the thymus plus ii) a generation model post selection (Marcou et al., 2018; Sethna et al., 2019). Other improvements of this model were brought, by allowing the inference of new generative models from custom repertoires and support from other animal model such as a complete mouse or chicken models (Marcou et al., 2018). The theoretical model does not take into account the TRBD1-TRBD2 anecdotal rearrangement (see 1.1.2.3), nor the

downregulation of TdT in foetal stages (Bodger et al., 1983), and might require fine tuning for some specific use case. Nevertheless, comparison of Pdata and Pgen in large datasets have shown the robustness of the approach (Murugan et al., 2012).

1.5.1.2 Curves that make you blush

T cell populations have been extensively studied through mathematics in a field called quantitative immunology, in various context such as thymic export, T cell proliferation or lymphocyte turnover (De Boer and Perelson, 2013; Sidorov and Romanyukha, 1993; Thomas-Vaslin et al., 2008). Mathematical models have also been used to describe AIRR, such as B and T repertoire. AIRR typically display an asymmetrical distribution where a few top clones will dominate and the majority of them will not be expanded. As such, clonotypes distribution are often modelled as a power-law, Zipf-like distribution (Desponds et al., 2016; Greiff et al., 2015a; Mora et al., 2010) or Poisson distribution (Laydon et al., 2015). Modelling of the distribution has many useful applications, especially for the simulation of new repertoires. Greiff's lab has developed ImmuneSim (Weber et al., 2020), an R package designed to generate random *in silico* AIRR repertoires with tuneable parameters for power-law or VDJ gene usages.

1.5.1.3 The centralised architecture of TCR networks

TCR repertoires are an ensemble of discrete values (TCR) that are independently produced and exported to the periphery. Nevertheless, the selected or expanded clones are not random, obeying to complex dynamics of competition for the pMHC niche (Freitas and Rocha, 2000; Stirk et al., 2008). Working on mice, Nir Friedman's group showed how TCR repertoires are structured in networks of CDR3s. By studying public (shared) versus private (non-shared) CDR3β in mice, the group showed how public CDR3β are found at higher frequencies in individuals (Madi et al., 2014). By clustering together CDR3β distant by at most 1 amino acid, they further demonstrated how public CDR3β wtih similar specificities were linked in the same clusters, and obtained similar results in human cohorts (Madi et al., 2017). The immune repertoire is imprinted with this network structure during its differentiation in the thymus. Quinou *et al*. showed how $CD8^+$ mature thymic T cells selection drives the selection of related CDR3s (Quiniou et al., 2023).

Far from being a theory or an abstract relationship, the paratope network has been directly linked to functionality of T cells. First evidences come from studies in mice. After infecting

mice with *Listeria monocytogenes* Jenkin's lab demonstrated how CD4 T cells differentiation into effectors subsets was biased based on T cell specificity. When presented two peptides derived from *Listeria monocytogenes* , $CD4^+$ T cells specific to an $I\text{-}A^a$ MHC-II restricted peptide would differentiate into a Th1, while $I\text{-}A^b$ would be biased towards T follicular helper profile (Tubo et al., 2013). In an extensive research paper, Zhang and colleagues further complimented this work, similar to the Hebbian theory of neural networks, *cells that wire together, fire together*. By modelling together TCR CDR3 and transcriptional profiles, they showed T cell networks of closely related TCRs were correlated to transcriptional profile (J. Zhang et al., 2021)[1]. Moreover, they hinted toward a "*wisdom of the crowd*" mechanism of T cells, where the TCR clonotype that is closest to the 'average' of all the clonotypes, within the same network, have a better antigen targeting efficiency to a known antigen, and tend to be more expanded.

1.5.2 Specificity and the limits of modelling

1.5.2.1 Experimental efforts from a community

Predicting T cells specificity has been major challenge of the past decades due to its complexity, but the reports of T cell specificities in the literature are scarce. Specificity is often collected from different methods and reported by groups independently. Facing this, a strong push from the TCR scientific community has been deployed to provide databases of TCR-peptide-MHC complexes. In the form of open online resources, several datasets of curated, annotated TCR specificities were made public, from which we can distinguish the generic VDJdb (Shugay et al., 2018), the structural-focused iEDB (Dhanda et al., 2019) or the pathology focused Mc-PAS (Tickotsky et al., 2017). These databases contain thousands of reported TCR-peptide specificities, often obtained through unreliable methodologies, which can be filtered. These databases reflect the recent topics of interest, hence being biased woth high reports of β and viral specificities of a restricted pool of antigens (Figure 10). Although not being ground truth, these remain widely used by the community to study the TCR repertoire.

[1] (Schattgen et al., 2022) would later report similar results on the correlation between TCR similarity and gene expression profiles of T cells

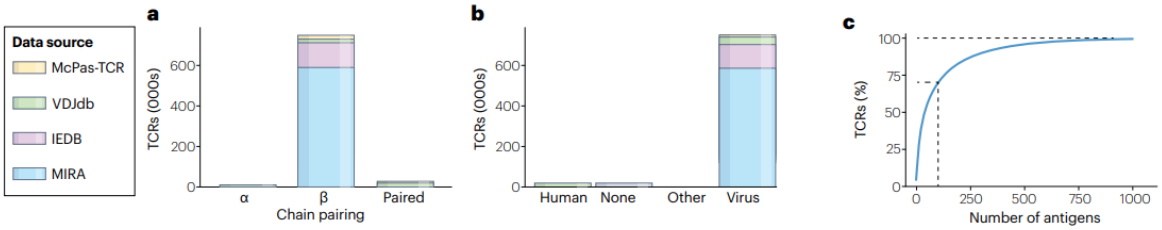


Figure 10: Can we predict T cell specificity with digital biology and machine learning? ***a****, Number of T cell receptors (TCRs) containing α-chains, β-chains or paired chains, showing variation in numbers according to the data set (manually curated catalogue of pathology-associated TCR sequences (McPas-TCR), VDJ database (VDJdb), Immune Epitope Database (IEDB) and multiplex identification of TCR antigen specificity (MIRA)).* ***b****, Number of TCRs per antigen species of origin, showing that the majority of all antigens reported as binding a TCR are of viral origin.* ***c****, Cumulative frequency of antigens, showing that a group of 100 antigens makes up 70% of TCR–antigen pairs. From Hudson et al, 2023.*

1.5.2.2 In silico prediction of HTS

With thousands of T cells reported in the databases, and with 10^{18} possible TCRs, several groups have proposed to infer the specificity of unknown T cells based on already described known. Leveraging the fact that TCR sharing similar properties do share similar specificities, an increasing number of approaches have been published to tackle the issue of specificity inference. Levenstein distance and its length restricted version, the Hamming distance, have been used to cluster CDR3 and infer specificity from neighbors (Chronister et al., 2021; Madi et al., 2017; Valkiers et al., 2021). Physico-chemical properties of amino acid in the CDR3 have been extrapolated to simulate similar docking to pMHC complex (Beshnova et al., 2020; Chronister et al., 2021; De Neuter et al., 2018; Gielis et al., 2019; Ostmeyer et al., 2019; Sidhom et al., 2021; Tong et al., 2020; Xu et al., 2022; Zhang et al., 2020), or a combination of both approaches (Glanville et al., 2017; Huang et al., 2020; Mayer-Blackwell et al., 2022). Other use structural data from the epitope to predict specificity of unknown T cell (H. Zhang et al., 2021) or leverage multi-omics from single-cell datasets (Fischer et al., 2020; Schattgen et al., 2022; Z. Zhang et al., 2021).

The plurality of tools highlights the complexity of the TCR specificity, but also how this allows different approaches from modelling. GLIPH2 (Huang et al., 2020) was originally developed to identify sets of TCR specific to *Mycobacterium tuberculosis (Mtb)* epitopes. The original paper also described how single amino acid mutations of key positions lead to drastically decreased affinity for binding, as predicted by GLIPH2 and confirmed *in vitro*. Studying the circulating T cells from two groups of *Mtb* infected patients, progressor and non-progressor, the group associated GLIPH2 with an experimental epitope screening approach (Musvosvi et al., 2023). They identified a cluster of CDR3 harboring a common pattern "**S%LAAGQET**", and associated to the Mtb-progressor group. *In vitro* cloning of TCR and

challenge with *Mtb* lysates confirmed the specificity of T cells. These results demonstrate how *in silico* modelling has become a viable method to extract epitope-specific TCRs from large and noisy datasets.

2 CHAPTER 2: OBJECTIVE AND EXPERIMENTS

2.1 OBJECTIVES

Cardiovascular diseases are more than just inflammatory diseases, but are also at the centre of active immune adaptive responses, whether it is at the initiation, or the worsening of the disease. The nascent interest for immuno-cardiology has proven that T cells were heavily involved in the pathophysiology of most CVD and that they represent a promising reservoir of biomarkers. Hence, there is an unmet need for biomarkers to monitor the ongoing myocardial repair before it reaches a critical point. Moreover, harnessing T cells and their specificities in CVD is a relevant target to develop new therapies and enhance cardiac repair.

However it is difficult to access to the damaged tissues in living humans, whether is it because the organs are hardly accessible such as the heart, or procedures are restricted to particular clinical situations. Therefore, there is a strong benefit to develop methods that can deliver an accurate picture of the cardiac repair without sampling the actual myocardium. There is compelling evidence that the T cell compartment foster myocardial repair and circulating T cells can reflect the tissue specific response happening in the heart. Circulating T cells, easily accessible from a blood sampling and in abundant numbers, appears as a promising tool for the diagnostic and exploration of the underlying immune processes in a quick, reliable, and non-invasive procedure.

The general objectives of this thesis are to i) identify biomarkers and ii) better understand the pathophysiology of CVD using TCR repertoire modelling of circulating T cells in CVD To reach these objectives, we defined specific objectives;

1. Establish a quality control protocol to ensure high-dimensional TCR data reliability – Objective 1
2. Promises of TCR sequencing in pathophysiology understanding - an application to MISC –Objective 2
3. Identification of the minimal TCR feature that could serve as biomarker –Objective 3

Objective 1: TCR repertoire sequencing is a recent field and very few tools exist to assess the quality of downstream data alignment. By using new generation sequencing

technologies, my work first consisted in evaluating and developing the robustness of highly multiplexed data from massive sequencing. More precisely, my work aimed at designing and developing an extensive control quality tool tailored for TCR-seq data, and determine the impact of sequencing platform changes imposed by providers' race for speed, depth and cost. This has been possible, thanks to a massive dataset of 1762 samples obtained by the host laboratory using with two sequencing platforms, HiSeq and NovaSeq. Chapter 2 of this manuscript will be dedicated to the description of the R package Quality control for TCR Repertoires (QtCR) and how it led to strategies to identify reads contamination biases and correct them.

Objective 2: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a viral agent with a pulmonary tropism. In some rare paediatric cases, infection leads to an acute multi-organ inflammatory syndrome with similar characteristics with Kawasaki disease and toxic shock syndrome, two diseases mediated by superantigens. From a paediatric cohort of patient with Multisystem Inflammatory Syndrome in Children (MIS-C), we used non-invasive whole blood sequencing to assess the polyclonality of TRBV12-3 expansions, confirming its relevance with partial results obtained with cytometry and multiplex nucleic acid hybridisation methods. This work, covered in Chapter 3, highlighted that COVID-19 elicited a superantigen T cell immune response in patients with recent SARS-CoV2 infection.

Objective 3: Results obtained from objective 2 showed how TCR sequencing can allow to confirm the pathophysiology mechanism in a context of superantigen stimulation. However, this is one specific case with a strong impact of TCR repertoire perturbation (%TRBV11-2 usage and contraction). As introduced in section 1.2.4, CVD are associated with different type of antigens, some of them being directed against the self. Therefore, the alterations of the repertoire before or after the CVD development might be subtle, and TCR sequencing promise for pathophysiology understanding might be challenging. To start tackling this challenge and determine whether TCR sequencing could be used as a biomarker of CVD, we analysed the TCR sequencing of 28 patients from frozen whole blood, we describe an innovative method to identify a signature of CDR3 that classifies our patients based on their cardiac repair recovery outcome. We confirmed our results on control patients and demonstrate how a distinct set of CDR3 is enriched in the circulating repertoire of patients with good healing outcome.

Objective 4: Chapter 3 identified a circulating T cell signature of myocardial repair associated with good healers. In this part, we sought to find this T cell signature on two other cohorts. First, in cardiac biopsies of patients who died from myocardial infarction. Second, on a cohort of 69 patients with type II diabetes that just suffered myocardial infarction. Here, we show how clustering approaches can be leveraged for repertoires with limited amount of material.

2.2 METHODS AND COHORTS

2.2.1 Cohorts

2.2.1.1 TRiPoD project

The Treg repertoire in Physiology or Diseases (TRiPoD) is a dataset of 1872 murine repertoire from 4 T-cell population and 8 different organs. Samples were all prepared by our lab using the UMI-free SMARTer Human TCR a/b Profiling Kit v1 (TaKaRa Bio). n= 509 samples were sequenced using HiSeq 2500 with 300bp, single-end sequencing. n= 1363 using NovaSeq 6000 with 250bp, paired-end sequencing (see table). Demultiplexing was performed by the facilities and raw FASTQ were aligned using MiXCR on our servers. TriPod murine project represent more than 1700 sequenced TCR libraries, sequenced on 23 batches, spanning an 8 years period. As NovaSeq uses a different chemistry (ExAmp, see 1.3.2), read layout (single-end vs paired-end), the murine TriPoD project was used as a robust benchmark to ensure the reproducibility of sequencing across different platforms.

	HiSeq 2500	**NovaSeq**
Read length (bp)	300	250
Read layout	single-end	paired-end
Sequencing depth	200GB	400GB
Flow Cell splitting	Yes (2)	Yes (2)
Samples processed	192	192
Theoretical Reads/sample	~1 x 10^6	~2 x 10^6

Table 1: Comparison of HiSeq and NovaSeq parameters.

2.2.1.2 The AIR-MI Project

The autoimmune repertoire in myocardial infarction (AIR-MI) is a transnational projects between French (Leader Dr Mariotti-Ferrandiz), Austrian (Pr Peter Rainer) and German laboratories (Dr Gustavo Ramos). AIR-MI project is built on the hypothesis that T cell shaped

adaptive immune processes crucially regulate cardiac repair after injury. The project framework allowed to exploit data from 3 cohorts:

The Etiology, Titre-Course, and Survival (ETiCS) study is an investigator-initiated, prospective multicentre diagnostic study under the auspices of the CNHF (Deubner et al., 2010b). Among the 400 patients, we had access to 150 patients with initial acute myocardial infarction. Patients were followed for 12 months during 4 visits: initial inclusion, 3-, 6- and 12-months. Participants were defined into STEMI or non STEMI according to the observation of acute ST elevation at inclusion following guidelines (Antman et al., 2004). At each timepoints, a comprehensive cardiac assessment is collected. Baseline assessment includes clinical status, results from cardiac catheterisation (left and right heart pressures / haemodynamic), echocardiography, electrocardiogram (ECG), Holter-ECG and blood sampling.

We selected patients with a sufficient initial infarct to assess the clinical recovery, with baseline EF < 50%. Patients were further grouped into two outcomes, good healer and poor healers depending out their cardiac recovery outcome. Patients with %δLVEF < 13% between index hospitalisation and 12-month follow-up were considered poor healers, while %δLVEF > 13% is classified as good healer. Healthy volunteers obtained from Etablissement Français du Sang were collected, with regard to age and sex. Stratification strategy and demographics data are presented in Figure 11.

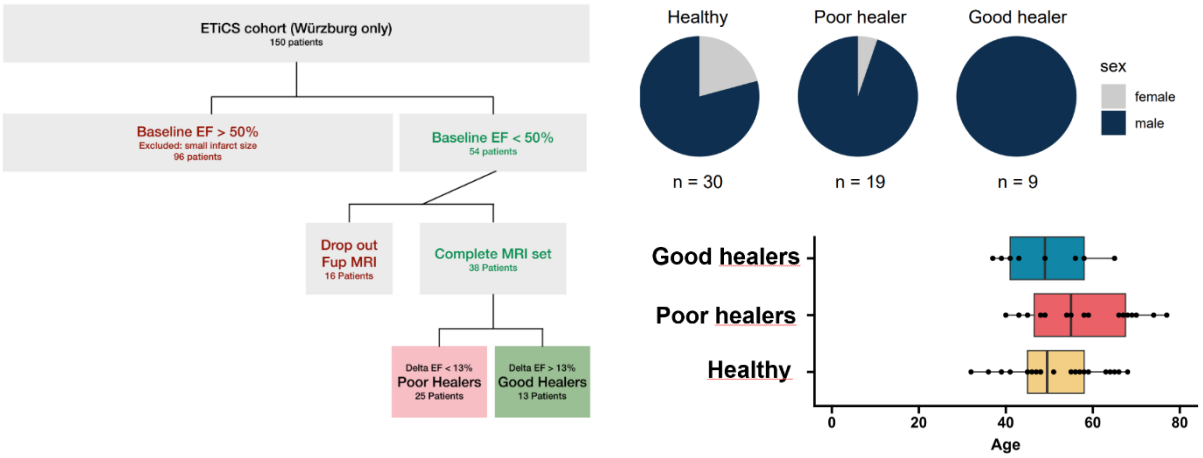


Figure 11: (Left) Patient stratification as good or bad healers from the ETiCS cohort. Post-hoc analysis/stratification scheme of Würzburg patients according to the longitudinal change in their LVEF within 12 months after MI (ΔEF). (Top Right) Sex ratio among patients in cohort. (Bottom right)Age distribution among groups.

In addition to the ETiCS cohort, we had access to n=30 cardiac biopsies from Institute of Pathology biobank (Medical University of Graz), with matched non control tissue (n=30), in

this case lungs. Biopsies were collected from post-MI patients over the last 10 years, and preserved in formalin-fixed paraffin embedding (FFPE). Cardiac biopsies represent fatal early MI events, with matched lungs. Data were prepared and sequenced by collaborators. The role of these samples is not to stratify outcomes of infiltrating T cells, but rather explore the overlap between the circulating cardiac signatures obtained with the ETiCS cohort for comparison.

Lastly, I determined whether our signature obtained with the ETiCS cohort could be generalisable to other patients. To this end, we had access to frozen blood samples of patients from the Empagliflozin in patients with acute Myocardial infarction (EMMY) (Tripolt et al., 2020; von Lewinski et al., 2022). Participants in the EMMY trial either received a 10mg Empagliflozin daily dose, or a placebo (NCT03087773). Similarly, to ETiCS, patient's blood was collected at hospital admission, and cardiac functions were assessed. At 6-month follow-up (Visit 4, furthest timepoint), we determined the change in %δLVEF and stratified patients in poor and good healers groups. Among the 199 samples sent, we sequenced 116 samples with sufficient quality on MiniSeq (Illumina), 300bp single-end.

2.2.1.3 MIS-C paediatric patients

As a part of a national collaboration emerging from the rise of Kawasaki-like syndromes after COVID-19, we investigated the circulating T cell repertoire of paediatric patients. Paediatric patients with north African ethnicities were reporting Kawasaki-like manifestations after COVID-19 infection, called MIS-C, prompting suspicion of a similar superantigen effect. Our collaborators recruited of a cohort of patients with Kawasaki-like clinical symptoms and we performed deep sequencing of their T cell receptor at baseline hospitalisation and ~6 weeks after. We obtained n=16 MIS-C patients' repertoires, with 1- or 2 time-points (hospitalisation or 6-8 week follow-up). In parallel, we obtained the cytometry Vβ profiling or circulatory T cells at hospitalisation, along with Nanostring TRBV profiling.

2.2.2 Methods

2.2.2.1 RNA extraction from frozen whole blood

Cryopreserved whole blood is a very easy way to collect material in patients which comes with great challenges in its handling. Whole blood is a tricky material to work with because of the amount of RNA degrading enzymes (RNAse) it contains, and the labile feature of some

immune cells, such as neutrophils, who tend to die rapidly after collection, releasing RNAse and inducing cell death. Multiple commercial methods have been developed to directly address this issue. One the most popular, PAXgene tubes (PreAnalytix), contains reagents that directly stabilise intracellular RNA upon collection and reduce the risks of losing material (Chai et al., 2005). Other methods consist of separating peripheral blood mononucleated cells (PBMC) and freezing them, to isolate immune cells from granulocytes (neutrophils, basophils and eosinophils) and blood components (erythrocytes and platelets). This prevent the degradation of sample RNA, and enrich blood in T cells. Indeed, T cells make up around 20% of circulating cells in humans, with about 1 x 10^6 T lymphocytes per mL of blood in healthy adults, although it can vary greatly during life events such as active infections or treatments.

Red blood cells globin can take up to 76% of of total mRNA transcripts. A whole trancriptome study from the Genotype-Tissue Expression (GTEx) Consortium have shown red blood cells globin expression from blood samples take ~60% of the transcriptome (Melé et al., 2015). Studies focusing on transcripts have found that β-globins could take up to 76% of mRNA in total blood (Mastrokolias et al., 2012). Benchmarks performed in the lab have shown that β-globin depletion does not affect TCR recovery, or downstream repertoire diversity (results not shown).

Moreover, TCR genes are not among the most expressed genes of T cells. As shown in Figure 12, constant beta chain of the TCR (TRBC2) is expressed between 200 and 400 transcripts per million (TPM) in mature peripheral T cells (Adams et al., 2012), lower than actin-β gene with 700-4000 TPM. As a comparison, β-globin (HHB gene) is expressed at 14200 TPM in erythrocytes in this dataset. This is in-par with other's estimation made of about 500TPM in "pure lymphocyte population" (Brown et al., 2015). This stresses the need to collect RNA of the best quality possible, as degradation will drastically influence the capture of T cell diversity in a sample, both quantitatively and

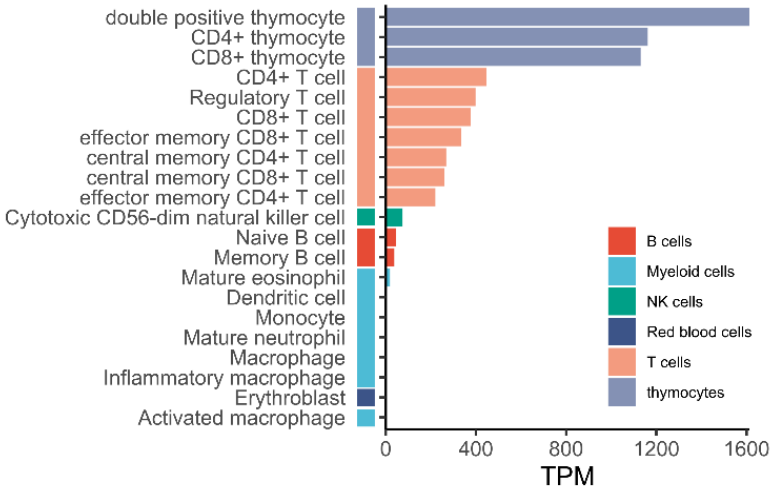


Figure 12: Expression levels of TRBC1-2 gene in circulating populations, expressed in transcript per millions (TPM). Gene expression from BLUEPRINT dataset (Adams et al., 2012)

qualitatively. Indeed, increasing state of degradation of RNA samples lead to poorer performance of RT and amplification steps, and hence recovery of full length TCR prior to alignment (Shen et al., 2018). Genolet *et al.* also have shown how RNA degradation preferentially affects TCRα, hence biasing poor quality library towards TCRβ recovery (Genolet et al., 2023).

RNA quality is computed using the RNA Integrity Number (RIN), assigning integrity value to RNA peaks measurements of ribosomal peaks 18S and 28S (Schroeder et al., 2006). ETiCS cohort was used to benchmark our ability to capture T cell receptor in the worse conditions. Indeed, samples were collected more than 10 years ago, and whole blood was directly stored at -80°C without any preparation. Thawing samples resulted in a very low RNA quality. Thawing cells ruptures their plasma membrane due to ice crystal formation (Pegg, 2010), resulting in the release of RNA of interest in the extracellular medium, where there is an intense RNAse activity. Several studies have benchmarked the best kits and methods to extract RNA from this kind of samples (Kim et al., 2014; Yamagata et al., 2021). They concluded that fast thawing and NucleoSpin RNA blood kit (Macherey-Nagel) were the best combination for yield and RNA purity. Healthy volunteers obtained from EFS were prepared in a similar fashion limit bias, and whole blood was stored at -80°C upon collection for a year prior to library preparation. Despite similar preparation, we obtained low RIN for samples from biobanks. Distribution of RIN numbers obtained from this protocol is shown in Figure 13. RIN scores and dosage were obtained from BioAnalyser 2100 (Agilent) or TapeStation 4200 (Agilent).

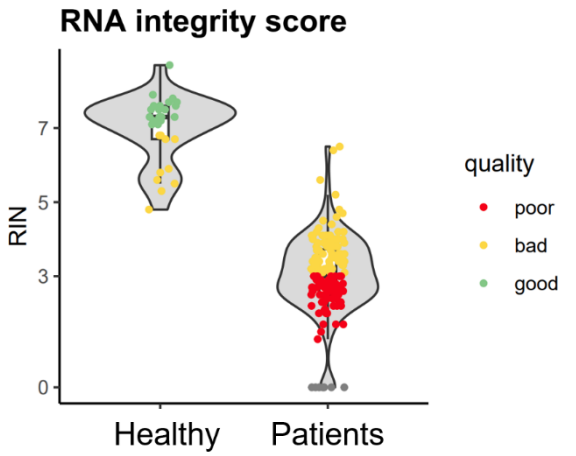


Figure 13: RNA Integrity Numbers obtained from extracted RNA of healthy volunteers or samples of patients. RIN thresholds: poor <3, 3<bad<7, good >7

2.2.2.2 Preparation of DNA TCR libraries and sequencing and validation

300ng of RNA template was used for the preparation of libraries from whole blood TCR. TCR amplification and barcoding was made with SMARTer Human TCR a/b Profiling Kit version 1 (TakaraBio) following their protocol. Experiments were automatised on pipetting robot, when possible, to ensure reproducibility. One minor optimisation was made to the protocol, using twice more DNA template than recommended during second round of PCR

(barcoding step, 2µL of DNA instead of 1). Individual libraries were purified using AMPure XP beads (Agencourt), before qualification on TapeStation 4200. Libraries sequenced in Miniseq were further purified on agarose gel electrophoresis migration (1% agarose), and DNA purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel).

We stress on the purification part, as it is a crucial and often overlooked step before sequencing. We noticed striking improvements when performing pool purification before sequencing, in all systems used in this thesis (MiniSeq, HiSeq or NovaSeq). Pool purification, although not performed usually in our pipeline, showed important improvement of yield when done with care (results not shown).

Equimolar pooling of libraries was done before sequencing on dedicated platforms (MiniSeq, HiSeq or NovaSeq). Demultiplexing was done by platforms (LIGAN, Institut du Cerveau et de la Moelle, or Institut Pasteur). Raw sequence quality was assessed using FastQC (Andrews, 2010). TCR sequence alignment was performed using MiXCR v3.0.13 (Bolotin et al., 2015) using recommended parameters for TaKaRa 5' RACE kit strategy. Validation of TCR data in accordance with initial material is performed using our QtCR package, developed in Chapter 0.

3 RESULTS

3.1 CHAPTER 3: IDENTIFICATION AND CORRECTION OF INDEX HOPPING THROUGH (DEEP) QUALITY CONTROL OF TCR REPERTOIRES

The TCR sequencing field is in constant evolution with rapidly evolving methods. The race to increasing multiplexing, deeper sequencing is often at the cost of a thorough quality control (QC). Indeed, QC is a critical step in data analysis to avoid drawing compromised conclusion based on experimental biases. In TCR-seq, poor data quality can be attributed to initial sample retrieval, library preparation or sequencing parameters. So far, there is no dedicated tool to integrate these information. To this end, I have developed Quality Control for T cell Repertoires (QtCR), an integrative pipeline that aggregates the relevant information from all steps prior to sequencing to identify discrepancies and batch effects.

A recent benchmark published by our team showed how the different commercially available methods for TCR repertoire library preparation could yield biases in downstream analysis (Barennes et al., 2021). Moreover, another critical step during repertoire data production is the sequencer. Each generation of sequencer brought deeper sequencing. For instance, Illumina's ExAmp technology came with their NovaSeq series (Shen et al., 2014) . However, most datasets produced in the laboratory were sequenced using HiSeq 2500, which featured different read length, non-paired data and a lower depth.

To evaluate how these parameters could affect gene mapping and downstream analysis such as diversity metrics, I had access to a dataset of a murine atlas of more than 1762 TCR samples produced by the laboratory in the last 6 years. Libraries were all produced using the same protocol, but were sequenced on two different platforms: HiSeq 2500 and NovaSeq 6000, which featured the ExAmp technology.

The QtCR tool and its demonstration are detailed as a manuscript submitted for publication. In brief, we showed how different depth of sequencing did not affect global clustering of sample nor diversity metrics in this dataset. In contrast, we demonstrated how single- and paired-end read layouts affected the MiXCR gene alignment, along with the different reads. Applying a normalisation of read length and layouts, we showed that the observed batch effects were corrected. Finally, the analysis of samples sequenced with NovaSeq 6000 showed unexpected clustering of samples sharing same sequencing index. This

observation suggested a well-known phenomenon associated with the recent Illumina sequencing chemistry, namely index hopping, that lead us develop a novel method to remove such contamination in batches of samples produced with the NovaSeq sequencer.

Identification and correction of index hopping through (deep) quality control of TCR repertoires

Kenz Le Gouge[1], Paul Stys[1,*], Nicolas Coatnoan[1,2,*], Gwladys Fourcade[1,*], Vanessa Mhanna[1,2], Pierre Barennes[1,2], Hélène Vantomme[1,2], Vimala Diderot[1], Otriv Nguekap-Tchoumba[1,2], Nicolas Tchitchek[1], Adrien Six[1], David Klatzmann[1,2], Encarnita Mariotti-Ferrandiz[1,3,§]

[1]Sorbonne Université, INSERM, Immunology-Immunopathology-Immunotherapy (i3), F-75005 Paris, France.

[2]AP-HP, Hôpital Pitié-Salpêtrière, Clinical Investigation Center for Biotherapies (CIC-BTi) and Immunology-Inflammation-Infectiology and Dermatology Department (3iD), Paris, France

[3] Institut Universitaire de France

§ **Corresponding author**

***Equal contribution**

Encarnita Mariotti-Ferrandiz
UMRS959, Immunology-Immunopathology-Immunotherapy lab
Sorbonne Université/INSERM
83 boulevard de l'Hôpital
75013 Paris
France
Phone: +33 1 42 17 74 68
Email: encarnita.mariotti@sorbonne-universite.fr

Contributions:

KLG: conceived, developed and implemented QtCR, wrote the manuscript draft

NC : supervised the libraries' preparation and set-up the molecular experimental protocols

GF: supervised and performed the cell biology experiments, collected and curated the metadata

VM, KLG, PB, HV, VD, ONG: performed the cell biology experiments and produced the libraries

PS: collected and standardised the metadata

NT: provided feedback in the package design

AS: supervised the metadata collection and curation

DK: supervised the experiments and obtained funding

EMF: conceived and supervised the study, obtained funding, wrote the manuscript draft

3.1.1 Abstract

Integration of high quality T cell receptor sequencing (TCR-seq) datasets is necessary to build large-scale atlases of immune receptors. Yet, there is no specialized tools ensuring a thorough quality control (QC) of the data. We thus developed QtCR, an integrative R package designed to identify and quantify discrepancies in TCR-seq data on a multi-scale level. We tested QtCR in a murine dataset of 1762 samples generated with variations in the protocol due to technology evolution. We found that read length, but not sequencing depth, introduces strong batch effects that can be mitigated by harmonizing sequencing and alignment layouts. We also showed that NovaSeq 6000 sequencing introduces index hopping contamination and how it affect on sample similarity. To tackle this issue, we developed a prior-free algorithm to filter signal from noise and established its performance by restoring a biologically relevant clustering in highly contaminated data.

3.1.2 Introduction

The widespread use of high throughput sequencing and the ever-increasing depth of the new sequencers have recently revolutionised the field of adaptive immune receptor repertoire (AIRR). This increase in popularity has been pushed with the help of the development and commercialization of ready-to-use kits with efficient primers, which allowed relatively robust and consistent results[1]. Recent developments in sequencing platforms offer multiplexing through dual indexing of reads, which allowed an increasing number of simultaneously sequenced samples in a single run, with up to 96 for most of the TCR kits. The constant evolution of how AIRR sequencing (AIRR-seq) data are produced, processed and analysed [2] stresses the need to harmonise pipelines of data production through robust benchmarking[1]. As such, a scientific society, the AIRR community, has been founded and is leading several studies to compare and set guidelines on the good practices of AIRR-seq data[3–5]

However, these good practices may not alleviate the main issue of the reliability of the data production. AIRR-seq, just like any other sequencing approach, is very sensitive to data production issues[6,7]. Sample collection, extraction, and preparation before sequencing are often carried out by multiple experimenters, who can also be different from the ultimate end-user (the bioinformatician) analysing the data. This compartmentalisation of data production and analysis further complexifies the identification of library preparation issues. To mitigate these issues, authors have proposed to use technical replicates [8] or spike-ins controls[9,10], but this comes at a prohibitive cost, especially for large cohorts. Thorough quality control and data collection at every step of the library production then appear as a critical step towards data analysis to avoid drawing compromised conclusions on technical biases[4].

To date, very few solutions exist to assess the quality of AIRR-seq. FastQC is a widely used tool to check for poor raw output from the sequencer but was not developed to handle alignment and mapping metrics[11]. On the other hand, the most popular AIRR-focused alignment tool, MiXCR, was recently updated and offers AIRR-tailored QC functions related to the mapping and alignment[12]. However, these tools do not take into account the many technical and experimental parameters that may have influenced these metrics. Notably, contamination is a very important issue that is hardly solved, as many factors can lead to it, whether it happens during collection, storage, library preparation or during sequencing[6,13–15]. The latter is often due to a recently introduced phenomenon called index hopping[16], where sequencing reads are not correctly attributed to samples due to misassigned indexes and identified in many other applications[17–20]. A widely used technique is singleton removal[21], accounting for errors in

sequencing, but remains limited for mid- or high-level contamination. Although a recent approach has been developed to tackle the issue of T-cell receptor sequencing (TCR-seq), it requires extensive *a priori* knowledge and assumptions of the contamination, such as control "clean" datasets, which hinders its potential uses[22].

Therefore, given the widespread use of AIRR-seq for pathophysiology study and disease diagnosis, we believe it is of utmost importance to have efficient quality control tools that should enable the detection of contamination, but also technical issues to identify technical outliers from biological outliers. To this end, we first focused on TCR-seq as our main field of research and developed the Quality control for T cell receptor Repertoire, QtCR, an integrative tool to aggregate relevant data from the library preparation to sequencing output of TCR-seq data. QtCR aggregates all the data collected by experimenters during the library preparation, along with alignment and TCR-seq specific metrics. QtCR offers a wide range of QC metrics with absolute and relative thresholds to identify abnormal data and outliers about the input material identifying source of contamination, batch effects or meta-batch effects.

We validated the performance of our tool with a large dataset of a murine TCR atlas covering eight years of continuous production, multiple alignment references version, different sequencer platforms and sequencing facilities, as well as various mouse genetic backgrounds and sorted T cell subsets.

3.1.3 Methods

Mouse experiments

Six- to 80-week-old male and female C57BL/6- and NOD-Foxp3-EGFP transgenic mice expressing GFP under the control of the Foxp3 gene promoter were, respectively, provided by B. Malissen (Luminy, Marseille, France) and V. Kuchroo (Brigham and Women's Hospital, Boston, MA). All animals were maintained at the Sorbonne Université Centre d'Expérimentation Fonctionnelle animal facility under specific pathogen-free conditions in agreement with the current European legislation on animal care, housing, and scientific experimentation (agreement number A751315). All procedures were approved by the local animal ethics committee (Paris, France)

Cell preparation

As detailed in a previous study[23], cells from various lymphoid organs were harvested and incubated with fluorescent antibodies prior to being sorted on a FACSAria II cytometer (BD

Biosciences) with >95% purity into the four subsets: CD4+ FoxP3+ (Tregs), CD4+ FoxP3- (Teffs), CD8+ FoxP3- (CD8 T cells) and thymic CD4+ CD8+ (precursors). RNA was extracted using the RNAqueous Total RNA Isolation Kit (Invitrogen).

Library preparation

Starting from 100 ng of RNA quantified with NanoDrop (ThermoFisher Scientific), 1762 TCR librairies were prepared with the SMARTer Mouse TCRa/b Profiling Kit V1 (Takara Bio), assisted with a pipetting robot Ascia (PRIMADIAG) allowing the simultaneous preparation of 96 libraries on PCR plates. Importantly, library position on 96-well plates were randomised prior to library preparation. Libraries were then quantified prior to sequencing using BioAnalyzer 2100 (Agilent) or TapeStation 4200 (Agilent).

Sequencing and pre-processing of TCR seq data

TCR libraries were sequenced with either HiSeq 2500 single-end (300 bp) (Illumina) or NovaSeq 6000 paired-end (250 bp) (Illumina) and demultiplexed by the different facilities.

First, we used FastQC[11] v0.11.9 to assess the reads and overall runs quality. Qualified raw FASTQ files were originally aligned using MiXCR[12] v3.0.3 (HiSeq 2500) or MiXCR v3.0.13 (NovaSeq 6000), reflecting evolution of tools over time. Non-productive or ambiguous TCR were filtered out, along with TCRs with CDR3 between 14 ± 8 amino-acid lengths. When mentioned, clonotypes with counts of 1 (singletons) were removed.

Sample metadata

We collected and compiled several variables for each samples for the investigation of biases in murine models that can be characterized in four large categories: murine models (genetic background, intervention), biological material (organ, cell type, number of sorted cells), library production (barcodes, RNA concentration, cDNA concentration), or sequencer related (type of sequencer, run, flowcell lane).

Generation of the QtCR objects

Original metadata file, FastQC files, MiXCR alignment reports and MiXCR alignment files were loaded into QtCR. For all samples, sequenced reads, aligned reads, percentage of aligned reads were parsed from MiXCR reports. Similarity matrices (Jaccard and Morisita-Horn) and Renyi profiles were computed from aligned files using the *vegan* package[24] based on clonotypes

incidence matrix, along with TRV and TRJ gene usage. Principal Component Variance Analysis (PVCA) was computed using a slightly adapted code from the *pvca* Bioconductor package[25] to make it more flexible. Principal Variance Component of indicated meta-variables are computed from raw matrices of either similarity or gene usage. Matrices are then centred but not scaled, and PVCA is performed with a variance threshold of 0.3 We used either the non-linear uniform manifold approximation and projection (UMAP) algorithm or the linear principal component analysis (PCA) for reduction of high dimensions matrices into two dimensions. Low dimension projections on the 2D space are computed from TRV and TRJ gene usage matrices, Jaccard and Morisita-Horn similarity matrices for TRA and TRB chain separately and plotted using *ggplot2*[26].

Detection of diversity outliers

Outliers are detected using the Local Outlier Factor (LOF), an algorithm identifying distant points in high dimensional space. LOF is determined as the mean reachability of a point A across its k nearest neighbours[27]. Single LOF score interpretation is limited, so we summed it across multiple values of *k* to encompass multiple scales of seclusion. For a set of *n* samples, the total LOF score of a point *A* is as follows:

$$LOF_{total}(A) = \sum_{k=3}^{[\sqrt{n}]} LOF_k(A)$$

Using Tukey definition of an outlier, we used the 1st quartile and interquartile range to define the outlier threshold

$$Outlier\ threshold = Q1 + 1.5 * IQR$$

Samples for which $LOF_{total}(A) > Outlier\ threshold$ are considered outliers.

Decontamination of TCR repertoires using DeconTCR

DeconTCR is an original method that uses alignment data to find enriched clonotypes in samples sharing a feature (here, indexes). Two clonotypes with different nucleotide sequences can lead to the same amino acid sequence. This creates similarity between samples despite sequences arising from different clones. We define this process as collision, which generates false positives during decontamination. Collision is less likely to happen using nucleotide sequences rather than amino acid. Indeed, the 14 bases long example CDR3 "**CASSALASLNEQFF**" can be derived from 2 *4* 6* 6* 4* 6* 4* 6* 6* 2* 2* 2* 2* 2 =

31,850,496 distinct nucleotide combinations. To mitigate false positives due to collision, a clonotype is defined as TRV-CDR3nt-TRJ. The DeconTCR method first selects clonotypes shared by at least 30% of samples in a given index. It then computes the enrichment of each clonotype based on presence/absence per index using a Fisher-test score. For each enriched clonotype, a Shannon filter threshold is applied to its count distribution among all other samples of the same index. Samples for which their counts are considered as noise are discarded. DeconTCR performs a benchmark across different thresholds for the sharing fraction and fisher enrichment score. An extensive comparison of each combination is evaluated by the user, which defines the best trade-off between sensitivity and specificity of the filters.

DeconTCR Shannon filters were computed using the vegan package[24]. Determination of the contamination threshold is done by computing the Shannon entropy (H'). For a given clonotype X shared by n samples with p counts, the formula is as follows.

$$H'(X) = \sum_{i}^{n} ln(p_i^{pi})$$

Computing the exponential of this formula gives the rank of the threshold to discriminate signal from noise. A list of dummy distributions is shown in **Supplemental figure 8** to show its performance.

Decontamination performance is assessed on the amino acid clonotype definition to compare the results with previous analysis. Jaccard and Morisita-Horn similarity matrices were computed, along with their low-dimension projections as described in the "Generation of QtCR objects" methods section. The number of clonotypes filtered and corresponding number of counts filtered for each sample are collected to assess performance.

Statistical tests

Parametric student *t-test* and non-parametric Wilcoxon *U-test* p-values were computed using R. P-values were considered significant for $p \leq 0.05$. For effect size, parametric Cohen's *d* and non-parametric Cliff's delta (δ) were used to assess the magnitude of differences between conditions.

3.1.4 Results (1600 words)

3.1.4.1 Dataset presentation

A comprehensive atlas of 1872 TCR-seq repertoires was used for this study. Due to sequencing platforms evolution, 509 (27%) repertoires were sequenced with HiSeq 2500 and 1363 (73%) with NovaSeq 6000 (Illumina) on 22 batches of up to 96 samples (**Fig. 1**). Samples contained alpha (TRA) and beta (TRB) sequences from 2 genetic backgrounds (NOD, C57Bl/6) and 16 projects (combination of genetic backgrounds, sex, age and treatment), sorted T cells in 4 main subsets, $CD4^+$ T cells (Teff), CD8 T cells, CD4 regulatory T cells (Tregs) and thymic immature T cells (precursors). For each sample, genetic background, organ of origin, sorted cell number, nucleic acid concentration after extraction and library preparation as well as library preparation and sequencing batches, plate position and Illumina indexes, sequencer platform, and finally experimenter identification for each of the steps were compiled into a metadata file (**Table 2**). In parallel, TCR-seq data were aligned as described in the material section. Both metadata file and TCR aligned files were merged using QtCR. Among the additional relevant parsed parameters, FastQC sample quality, MiXCR version were collected and used in further analysis.

3.1.4.2 Differences of sequencing depth does not impact sample clustering

When using this massive dataset with samples collected and processed over a span of several years (**Supp. Fig. 1**), we initially assessed the reliability of our approach between the two sequencer platforms. Indeed, HiSeq 2500 was discontinued in 2021, and we switched to NovaSeq 6000, which offered twice more reads per sample but with a different read length (**Table 1**). We compared the amount of sequenced and processed material between HiSeq and NovaSeq (**Fig. 2A**). As expected, NovaSeq yielded significantly more raw reads and resulted in more reads aligned ($p < 0.001$, Cohen's $d = 0.58$). Interestingly, the better yield of reads was accompanied by a significantly lower percentage of aligned reads and strong effect size of the aligner for the NovaSeq ($p < 0.001$, Cohen's $d = 0.41$) (**Fig. 2B**). This quantitative difference was not associated with a qualitative imbalance between TRA and TRB mapped reads (δ Cliff = 0.10). We looked further into the difference between the alignment performance and observed that the percentage of aligned sequences varied between batch, suggesting external

factors associated with the library preparation or the sequencing run? (**Fig. 2C**). To assess whether this quantitative difference was associated with differences in generated data, we compared the diversity of TRA and TRB chain rearrangements and found no major differences for both chains when comparing the Shannon entropy (δ Cliff = 0.11 and 0.13 respectively) (**Fig. 2D**). As gene usage is highly linked to cell subsets and genetic backgrounds, we investigated whether the differences of technologies resulted in a sequencer-driven clustering. To measure this, we used principal variance component analysis (PVCA) on the gene frequency matrices to quantify the amount of variance associated with a set of variables from the QtCR-compiled metadata (meta-variable). Sequencers accounted for less than 5% of the weighted variance, confirming that at the global scale of our dataset, the sequencer used had no impact on clustering (**Fig. 2E**). Interestingly, we found that the genetic background was the major driver of variance for the TRA chain, whereas cell type was the main contributor for the TRB chain. We confirmed this result by projecting the TRBV usage on a 2-dimension space using principal component analysis (PCA) or UMAP projection, confirming no visible bias (**Fig. 2F**).

To evaluate the impact of the sequencer platform on the clone distribution and sharing, we computed the Jaccard and Morisita-Horn (MH) similarity indices on the list of clonotypes, which respectively measure the overlap and similarity of distribution of a variable, here a clonotype, between series of samples. Again, we found no association between sample similarity and the sequencing technology for both chains (**Fig. 2G-H**).

3.1.4.3 Read length alignment partially drives T cell subsets' clustering

We then investigated the sequencer impact on single population of cells. If the sequencer had no impact on the global variance of four populations and two genetic backgrounds, it might have been "diluted" by the samples' heterogeneity. To observe more subtle differences the sequencer might have introduced, a subset of a homogeneous population of $CD8^+$ T cells of C57Bl/6 was selected from the whole dataset (**Fig. 3A**). Local Outlier Factor (LOF) was used to remove samples with abnormal diversity, which might introduce unwanted variation (**Fig. 3B**). This method accurately detected and removed under-sequenced samples (**Supp. 2**). Only samples within boundaries in both TRA and TRB chains were used for the subsequent analysis. We then computed PVCA on three experimental variables associated with the dataset,:the

harvested organ, the lane and the sequencer platform. Strikingly, almost 100% of TRA Morisita-Horn and TRAV gene usage variability was attributed to the sequencer platform variable (**Fig. 3C**). This has been further confirmed on a 2D projection of those two matrices by PCA (**Supp. 3A-B**). We obtained similar results in the NOD background and in the Teff subset (**Supp. S4**). Given the specific impact of sequencer platform on the TRA rearrangements, we hypothesised that incorrect TRAV genes mapping lead to the high differences in MH index, and was likely due to the alignment parameters. As the main difference between the two sequencers parameters is the read layout (300bp single-end for HiSeq, 250bp paired-end for NovaSeq), we realigned all the samples by varying the read length, using paired-end reads or single-end reads and two distinct version of the aligner. As such, the NovaSeq 250bp paired-end read samples (n=21) with were realigned as 250bp single-end reads using MiXCR version 3.0.13 (SE_250_NS_3.0.13). The 300bp HiSeq single-end read samples (n=32) were realigned with the newer aligner version 3.0.13 of MiXCR (SE_300_HS_3.0.13), or trimmed single-end read to 250bp (SE_250_HS_3.0.13) (**Fig. 3D**). We compared these conditions with the original ones, namely 250bp paired-end NovaSeq (PE_250_NS_3.0.13) and 300bp single-end HiSeq (SE_300_HS_3.0.3). We computed the pairwise Jensen-Shannon divergence (JSD) matrix between TRAV genes usage in these different conditions to find the most robust method for both NovaSeq and HiSeq data and projected samples on the UMAP low dimension space (**Fig. 3E**). Paired-end layout features the most dissimilar TRAJ usage of all methods. Interestingly, both 250bp-SE sets of samples showed the most similarity, even more similar than the duplicated 300bp vs $250_{trimmed}$ bp HiSeq samples. To measure how overlapping the distributions were, we defined the identity score of the TRAV distributions by computing the sum of the absolute frequencies' differences between our reference (SE_250_HS_3.0.13) and all the others (**Fig. 3G**). NovaSeq 250bp SE alignment scored the highest identity (92% identity), confirming its relevance. Paired-end alignment not only has the lowest identity score, this analysis also revealed that paired-end alignment identified TRA genes that were not observed in the other methods. These genes can be seen at the tail of the gene usage distribution. We computed the gene usage and similarity PVCA on the HiSeq and NovaSeq samples aligned with SE-250bp (**Fig. 3G**). We completely removed the sequencer effect on this homogeneous set of samples, without affecting much of the other meta-variables. These results show that different sequencing technologies with different read

layouts lead to unexpected gene mapping, which can be circumvented by realigning raw repertoires on the harmonised layouts.

3.1.4.4 Index hopping introduced by NOVAseq is detected by QtCR

To determine the extent of individual batch effects, we computed the PVCA of each batch separately on a set of relevant variables: project, cell type, and the Illumina reverse and forward indexes corresponding to the lines and column of our plates (**Fig. 4A**). As expected, HiSeq batches showed a very strong contribution of the cell type to the clustering. However, NovaSeq rather exhibits a concerning domination of the reverse index impact on the variance, with up to 80% of the total Jaccard matrix variance, but not in MH. This result suggested that a high amount of low-counts clonotypes were shared between samples with the index, as Jaccard gives the same weight to all clonotypes, no matter their abundance. To better understand these results, we focused on the Batch 19, featuring almost 80% of cumulative index-attributed variance for the Jaccard TRB index (**Fig. 4B**) and TRA chain (**Supp. 5**). Hierarchical clustering of the TRB Jaccard index showed a clearly defined reverse clustering, confirming the results from the PVCA (**Fig. 4C**). These patterns of high similarity between the reverse index were not found in the MH matrix (**Fig. 4D**), where sample clustering were more likely influenced by the cell subset of origin. Sequencing plate layout was designed randomly ruling out that observed similarity was due to comparable samples being placed on the same lines or column (**Fig. 4E**). However, when looking at the individual Jaccard scores of a random sample on a plate (*e.g.*, position E11), we observed a very characteristic crosshair pattern (**Fig. 4F**) described by other groups and attributed to index hopping[28,29]. We also confirmed the PVCA results by showing that samples from the same reverse (8/8) or forward (11/12) had a significantly higher score than samples from different indexes (**Fig. 4G**). Altogether, these results are consistent with an index hopping contamination due to the switch from HiSeq to NovaSeq sequencing. As index hopping is a random phenomenon, it should be correlated with the number of reads[17,29]. To confirm this, we computed the correlation between raw sequenced reads and Jaccard index between two repertoires, and found that repertoires sharing a sequencing index had positively correlated Jaccard index (same reverse: R = 0.47, same forward: R = 0.55, no common index: R = 0.18) (**Fig. 4H**). These results could be extended to all batches, as the mean contribution of reverse is positively correlated to the amounts of reads on NovaSeq (R=0.45) (**Supp. 6**).

3.1.4.5 Correction of index hopping using information theory

Index hopping appeared to be a random process at the sequence/read/template level leading to the index being exchanged between reads. In the case of redundant dual indexing (RDI), it means assigning them to another sample on the same line or column (**Supp 7**). Thus, we hypothesised that the more common a read is, the more likely it is to be found on all the samples of a same line or a same column, with the fold change equal to the probablility of index hopping (P_{hop}) * the number of indexes. Estimations of P_{hop} are in the range of 1/10 and 1/1000 [16,29]. In line with these estimates, in our data, the P_{hop} was estimated at about 1% (**Supp 11**). This contamination results in high frequency clonotypes more likely to be spread among indexes, with several logs of magnitude lower than in the original sample. Index hopping can then be assimilated to noise, spreading from an original source of signal. In order to remove artificial similarity between samples due to index-hopping, we developed DeconTCR, an algorithm that relies on information theory entropy to identify noise from the original signal (**Fig. 5A**). First, we identified the TCRs that were significantly enriched for a reverse or a forward, and shared by at least 30% of all samples with the same index. Each enriched clone count is then ranked against all the other clones of the same index. To determine a filter threshold, Shannon entropy is computed on the count distributions. Decontamination performance was then benchmarked based on the balance diminution of same index similiarity and the number of sequences filtered. This allowed to determine that a filter threshold of 0.01 was an optimal trade-off (**Supp. 9A-D**).

Applied on the 95 samples of "Batch 19", we show that the original UMAP projection clustering was driven by reverse indexing (**Fig. 5C-D)**. Removing singleton did not affect clustering. A combination of DeconTCR + singletons removal offers the best performance to restore the expected cell type clustering**, Supp 10**).

3.1.5 Discussion

In this work, we described how QtCR, an integrative quality control pipeline featuring tools for identifying variables contributing to variance, is used to identify biases from data of different production source in TCR-seq sequences. From a set of samples of one cell population, to 96 multiplexed samples in a single run of multiple T subsets, to an aggregation of thousands of repertoires from different runs, QtCR efficiently identified expected and unexpected technical impacts of sequencing.

Currently, there is no dedicated tool to measure and assess biases introduced by technical and experimental factors. Here we provide a toolbox to analyse the impact of metadata on data. The package can be used on the command line for the most advanced functions, but it also features a user-friendly web interface. Metadata MiAIRR[5,30,31] compliant and can also accept any other type of information, such as plate position or the nucleic acid concentration of the sample. The detailed information of how this is processed by QtCR is described in the online documentation.

Evaluated on 1872 samples sequenced in 96 multiplexed-sample batches, we showed how differences in sequencing depth did not affect large scale analysis between repertoires. Although HiSeq 2500 featured fewer reads than NovaSeq 6000, when we computed high granularity metrics such as clonotypes diversity, TCR gene usage or similarity distances at the scale of a large dataset, we found that integrating data obtained from both sequencers did not critically impact the sample clustering.

When then investigating in more detail how the read length difference between HiSeq 2500 (250bp, single-end) and NovaSeq (300bp, paired-end) could influence the downstream analysis. We first showed how QtCR accurately detected and removed outliers. We then demonstrated on the curated dataset of CD8 T cells how murine TRAV gene mapping was flawed between paired-end and single-end read pairing. A benchmark of read modifications before alignment and MiXCR[12] versions demonstrated that an harmonisation of alignment methods dramatically reduces the gap between the repertoires. This result highlights how a consistent alignment procedure is required for clonotype mapping, and stresses the need for an open alignment method description to ensure robust reproducibility and future reuse of data by other scientists.

Finally, QtCR identified a pattern of contamination due index hopping in our dataset. Index hopping is not a new issue, and recent kits were developed with unique dual indexing (UDI) to drastically reduce the possibilities of hopping through impossible combinations[32]. However, any dataset that has been produced using redundant dual indexing (RDI) needs to address these issues, especially for reuse in further studies. Single-cells datasets are also on the hook. Indeed, most commercial kits use 96 UDI, but it does not completely alleviate the problem and leads to “phantom” cells[33], which can also lead to the false discovery of TCRs. This is of the utmost

concern as most of the new strategies developed heavily rely on the sharing of TCR to infer clinical conclusions[34].

In order to mitigate with index hopping, we designed DeconTCR, an unbiased method for decontaminating data based on shared features. Using this novel and efficient method, we showed how DeconTCR restored a biologically accurate clustering after removing the index-induced contamination with minimal alteration of the data. DeconTCR works best in a random distribution of samples to limit possible confounding factors; hence, it might not be as efficient with all datasets. Randomising samples is a recommended practice for any experimental design[35–38], and our method confirms that TCR multi-batch projects also need careful planning.

Our work focused on TCR-seq, but can also be ported to any AIRR-seq. BCR-seq shares many common characteristics with TCR-seq, such as the need to capture a full-length receptor through specific primers, an immense diversity of possible receptors generated, and very similar analysis methods[2,10,21,39]. BCR-seq and TCR-seq are often pre-processed in a comparable fashion [12,40,41], thus QtCR could be adapted and optimised to support BCR-QC with minimal modifications.

This study emphasises the importance of providing a comprehensive metadata with publicly available datasets. We also stress the importance of specifying the parameters used for alignment. AIRR community Minimal Standards Working Group set standards for metadata in collaboration with NCBI to facilitate the sharing of such information[30]. This practice is imperative for upcoming research initiatives as it will enhance our capacity to comprehend and effectively compare NGS derived datasets. Our findings strongly encourage thorough quality control before any AIRR-seq integration of data from different origins (batches, protocols, labs…).

Our investigations show the importance of collecting experimental metadata, since repertoires produced with different sequencing methods or aligned differently cannot be compared without adjustment. Metadata should be as standardised as possible to facilitate data sharing and comparisons[5]. However, reality shows that overly comprehensive or technical standards are a barrier to correct metadata filling (**Supp 12**). This points to the need for a simple tool, accessible to all, to fill in metadata as experiments progress (and thus avoid disconnection between those who produce the data and those who analyse it[42].

Furthermore, with QtCR we identified the impact of the alignment procedure on data comparison.. Therefore, although obvious, these results emphasize the critical need of sharing raw data rather than processed ones. This will not only allow evaluating how comparable datasets could be, but also ensure accurate comparison especially when gene reference change.

Altogether, QtCR is a user-friendly and flexible tool that should enable advanced quality control, identification of real outliers and allow intervention when contaminations are suspected. The flexibility of the tool should allow it to identify other sources of bias or contaminations not covered in this study simply by adding additional variables in the metadata. This tool should permit a better integration of AIRR datasets from various platforms and experimental design by determining the impact of technical bias on biological interpretations, opening data sharing to a new era of data reuse and comparison.

3.1.6 Figures and legends

3.1.6.1 Legends

Figure 1 : QtCR workflow and dataset presentation
A : QtCR workflow overview. Experimental metadata is collected during all all TCR libraries production steps. We use QtCR to aggregate all the data and compute TCR-related metrics. These metrics are then compared to experimental metadata to identify i) outliers based on discrepancies between aligned and expected output, and ii) experimental or biological features associated with sample clustering.

Figure 2 : Differences of sequencing depth does not impact sample clustering
A : Violin plots of the amounts of raw reads (left) or aligned reads (right) between HiSeq 2500 (n=509) and NovaSeq 6000 (n=1363) TCR-seq samples.
B : Violin plot of the percentage of aligned reads out of the sequenced reads (left), or the proportion of TRB reads (right) between HiSeq 2500 and NovaSeq 6000 TCR-seq samples.
C : Violin plot of the percentage of aligned reads per batch.
D : Shannon index diversity between HiSeq and NovaSeq samples, TRA (left) and TRB (right).
E : proportion of gene usage variance explained by each meta-variable.
F : PCA (left) and UMAP (right) projection of sample TRBV gene usage.
G : Proportion of similarity matrix variance explained by each meta-variable.
H : PCA (left) and UMAP (right) projection of sample TRA jaccard index.

Figure 3 : Read length alignment partially drives T cell clustering
A : Highlighted C57Bl/6 young females CD8 T cells repertoires PCA projection of the complete murine dataset.
B : Distribution of TRA and TRB LOF scores of C57Bl/6 young females CD8 T cells repertoires with inliers (green) and outliers (red). Dashed red lines correspond to the outlier threshold.
C : Proportion variance explained by each meta-variable in similarity matrix and genes usage.
D : Jensen-Shannon Divergence of TRAV genes usages between conditions.
E : Identity score between TRAV usage distributions compared to HiSeq 2500 samples aligned with 250pb, single-end and 3.0.13 version of MiXCR. TRAV genes are ordered by decreasing frequencies found in condition SE_250_HS_3.0.13 (green).

Figure 4 : Unexpected index hopping introduced by NOVAseq is detected by QtCR
A : PVCA contribution of meta-variables among all batches, for Jaccard and MH indexes. Batches with less than X samples were discarded as we could not properly evaluate dual indexing impact.
B : PVCA scores of “Batch 19”.
C-D : Heatmap of the scores of Jaccard (C) and MH (D) index between the Batch 19 samples (n=95)
E : Distribution of cell subsets in the original plate sent to sequencing.

F : Jaccard overlap scores of Batch 19 compared to sample S_0 (red).
G :Jaccard means score of Batch 19 sample sharing a reverse (top) or a forward index bottom). Confidence intervals expressed in standard deviation. Wilcoxon U-test.
H : Amount of reads in a sample and associated Jaccard score

Figure 5 : Correction of index hopping using DeconTCR
A : Index hopping mechanism. On redundant dual indexing (RDI), reverse index (i5) and forward (i7) are used to individually barcode each sample. In a 96 well plate, the 8 reverse and 12 forward correspond to lines and columns, respectively. During sequencing, non-specific annealing of barcodes happen at a random rate and introduce incorrect read assignments.
B : DeconTCR workflow. 1) Reads enriched for each barcode are identified. 2)We use the Shannon entropy to identify the source of contamination (true signal) from contaminated samples (noise). This process is repeated for clonotype across all indexes. 3) Freshly filtered samples are then reanalysed to qualify the amount of contamination eliminated and improvement of biologically-related clustering.
C : Projection of the contaminated dataset after removing singletons, going through DeconTCR, or both. Top panels indicate cell type clustering, bottom panels indicate reverse clustering.
D : Jaccard similarity matrix of Batch 19 prior and after going through DeconTCR and singletons removal. Samples were arranged either based on their cell type (top) or reverse (bottom).
E : Percentage of clone counts remaining in samples after using DeconTCR for decreasingly stringent fisher thresholds compared to the original unfiltered dataset.
F : Percentage of clone counts remaining in samples after using DeconTCR for decreasingly stringent fisher thresholds compared to the original unfiltered dataset.
G-H : For each reverse (G) or forward (H), we computed the mean intra- and extra-cluster distances (respectively dotted and full lines) for the Jaccard TRB matrix.

3.1.6.2 Figures

Figure 1

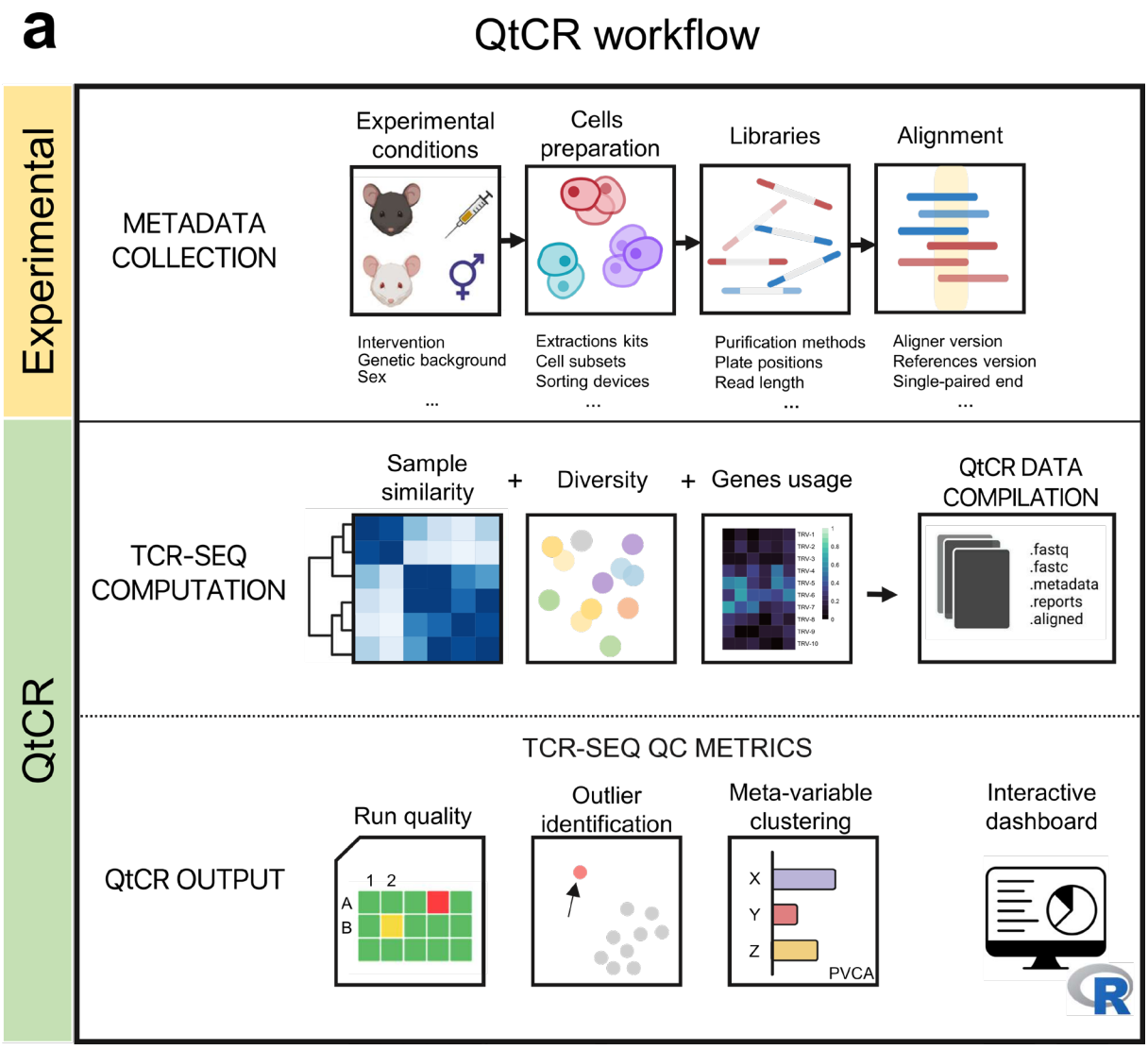


Figure 2

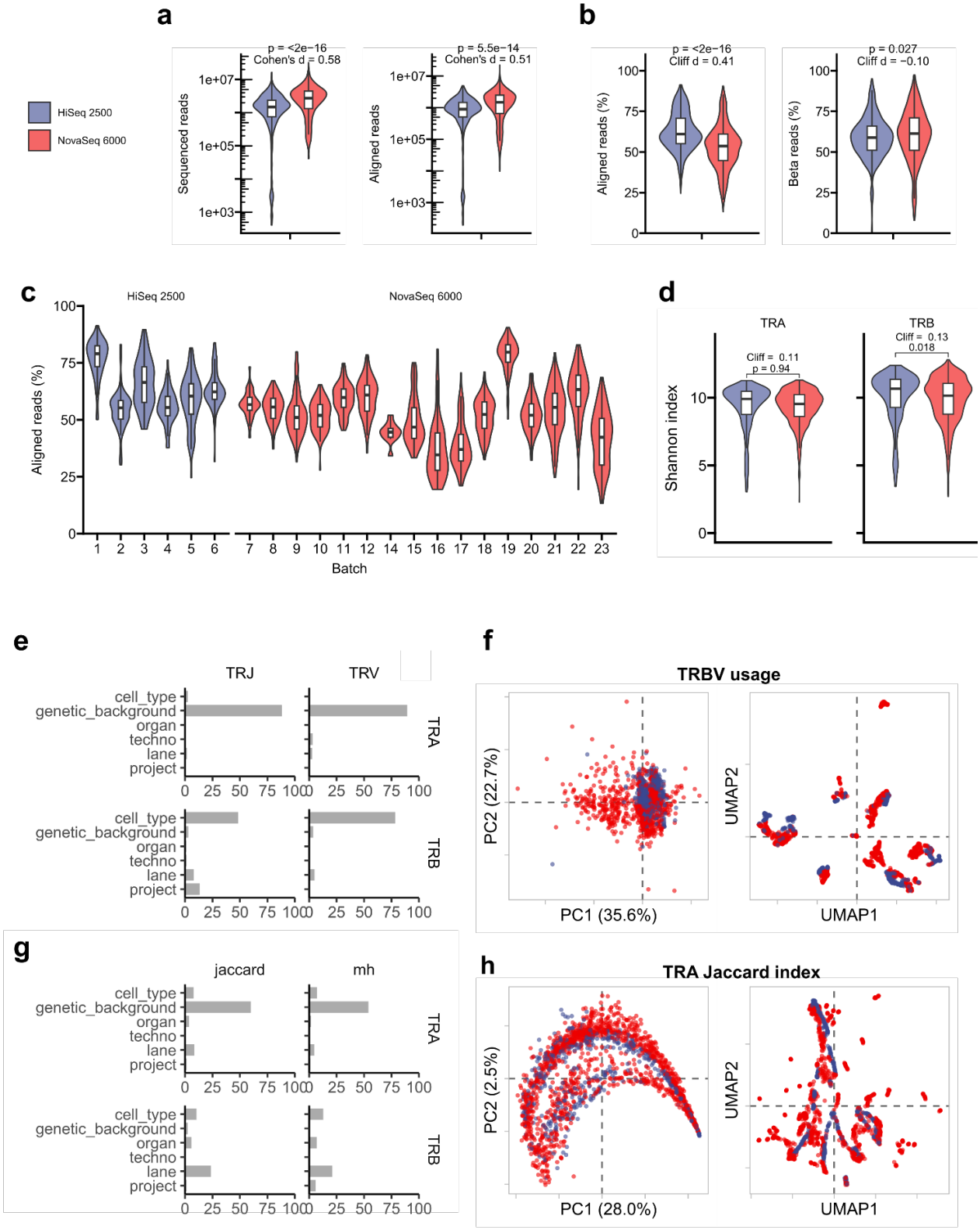

a
b
HiSeq 2500
NovaSeq 6000
Sequenced reads
Aligned reads
Aligned reads (%)
Beta reads (%)
p = <2e−16
Cohen's d = 0.58
p = 5.5e−14
Cohen's d = 0.51
p = <2e−16
Cliff d = 0.41
p = 0.027
Cliff d = −0.10
c
d
HiSeq 2500
NovaSeq 6000
Aligned reads (%)
Batch
TRA
TRB
Shannon index
Cliff = 0.11
p = 0.94
Cliff = 0.13
0.018
e
f
TRJ
TRV
TRA
TRB
cell_type
genetic_background
organ
techno
lane
project
TRBV usage
PC2 (22.7%)
PC1 (35.6%)
UMAP2
UMAP1
g
h
jaccard
mh
TRA Jaccard index
PC2 (2.5%)
PC1 (28.0%)

Figure 3

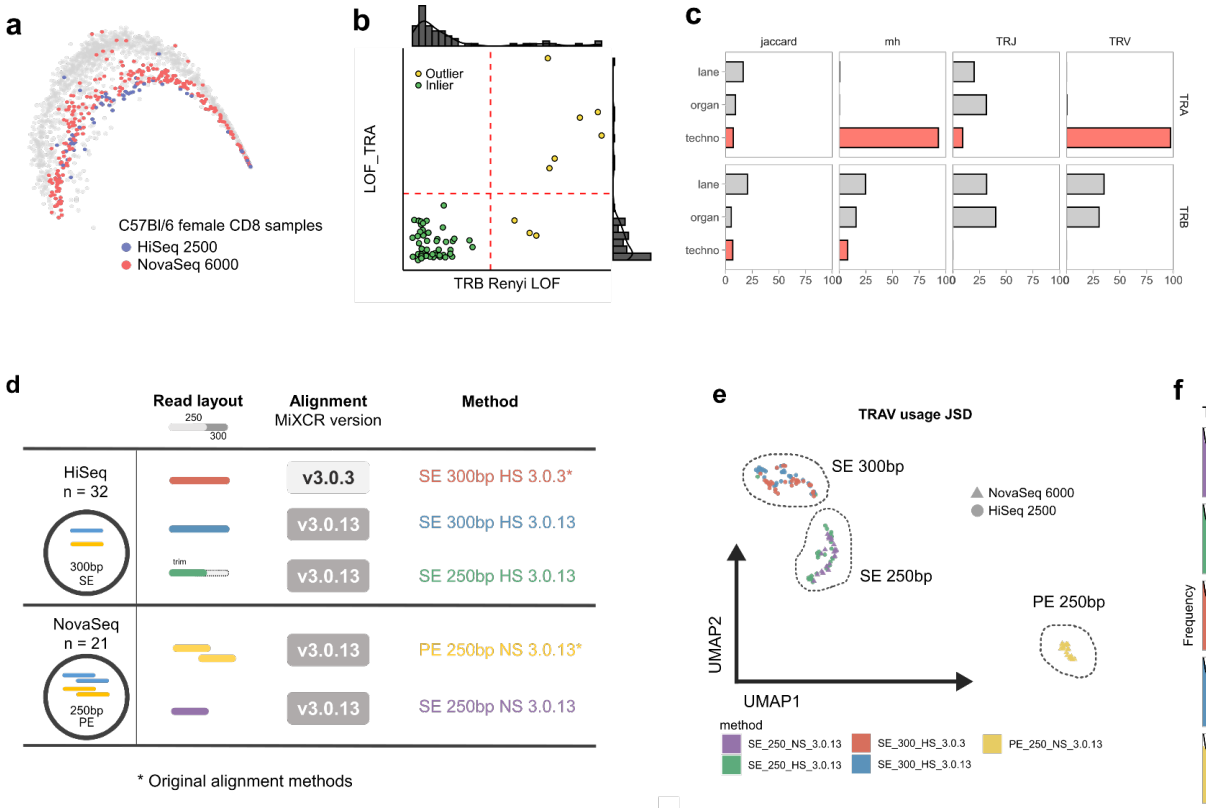

a
C57Bl/6 female CD8 samples
HiSeq 2500
NovaSeq 6000
b
Outlier
Inlier
LOF_TRA
TRB Renyi LOF
c
jaccard
mh
TRJ
TRV
lane
organ
techno
TRA
TRB
0 25 50 75 100
d
Read layout
250
300
Alignment
MiXCR version
Method
HiSeq
n = 32
300bp
SE
trim
v3.0.3
SE 300bp HS 3.0.3*
v3.0.13
SE 300bp HS 3.0.13
v3.0.13
SE 250bp HS 3.0.13
NovaSeq
n = 21
250bp
PE
v3.0.13
PE 250bp NS 3.0.13*
v3.0.13
SE 250bp NS 3.0.13
* Original alignment methods
e
TRAV usage JSD
SE 300bp
SE 250bp
PE 250bp
NovaSeq 6000
HiSeq 2500
UMAP2
UMAP1
method
SE_250_NS_3.0.13
SE_300_HS_3.0.3
PE_250_NS_3.0.13
SE_250_HS_3.0.13
SE_300_HS_3.0.13
f
TRAV frequency across alignments
1.00
0.92
0.87
0.87
0.77
Frequency
TRAV
g
jaccard
mh
TRJ
TRV
organ
lane
techno
TRA
TRB
0 25 50 75 100

Figure 4

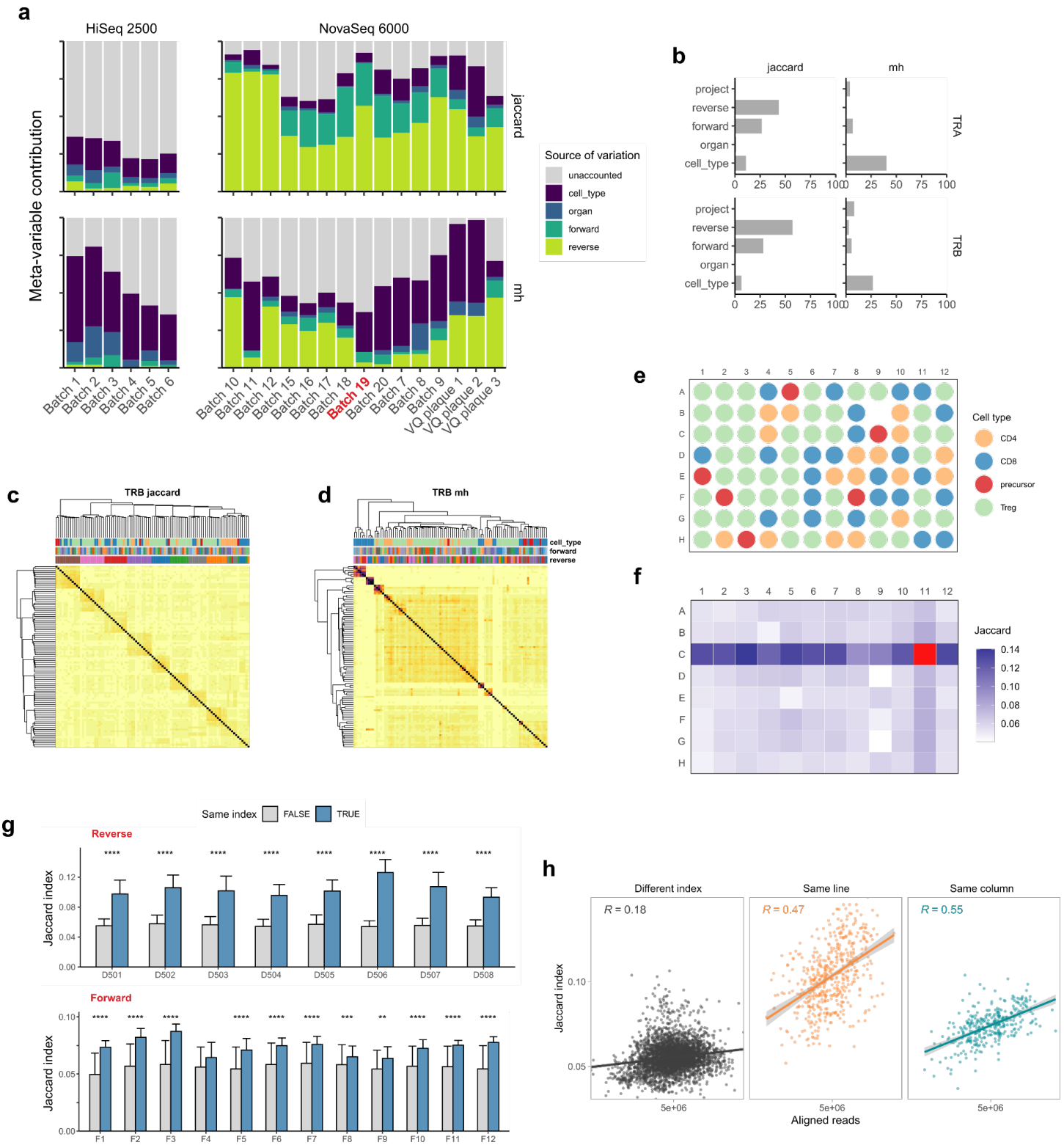
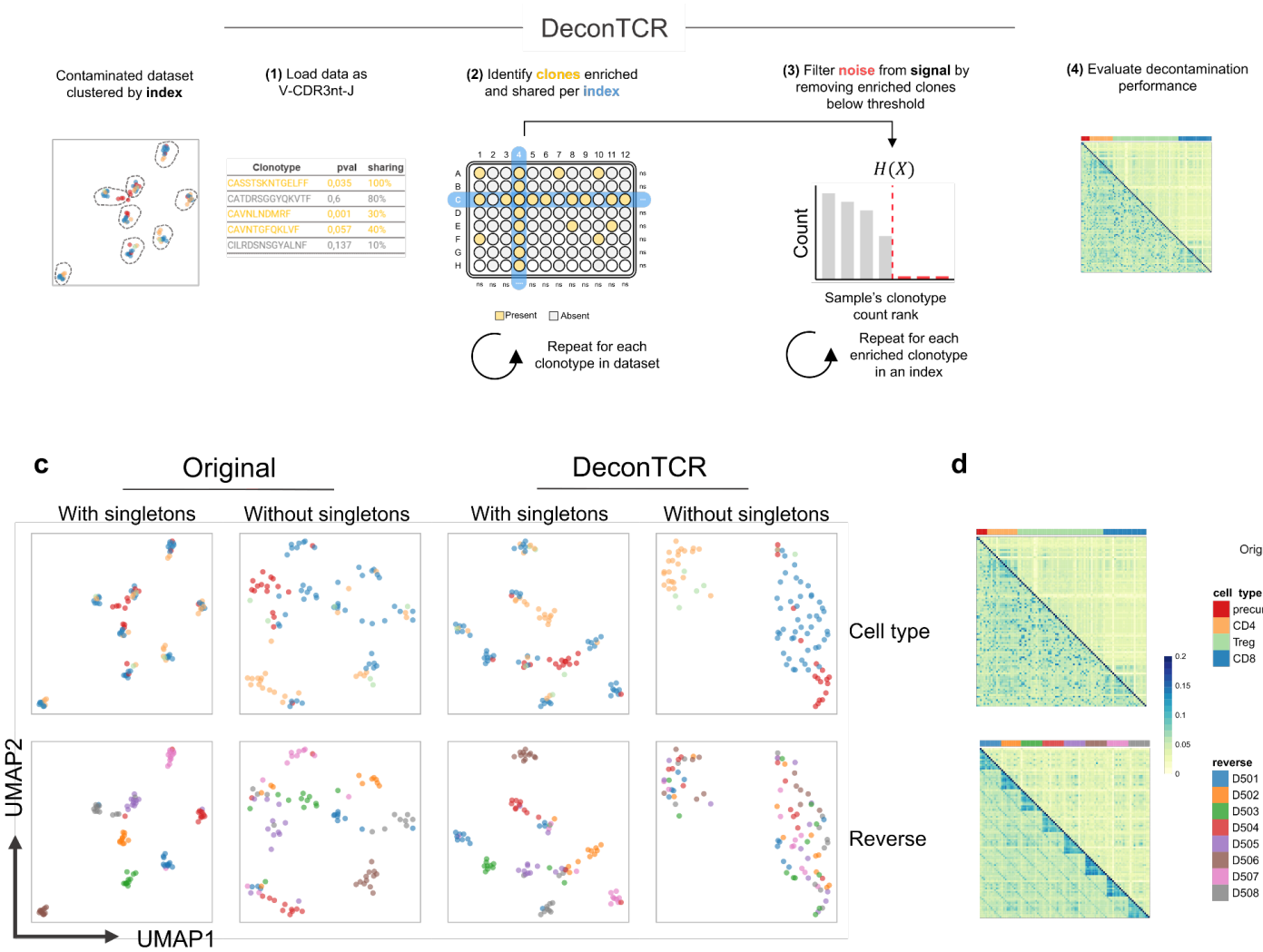

a
HiSeq 2500
NovaSeq 6000
Meta-variable contribution
jaccard
mh
Source of variation
unaccounted
cell_type
organ
forward
reverse
Batch 1
Batch 2
Batch 3
Batch 4
Batch 5
Batch 6
Batch 10
Batch 11
Batch 12
Batch 15
Batch 16
Batch 17
Batch 18
Batch 19
Batch 20
Batch 7
Batch 8
Batch 9
VQ plaque 1
VQ plaque 2
VQ plaque 3
b
jaccard
mh
project
reverse
forward
organ
cell_type
TRA
TRB
0
25
50
75
100
c
TRB jaccard
d
TRB mh
cell_type
forward
reverse
e
Cell type
CD4
CD8
precursor
Treg
f
Jaccard
0.14
0.12
0.10
0.08
0.06
g
Same index
FALSE
TRUE
Reverse
Forward
Jaccard index
D501
D502
D503
D504
D505
D506
D507
D508
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
F11
F12
h
Different index
Same line
Same column
R = 0.18
R = 0.47
R = 0.55
Jaccard index
Aligned reads
5e+06

Figure 5


DeconTCR
Contaminated dataset clustered by index
(1) Load data as V-CDR3nt-J
Clonotype pval sharing
CASSTSKNTGELFF 0,035 100%
CATDRSGGYQKVTF 0,6 80%
CAVNLNDMRF 0,001 30%
CAVNTGFQKLVF 0,057 40%
CILRDSNSGYALNF 0,137 10%
(2) Identify clones enriched and shared per index
Present
Absent
Repeat for each clonotype in dataset
(3) Filter noise from signal by removing enriched clones below threshold
H(X)
Count
Sample's clonotype count rank
Repeat for each enriched clonotype in an index
(4) Evaluate decontamination performance
c
Original
DeconTCR
With singletons
Without singletons
With singletons
Without singletons
Cell type
Reverse
UMAP2
UMAP1
d
Filtered
Original
cell type
precursor
CD4
Treg
CD8
reverse
D501
D502
D503
D504
D505
D506
D507
D508

3.1.6.3 Tables

Table 1: Specifications of sequencers HiSeq 2500 and NovaSeq 6000 used in this study

	HiSeq 2500	**NovaSeq**
Read length (bp)	300	250
Read layout	single-end	paired-end
Sequencing depth	200GB	400GB
Flow Cell splitting	Yes (2)	Yes (2)
Samples processed	192	192
Theoretical Reads/sample	~1 x 10^6	~2 x 10^6

Table 2: Description of metavariables used in the murine dataset analysis in QtCR

sample_id	unique identifying name for each library
project	Category combining both intervention and genetic background
run	sequencing run unique ID
lane	sequencing run unique ID + flowcell lane (1 or 2)
species	Mus Musculus
reverse	Illumina D50X index
forward	Illumina D7XX index
techno	Sequencer used (HiSeq 2500 or NovaSeq 6000)
organ	Organ from which cells were sorted
cell_number	Number of sorted cell before lysis
genetic_bakground	Mouse genetic background
rna_conc	Concentration of libraries before preparation
cell_type	Sorted cell types were aggregated into precursors, CD4, CD8 and Treg

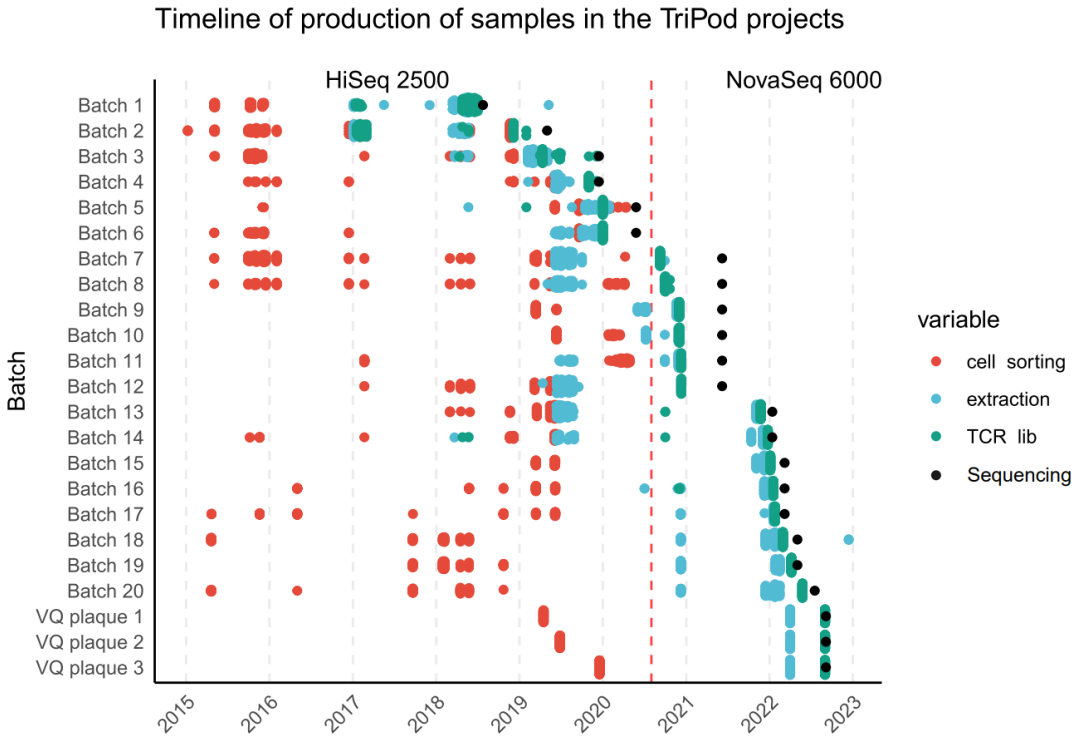
Supplementary data

Acknowledgements
The authors thank UMR 8199 LIGAN-PM Genomics platform (Lille, France), which belongs to the “Federation de Recherche” 3508 Labex European Genomics Institute for Diabetes (ANR-10-LABX-46) and was supported by the ANR Equipex 2010 session, and the iGenSeq platform of the Institut du Cerveau (Paris, France) for performing the sequencing of our libraries.
Funding
KLG/EMF: KLG doctoral fellowship was supported by ERANET-CVD JCT2018 (ANR-18-ECVD-0001) and additional support from Sorbonne Université. EMF work was supported by the iReceptorPlus (H2020 Research and Innovation Programme 825821) and SirocCo (ANR-21-CO12-0005-01) grants and the Institut Universitaire de France. The data used in this study have been funded by the TriPoD European Research Council Advanced EU (322856).
To be completed

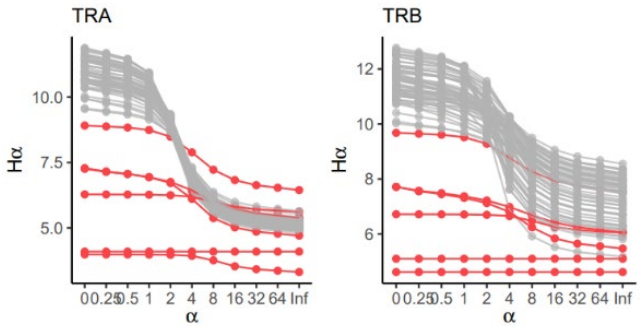
Data availability
Not available until paper released.
A part of the dataset is already available as it was part of a prior study(Mhanna et al., 2021), where FASTQ data were deposited in the National Center for Biotechnology Information Sequence Read Archive repository under the BioProject identification number PRJNA635928. As a large portion of the dataset is still under active analysis, the remaining part of the dataset will be made available along with the publications.

Code availability
Code for the QtCR package is available at https://github.com/kenzlegouge/QtCR
Code for the DeconTCR package is available at https://github.com/kenzlegouge/DeconTCR

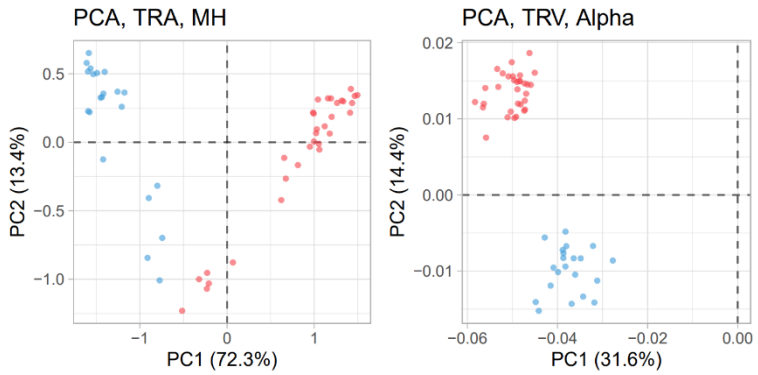
3.1.6.4 Supplementary figures



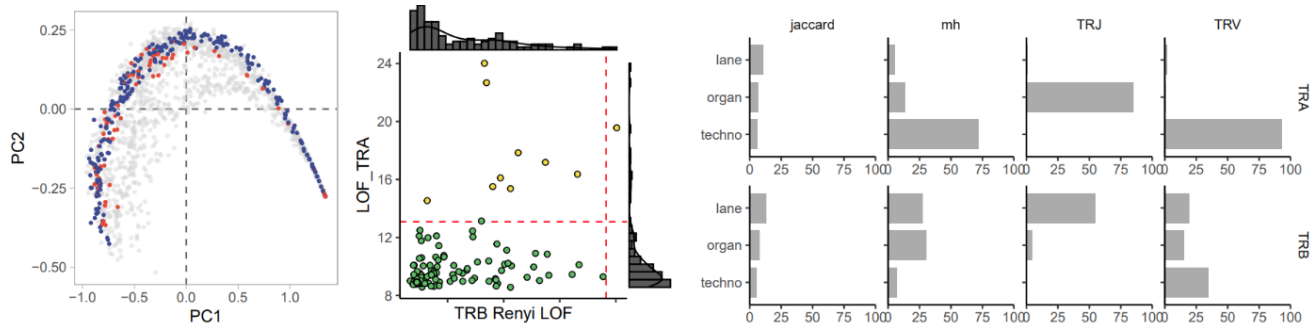
Supplemental 1 : Timeline of production of Tripod samples. The dataset we used features cells that were sorted back in 2015 and sequenced until mid 2022. Most samples were sorted before 2020, year at which the facility stopped supporting HiSeq 2500 sequencers and swithed to NovaSeq 6000 (red line).



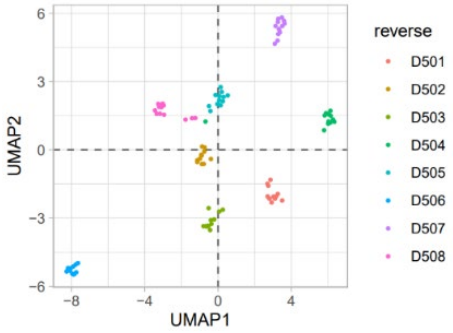
Supplemental 2 : Outliers based on the TCR diversity Renyi profile that were further removed from the analysis of Figure 2. QtCR uses Local Outlier Factor (LOF) to detect samples that are isolated to other when projected onto a space. Please refer to the Methods section for more details on the approach.



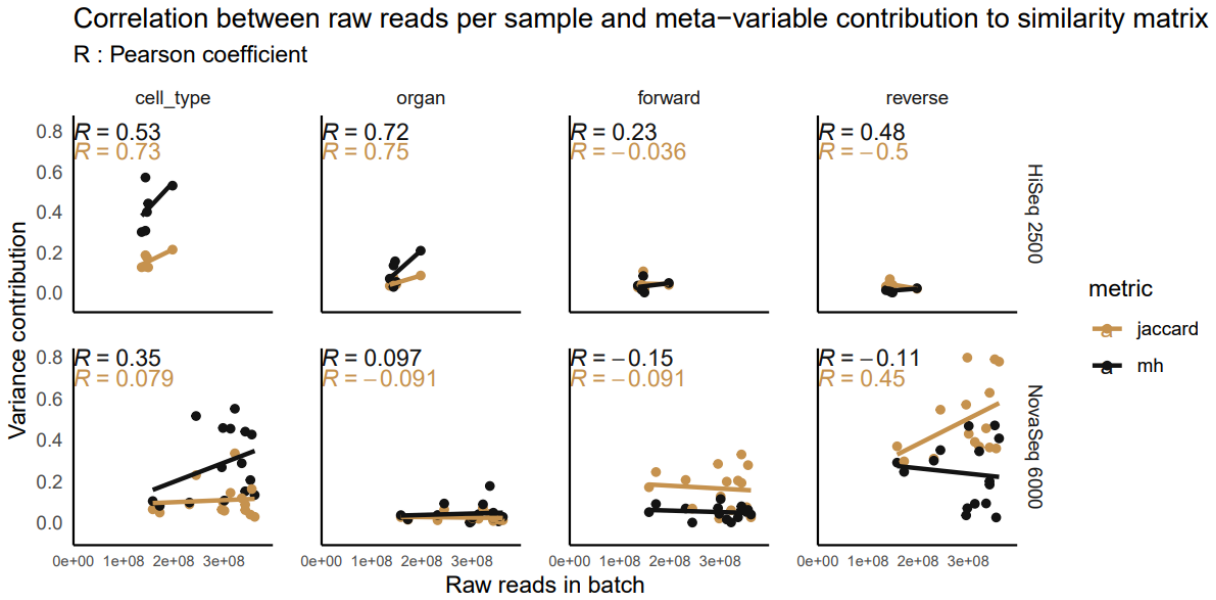
Supplemental 3: Principal Component Analysis of C57Bl/6 young females CD8 cells
A : PCA of the Morisita-Horn similarity of TRA clonotypes on samples sequenced using HiSeq 2500 (blue) or NovaSeq (red)
B : PCA of the TRAV usage on samples sequenced using HiSeq 2500 (blue) or NovaSeq (red)



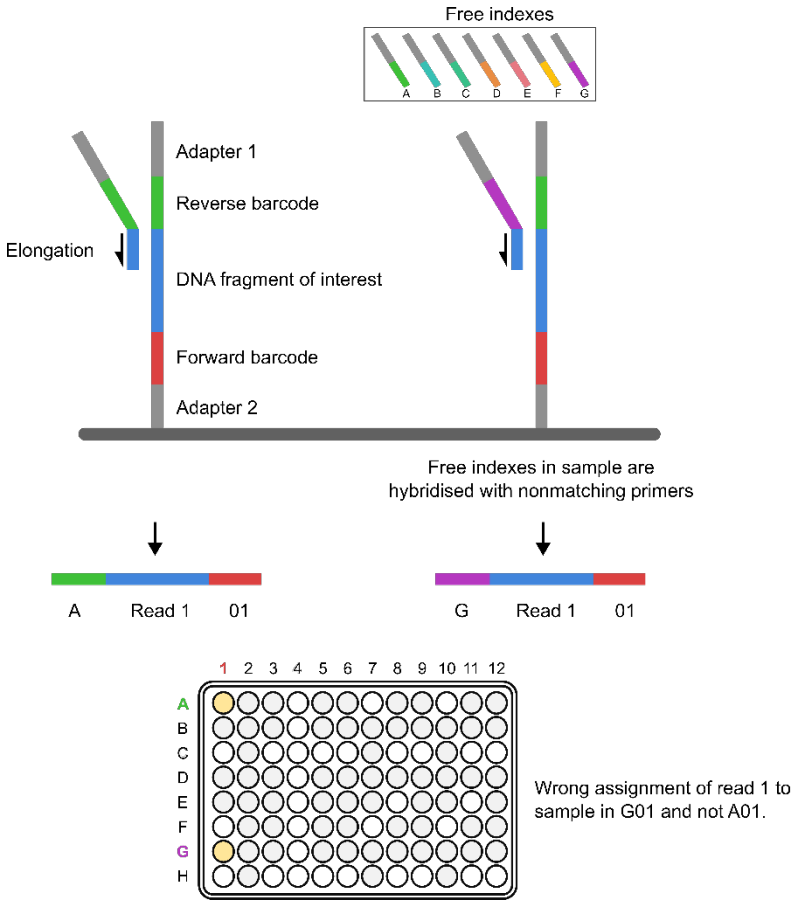
Supplemental 4 : **A** : Highlighted C57Bl/6 young CD4 T cells repertoires PCA projection of the complete murine dataset.
B : Distribution of TRA and TRB LOF scores of C57Bl/6 young females CD4 T cells repertoires with inliers (green) and outliers (red). Dashed red lines correspond to the outlier threshold.
C : Proportion of variance explained by each meta-variable in similarity matrix and genes usage



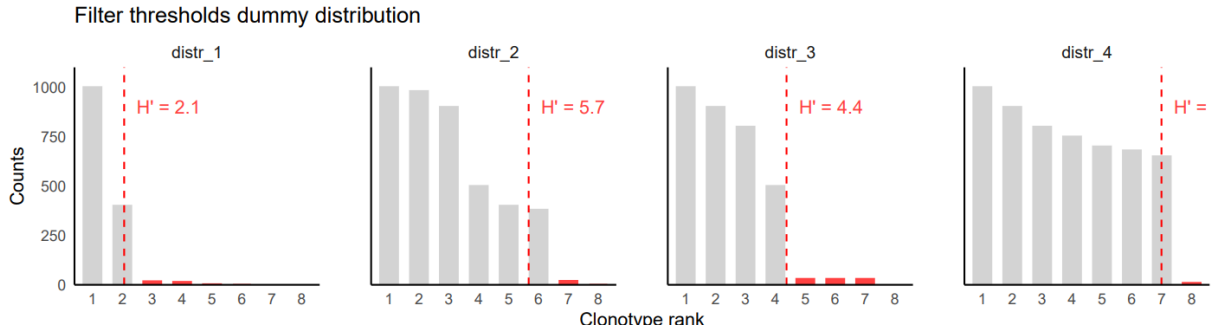
Supplemental 5 : UMAP of Jaccard index, TRB. Clustering of samples is driven by the reverse index. Samples from TRA chain are also clustered by reverse.



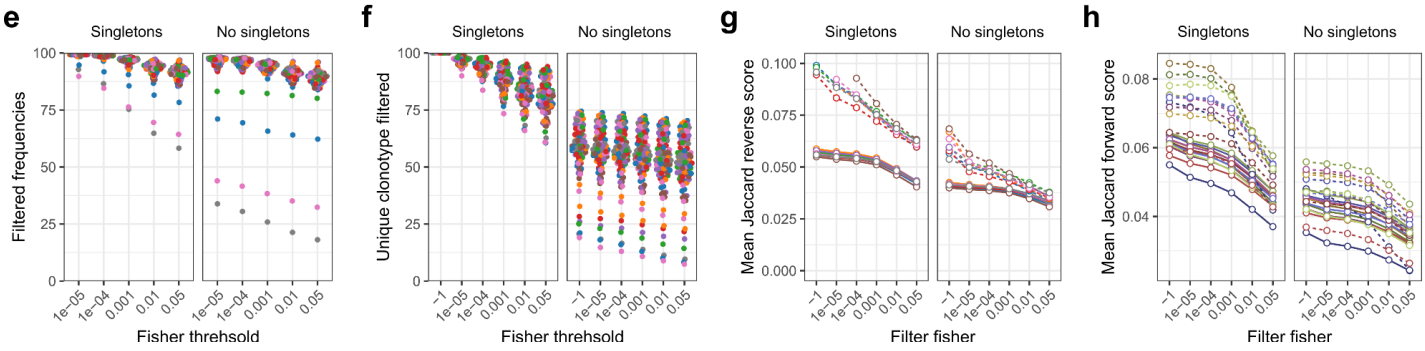
Supplemental 6 : Correlation between the total amounts of raw reads in a batches compared to PVCA contribution of each meta-variable (cell type, organ, forward and reverse) to the overall variance of Jaccard or Morisita Horn similarity Matrix. Pearson coefficient.



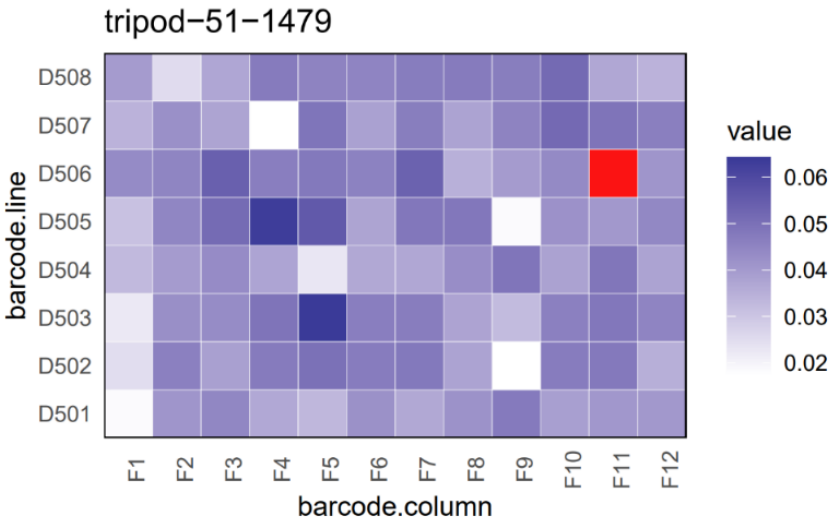
Supplemental 7: Mechanism for dual indexing hopping and line/column contamination. In this example, a DNA fragment has been hybridised to the flowcell. (left) Priming of the reverse results in elongation of the read and correct barcode sequencing. (right) Free indexes remaining in the sample after poor purification of libraries or non-specific hybridisation results in elongation of a read with another barcode. During multiplexing, this will lead to the wrong assignment of reads. In this example, this lead to initial read in A01 to be assigned the sample in G01.



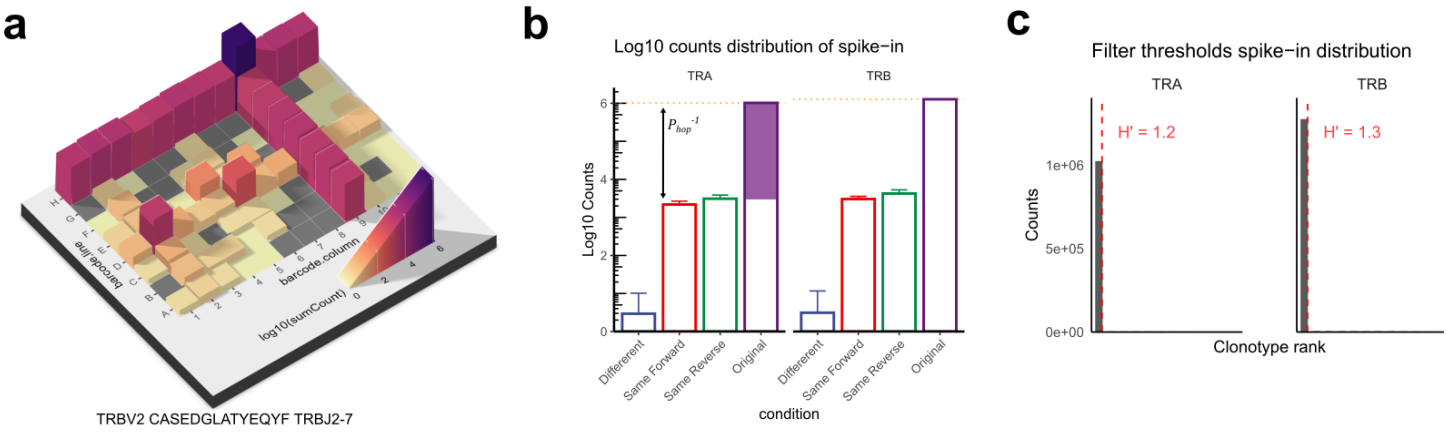
Supplemental 8 : 4 examples of counts distribution and the Shannon threshold defined to filter noise from signal. In all examples, the Shannon threshold was able to correctly distinguish between a strong signal and background noise, even when confronted to more subtle differences (distr4).



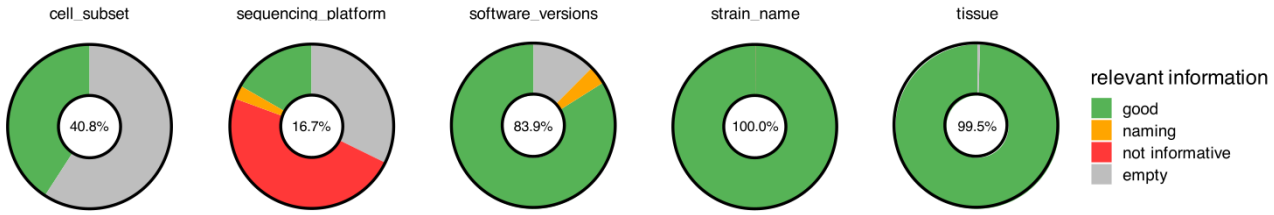
Supp 9: Evaluation of different threshold performance on decontamination of Batch 19 (n=95 samples). X axis represent the Fisher test pvalue used for considering significant contamination. A higher pvalue indicates less stringent threshold, and thus more read to be filtered. For each analysis, we plotted the results with (left) or removed singletons (right) after decontamination.
E: Amount of the cumulative frequencies of clonotypes filtered per index. Each point is a sample, colored by reverse.
F: Percentage of unique clonotypes remaining in samples after decontamination. Each point is a sample, colored by reverse.
G: Mean Jaccard reverse score across samples. Jaccard distance between samples sharing a same reverse (solid) or using different (dashed) reverse were plotted. Ideally, there should not be a difference between those two lines, as samples are placed randomly. We see an optimal of clustering at 0.01
H: Similar plot as G, but using forward primers.



Supplemental 10 : Sample "tripod-51-1479" similarity after decontamination of Batch 19 with DeconTCR. Similarity was performed using Jaccard index.



Supplemental 11: Index hopping probability calculation. To estimate P_{hop}, we used a spiked-in library from a 96 well batch sequenced with NovaSeq. On position H9, we sequenced a unique clone (`TRBV2 CASEDGLATYEQYF TRBJ2-7`) with 10^6 counts associated.
A : distribution of counts in the plate. A very typical cross pattern is observed in line H and column 9.
B: Assuming this is only due to contamination, the ratio between contamination and original signal can be estimated to P_{hop}.
C : we confirmed DeconTCR performance in identifying noise. DeconTCR confidently identified that only the 1st clone was true signal.



Supplemental 12 : iReceptor metadata investigation reveals incomplete filling.
We analyzed the user-filled metadata from online iReceptor database. The percentage indicates the proportion of correct information for each data item. We classified information as good if they were both correctly named and relevant. In the 5 columns analogous of our study, we found that only 16% of samples have correctly filled sequencing platform column. More surprisingly, for 59.2% of samples, the cell population was not specified.

Database was downloaded on 2023-09-07. It includes 9876 samples from 91 studies.

3.1.7 References

1. Barennes, P. et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. Nat Biotechnol 39, 236–245 (2021).
2. Six, A. et al. The Past, Present, and Future of Immune Repertoire Biology – The Rise of Next-Generation Repertoire Analysis. Front Immunol 4, (2013).
3. Eugster, A. et al. AIRR Community Guide to Planning and Performing AIRR-Seq Experiments. in Immunogenetics: Methods and Protocols (ed. Langerak, A. W.) 261–278 (Springer US, 2022). doi:10.1007/978-1-0716-2115-8_15.
4. Trück, J. et al. Biological controls for standardisation and interpretation of adaptive immune receptor repertoire profiling. eLife 10, e66274 (2021).
5. Vander Heiden, J. A. et al. AIRR Community Standardised Representations for Annotated Immune Repertoires. Front Immunol 9, 2206 (2018).
6. Best, K., Oakes, T., Heather, J. M., Shawe-Taylor, J. & Chain, B. Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. Sci Rep 5, 14629 (2015).
7. Kebschull, J. M. & Zador, A. M. Sources of PCR-induced distortions in high-throughput sequencing data sets. Nucleic Acids Research 43, e143 (2015).
8. Robasky, K., Lewis, N. E. & Church, G. M. The Role of Replicates for Error Mitigation in Next-Generation Sequencing. Nat Rev Genet 15, 56–62 (2014).
9. Dahal-Koirala, S. et al. TCRpower: quantifying the detection power of T-cell receptor sequencing with a novel computational pipeline calibrated by spike-in sequences. Briefings in Bioinformatics 23, bbab566 (2022).
10. Friedensohn, S. et al. Synthetic Standards Combined With Error and Bias Correction Improve the Accuracy and Quantitative Resolution of Antibody Repertoire Sequencing in Human Naïve and Memory B Cells. Front Immunol 9, 1401 (2018).
11. Andrews, S. FASTQC. A quality control tool for high throughput sequence data. (2010).
12. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. Nature Methods 12, 380–381 (2015).
13. Minich, J. J. et al. Quantifying and Understanding Well-to-Well Contamination in Microbiome Research. mSystems 4, 10.1128/msystems.00186-19 (2019).
14. Olarerin-George, A. O. & Hogenesch, J. B. Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive. Nucleic Acids Res 43, 2535–2542 (2015).
15. Salter, S. J. et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology 12, 87 (2014).
16. Illumina. Effects of Index Misassignment on Multiplexing and Downstream Analysis. 4 (2017).
17. Costello, M. et al. Characterisation and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. BMC Genomics 19, 332 (2018).
18. Guenay-Greunke, Y., Bohan, D. A., Traugott, M. & Wallinger, C. Handling of targeted amplicon sequencing data focusing on index hopping and demultiplexing using a nested metabarcoding approach in ecology. Sci Rep 11, 19510 (2021).
19. Jia, Y. et al. Sequencing introduced false positive rare taxa lead to biased microbial community diversity, assembly, and interaction interpretation in amplicon studies. Environmental Microbiome 17, 43 (2022).

20. Larsson, A. J. M., Stanley, G., Sinha, R., Weissman, I. L. & Sandberg, R. Computational correction of index switching in multiplexed sequencing libraries. Nat Methods 15, 305–307 (2018).
21. Greiff, V., Miho, E., Menzel, U. & Reddy, S. T. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. Trends in Immunology 36, 738–749 (2015).
22. Li, R. et al. A novel statistical method for decontaminating T-cell receptor sequencing data. Briefings in Bioinformatics bbad230 (2023) doi:10.1093/bib/bbad230.
23. Mhanna, V. et al. Impaired Activated/Memory Regulatory T Cell Clonal Expansion Instigates Diabetes in NOD Mice. Diabetes 70, 976–985 (2021).
24. Dixon, P. VEGAN, A Package of R Functions for Community Ecology. Journal of Vegetation Science 14, 927–930 (2003).
25. Li, J., Bushel, P. R., Chu, T.-M. & Wolfinger, R. D. Principal Variance Components Analysis: Estimating Batch Effects in Microarray Gene Expression Data. in Batch Effects and Noise in Microarray Experiments 141–154 (John Wiley & Sons, Ltd, 2009). doi:10.1002/9780470685983.ch12.
26. Wickham, H. Ggplot2: elegant graphics for data analysis. (Springer, 2009).
27. Breunig, M. M., Kriegel, H.-P., Ng, R. T. & Sander, J. LOF: identifying density-based local outliers. SIGMOD Rec. 29, 93–104 (2000).
28. Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. L. & Marioni, J. C. Detection and removal of barcode swapping in single-cell RNA-seq data. Nat Commun 9, 1–6 (2018).
29. Sinha, R. et al. Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. http://biorxiv.org/lookup/doi/10.1101/125724 (2017) doi:10.1101/125724.
30. Rubelt, F. et al. Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data. Nat Immunol 18, 1274–1278 (2017).
31. Breden, F. et al. Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data. Frontiers in Immunology 8, (2017).
32. MacConaill, L. E. et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics 19, 30 (2018).
33. Farouni, R., Djambazian, H., Ferri, L. E., Ragoussis, J. & Najafabadi, H. S. Model-based analysis of sample index hopping reveals its widespread artifacts in multiplexed single-cell RNA-sequencing. Nat Commun 11, 2704 (2020).
34. Emerson, R. O. et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat Genet 49, 659–665 (2017).
35. Auer, P. L. & Doerge, R. W. Statistical Design and Analysis of RNA Sequencing Data. Genetics 185, 405–416 (2010).
36. Burger, B., Vaudel, M. & Barsnes, H. Importance of Block Randomization When Designing Proteomics Experiments. J. Proteome Res. 20, 122–128 (2021).
37. Fisher, R. A. The design of experiments. xi, 251 (Oliver & Boyd, 1935).
38. Kang, M., Ragan, B. G. & Park, J.-H. Issues in Outcomes Research: An Overview of Randomization Techniques for Clinical Trials. J Athl Train 43, 215–221 (2008).
39. Yaari, G. & Kleinstein, S. H. Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med 7, 1–14 (2015).

40. Vander Heiden, J. A. et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 30, 1930–1932 (2014).
41. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. Nat Methods 18, 627–630 (2021).
42. Christley, S. et al. The ADC API: A Web API for the Programmatic Query of the AIRR Data Commons. Frontiers in Big Data 3, (2020).

3.2 CHAPTER 4: POLYCLONAL EXPANSION OF TCR VBETA 21.3+ CD4+ AND CD8+ T CELLS IS A HALLMARK OF MULTISYSTEM INFLAMMATORY SYNDROME IN CHILDREN

The second aim of my PhD was to reveal the promises of TCR sequencing in pathophysiology understanding. This objective took place during the 2019 SARS-CoV-2 pandemic. Early reports showed how children were as much subject to COVID-19 symptoms compared to adults (W. Liu et al., 2020; Lu et al., 2020). Children displayed typically milder symptoms and lower number of cases. However, clinicians reported a large outbreak of children with symptoms analogous to Kawasaki or toxic shock syndrome, diseases mediated by superantigen effects. This acute disease was named multiple inflammatory symptom disease in children, or MIS-C, due to the clinical manifestations in patients.

MIS-C patients harboured high percentage of circulating T cells with some Vβ as shown in cytometry. As part of a collaboration between French hospitals to investigate this emerging disease, we sought to confirm the role of SARS-CoV-2 as a superantigen. To this end, we had access to peripheral blood repertoires of MIS-C paediatric patients. We performed massive parallel sequencing on the TCR genes to assess the clonal expansion level of TCRs using the TRBV12 genes, identified from cytometry.

As part of this collaborative work, I was dedicated to the production and analysis of TCR sequencing data from whole blood paediatric patients. In the results presented as a published article in Science Immunology, we demonstrated how TRBV12 expansions were oligclonal, and did not persist in time. In line with this result, we confirmed that TRBV11-2 expansions were not preferentially associated to a particular TRVJ gene. We show here how TCR-seq can complement cytometry or gene expression profiling approaches to confirm the superantigen effect, and provide additional information on the non-specificity of the response.

CORONAVIRUS

Polyclonal expansion of TCR Vβ 21.3$^+$ CD4$^+$ and CD8$^+$ T cells is a hallmark of multisystem inflammatory syndrome in children

Marion Moreews[1], Kenz Le Gouge[2†], Samira Khaldi-Plassart[3,4†], Rémi Pescarmona[1,4,5†], Anne-Laure Mathieu[1†], Christophe Malcus[6,7], Sophia Djebali[1], Alicia Bellomo[1], Olivier Dauwalder[1,8], Magali Perret[1,5], Marine Villard[1,5], Emilie Chopin[9], Isabelle Rouvet[9], Francois Vandenesch[1,8], Céline Dupieux[1,8], Robin Pouyau[10], Sonia Teyssedre[10], Margaux Guerder[10], Tiphaine Louazon[11], Anne Moulin-Zinsch[12], Marie Duperril[13], Hugues Patural[13,14], Lisa Giovannini-Chami[15,16], Aurélie Portefaix[17], Behrouz Kassai[17], Fabienne Venet[1,6], Guillaume Monneret[6,7], Christine Lombard[5], Hugues Flodrops[18], Jean-Marie De Guillebon[19], Fanny Bajolle[20], Valérie Launay[21], Paul Bastard[22,23], Shen-Ying Zhang[22,23,24], Valérie Dubois[25], Olivier Thaunat[1,25,26,27], Jean-Christophe Richard[28,29], Mehdi Mezidi[28,29], Omran Allatif[1], Kahina Saker[1,30], Marlène Dreux[1], Laurent Abel[22,23,24], Jean-Laurent Casanova[22,23,24,31], Jacqueline Marvel[1], Sophie Trouillet-Assant[1,30], David Klatzmann[2,32], Thierry Walzer[1]*, Encarnita Mariotti-Ferrandiz[2,32†], Etienne Javouhey[7,10†], Alexandre Belot[1,6]*

Multisystem inflammatory syndrome in children (MIS-C) is a delayed and severe complication of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection that strikes previously healthy children. As MIS-C combines clinical features of Kawasaki disease (KD) and toxic shock syndrome (TSS), we aimed to compare the immunological profile of pediatric patients with these different conditions. We analyzed blood cytokine expression and the T cell repertoire and phenotype in 36 MIS-C cases, which were compared with 16 KD, 58 TSS, and 42 coronavirus disease 2019 (COVID-19) cases. We observed an increase of serum inflammatory cytokines (IL-6, IL-10, IL-18, TNF-α, IFN-γ, sCD25, MCP1, and IL-1RA) in MIS-C, TSS, and KD, contrasting with low expression of HLA-DR in monocytes. We detected a specific expansion of activated T cells expressing the Vβ21.3 T cell receptor β chain variable region in both CD4 and CD8 subsets in 75% of patients with MIS-C and not in any patient with TSS, KD, or acute COVID-19; this correlated with the cytokine storm detected. The T cell repertoire returned to baseline within weeks after MIS-C resolution. Vβ21.3$^+$ T cells from patients with MIS-C expressed high levels of HLA-DR, CD38, and CX3CR1 but had weak responses to SARS-CoV-2 peptides in vitro. Consistently, the T cell expansion was not associated with specific classical HLA alleles. Thus, our data suggested that MIS-C is characterized by a polyclonal Vβ21.3 T cell expansion not directed against SARS-CoV-2 antigenic peptides, which is not seen in KD, TSS, and acute COVID-19.

[1]CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111, Université Claude Bernard, Lyon 1, CNRS, UMR5308, ENS de Lyon, Lyon F-69007, France. [2]Sorbonne Université, UPMC Univ Paris 06, INSERM UMRS 959, Immunology Immunopathology-Immunotherapy (i3), Paris, France. [3]Pediatric Nephrology, Rheumatology, Dermatology Unit, Hôpital Femme Mère Enfant, Hospices Civils de Lyon (RAISE), Lyon, France. [4]National Referee Centre for Rheumatic and AutoImmune and Systemic diseases in childrEn, Lyon, France. [5]Immunology Laboratory, Hospices Civils de Lyon, Lyon Sud Hospital, Pierre-Bénite, France. [6]Hospices Civils de Lyon, Edouard Herriot Hospital, Immunology Laboratory, 69437 Lyon, France. [7]EA 7426 "Pathophysiology of Injury-Induced Immunosuppression" (Université Claude Bernard Lyon 1 - Hospices Civils de Lyon - bioMérieux), Joint Research Unit HCL-bioMérieux, 69003 Lyon, France. [8]Centre National de Référence des Staphylocoques, Institut des Agents Infectieux, Hospices Civils de Lyon, F-69004 Lyon, France. [9]Cellular Biotechnology Department and Biobank, Hospices Civils de Lyon, Lyon, France. [10]Réanimation Pédiatrique Hôpital Femme-Mère-Enfant Hospices Civils de Lyon, Bron, France. [11]Service de pédiatrie, Centre Hospitalier de Valence, Valence, France. [12]Unité medico-chirurgicale des cardiopathies congénitales, hôpital Louis-Pradel, hospices civils de Lyon, 69677 Bron, France. [13]Pediatric intensive care unit, University hospital of Saint-Étienne, Saint-Priest-en-Jarez, France. [14]U1059 INSERM - SAINBIOSE - DVH – Université de Saint-Étienne, 42055 Saint-Étienne, France. [15]Pediatric Pulmonology and Allergology Department, Hôpitaux pédiatriques de Nice CHU-Lenval, Nice, France. [16]Université Côte d'Azur, Nice, France. [17]Center of Clinical Investigation, Lyon University Hospital, Bron, France. [18]Service de Pédiatrie, Groupe Hospitalier Sud Réunion, CHU de La Réunion, Saint Pierre, La Réunion, France. [19]Service de Néphrologie, Rhumatologie pédiatrique, Hôpitaux pédiatriques de Nice CHU-Lenval, Nice, France. [20]Hôpital Necker Enfants Malades, Centre de référence M3C, AP-HP, Paris, France. [21]Urgences pédiatriques, Hôpital femme Mère Enfant, Hospices Civils de Lyon, Bron, France. [22]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hospital for Sick Children, Paris, France. [23]University of Paris, Imagine Institute, Paris, France. [24]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA. [25]EFS Auvergne Rhône Alpes, laboratoire Histocompatibilité, 111, rue Elisée-Reclus, 69150 Décines, France. [26]Department of Transplantation, Nephrology, and Clinical Immunology, Edouard Herriot University Hospital, Lyon, France. [27]Lyon-Est Medical Faculty, Claude Bernard University (Lyon 1), 8, avenue Rockfeller, 69373 Lyon, France. [28]Médecine Intensive-Réanimation, Hôpital de la Croix-Rousse, Hospices Civils de Lyon, Lyon, France. [29]Lyon University, Lyon, France. [30]Laboratoire de Virologie, Institut des Agents Infectieux, Laboratoire associé au Centre National de Référence des virus des infections respiratoires, Hospices Civils de Lyon, Lyon, France. [31]Howard Hughes Medical Institute, New York, NY, USA. [32]Assistance Publique - Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, Biotherapy and Département Hospitalo-Universitaire Inflammation-Immunopathology-Biotherapy (i2B), Paris, France.
*Corresponding author. Email: alexandre.belot@chu-lyon.fr
†These authors contributed equally to this work.

INTRODUCTION

At the end of April 2020, European clinicians warned public health agencies about an abnormal increase of Kawasaki-like diseases and myocarditis requiring critical care support in the context of the ongoing coronavirus disease 2019 (COVID-19) epidemic in children (*1*–*3*). American clinicians also reported a large outbreak of severe inflammation in children after COVID-19 infection, a condition that is now named pediatric inflammatory multisystemic syndrome or MIS-C (*4*–*6*). The clinical phenotype of this emerging disease is broad and encompasses features of Kawasaki disease (KD) and toxic shock syndrome (TSS). Many cases require intensive care support, making MIS-C one of the most severe manifestation of COVID-19 in children. MIS-C occurs 3 to 4 weeks after acute COVID-19 in children (*3*, *5*–*7*).

To date, reports on MIS-C show slight differences in cytokine profiling and immunophenotype between MIS-C and KD or pediatric COVID-19 (*8*, *9*). Analysis of T cells reveals a lower number of T cells in MIS-C with no or subtle signs of activation (*10*). Multidimensional immune profiling on small numbers of patients shows differences between acute COVID-19 and prepandemic KD (*8*, *11*). A subset of activated CD8 T cells expressing the CX3C chemokine receptor (CX3CR1) is observed in MIS-C (*12*), and both CD8 and natural killer (NK) cells demonstrate an elevated expression of cytotoxicity genes (*13*). Anti–severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) antibodies (Abs) are equally produced in pediatric COVID-19 and MIS-C. Autoantibodies are uniquely found during MIS-C or KD, which supports the contribution of the humoral response to both diseases (*8*, *11*). Last, a role for genetic factors is evocated in MIS-C pathogenesis because it occurs more frequently in Hispanic or African children (*14*–*16*). Despite these pioneer studies, the immunological mechanism underlying MIS-C remains unknown.

To address this question, we compared the immune profile of patients with MIS-C with that of patients with COVID-19 and that of patients with other clinically similar entities such as KD and TSS. For this, we explored the cytokine and cellular immune profile using different techniques. Using flow cytometry and transcriptomic analyses, we uncovered a specific Vβ21.3$^+$ T cell expansion in 24 of 32 tested patients with MIS-C when assessed in the first month after onset. T cell receptor (TCR) sequencing revealed the polyclonal nature of the Vβ21.3$^+$ expansion. No specific human leukocyte antigen (HLA) bias was identified in patients, but we found a specific activation profile within Vβ21.3$^+$ T cells. This activation was transient with a normalization of the repertoire within days to weeks after the inflammatory episode. Together, our findings provide an immunological signature in MIS-C with potential implication in the diagnosis and treatment of this rare disease.

RESULTS

MIS-C presentation overlapped with TSS and KD

We took a cohort of 36 children with MIS-C and compared them with 16 KD cases diagnosed during and before the pandemic, 58 retrospective cases of patients with TSS, and 42 patients with acute COVID-19 (11 children and 31 adults). This comparison was motivated by previous descriptions of MIS-C in Europe and in the United States, showing a clinical overlap between staphylococcal toxin–mediated TSS and KD in patients with MIS-C (*1*–*3*). Figure 1A outlines the study flowchart and the clinical and biological parameters we evaluated. Patients diagnosed for MIS-C, classical KD, TSS, or acute COVID-19 were included. Patients were then subjected to deep immunological analyses combining cytokine profiling, TCR Vβ analysis, and T cell stimulation assays (Fig. 1A). We confirmed the strong clinical overlap between MIS-C, TSS, and KD. Many patients in the MIS-C group also fulfilled some of the five major criteria for TSS and KD, respectively (Fig. 1B). Considering the clinical parameters, the most frequent features of patients with MIS-C in our cohort were fever, cardiac dysfunction, gastrointestinal symptoms, coagulopathy, and systemic inflammation (table S1). Additional clinical data are presented in table S2 for KD, TSS, and acute COVID-19 and in table S3 for all patients. Moreover, table S4 gives a list of the patients analyzed in each of the following figure panels.

High levels of proinflammatory cytokines in MIS-C contrasted with lymphopenia and low HLA-DR expression in monocytes

SARS-CoV-2 can cause fatal acute respiratory distress syndrome in patients at risk. This manifestation is caused by delayed and poorly controlled immune responses, with a deleterious role of inflammatory cytokines. Moreover, we and others have identified a subgroup of severe COVID-19 patients with impaired type I interferon (IFN) production (*17*–*20*). Thus, a regulated production of cytokines is paramount for control of SARS-CoV-2 infection. This prompted us to investigate how cytokines could contribute to MIS-C pathogenesis. We compared the serum levels of IFN-α, IFN-γ, tumor necrosis factor–α (TNF-α), interleukin 10 (IL-10), soluble CD25 (sCD25), monocyte chemoattractant protein 1 (MCP1), IL-1 receptor antagonist (IL-1Ra), IL-6, and IL-18 between healthy controls and MIS-C, KD, TSS, and different forms of COVID-19 (mild pediatric, mild, or severe adult-onset COVID-19; see table S2 for a list of clinical features in the different patients' groups).

The expression of IFN-stimulated genes (ISGs) in blood cells was significantly higher in MIS-C compared with controls but rather low compared with patients with COVID-19 (Fig. 2, A to C). The level of serum IFN-α2 followed the same trends, whereas serum IFN-γ was variable among patients with MIS-C, with very high levels in a few patients. The expression of the other cytokines measured (IL-6, IL-10, IL-18, TNF-α, MCP1, IL-1RA, and sCD25) was very high in patients with MIS-C compared with controls and very similar to that of patients with KD, TSS, and severe COVID-19 (Fig. 2, B and C). The level of sCD25 was significantly higher in patients with TSS than in patients with MIS-C and significantly lower in patients with severe COVID-19 than in patients with MIS-C (Fig. 2, B and C). A previous study found higher levels of serum IL-6 in patients with KD than in patients with MIS-C, contrasting with our data (*8*).

To further explore the MIS-C immunological profile, we quantified the number of peripheral lymphocytes of different types and the expression of HLA-DR in patients' monocytes. T and NK cell counts were, on average, very low in patients with MIS-C and KD, whereas B cell counts were normal (Fig. 2D and fig. S1). We found a decreased expression of HLA-DR in monocytes in both patients with KD and MIS-C compared with controls (Fig. 2E and fig. S1). Together, our data show a strong similarity in cytokine profiles between MIS-C, KD, and TSS and highlight the decreased lymphocyte counts and low HLA-DR expression in monocytes in patients with MIS-C compared with controls.

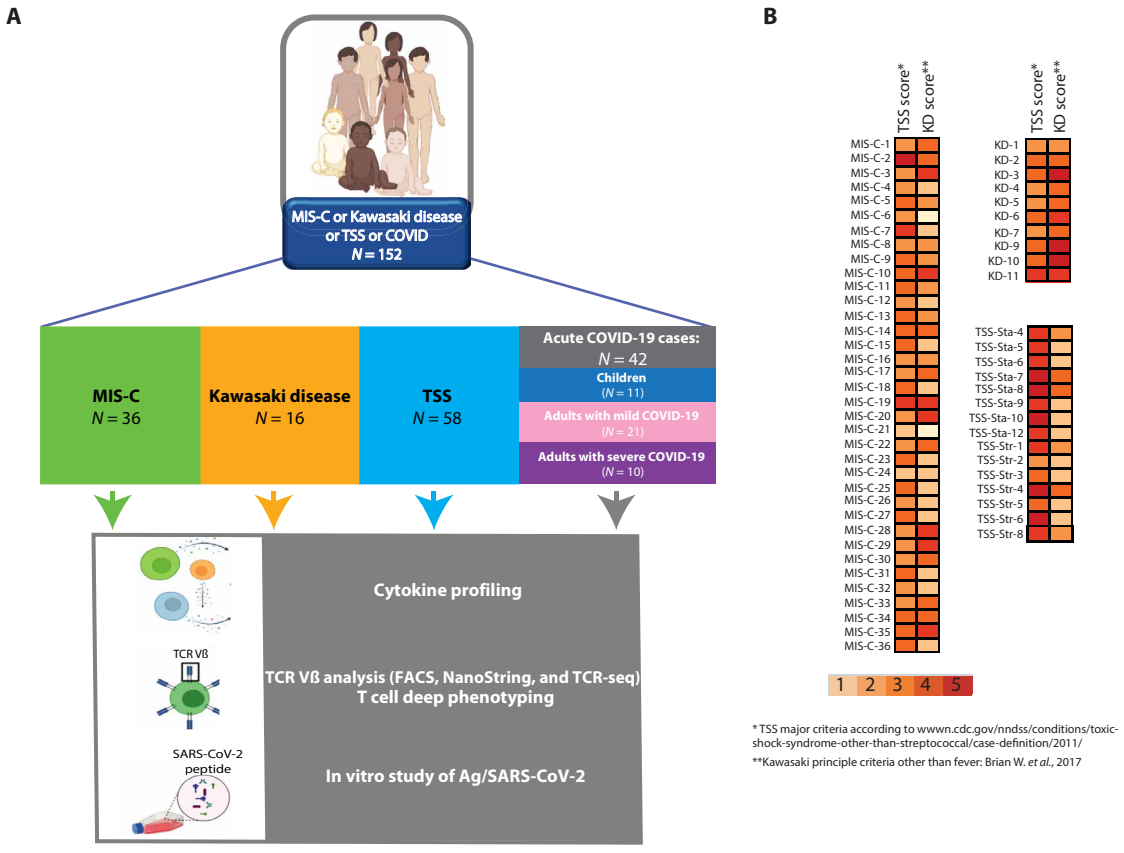


Fig. 1. Study design and clinical features of patients with MIS-C. (**A**) Outline of the study including patients with MIS-C, KD, TSS, and acute COVID-19 and the immunological investigation workflow. (**B**) Heatmap showing the number of major criteria of TSS or KD for patients with TSS, MIS-C, and KD included in our study, considering the following case definition criteria: five clinical items for TSS (fever, rash, desquamation, hypotension, and multisystem involvement) and five clinical items for KD in addition to fever (rash, cervical lymphadenopathy, bilateral conjunctivitis, oral mucosal changes, and peripheral extremity changes).

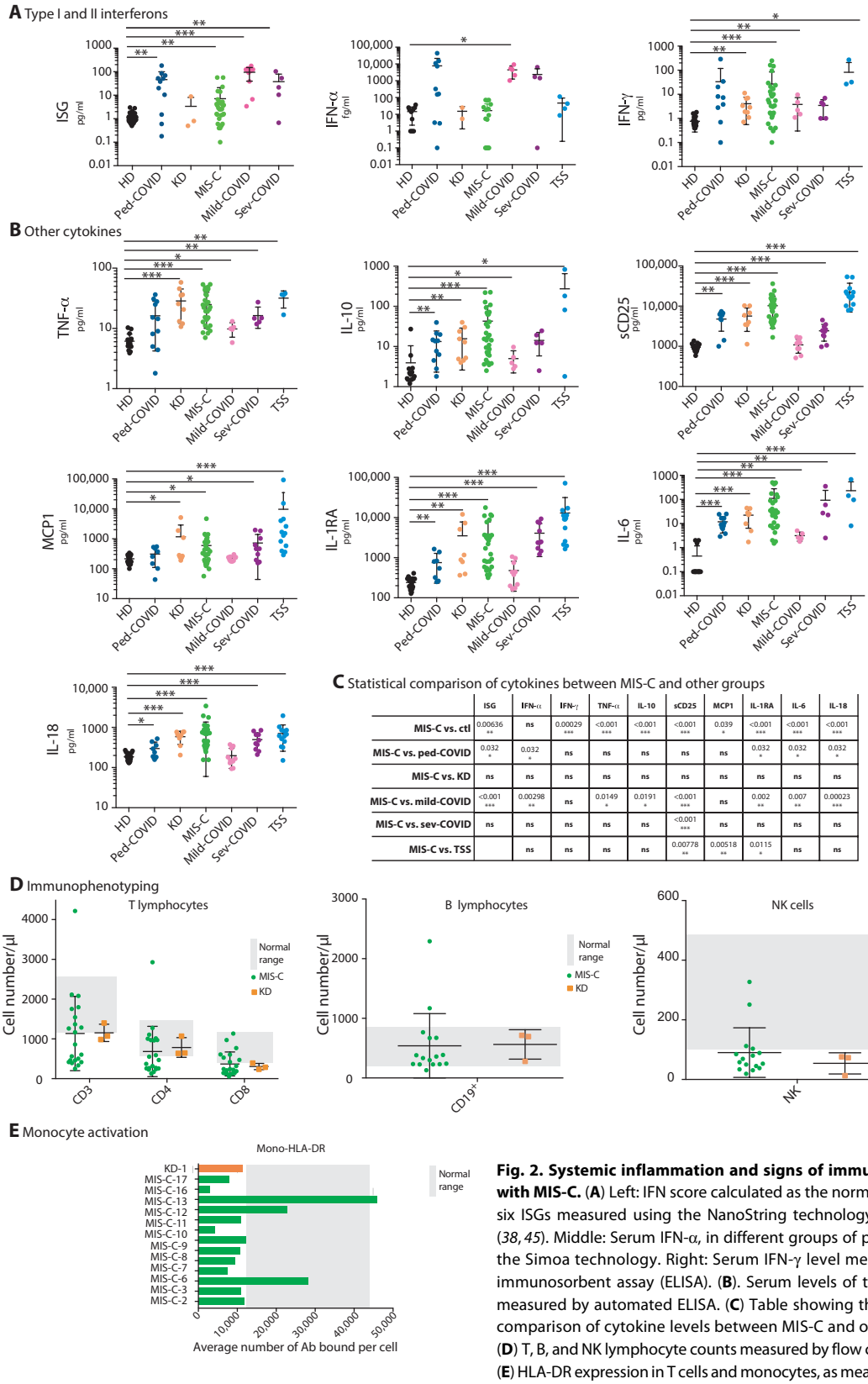
Expansion of Vβ21.3$^+$ peripheral T cells in a large fraction of patients with MIS-C

TSS toxin 1 (TSST1)–related TSS is associated with a skewing of the T cell repertoire toward Vβ2 as a result of TSST1 superantigen–induced proliferation of Vβ2$^+$ T cells (*21*). Every other *Staphylococcus aureus* superantigenic toxin induces the expansion of specific TCR Vβ subsets, i.e., Vβ 5.2, 5.3, 7.2, 9, 16, 18, and 22 for staphylococcal enterotoxin A (SEA) or Vβ 3, 12, 13.2, 14, 17, and 20 for SEB (*22*). Given the similarities between TSS and MIS-C, we explored the possibility that MIS-C was also associated with specific T cell expansions. To explore the T cell repertoire in MIS-C, we first used flow cytometry to assess the distribution of Vβ subunits in T cells from patients with MIS-C, in comparison with patients with KD, TSS, and COVID-19 (Fig. 3A and fig. S2A). As expected, patients with TSS displayed the hallmark expansion of the Vβ2$^+$ subset. Several Vβ-specific expansions were also visible in patients with MIS-C and, in most cases, Vβ21.3$^+$ expansions (Fig. 3A) in both CD4 and CD8 T subsets (fig. S3A). These expansions had similar amplitudes as the Vβ2$^+$ expansions in TSS (Fig. 3A). A principal components analysis (PCA) of the Vβ distribution in CD4 and CD8 T cells showed that the main parameters separating the different patients were the frequency of Vβ2$^+$ and of Vβ21.3$^+$ cells (fig. S3, B and C). Overall, the expansion of Vβ21.3$^+$ T cell subsets was seen in 15 of 26 (58%) patients with MIS-C and in none of the other conditions analyzed by flow cytometry, i.e., KD, TSS, and COVID-19 (Fig. 3A). Next, we wanted to use a different technique to test the specificity of this expansion, and we therefore performed transcriptomic analyses of Vβ expression in peripheral blood mononuclear cells (PBMCs) using the NanoString technology. This technique also requires much less material than flow cytometry, which allowed us to run lymphopenic samples from severe COVID-19 cases. This transcriptomic analysis firmly established that the Vβ21.3$^+$ T cell expansion is a hallmark of MIS-C because it was seen in 18 of 23 patients with MIS-C tested (fig. S3D). Thus, taking together flow cytometry and NanoString analyses, we found that 24 of 32 (75%) patients with MIS-C and none in the other clinical groups displayed *T cell receptor beta variable genes 11-2 (TRBV11-2)*/Vβ21.3$^+$ expansions.

We then compared the level of serum cytokines between MIS-C patients with and without Vβ21.3$^+$ T cell expansions at the time of the acute episode. The levels of IL-18 and IL-1RA (Fig. 3, C and D) were associated with the polyclonal Vβ21.3 expansions, but not those of the other cytokines tested (fig. S4, A and B), suggesting that Vβ21.3$^+$ T cells were associated with the cytokine storm.

TCR sequencing highlighted the polyclonal nature of TCR Vβ21.3 expansions

To investigate the clonality of Vβ21.3$^+$ expanded cells, we analyzed the TCR repertoire of 11 patients with MIS-C for whom whole-blood



	ISG	IFN-α	IFN-γ	TNF-α	IL-10	sCD25	MCP1	IL-1RA	IL-6	IL-18
MIS-C vs. ctl	0.00636 **	ns	0.00029 ***	<0.001 ***	<0.001 ***	<0.001 ***	0.039 *	<0.001 ***	<0.001 ***	<0.001 ***
MIS-C vs. ped-COVID	0.032 *	0.032 *	ns	ns	ns	ns	ns	0.032 *	0.032 *	0.032 *
MIS-C vs. KD	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
MIS-C vs. mild-COVID	<0.001 ***	0.00298 **	ns	0.0149 *	0.0191 *	<0.001 ***	ns	0.002 **	0.007 **	0.00023 ***
MIS-C vs. sev-COVID	ns	ns	ns	ns	ns	<0.001 ***	ns	ns	ns	ns
MIS-C vs. TSS		ns	ns	ns	ns	0.00778 **	0.00518 **	0.0115 *	ns	ns

Fig. 2. Systemic inflammation and signs of immune paralysis in patients with MIS-C. (**A**) Left: IFN score calculated as the normalized mean expression of six ISGs measured using the NanoString technology, as previously described (*38*, *45*). Middle: Serum IFN-α, in different groups of patients, as measured with the Simoa technology. Right: Serum IFN-γ level measured by enzyme-linked immunosorbent assay (ELISA). (**B**). Serum levels of the indicated cytokines as measured by automated ELISA. (**C**) Table showing the statistical results of the comparison of cytokine levels between MIS-C and other groups, as indicated. (**D**) T, B, and NK lymphocyte counts measured by flow cytometry in MIS-C and KD. (**E**) HLA-DR expression in T cells and monocytes, as measured by flow cytometry in MIS-C. Gray shading indicates the derived central 95% HD (healthy donor) reference interval (D and E). See table S4 for subject numbers per panel. Statistical test: Kruskal-Wallis test between healthy donors and all other groups with adjustment for multiple comparisons using Benjamini-Hochberg correction (A and B) or between MIS-C and all other groups (C) with the same strategy. *$P < 0.05$, **$P < 0.01$, and ***$P < 0.001$. ns, not significant.

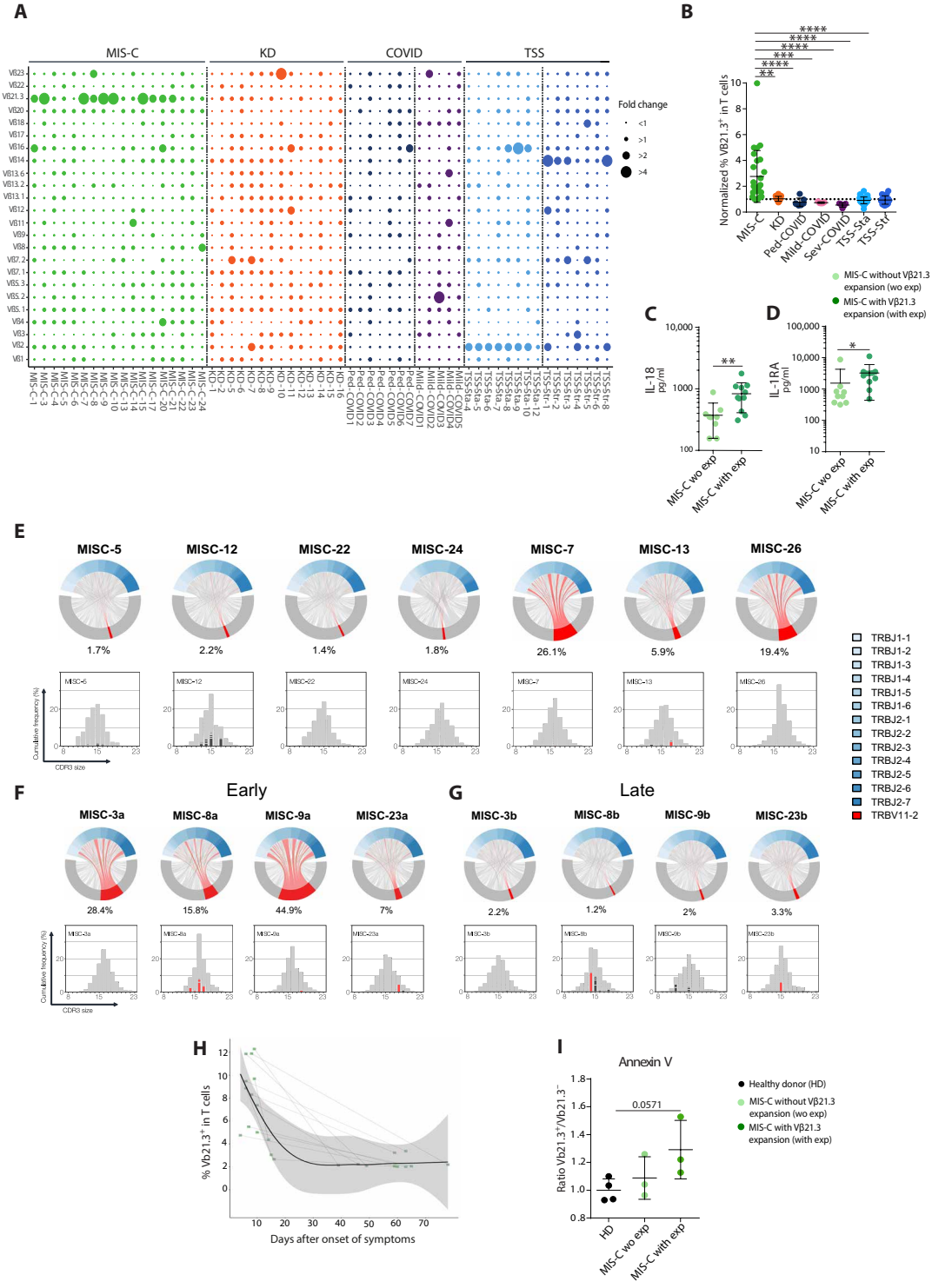


Fig. 3. Polyclonal Vβ21.3$^+$ T cell expansion in patients with MIS-C. (**A**) Frequency of total CD3$^+$ T cells expressing the indicated Vβ chains, as measured by flow cytometry using specific Abs against the corresponding Vβ within PBMCs of patients of the indicated group. Patients with TSS, mild COVID-19, pediatric COVID-19 (ped-COVID), KD, and MIS-C are colored in blue, pink, dark blue, orange, and green, respectively. Bubbles represent the normalized individual Vβ frequency reported to the mean frequency for each Vβ in the general adult population. (**B**) Normalized frequency of Vβ21.3$^+$ T cells in different clinical conditions, as indicated. (**C** and **D**) Serum IL-18 (C) and IL-1RA (D) levels in MIS-C patients with or without (wo) Vβ21.3$^+$ T cell expansions (exp). (**E** to **G**) Chord diagrams of the TRBV (bottom, gray) and TRBJ (top, blue) combinations assessed by TCR sequencing of TCRαβ chains in whole blood of patients with MIS-C. The relative frequency of all TRBVBJ combinations have been calculated per sample on the full TRB repertoire data. Combinations using TRBV11-2 are highlighted in red. Each red line indicates pairing with a given TRBJ, and the thickness indicates the frequency of this pairing. The percentage values under each chart indicate the percentage of clonotypes composed with the TRBV11-2 gene. In (E) to (G), the CDR3 length distribution of clonotypes using TRBV11-2 is shown as a histogram graph. Each clonotype is represented as a gray line. The thickness of the line represents the frequency of the clonotype within each repertoire. Because most of the clonotypes are not abundant, all the gray lines are stacked together and appear as a unique gray bar, which reflects the lack of expansion. Expanded clonotypes identified as detailed in Materials and Methods are shown in red. In (F) and (G), the same four patients are shown during the MIS-C episode (F) and after resolution (G). (**H**) Frequency of Vβ21.3$^+$ T cells at different time points during and after the MIS-C episode in different patients, as assessed by flow cytometry. (**I**) Annexin-V staining of T cells in the indicated patients' groups. Results show the ratio of the annexin-V fluorescence in Vβ21.3$^+$ versus Vβ21.3$^-$ T cells. See table S4 for subject numbers per panel. (B) Statistical test: Kruskal-Wallis test between MIS-C and all other groups with adjustment for multiple comparisons using Benjamini-Hochberg correction and (C, D, and I) unpaired Wilcoxon test comparing two groups. *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

RNA was available by TCR sequencing. We analyzed the composition of the TCR β rearrangements involving the TRBV11-2 gene (which corresponds to Vβ21.3). First, by representing the TRBV11-2/T cell receptor beta joining genes (TRBJ) combination usage as chord diagrams (Fig. 3, E and F), we confirmed the expansion of T cells using TRBV11-2 in 7 of 11 patients. These TRBV11-2 rearrangements were associated with multiple TRBJ genes, suggesting the polyclonal nature of the expansions. To further evaluate polyclonality, we analyzed the hypervariable sequence CDR3 length distribution of TRBV11-2 clonotypes (bar plots, Fig. 3, E to G). The CDR3 size distributions showed a bell-shaped Gaussian distribution as expected in polyclonal repertoires (*23*–*25*). To evaluate the degree of polyclonality, we identified the expanded clonotypes by setting a threshold based on the binomial distribution of the clonotype frequencies per sample (see Materials and Methods and fig. S5A). No major monoclonal expansions (red lines in the CDR3 spectratypes) explaining the global TRBV11-2 expansion were detected. Instead, most of the clonotypes were found at low frequencies (gray lines), typical of a polyclonal diverse repertoire. The percentages of expanded clonotypes were not significantly different between patients with or without TRBV11-2. We calculated the cumulative frequencies of these expanded clonotypes within the full repertoire and found that they were always far below the frequency of the full TRBV11-2 expansion in patients with expansions, representing, on average, 0.51% of the total repertoire. Last, these limited expansions represented, on average, 4.47% of the TRBV11-2 repertoire in patients with TRBV11-2 expansions and 6.31% in patients without TRBV11-2 expansions (table S6 and fig. S5B). To confirm the polyclonality of the TRBV11-2 expansion, we computed the Berger-Parker index (BPI) on TRBV11-2 clonotype for patients with MIS-C harboring or not TRBV11-2 expansions (fig. S5C). This index measures the proportional abundance of the most frequent clonotypes within TRBV11-2 clonotypes. There were no significant differences when we compared the BPI on TRBV11-2 clonotypes between patients with or without TRBV11-2 expansions, further confirming that TRBV11-2 expansions in the seven patients were not explained by monoclonal expansions.

Next, to address whether the Vβ21.3$^+$ T cell expansion persisted overtime, we repeated the TCR sequencing and the flow cytometry Vβ analyses in a group of patients for which blood samples were available during and after the acute inflammatory episode. As shown in Fig. 3 (F to H), the Vb21.3/TRBV11-2 distributions for all the patients returned to normal within days to weeks after MIS-C. When we compared the CDR3 length distributions by calculating the perturbation score using the ISEApeaks tool between repertoires obtained during and after the acute response, we found no differences between the two groups, further supporting the polyclonal expansion profile of TRBV11-2 during the acute response (fig. S5D). Last, this transient expansion suggested a proapoptotic phenotype of Vβ21.3$^+$ T cell. To test this hypothesis, we stained PBMCs from patients with MIS-C with annexin-V that marks early apoptotic cells. A higher fraction of Vβ21.3$^+$ compared with Vβ21.3$^-$ T cells were stained with annexin-V in MIS-C patients with Vβ21.3$^+$ expansions (Fig. 3I and fig. S2B), which substantiated our hypothesis.

Vβ21.3$^+$ T cells had an activated phenotype but did not react against SARS-CoV-2 peptides

Because Vβ21.3$^+$ T cells expand in patients with MIS-C, we investigated their activation status and the mechanisms underlying their proliferation. We found that the activation markers HLA-DR and CD38 were expressed at high levels in both CD4 and CD8 T cells from MIS-C patients with Vβ21.3$^+$ expansions compared with those without expansions and to healthy controls (Fig. 4, A and B). This was due to a specific up-regulation of CD38 and HLA-DR in Vβ21.3$^+$ CD4 and CD8 T cells in MIS-C patients with Vβ21.3 expansions compared with those without Vβ21.3 expansions (Fig. 4, C and D). A recent paper reports a specific activation of CX3CR1$^+$ CD4 and CD8 T cells in patients with MIS-C, as assessed by HLA-DR/CD38 levels (*12*). This prompted us to measure CX3CR1 levels in Vβ21.3$^+$ T cells. As shown in Fig. 4E and fig. S6A, Vβ21.3$^+$ T cells overexpressed CX3CR1 in both CD4 and CD8 T cells in MIS-C patients with Vβ21.3$^+$ expansions compared with those without expansions, although the percentage of CX3CR1$^+$ cells was not higher in MIS-C than in control patients (fig. S7A). Moreover, in patients with MIS-C, a large frequency of non-naïve CX3CR1$^+$ CD4 and CD8 T cells had an activated phenotype as previously reported (fig. S7B) (*12*).

Given that MIS-C came about weeks after COVID-19, we wondered whether Vβ21.3$^+$ T cells were raised against SARS-CoV-2 antigens. To test this possibility, we stimulated PBMCs from patients with MIS-C or convalescent COVID-19 with a commercial cocktail of SARS-CoV-2 peptides spanning S, N, and M viral proteins. T cells from patients with MIS-C responded poorly to stimulation with viral peptides, regardless of Vβ21.3 expansion, compared with T cells from patients with convalescent COVID-19 that responded well (Fig. 4, F and G, and fig. S6, B and C). This was not due to a lack of adaptive anti–SARS-CoV-2 response, because all patients with MIS-C tested had high SARS-CoV-2–specific Ab levels (fig. S7, C and E). Last, we could not identify any specific allele nor mutations of classical HLA class I or class II genes associated with TRBV11-2 expansions by genomic sequencing of the HLA loci of 13 patients with MIS-C (table S4). Together with the lack of Vβ21.3$^+$ expansion in patients with COVID-19, these data showed that Vβ21.3$^+$ T cells were not specific for HLA-restricted SARS-CoV-2 peptides. Together, these data revealed that the Vβ21.3$^+$ CD4 and CD8 T cell expansion were highly activated and expressed CX3CR1 but had poor responsiveness to SARS-CoV-2 antigens.

DISCUSSION

Here, we confirmed the strong overlap in clinical phenotype between KD, MIS-C, and TSS; MIS-C and TSS had similar defining features, specifically cardiac dysfunction, hypotension, maculopapular skin rash, and conjunctivitis. We recently identified the critical importance of early steroid therapy in the management of MIS-C, similarly to what has been previously shown in TSS (*26*, *27*). MIS-C and TSS are obviously linked to infections, whereas many KD features suggest an infectious cause for KD as well (*28*). The epidemic of a novel coronavirus in 2005 (New Haven) was associated to KD and linked the viral infection to vascular inflammation (*29*).

We found important similarities in terms of cytokine expression between MIS-C, TSS, and KD, such as high TNF-α, IL-6, IL-18, and IL-1Ra levels. A previous study noted that a subgroup of patients with severe MIS-C had higher levels of IFN-γ, IL-18, GM-CSF (granulocyte-macrophage colony-stimulating factor), RANTES (regulated upon activation, normal T cell expressed and secreted), IP-10 (interferon gamma-induced protein 10), IL-1α, and SDF-1 (stromal cell–derived factor 1) than patients with mild MIS-C or KD (*30*). We also observed a subset of patients with MIS-C with high serum

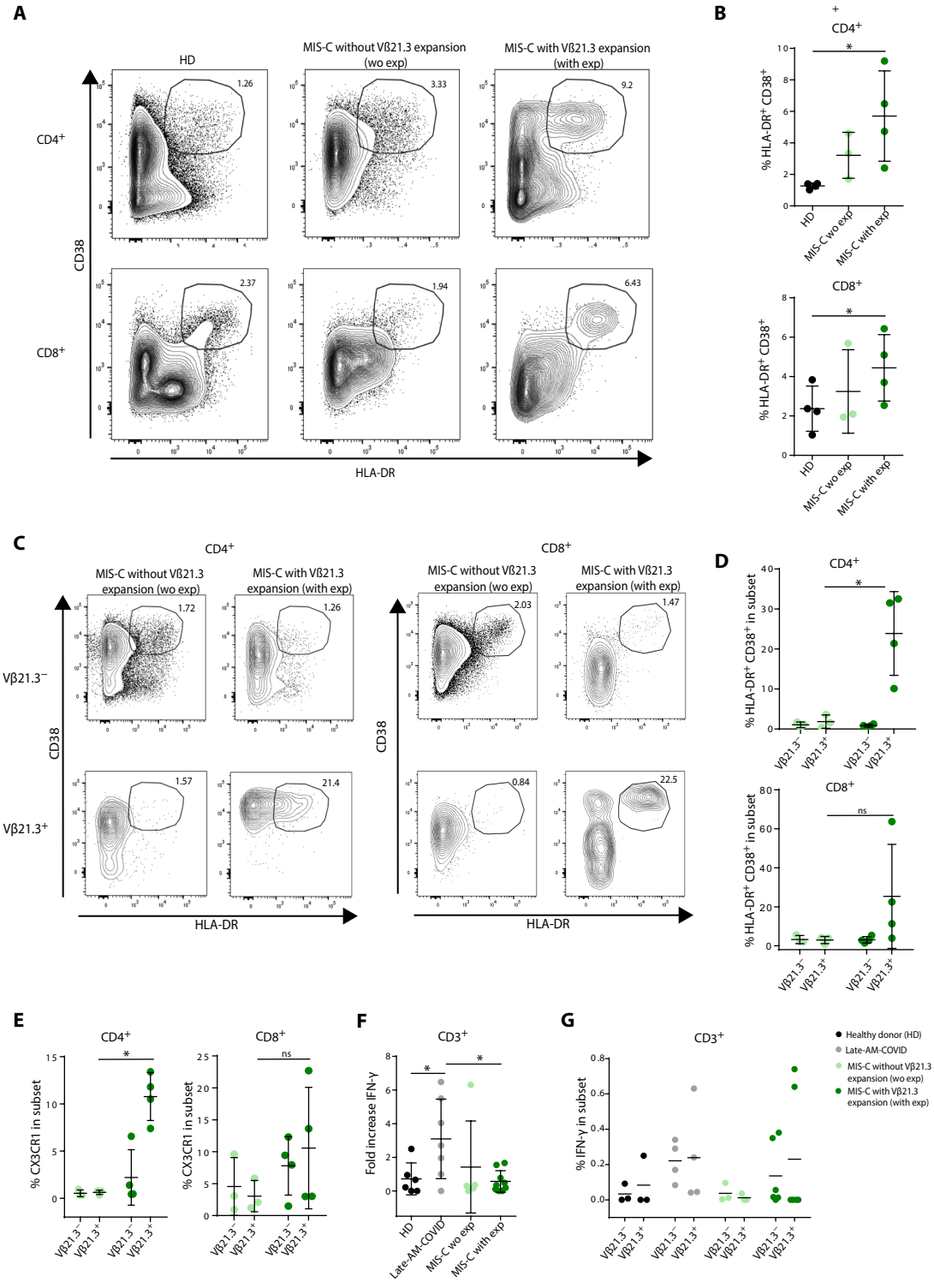


Fig. 4. T cell activation within Vβ21.3 and stimulation of T cells with viral peptides in vitro. (**A** to **D**) Flow cytometry analysis of CD38 and HLA-DR expression in CD4 or CD8 T cells from the indicated patients' groups (exp: Vβ21.3$^+$ T cell expansion). (A) A representative staining and (B) the mean ± SD frequency of CD38$^+$HLA-DR$^+$ CD4 (top) and CD8 (bottom) T cells. (C and D) A Vβ21.3$^+$ Ab was also included in the flow cytometry panel used in (A) and (B), allowing a specific comparison of the Vβ21.3$^-$ and Vβ21.3$^+$ T cells in patients with MIS-C. (C) A representative dot plot of CD38 and HLA-DR expression in the indicated subsets; (D) means ± SD frequency of CD38$^+$HLA-DR$^+$ in the indicated CD4 (top) and CD8 (bottom) T cell subsets. (**E**) Frequency of CX3CR1$^+$ cells in gated Vβ21.3$^-$ and Vβ21.3$^+$ CD4$^+$ (left) and CD8$^+$ (right) T cells in MIS-C without and MIS-C with expansion. (**F**) PBMCs from control, patients with COVID-19 (adults, 6 months after infection, Late-AM-COVID), or patients with MIS-C (with or without Vβ21.3$^+$ T cell expansions) were stimulated for 6 hours with a commercial cocktail of synthetic peptides from S, N, and M SARS-CoV-2 proteins in the presence of Golgi secretion inhibitors for the last 5 hours. Intracellular IFN-γ expression was then measured in T cells by flow cytometry. The fold increase was calculated as the ratio between the stimulated and the unstimulated conditions. (**G**) The frequency of Vβ21.3$^+$ and Vβ21.3$^-$ T cells expressing IFN-γ after stimulation with S, N, and M SARS-CoV-2 peptides in the different patient groups as indicated (one dot, one patient). See table S4 for subject numbers per panel. (B) Kruskal-Wallis test between three groups with adjustment for multiple comparisons using Benjamini-Hochberg correction and (D to G) Wilcoxon test comparing two groups. *$P < 0.05$.

IFN-γ, IL-18, and sCD25. These observations confirm previous reports showing a clinical and biological overlap between MIS-C and macrophage activation syndrome (*3*) and suggest the importance of IFN-γ in the disease.

Here, we reported the expansion of a TCR Vβ21.3$^+$ T cell subset with an activated phenotype in as many as 75% of patients with MIS-C. Vβ21.3$^+$ T cell expansions were also reported in smaller numbers of patients with MIS-C in two recent studies (*13*, *31*). In both Porritt *et al.* (*31*) and our study, Vβ21.3$^+$ T cell expansions appeared polyclonal as judged by the large number of TRBJ gene segments associated with TRBV11.2 and by the even distribution of the CDR3 domain. Our study further showed that Vβ21.3$^+$ CD4 and CD8 T cell expansions are a discriminating feature of patients with MIS-C compared with patients with KD, TSS, and COVID-19.

We observed a correlation between Vβ21.3$^+$ T cell expansions and the level of serum cytokines IL-18 and IL-1RA from matching samples, confirming a previous study (*31*) and indicating that Vβ21.3$^+$ T cell expansions were associated with the cytokine storm. Our data also showed that Vβ21.3$^+$ T cells have an activated phenotype, with high HLA-DR and CD38 expression, and that activated Vβ21.3$^+$ T cells expressed high levels of CX3CR1, a marker of patrolling monocytes and of cytotoxic lymphocytes. CX3CR1 binds to CX3CL1, a membrane-bound chemokine induced on vascular endothelial cells upon inflammation (*12*). The CX3CL1-CX3CR1 axis is thought to have an important role in vascular inflammation in different inflammatory diseases (*32*) and could contribute to MIS-C pathogenesis. This interaction could promote the cytotoxic action of different lymphocyte populations, which fits with the reported elevated expression of cytotoxicity genes in NK and CD8$^+$ T cells in patients with MIS-C (*13*).

We demonstrated that both TSS and MIS-C were marked by the polyclonal proliferation of a specific Vβ subset, i.e., Vβ2$^+$ cells for TSS related to TSST1 and Vβ21.3$^+$ cells for MIS-C. The amplitude of the expansion was also similar in both syndromes. Considering the other clinical phenotype similarities between MIS-C and TSS shown in this study, cytokine production and treatment, this raises the hypothesis that Vβ21.3$^+$ cell expansions are caused by a superantigen structure in MIS-C. The term superantigen has been coined by Kappler and Marrack as an operational definition of various T cell–activating substances with specificity for T cell antigen receptors Vβ subunits regardless of the rearrangement and antigen specificity (*33*). Superantigens bind external regions of TCR and MHC molecules (*34*) and can induce massive expansions of T cells expressing one specific TCR Vβ chain, whereas classical antigens induce the expansion of T cells bearing different Vβ. Previous papers have suggested that the SARS-CoV-2 spike protein could behave as a superantigen structure (*35*). Using in silico modeling, Porritt *et al.* identified a putative interaction between Vβ21.3 and a superantigen-like motif in spike. However, Vβ21.3$^+$ T cell expansions occur in a delayed manner relative to SARS-CoV-2 infection, and the virus is often undetectable in patients with MIS-C at the time of the acute inflammation. The kinetics of MIS-C relative to COVID-19 is compatible with a causal role of anti–SARS-CoV-2 Abs. One can hypothesize that immune complexes composed of SARS-CoV-2 bound to Abs may act as superantigen structures. However, a previous study failed to detect these immune complexes in patients with MIS-C (*30*). In addition, Vβ-restricted T cells adhere to endothelial cells after superantigen activation (*36*), and, thus, the CX3CR1$^+$ Vβ21.3–expanded T cells may play a role in vascular injury in MIS-C. Alternative mechanisms may be put forward, such as secondary autoimmune reactions. Several studies have indeed reported the appearance of autoantibodies in patients with MIS-C, some of which directed against endothelial antigens (*8*, *11*), whereas others have reported immune events consistent with autoimmunity, such as the expansion of proliferating plasmablasts (*13*) or the persistence of functional SARS-CoV-2–specific monocyte-activating Abs (*37*). How B cell–mediated autoimmunity would be linked to Vβ-specific T cell expansions is, however, unclear. One could speculate that immune complexes composed of autoantibodies and endogenous antigens could behave as superantigens.

Last, given the rarity of MIS-C, there could be a genetic susceptibility to this postinfectious disease, promoting hyperinflammatory reaction of adaptive immunity in response to SARS-CoV-2 (*16*). We limited our analysis to classical HLA alleles but did not find any significant association, although a previous study reported an HLA-I bias in a smallest group of patients with MIS-C (*31*). Our MIS-C samples were obtained in most of cases after anti-inflammatory treatments (see table S3), and it is likely that those treatments affect the level of serum cytokines, which could have affected the comparisons we made between clinical conditions and the associations between cytokines and T cell expansions. Together, MIS-C shared clinical and immunological anomalies with KD and TSS but was specifically characterized by a polyclonal Vβ21.3 expansion in CD4 and CD8 T cells associated to activation and CX3CR1 expression.

MATERIALS AND METHODS

More information for all of these protocols can be found in the Supplementary Materials.

Study design and human subjects

The immunological profiles of 36 MIS-C, 16 KD, 58 TSS, and 42 non–MIS-C COVID-19 cases were included (Fig. 1A). Samples were collected within the first week of symptoms and analyzed for cytokine immunoprofiling, standard immunophenotyping, Vβ expression, TCR sequencing, and SARS-CoV-2–dependent T cell response. Because of low-volume sampling of pediatric patients, we did not have the same availability for research blood draws. The samples used for each experiment are detailed in table S4. The main clinical features are summarized in tables S1 to S3. Written informed consent was obtained for all data collection and blood sampling as detailed in the Supplementary Materials.

Immunological analyses

Cytokines and IFN score assessment

Plasma concentrations of IL-6, TNF-a, IFN-γ, IL-10, IL-18, MCP-1, IL-1RA, and sCD25 were measured by Simple Plex technology using an enzyme-linked lectin assay instrument (ProteinSimple). Plasma IFN-α concentrations were determined by single-molecule array (Simoa) on an HD-1 analyzer (Quanterix) using a commercial kit for IFN-α2 quantification (Quanterix). RNA was extracted from whole blood, and IFN score was obtained using nCounter analysis technology (NanoString Technologies) by calculating the median of the normalized count of six ISGs as previously described (*38*).

T cell Vβ repertoire analysis and immunophenotyping

PBMCs were stained with surface markers, CD3, CD4, CD8, CD14, CD16, CD19, CCR7, CD38, Vβ21.3, HLA-DR, CX3CR1, and CD45RA (further details on these stains are included in the Supplemental Materials). Cell apoptosis was assessed by annexin-V staining. All samples were acquired on a BD LSRFortessa (BD Biosciences) flow cytometer and analyzed using FlowJo version 10 software. Monocyte HLA-DR expression was determined on EDTA-anticoagulated peripheral whole blood as previously described (*39*). The phenotypic analysis of T cell Vβ repertoire was performed on whole blood sample using the IOTest Beta Mark kit (Beckman Coulter). Whole-blood cells were stained for CD3, CD4, CD8, and each combination of three fluorescein isothiocyanate (FITC)–, phycoerythrin (PE)–, and FITC/PE–conjugated anti–Vβ monoclonal Abs (mAbs) (Beckman Coulter) in eight sample tubes. Expansions were defined, respectively, for values above the mean + 2 SD or below the minimum reference values of the corresponding family. Additional samples were analyzed for TRBV from total RNA with NanoString technology (Supplementary Materials).

TCR sequencing and analysis

TCR α/β libraries were prepared from 300 ng of RNA from each sample with the SMARTer Human TCR a/b Profiling Kit (Takara Bio) (*40*). The sequencing was carried out on a MiSeq Illumina sequencer using the MiSeq v3 PE300 protocol at the Biomics Platform (Institut Pasteur, Paris, France). Single-end sequences were aligned and annotated using MiXCR 3.0.13 (*41*), providing a list of clonotypes. Analyses were performed in R 4.0.3 on the TRB clonotype lists obtained with MiXCR. For each clonotype, read count was recorded. Frequencies for TRBV, TRBJ, and clonotypes were calculated on the basis of the total read counts per sample. Chord diagrams were made using the circlize package (*42*) on TRBVBJ frequencies, and CDR3 length spectratypes were made using ggplot2 (*43*) using clonotype frequencies. To identify TRBV11-2 expanded clonotypes, the first (Q1) and third (Q3) quartiles and the interquartile range (IQR) were computed for all patients without TRBV11-2 expansion. Expanded clonotypes are defined as those with counts superior to Q3 + (1.5*IQR).

Stimulation with SARS-CoV-2 overlapping peptide pools and flow cytometry

PBMCs were stimulated with SARS-CoV-2 PepTivator pooled S, N, and M peptides (Miltenyi Biotec) at a final concentration of 2 μg ml^{-1} for 1 hour in the presence of 2 μg ml^{-1} of mAbs CD28 and CD49d and then for an additional 5 hours with GolgiPlug and GolgiStop (BD Biosciences). Similar surface markers were stained. Cells were then washed, fixed with Cytofix/Cytoperm (BD Biosciences), and stained with V450-conjugated anti–IFN-γ. All samples were acquired on a BD LSRFortessa (BD Biosciences) flow cytometer and analyzed using FlowJo version 10 software.

Serology

Serum samples were tested with three commercial assays: the Wantai Ab assay detecting total Abs against the receptor binding domain (RBD) of the S protein, the bioMérieux VIDAS assay detecting immunoglobulin G (IgG) to the RBD, and the Abbott ARCHITECH assay detecting IgG to the N protein.

Statistical analyses

All tests were performed two sided with a nominal significance threshold of $P < 0.05$. We used nonparametric tests appropriated to the low number of observations in each of our experimental conditions, i.e., the Wilcoxon or Kruskal-Wallis test depending on whether we have two or more conditions to compare, respectively. Multiple comparisons performed with the Dunn's all-pairs comparison for Kruskal-type ranked data were corrected by the false discovery rate method of Benjamini-Hochberg (*44*). PCA was made in R with stats package and visualized with ggplot2 (*43*) for Vβ frequencies obtained by flow cytometry. All statistical analyses were performed using GraphPad with the help of a professional biostatistician.

SUPPLEMENTARY MATERIALS

immunology.sciencemag.org/cgi/content/full/6/59/eabh1516/DC1
Methods
Figs. S1 to S7
Tables S1 to S7

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. S. Riphagen, X. Gomez, C. Gonzalez-Martinez, N. Wilkinson, P. Theocharis, Hyperinflammatory shock in children during COVID-19 pandemic. *Lancet* **395**, 1607–1608 (2020).
2. L. Verdoni, A. Mazza, A. Gervasoni, L. Martelli, M. Ruggeri, M. Ciuffreda, E. Bonanomi, L. D'Antiga, An outbreak of severe Kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: An observational cohort study. *Lancet* **395**, 1771–1778 (2020).
3. A. Belot, D. Antona, S. Renolleau, E. Javouhey, V. Hentgen, F. Angoulvant, C. Delacourt, X. Iriart, C. Ovaert, B. Bader-Meunier, I. Kone-Paut, D. Levy-Bruhl, SARS-CoV-2-related paediatric inflammatory multisystem syndrome: An epidemiological study, France, 1 March to 17 May 2020. *Eurosurveillance* **25**, 2001010 (2020).
4. M. Levin, Childhood multisystem inflammatory syndrome—A new challenge in the pandemic. *N. Engl. J. Med.* **383**, 393–395 (2020).
5. E. M. Dufort, E. H. Koumans, E. J. Chow, E. M. Rosenthal, A. Muse, J. Rowlands, M. A. Barranco, A. M. Maxted, E. S. Rosenberg, D. Easton, T. Udo, J. Kumar, W. Pulver, L. Smith, B. Hutton, D. Blog, H. Zucker; New York State and Centers for Disease Control and Prevention Multisystem Inflammatory Syndrome in Children Investigation Team, Multisystem inflammatory syndrome in children in New York State. *N. Engl. J. Med.* **383**, 347–358 (2020).
6. L. R. Feldstein, E. B. Rose, S. M. Horwitz, J. P. Collins, M. M. Newhams, M. B. F. Son, J. W. Newburger, L. C. Kleinman, S. M. Heidemann, A. A. Martin, A. R. Singh, S. Li, K. M. Tarquinio, P. Jaggi, M. E. Oster, S. P. Zackai, J. Gillen, A. J. Ratner, R. F. Walsh, J. C. Fitzgerald, M. A. Keenaghan, H. Alharash, S. Doymaz, K. N. Clouser, J. S. Giuliano Jr., A. Gupta, R. M. Parker, A. B. Maddux, V. Havalad, S. Ramsingh, H. Bukulmez, T. T. Bradford, L. S. Smith, M. W. Tenforde, C. L. Carroll, B. J. Riggs, S. J. Gertz, A. Daube, A. Lansell, A. Coronado Munoz, C. V. Hobbs, K. L. Marohn, N. B. Halasa, M. M. Patel, A. G. Randolph; Overcoming COVID-19 Investigators; CDC COVID-19 Response Team, Multisystem inflammatory syndrome in U.S. children and adolescents. *N. Engl. J. Med.* **383**, 334–346 (2020).
7. A. Belot, D. Levy-Bruhl, Multisystem inflammatory syndrome in children in the United States. *N. Engl. J. Med.* **383**, 1793–1794 (2020).
8. C. R. Consiglio, N. Cotugno, F. Sardh, C. Pou, D. Amodio, L. Rodriguez, Z. Tan, S. Zicari, A. Ruggiero, G. R. Pascucci, V. Santilli, T. Campbell, Y. Bryceson, D. Eriksson, J. Wang, A. Marchesi, T. Lakshmikanth, A. Campana, A. Villani, P. Rossi; CACTUS Study Team, N. Landegren, P. Palma, P. Brodin, The immunology of multisystem inflammatory syndrome in children with COVID-19. *Cell* **183**, 968–981.e7 (2020).
9. C. Diorio, S. E. Henrickson, L. A. Vella, K. O. McNerney, J. Chase, C. Burudpakdee, J. H. Lee, C. Jasen, F. Balamuth, D. M. Barrett, B. L. Banwell, K. M. Bernt, A. M. Blatz, K. Chiotos, B. T. Fisher, J. C. Fitzgerald, J. S. Gerber, K. Gollomp, C. Gray, S. A. Grupp, R. M. Harris, T. J. Kilbaugh, A. R. O. John, M. Lambert, E. J. Liebling, M. E. Paessler, W. Petrosa, C. Phillips, A. F. Reilly, N. D. Romberg, A. Seif, D. A. Sesok-Pizzini, K. E. Sullivan, J. Vardaro, E. M. Behrens, D. T. Teachey, H. Bassiri, Multisystem inflammatory syndrome in children and COVID-19 are distinct presentations of SARS-CoV-2. *J. Clin. Invest.* **130**, 5967–5975 (2020).
10. M. J. Carter, M. Fish, A. Jennings, K. J. Doores, P. Wellman, J. Seow, S. Acors, C. Graham, E. Timms, J. Kenny, S. Neil, M. H. Malim, S. M. Tibby, M. Shankar-Hari, Peripheral immunophenotypes in children with multisystem inflammatory syndrome associated with SARS-CoV-2 infection. *Nat. Med.* **26**, 1701–1707 (2020).
11. C. N. Gruber, R. S. Patel, R. Trachtman, L. Lepow, F. Amanat, F. Krammer, K. M. Wilson, K. Onel, D. Geanon, K. Tuballes, M. Patel, K. Mouskas, T. O'Donnell, E. Merritt, N. W. Simons, V. Barcessat, D. M. Del Valle, S. Udondem, G. Kang, S. Gangadharan, G. Ofori-Amanfo, U. Laserson, A. Rahman, S. Kim-Schulze, A. W. Charney, S. Gnjatic, B. D. Gelb, M. Merad, D. Bogunovic, Mapping systemic inflammation and antibody responses in multisystem inflammatory syndrome in children (MIS-C). *Cell* **183**, 982–995.e14 (2020).
12. L. A. Vella, J. R. Giles, A. E. Baxter, D. A. Oldridge, C. Diorio, L. Kuri-Cervantes, C. Alanio, M. B. Pampena, J. E. Wu, Z. Chen, Y. J. Huang, E. M. Anderson, S. Gouma, K. O. McNerney, J. Chase, C. Burudpakdee, J. H. Lee, S. A. Apostolidis, A. C. Huang, D. Mathew, O. Kuthuru, E. C. Goodwin, M. E. Weirick, M. J. Bolton, C. P. Arevalo, A. Ramos, C. J. Jasen, P. E. Conrey, S. Sayed, H. M. Giannini, K. D'Andrea; UPenn COVID Processing Unit, N. J. Meyer, E. M. Behrens, H. Bassiri, S. E. Hensley, S. E. Henrickson, D. T. Teachey, M. R. Betts, E. J. Wherry, Deep immune profiling of MIS-C demonstrates marked but transient immune activation compared to adult and pediatric COVID-19. *Sci. Immunol.* **6**, eabf7570 (2021).
13. A. Ramaswamy, N. N. Brodsky, T. S. Sumida, M. Comi, H. Asashima, K. B. Hoehn, N. Li, Y. Liu, A. Shah, N. G. Ravindra, J. Bishai, A. Khan, W. Lau, B. Sellers, N. Bansal, P. Guerrerio, A. Unterman, V. Habet, A. J. Rice, J. Catanzaro, H. Chandnani, M. Lopez, N. Kaminski, C. S. Dela Cruz, J. S. Tsang, Z. Wang, X. Yan, S. H. Kleinstein, D. van Dijk, R. W. Pierce, D. A. Hafler, C. L. Lucas, Immune dysregulation and autoreactivity correlate with disease severity in SARS-CoV-2–associated multisystem inflammatory syndrome in children. *Immunity* **54**, 1083–1095.e7 (2021).
14. A. Schvartz, A. Belot, I. Kone-Paut, Pediatric inflammatory multisystem syndrome and rheumatic diseases during SARS-CoV-2 pandemic. *Front. Pediatr.* **8**, 833 (2020).
15. J.-L. Casanova, H. C. Su, A global effort to define the human genetics of protective immunity to SARS-CoV-2 infection. *Cell* **181**, 1194–1199 (2020).
16. V. Sancho-Shimizu, P. Brodin, A. Cobat, C. M. Biggs, J. Toubiana, C. L. Lucas, S. E. Henrickson, A. Belot; MIS-C@CHGE, S. G. Tangye, J. D. Milner, M. Levin, L. Abel,

D. Bogunovic, J.-L. Casanova, S.-Y. Zhang, SARS-CoV-2–related MIS-C: A key to the viral and genetic causes of Kawasaki disease? *J. Exp. Med.* **218**, e20210446 (2021).
17. S. Trouillet-Assant, S. Viel, A. Gaymard, S. Pons, J.-C. Richard, M. Perret, M. Villard, K. Brengel-Pesce, B. Lina, M. Mezidi, L. Bitker, A. Belot, Type I IFN immunoprofiling in COVID-19 patients. *J. Allergy Clin. Immunol.* **146**, 206–208.e2 (2020).
18. J. Hadjadj, N. Yatim, L. Barnabei, A. Corneau, J. Boussier, N. Smith, H. Péré, B. Charbit, V. Bondet, C. Chenevier-Gobeaux, P. Breillat, N. Carlier, R. Gauzit, C. Morbieu, F. Pène, N. Marin, N. Roche, T.-A. Szwebel, S. H. Merkling, J.-M. Treluyer, D. Veyer, L. Mouthon, C. Blanc, P.-L. Tharaux, F. Rozenberg, A. Fischer, D. Duffy, F. Rieux-Laucat, S. Kernéis, B. Terrier, Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
19. P. Bastard, L. B. Rosen, Q. Zhang, E. Michailidis, H.-H. Hoffmann, Y. Zhang, K. Dorgham, Q. Philippot, J. Rosain, V. Béziat, J. Manry, E. Shaw, L. Haljasmägi, P. Peterson, L. Lorenzo, L. Bizien, S. Trouillet-Assant, K. Dobbs, A. A. de Jesus, A. Belot, A. Kallaste, E. Catherinot, Y. Tandjaoui-Lambiotte, J. L. Pen, G. Kerner, B. Bigio, Y. Seeleuthner, R. Yang, A. Bolze, A. N. Spaan, O. M. Delmonte, M. S. Abers, A. Aiuti, G. Casari, V. Lampasona, L. Piemonti, F. Ciceri, K. Bilguvar, R. P. Lifton, M. Vasse, D. M. Smadja, M. Migaud, J. Hadjadj, B. Terrier, D. Duffy, L. Quintana-Murci, D. van de Beek, L. Roussel, D. C. Vinh, S. G. Tangye, F. Haerynck, D. Dalmau, J. Martinez-Picado, P. Brodin, M. C. Nussenzweig, S. Boisson-Dupuis, C. Rodríguez-Gallego, G. Vogt, T. H. Mogensen, A. J. Oler, J. Gu, P. D. Burbelo, J. I. Cohen, A. Biondi, L. R. Bettini, M. D'Angio, P. Bonfanti, P. Rossignol, J. Mayaux, F. Rieux-Laucat, E. S. Husebye, F. Fusco, M. V. Ursini, L. Imberti, A. Sottini, S. Paghera, E. Quiros-Roldan, C. Rossi, R. Castagnoli, D. Montagna, A. Licari, G. L. Marseglia, X. Duval, J. Ghosn; HGID Lab; NIAID-USUHS Immune Response to COVID Group; COVID Clinicians; COVID-STORM Clinicians; Imagine COVID Group; French COVID Cohort Study Group; Milieu Intérieur Consortium; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort, J. S. Tsang, R. Goldbach-Mansky, K. Kisand, M. S. Lionakis, A. Puel, S.-Y. Zhang, S. M. Holland, G. Gorochov, E. Jouanguy, C. M. Rice, A. Cobat, L. D. Notarangelo, L. Abel, H. C. Su, J.-L. Casanova, Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
20. Q. Zhang, P. Bastard, Z. Liu, J. L. Pen, M. Moncada-Velez, J. Chen, M. Ogishi, I. K. D. Sabli, S. Hodeib, C. Korol, J. Rosain, K. Bilguvar, J. Ye, A. Bolze, B. Bigio, R. Yang, A. A. Arias, Q. Zhou, Y. Zhang, F. Onodi, S. Korniotis, L. Karpf, Q. Philippot, M. Chbihi, L. Bonnet-Madin, K. Dorgham, N. Smith, W. M. Schneider, B. S. Razooky, H.-H. Hoffmann, E. Michailidis, L. Moens, J. E. Han, L. Lorenzo, L. Bizien, P. Meade, A.-L. Neehus, A. C. Ugurbil, A. Corneau, G. Kerner, P. Zhang, F. Rapaport, Y. Seeleuthner, J. Manry, C. Masson, Y. Schmitt, A. Schlüter, T. L. Voyer, T. Khan, J. Li, J. Fellay, L. Roussel, M. Shahrooei, M. F. Alosaimi, D. Mansouri, H. Al-Saud, F. Al-Mulla, F. Almourfi, S. Z. Al-Muhsen, F. Alsohime, S. A. Turki, R. Hasanato, D. van de Beek, A. Biondi, L. R. Bettini, M. D'Angio, P. Bonfanti, L. Imberti, A. Sottini, S. Paghera, E. Quiros-Roldan, C. Rossi, A. J. Oler, M. F. Tompkins, C. Alba, I. Vandernoot, J.-C. Goffard, G. Smits, I. Migeotte, F. Haerynck, P. Soler-Palacin, A. Martin-Nalda, R. Colobran, P.-E. Morange, S. Keles, F. Çölkesen, T. Ozcelik, K. K. Yasar, S. Senoglu, Ş. N. Karabela, C. Rodríguez-Gallego, G. Novelli, S. Hraiech, Y. Tandjaoui-Lambiotte, X. Duval, C. Laouénan; COVID-STORM Clinicians; COVID Clinicians; Imagine COVID Group; French COVID Cohort Study Group; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort; NIAID-USUHS/TAGC COVID Immunity Group, A. L. Snow, C. L. Dalgard, J. D. Milner, D. C. Vinh, T. H. Mogensen, N. Marr, A. N. Spaan, B. Boisson, S. Boisson-Dupuis, J. Bustamante, A. Puel, M. J. Ciancanelli, I. Meyts, T. Maniatis, V. Soumelis, A. Amara, M. Nussenzweig, A. García-Sastre, F. Krammer, A. Pujol, D. Duffy, R. P. Lifton, S.-Y. Zhang, G. Gorochov, V. Béziat, E. Jouanguy, V. Sancho-Shimizu, C. M. Rice, L. Abel, L. D. Notarangelo, A. Cobat, H. C. Su, J.-L. Casanova, Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
21. Y. Choi, J. A. Lafferty, J. R. Clements, J. K. Todd, E. W. Gelfand, J. Kappler, P. Marrack, B. L. Kotzin, Selective expansion of T cells expressing V beta 2 in toxic shock syndrome. *J. Exp. Med.* **172**, 981–984 (1990).
22. D. Thomas, O. Dauwalder, V. Brun, C. Badiou, T. Ferry, J. Etienne, F. Vandenesch, G. Lina, *Staphylococcus aureus* superantigens elicit redundant and extensive human Vβ patterns. *Infect. Immun.* **77**, 2043–2050 (2009).
23. C. Pannetier, S. Delassus, S. Darche, C. Saucier, P. Kourilsky, Quantitative titration of nucleic acids by enzymatic amplification reactions run to saturation. *Nucleic Acids Res.* **21**, 577–583 (1993).
24. C. Fozza, F. Barraqueddu, G. Corda, S. Contini, P. Virdis, F. Dore, S. Bonfigli, M. Longinotti, Study of the T-cell receptor repertoire by CDR3 spectratyping. *J. Immunol. Methods* **440**, 1–11 (2017).
25. G. Gorochov, A. U. Neumann, A. Kereveur, C. Parizot, T. Li, C. Katlama, M. Karmochkine, G. Raguin, B. Autran, P. Debré, Perturbation of $CD4^+$ and $CD8^+$ T-cell repertoires during progression to AIDS and regulation of the $CD4^+$ repertoire during antiviral therapy. *Nat. Med.* **4**, 215–221 (1998).
26. N. Ouldali, J. Toubiana, D. Antona, E. Javouhey, F. Madhi, M. Lorrot, P.-L. Léger, C. Galeotti, C. Claude, A. Wiedemann, N. Lachaume, C. Ovaert, M. Dumortier, J.-E. Kahn, A. Mandelcwajg, L. Percheron, B. Biot, J. Bordet, M.-L. Girardin, D. D. Yang, M. Grimaud, M. Oualha, S. Allali, F. Bajolle, C. Beyler, U. Meinzer, M. Levy, A.-M. Paulet, C. Levy, R. Cohen, A. Belot, F. Angoulvant; French Covid-19 Paediatric Inflammation Consortium, Association of intravenous immunoglobulins plus methylprednisolone vs immunoglobulins alone with course of fever in multisystem inflammatory syndrome in children. *JAMA* **325**, 855–864 (2021).
27. J. K. Todd, M. Ressman, S. A. Caston, B. H. Todd, A. M. Wiesenthal, Corticosteroid therapy for patients with toxic shock syndrome. *JAMA* **252**, 3399–3402 (1984).
28. D. Y. M. Leung, C. Meissner, D. Fulton, P. M. Schlievert, The potential role of bacterial superantigens in the pathogenesis of Kawasaki syndrome. *J. Clin. Immunol.* **15**, S11–S17 (1995).
29. F. Esper, E. D. Shapiro, C. Weibel, D. Ferguson, M. L. Landry, J. S. Kahn, Association between a novel human coronavirus and Kawasaki disease. *J. Infect. Dis.* **191**, 499–502 (2005).
30. A. Esteve-Sole, J. Anton, R. M. Pino-Ramirez, J. Sanchez-Manubens, V. Fumadó, C. Fortuny, M. Rios-Barnes, J. Sanchez-de-Toledo, M. Girona-Alarcón, J. M. Mosquera, S. Ricart, C. Launes, M. F. de Sevilla, C. Jou, C. Muñoz-Almagro, E. González-Roca, A. Vergara, J. Carrillo, M. Juan, D. Cuadras, A. Noguera-Julian, I. Jordan, L. Alsina, Similarities and differences between the immunopathogenesis of COVID-19-related pediatric multisystem inflammatory syndrome and Kawasaki disease. *J. Clin. Invest.* **131**, e144554 (2021).
31. R. A. Porritt, L. Paschold, M. Noval Rivas, M. H. Cheng, L. M. Yonker, H. Chandnani, M. Lopez, D. Simnica, C. Schultheiß, C. Santiskulvong, J. van Eyk, J. K. McCormick, A. Fasano, I. Bahar, M. Binder, M. Arditi, HLA class I-associated expansion of TRBV11-2 T cells in multisystem inflammatory syndrome in children. *J. Clin. Invest.* **131**, e146614 (2021).
32. Y. Tanaka, K. Hoshino-Negishi, Y. Kuboi, F. Tago, N. Yasuda, T. Imai, Emerging role of Fractalkine in the treatment of rheumatic diseases. *Immunotargets Ther.* **9**, 241–253 (2020).
33. J. Kappler, B. Kotzin, L. Herron, E. W. Gelfand, R. D. Bigler, A. Boylston, S. Carrel, D. N. Posnett, Y. Choi, P. Marrack, V beta–specific stimulation of human T cells by staphylococcal toxins. *Science* **244**, 811–813 (1989).
34. M. Hoffman, "Superantigens" may shed light on immune puzzle. *Science* **248**, 685–686 (1990).
35. M. H. Cheng, S. Zhang, R. A. Porritt, M. Noval Rivas, L. Paschold, E. Willscher, M. Binder, M. Arditi, I. Bahar, Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25254–25262 (2020).
36. P. A. Brogan, V. Shah, N. Klein, M. J. Dillon, V beta–restricted T cell adherence to endothelial cells: A mechanism for superantigen-dependent vascular injury. *Arthritis Rheum.* **50**, 589–597 (2004).
37. Y. C. Bartsch, C. Wang, T. Zohar, S. Fischinger, C. Atyeo, J. S. Burke, J. Kang, A. G. Edlow, A. Fasano, L. R. Baden, E. J. Nilles, A. E. Woolley, E. W. Karlson, A. R. Hopke, D. Irimia, E. S. Fischer, E. T. Ryan, R. C. Charles, B. D. Julg, D. A. Lauffenburger, L. M. Yonker, G. Alter, Humoral signatures of protective and pathological SARS-CoV-2 infection in children. *Nat. Med.* **27**, 454–462 (2021).
38. R. Pescarmona, A. Belot, M. Villard, L. Besson, J. Lopez, I. Mosnier, A.-L. Mathieu, C. Lombard, L. Garnier, C. Frachette, T. Walzer, S. Viel, Comparison of RT-qPCR and Nanostring in the measurement of blood interferon response for the diagnosis of type I interferonopathies. *Cytokine* **113**, 446–452 (2019).
39. J. Demaret, A. Walencik, M.-C. Jacob, J.-F. Timsit, F. Venet, A. Lepape, G. Monneret, Interlaboratory assessment of flow cytometric monocyte HLA-DR expression in clinical samples. *Cytometry B Clin. Cytom.* **84B**, 59–62 (2013).
40. J. Hadjadj, C. N. Castro, M. Tusseau, M.-C. Stolzenberg, F. Mazerolles, N. Aladjidi, M. Armstrong, H. Ashrafian, I. Cutcutache, G. Ebetsberger-Dachs, K. S. Elliott, I. Durieu, N. Fabien, M. Fusaro, M. Heeg, Y. Schmitt, M. Bras, J. C. Knight, J.-C. Lega, G. Lesca, A.-L. Mathieu, M. Moreews, B. Moreira, A. Nosbaum, M. Page, C. Picard, T. Ronan Leahy, I. Rouvet, E. Ryan, D. Sanlaville, K. Schwarz, A. Skelton, J.-F. Viallard, S. Viel, M. Villard, I. Callebaut, C. Picard, T. Walzer, S. Ehl, A. Fischer, B. Neven, A. Belot, F. Rieux-Laucat, Early-onset autoimmunity associated with SOCS1 haploinsufficiency. *Nat. Commun.* **11**, 5341 (2020).
41. D. A. Bolotin, S. Poslavsky, I. Mitrophanov, M. Shugay, I. Z. Mamedov, E. V. Putintseva, D. M. Chudakov, MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
42. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, *circlize* implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
43. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
44. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
45. G. I. Rice, G. M. A. Forte, M. Szynkiewicz, D. S. Chase, A. Aeby, M. S. Abdel-Hamid, S. Ackroyd, R. Allcock, K. M. Bailey, U. Balottin, C. Barnerias, G. Bernard, C. Bodemer, M. P. Botella, C. Cereda, K. E. Chandler, L. Dabydeen, R. C. Dale, C. De Laet,

C. G. E. L. De Goede, M. D. Toro, L. Effat, N. N. Enamorado, E. Fazzi, B. Gener, M. Haldre, J.-P. S.-M. Lin, J. H. Livingston, C. M. Lourenco, W. Marques Jr., P. Oades, P. Peterson, M. Rasmussen, A. Roubertie, J. L. Schmidt, S. A. Shalev, R. Simon, R. Spiegel, K. J. Swoboda, S. A. Temtamy, G. Vassallo, C. N. Vilain, J. Vogt, V. Wermenbol, W. P. Whitehouse, D. Soler, I. Olivieri, S. Orcesi, M. S. Aglan, M. S. Zaki, G. M. H. Abdel-Salam, A. Vanderver, K. Kisand, F. Rozenberg, P. Lebon, Y. J. Crow, Assessment of interferon-related biomarkers in Aicardi-Goutières syndrome associated with mutations in *TREX1*, *RNASEH2A*, *RNASEH2B*, *RNASEH2C*, *SAMHD1*, and *ADAR*: A case-control study. *Lancet Neurol.* **12**, 1159–1169 (2013).

Acknowledgments: This work is dedicated to the memory of T. Kawasaki. We thank the patients and families who contributed to this work. We also thank L. Ma and L. Lemée from the Biomics Platform C2RT, Institut Pasteur (Paris, France), supported by France Génomique (ANR-10-INBS-09-09) and IBISA. Human biological samples and associated data were obtained from NeuroBioTec (CRB HCL, Lyon France, Biobank BB-0033-00046), and we thank N. Duffay for advice and help for the collection. We acknowledge G. Oriol for technical advice. We thank C. Bailey for discussion and NanoString for providing kits. **Funding:** This work was supported by Hospices Civils de Lyon, Fondation Hospices Civils de Lyon, Square Foundation, Grandir-Fonds de solidarité pour l'enfance, and Olympique Lyonnais Foundation. K.L.G.'s work is supported by the AIR-MI grant (ANR-18-ECVD-0001). E.M.-F. is funded by AIR-MI (ANR-18-ECVD-0001), iReceptorPlus (H2020 Research and Innovation Programme 825821), and SirocCo (ANR-21-CO12-0005-01) grants. D.K.'s contributions are funded by iMAP (ANR-16-RHUS-0001), Transimmunom LabEX (ANR-11-IDEX-0004-02), and TriPoD ERC Research Advanced (Fp7-IdEAS-ErC-322856) grants. T.W. and J.M.'s lab was funded by IDEX Université de Lyon 1 grant. A. Belot, S.T.-A., and M.D. were funded by ANR ANR-20-COVI-0064.
Author contributions: A. Belot, T.W., D.K., J.M., and E.M.-F. designed and analyzed experiments. M. Moreews, K.L.G., A. Bellomo, C.M., R. Pescarmona, S.K.-P., S.D., A.-L.M., M.P., M.V., E.C., K.S., I.R., F. Vandenesch, M. Dreux, P.B., and S.-Y.Z. performed and analyzed experiments. C.M., G.M., and F.Venet conceptualized the fluorescence-activated cell sorting (FACS) analysis. C.M. and R. Pescarmona supervised cytokine experiments. E.J. performed inclusions, chaired the clinical investigation, and took care of all ethical committee agreements. A.P., E.J., and B.K. set up the clinical trial. R. Pouyau, S.T., M.G., T.L., F. Venet, A.M.Z., M. Duperril, H.P., L.G.-C., J.-C.R., M. Mezidi, O.D., J.-M.D.G., F.B., and V.L. provided clinical samples and clinical details for all cohorts. O.T. and V.D. explored HLA in patients with MIS-C, and J.-L.C. and L.A. supervised genetic inference exploration of HLA. O.A. reviewed all statistics. I.R. and E.C. provided biobanking and helped generate material for the study. T.W. and A. Belot supervised, designed, and funded this study. T.W. and A. Belot prepared the initial draft. All authors critically reviewed the paper and agreed on the final form. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All information and data are available upon request, and Fastq data for TRB sequencing were deposited in the NCBI Sequence Read Archive repository under the BioProject ID PRJNA727805. All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials.

Submitted 19 February 2021
Resubmitted 30 April 2021
Accepted 18 May 2021
Published First Release 25 May 2021
Final published 28 June 2021
10.1126/sciimmunol.abh1516

Citation: M. Moreews, K. Le Gouge, S. Khaldi-Plassart, R. Pescarmona, A.-L. Mathieu, C. Malcus, S. Djebali, A. Bellomo, O. Dauwalder, M. Perret, M. Villard, E. Chopin, I. Rouvet, F. Vandenesch, C. Dupieux, R. Pouyau, S. Teyssedre, M. Guerder, T. Louazon, A. Moulin-Zinsch, M. Duperril, H. Patural, L. Giovannini-Chami, A. Portefaix, B. Kassai, F. Venet, G. Monneret, C. Lombard, H. Flodrops, J.-M. De Guillebon, F. Bajolle, V. Launay, P. Bastard, S.-Y. Zhang, V. Dubois, O. Thaunat, J.-C. Richard, M. Mezidi, O. Allatif, K. Saker, M. Dreux, L. Abel, J.-L. Casanova, J. Marvel, S. Trouillet-Assant, D. Klatzmann, T. Walzer, E. Mariotti-Ferrandiz, E. Javouhey, A. Belot, Polyclonal expansion of TCR Vβ 21.3$^+$ CD4$^+$ and CD8$^+$ T cells is a hallmark of multisystem inflammatory syndrome in children. *Sci. Immunol.* **6**, eabh1516 (2021).

Science Immunology

Polyclonal expansion of TCR V# 21.3+ CD4+ and CD8+ T cells is a hallmark of multisystem inflammatory syndrome in children

Marion Moreews, Kenz Le Gouge, Samira Khaldi-Plassart, Rémi Pescarmona, Anne-Laure Mathieu, Christophe Malcus, Sophia Djebali, Alicia Bellomo, Olivier Dauwalder, Magali Perret, Marine Villard, Emilie Chopin, Isabelle Rouvet, Francois Vandenesch, Céline Dupieux, Robin Pouyau, Sonia Teyssedre, Margaux Guerder, Tiphaine Louazon, Anne Moulin-Zinsch, Marie Duperril, Hugues Patural, Lisa Giovannini-Chami, Aurélie Portefaix, Behrouz Kassai, Fabienne Venet, Guillaume Monneret, Christine Lombard, Hugues Flodrops, Jean-Marie De Guillebon, Fanny Bajolle, Valérie Launay, Paul Bastard, Shen-Ying Zhang, Valérie Dubois, Olivier Thaunat, Jean-Christophe Richard, Mehdi Mezidi, Omran Allatif, Kahina Saker, Marlène Dreux, Laurent Abel, Jean-Laurent Casanova, Jacqueline Marvel, Sophie Trouillet-Assant, David Klatzmann, Thierry Walzer, Encarnita Mariotti-Ferrandiz, Etienne Javouhey, and Alexandre Belot

Sci. Immunol. **6** (59), eabh1516. DOI: 10.1126/sciimmunol.abh1516

MIS-C's unique TCR repertoire

Multisystem inflammatory syndrome in children (MIS-C) is a severe complication that develops in children previously infected with SARS-CoV-2, with similar features to Kawasaki disease (KD) and toxic shock syndrome (TSS). It is still unclear what immunologic correlates differentiate MIS-C from KD and TSS. Here, Moreews *et al*. looked at the circulating T cell repertoire and phenotype of 36 patients with MIS-C, 16 with KD, 58 with TSS, and 42 with COVID-19. They found that 75% of patients with MIS-C, and none from the other groups, expressed the V#21.3 T cell receptor # chain variable region in both CD4 and CD8 T cells. These cells had an activated and vascular patrolling phenotype but were not specific to SARS-CoV-2. Together, this work shows unique T cell responses in patients with MIS-C after convalescence.

Science Immunology
AAAS

Supplementary Materials for

Polyclonal expansion of TCR Vβ 21.3$^+$ CD4$^+$ and CD8$^+$ T cells is a hallmark of multisystem inflammatory syndrome in children

Marion Moreews *et al.*

Corresponding author: Alexandre Belot, alexandre.belot@chu-lyon.fr

Sci. Immunol. **6**, eabh1516 (2021)
DOI: 10.1126/sciimmunol.abh1516

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Methods:

Study design and Human subjects

The 36 MIS-C patients were included from April 2020 to April 2021 from French participating centers (HPI COVID). We took advantage of previous collection of KD from Necker's Hospital and additional patients with KD or TSS previously included into a study on toxic shock syndrome (approved by the Ethical review board Sud Est IV, DC-2008-176). Acute COVID-19 patients were derived from either HPI project on pediatric COVID-19 (HPI COVID), n=11 or from two ongoing project on mild adult COVID-19 in health care providers (COVID-SER), n=21 or severe adult COVID-19 in critical care unit (COVID-Rea), n=10. All patients could not be included in all analysis, this information is provided in Table S1.

HPI COVID: Written informed consent was obtained for data collection and blood sampling relating to patients and healthy control subjects. The clinical study for children has been registered on ClinicalTrial.gov(NCT04376476) and approved by the national review board for biomedical research in April 2020 (Comité de Protection des Personnes Sud Méditerranée I, Marseille, France) (ID-RCB: 2020-A01102-37).

COVID-SER: For the mild adult COVID-19 cohort, the clinical study registered on ClinicalTrial.gov (NCT04341142) has been fully detailed 48. In the present study, only patients with mild symptoms of COVID-19 were included. Written informed consent was obtained from all participants and approval was obtained from the national review board for biomedical research in April 2020 (Comité de Protection des Personnes Sud Méditerranée I, Marseille, France; ID RCB 2020-A00932-37).

COVID-rea: For the severe adult COVID-19 cohort, the study was registered to the French National Data Protection Agency under the number 20-097 and was approved by an ethical committee for biomedical research (Comité de Protection des Personnes HCL) under the number N°20-41. In agreement with the General Data Protection Regulation (Regulation (EU) 2016/679 and Directive 95/46/EC) and the French data protection law (Law n°78-17 on 06/01/1978 and Décret n°2019-536 on 29/05/2019), we obtained consent from each patient or his next of kin.

Cytokines and IFN score assessment

Whole blood was sampled on EDTA tubes and plasma was frozen at -20°C within 4 hours following blood collection. Plasma concentrations of IL-6, TNF-a, IFN-g, IL-10, MCP-1, IL-1ra and CD25s were measured by Simpleplex technology using ELLA instrument (ProteinSimple), following manufacturer's instructions. Plasma IFN-α concentrations were determined by single-molecule array (Simoa) on a HD-1 Analyzer (Quanterix) using a commercial kit for IFN-α2 quantification (Quanterix). Whole blood was collected on PAXgene blood RNA tubes (BD Biosciences) or on EDTA tubes for IFN signature, RNA extraction was performed with the kit maxwell 16 LEV simply RNA blood associated with the Maxwell extractor (Promega) and quantified by absorbance (Nanovue). IFN score was obtained using nCounter analysis technology (NanoString Technologies) by calculating the mediane of the normalized count of 6 ISGs as previously described(*45*)

Lymphocytes immunophenotyping

CD3, CD4 and CD8 T lymphocyte subsets were enumerated on EDTA-anticoagulated peripheral whole blood by single-platform the fully automated volumetric single plateforme technology flow cytometer AQUIOS CL (Beckman-Coulter) as previously described(*46*). The phenotypic characterization of B, NK and T activated lymphocyte subsets were performed on EDTA-anticoagulated whole blood using the following combination of monoclonal antibodies: APC-Alexa Fluor 750-conjugated anti-CD3, Pacific Blue-conjugated anti-CD4, Krome Orange-conjugated anti-CD8, FITC-conjugated anti-HLA-DR, APC-conjugated anti-CD19, Krome Orange-conjugated anti-CD16 and ECD-conjugated anti-CD56 (Beckman-Coulter). The preparations were lysed and fixed by thoroughly mixing and incubating for 10 min successively with 500µL of OptiLyse C reagent (Beckman-Coulter) and 1mL of PBS. The cells were centrifuged for 5min at 400g, resuspended in 500µl of PBS and acquired on a NAVIOS flow cytometer (Beckman-Coulter).

Monocyte HLA-DR expression assessment

Monocyte HLA-DR expression was evaluated on EDTA-anticoagulated blood processed within 3 hours after withdrawal. The expression of mHLA-DR was determined using the Anti-HLA-DR/Anti-Monocyte Quantibrite assay (BD Biosciences, San Jose, USA) on a Navios flow cytometer (Beckman Coulter, Hialeah, FL) and flow data were analysed using Navios software (Beckman Coulter).

Total number of antibodies bound per cell (AB/C) were quantified using calibration with a standard curve determined with BD Quantibrite phycoerythrin (PE) beads (BD Biosciences) as described previously (*47*).

T-cell Vβ repertoire analysis and immunophenotyping

The phenotypic analysis of T-cell Vβ repertoire was performed on whole blood sample using the IOTest Beta Mark kit (Beckman-Coulter) containing 24 monoclonal antibodies (mAbs) identifying ~ 70% of the T cell repertoire. Whole blood cells were stained with APC-Alexa Fluor 750-conjugated anti-CD3, Pacific Blue-conjugated anti-CD4, Krome Orange-conjugated anti-CD8 and each combination of 3 FITC-, PE- and FITC/PE-conjugated anti-Vβ mAbs (Beckman-Coulter) in 8 sample tubes. Whole blood sample were lysed with OptiLyse C Lysing Solution (Beckman-Coulter), washed and fixed in 0.5% formaldehyde in PBS. 0.5 to 10^4 T cells were acquired on a NAVIOS flow cytometer and data were analyzed using NAVIOS software. Lymphocytes were first gated according to FSC/SSC parameter, then by selection of CD3+, CD4+ and CD3+CD4- positive cells. The proportion of each Vβ family was compared to the minimum and the mean+2SD of each reference values obtained from data from IOTest Beta Mark® kit to evaluate expanded or restricted Vβ family. Expansions or restrictions were defined respectively for values above the mean+2SD or below the minimum reference values of the corresponding family. Additional samples were analyzed for TRBV from total RNA with Nanostring technology (Supplemental Material).

Nanostring TCR expression analysis

Total RNA was extracted from PAXgene™ tubes using the Maxwell® 16 LEV simplyRNA Blood kit (Promega), following the manufacturer's guidelines. The RNA quantity was determined using a Nanodrop (Thermo Scientific). 200 ng total RNA were hybridized with the nCounter® T cell repertoire panel (Nanostring©, #LBL-10805-01) and counted on an nCounter® FLEX platform according to the manufacturer's guidelines. Raw counts were normalized using internal positive standards and 12 housekeeping genes. Raw counts of *TRBV* genes were expressed as a proportion among total *TRBV* gene counts for each patient and normalized using the median value from the healthy control group.

TCR-sequencing

RNA was extracted from whole blood as reported above. T cell receptor (TCR) alpha/beta libraries were prepared from 300ng of RNA from each sample with SMARTer Human TCR a/b Profiling Kit (Takarabio) following provider protocol as previously described (32). Briefly, the reverse transcription was performed using a mixture of TRBC and TRAC reverse primers and further extended with a template-switching oligonucleotide (SMART-Seq v4). cDNAs were then amplified following two semi-nested PCR: a first PCR with TRBC and TRAC reverse primers as well as a forward primer hybridizing to the SMART-Seq v4 sequence added by template-switching and a second PCR targeting the PCR1 amplicons with reverse and forward primer including Illumina Indexes allowing for sample barcoding. PCR2 are then purified using AMPUre XP beads (Beckman-Coulter). The sequencing was then carried out on a MiSeq Illumina sequencer using the MiSeq v3 PE300 protocol at the Biomics Platform (Institut Pasteur, Paris, France). Single end sequences were aligned and annotated using MiXCR 3.0.13 (*48*), providing a list of clonotypes, each of which is defined as a unique combination of one TRBV gene with one CDR3 amino-acid sequence and one TRBJ gene.

TCR-Seq repertoire analysis

Analyses were performed in R 4.0.3 on the TRB clonotype lists obtained with MiXCR. For each clonotype, read count was recorded. Frequencies for TRBV, TRBJ and clonotypes were calculated based on the total read counts per sample. Chord diagrams were made using the circlize package(*49*) on TRBVBJ frequencies, CDR3 length spectratypes were made using ggplot2 (*43*) using clonotype frequencies. To identify TRBV11-2 expanded clonotypes, first (Q1) and third (Q3) quartiles and the interquartile range (IQR) were computed for all patients without TRBV11-2 expansion. Expanded clonotypes are defined as those with counts superiors to Q3+(1.5*IQR).

Immunophenotyping of Vβ 21.3+ T cells

Thawed PBMC were labeled labeled using Fixable Viability Dye eFluor™ 506 from Thermo Fisher. PBMCs were stained with surface markers, APC-conjugated anti-CD3, BUV486- conjugated anti-CD4, PE-Cy7-conjugated anti-CD8, APC-Cy7-conjugated anti-CD14, APC-Cy7-conjugated anti-CD16, APC-Cy7-conjugated anti-CD19, BV711-conjuagetd anti-CCR7 and BV421-conjugated anti-CD38 (BioLegend),

FITC-conjugated anti-Vb21.3 (Miltenyi), Biotin-conjugated anti-HLA-DR, APC-conjugated CX3CR1 (Ebiosciences), BV605-conjugated anti-CD45RA and streptavidine-conjugated PE-texas Red (BD). Cells were then washed, fixed with PBS/Formalin 2%(Sigma-Aldrich). Cell apoptosis was assessed by annexin V staining with the PE Annexin V Apoptosis Detection Kit I (BD). All samples were acquired on a BD LSRFortessa (BD Biosciences) flow cytometer and analyzed using FlowJo version 10 software.

Stimulation with SARS-CoV-2 overlapping peptide pools and flow cytometry

Briefly, overnight-rested PBMCs were stimulated with SARS-CoV-2 PepTivator pooled peptides (Miltenyi Biotec) at a final concentration of 2 μg ml−1 for 1 h in the presence of 2 μg ml−1 monoclonal antibodies CD28 and CD49d, and then for an additional 5 h with GolgiPlug and GolgiStop (BD Biosciences). Dead cells were labeled using LIVE/DEAD Fixable near IR dye from Invitrogen. Surface markers, including APC-conjugated anti-CD3, BUV486- conjugated anti-CD4, PE-Cy7-conjugated anti-CD8, APC-Cy7-conjugated anti-CD14, APC-Cy7-conjugated anti-CD16 and APC-Cy7-conjugated anti-CD19 (BioLegend) and FITC-conjugated anti-Vb21.3 (Miltenyi) were stained. Cells were then washed, fixed with Cytofix/Cytoperm (BD Biosciences) and stained with V450- conjugated anti-IFNγ (eBioscience). Negative controls without peptide stimulation were run for each sample. All samples were acquired on a BD LSRFortessa (BD Biosciences) flow cytometer and analyzed using FlowJo version 10 software.

Serology

Serum samples were tested with three commercial assays: the Wantai Ab assay detecting total antibodies against the receptor binding domain (RBD) of the S protein, the bioMérieux Vidas assay detecting IgG to the RBD and the Abbott Architect assay detecting IgG to the N protein.

Supplementary Figures

fig.S1

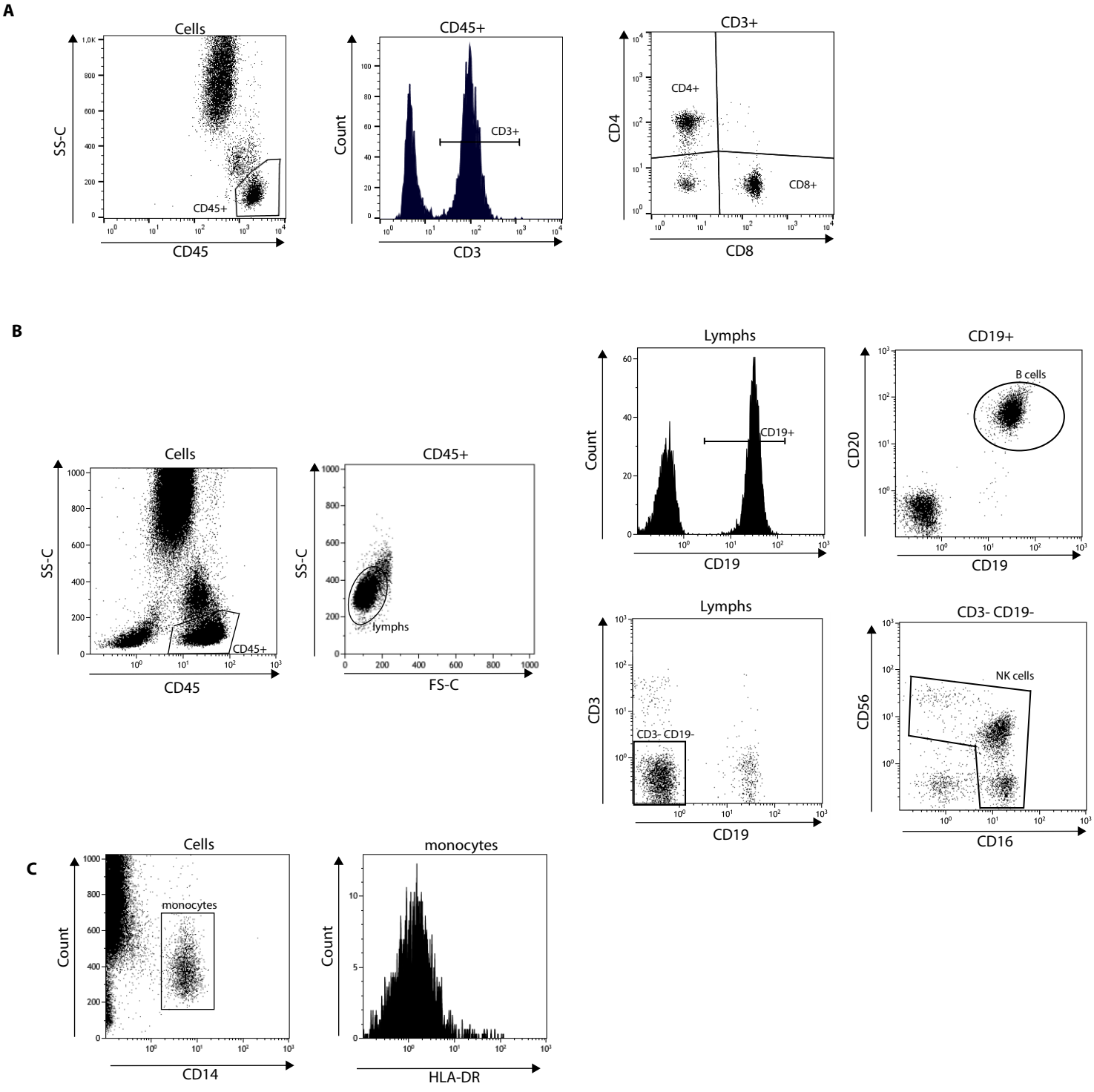


Figure S1: Assessment of T, B, NK cells and HLA-DR monocytes.

Gating strategy and representative flow cytometry plots for immune populations from whole blood stain of CD4+ and CD8+ T cells (A), NK cells and B cells (B) and for HLA-DR expression in monocytes (C).

fig.S2

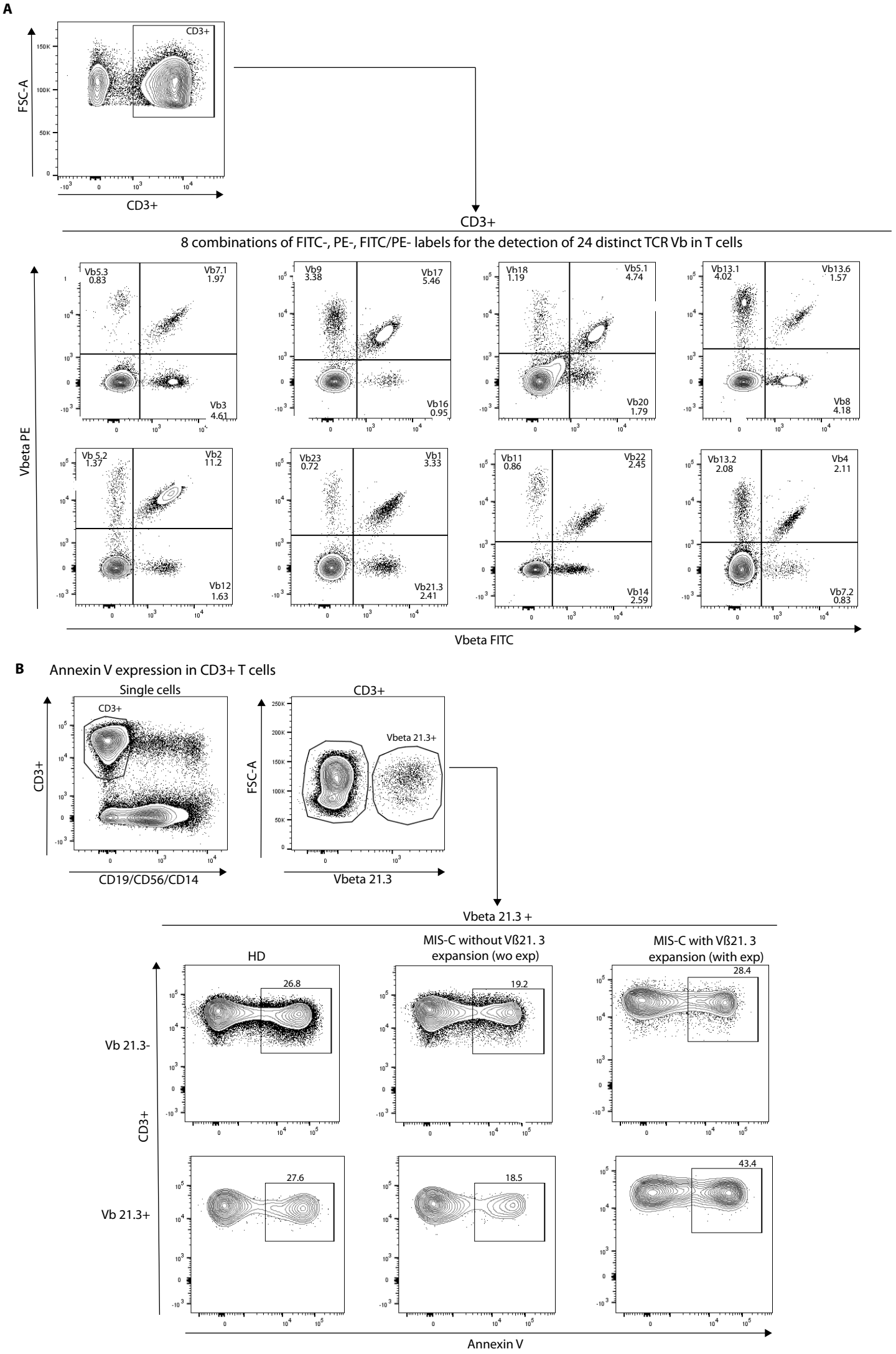

A
CD3+
FSC-A
CD3+
CD3+
8 combinations of FITC-, PE-, FITC/PE- labels for the detection of 24 distinct TCR Vb in T cells
Vbeta PE
Vbeta FITC
Vb5.3 0.83
Vb7.1 1.97
Vb3 4.61
Vb9 3.38
Vb17 5.46
Vb16 0.95
Vb18 1.19
Vb5.1 4.74
Vb20 1.79
Vb13.1 4.02
Vb13.6 1.57
Vb8 4.18
Vb 5.2 1.37
Vb2 11.2
Vb12 1.63
Vb23 0.72
Vb1 3.33
Vb21.3 2.41
Vb11 0.86
Vb22 2.45
Vb14 2.59
Vb13.2 2.08
Vb4 2.11
Vb7.2 0.83
B
Annexin V expression in CD3+ T cells
Single cells
CD3+
CD3+
CD19/CD56/CD14
CD3+
Vbeta 21.3+
FSC-A
Vbeta 21.3
Vbeta 21.3 +
HD
MIS-C without Vß21. 3 expansion (wo exp)
MIS-C with Vß21. 3 expansion (with exp)
Vb 21.3-
CD3+
Vb 21.3+
26.8
19.2
28.4
27.6
18.5
43.4
Annexin V

Figure S2: Assessment of T-cell receptor repertoire and T cell apoptosis by flow cytometry

(A) Representative flow cytometry plots of total CD3+ T cells expressing the indicated V-beta (Vβ) chains using specific antibodies against the corresponding Vβ within PBMCs of one patient (KD-12) as shown in Figure 3A.

(B) Annexin-V staining of T cells in the indicated patients' groups as determined by flow cytometry. Results show the the Annexin-V fluorescence in Vβ21.3+ vs Vβ21.3- T cells.

fig.S3

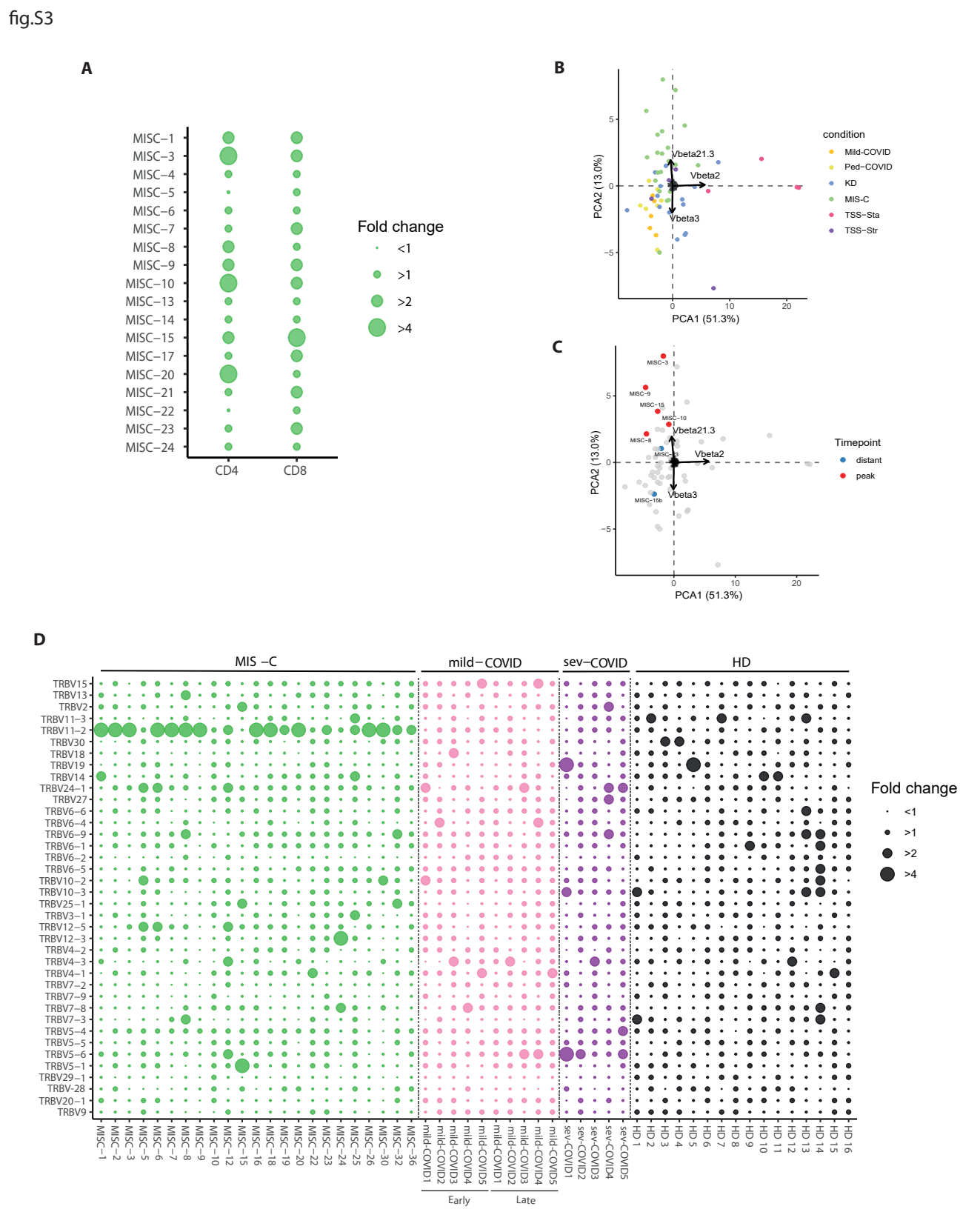
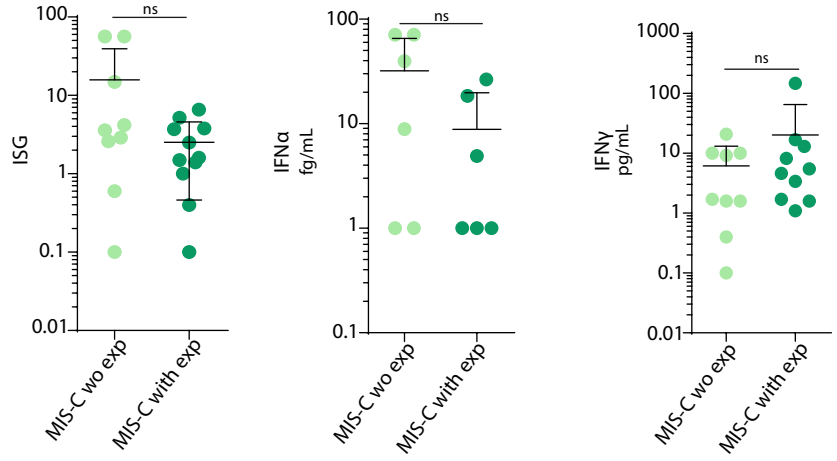

A
MISC−1
MISC−3
MISC−4
MISC−5
MISC−6
MISC−7
MISC−8
MISC−9
MISC−10
MISC−13
MISC−14
MISC−15
MISC−17
MISC−20
MISC−21
MISC−22
MISC−23
MISC−24
CD4
CD8
Fold change
<1
>1
>2
>4
B
PCA2 (13.0%)
PCA1 (51.3%)
Vbeta21.3
Vbeta2
Vbeta3
condition
Mild-COVID
Ped−COVID
KD
MIS-C
TSS−Sta
TSS−Str
C
PCA2 (13.0%)
PCA1 (51.3%)
Vbeta21.3
Vbeta2
Vbeta3
Timepoint
distant
peak
D
MIS −C
mild−COVID
sev−COVID
HD
TRBV15
TRBV13
TRBV2
TRBV11−3
TRBV11−2
TRBV30
TRBV18
TRBV19
TRBV14
TRBV24−1
TRBV27
TRBV6−6
TRBV6−4
TRBV6−9
TRBV6−1
TRBV6−2
TRBV6−5
TRBV10−2
TRBV10−3
TRBV25−1
TRBV3−1
TRBV12−5
TRBV12−3
TRBV4−2
TRBV4−3
TRBV4−1
TRBV7−2
TRBV7−9
TRBV7−8
TRBV7−3
TRBV5−4
TRBV5−5
TRBV5−6
TRBV5−1
TRBV29−1
TRBV-28
TRBV20−1
TRBV9
Early
Late
Fold change
<1
>1
>2
>4

Figure S3: Vβ TCR repertoire analysis

(A) Vβ21.3 TCR expansions in CD4 and CD8 T cells from MIS-C patients as determined by flow cytometry. (B-C) Principal component analysis of the Vβ distribution of CD3 T cells in the different clinical groups. Black vectors correspond to the contribution of Vβ2+, Vβ3+ and Vβ21.3+ cells to the representation. Percentages correspond to the variance captured by each axis. (B) highlights the different clinical conditions while (C) highlights the MIS-C patients during the acute episode (peak) and after (distant). One patient (MISC-15) for which 2 time points were available, was added to the same representation. (D) Transcriptomic analysis of TRBV genes expression in MIS-C patients, mild COVID19 patients at diagnosis and 6 months after infection, COVID19 patients hospitalized in intensive care units and healthy controls using the TCR diversity panel from Nanostring. Results are expressed relative to the median of healthy controls for each gene expression. Red spikes represent TRBV genes with ratio > 4. N=23 in MIS-C patients, N=5 in mild COVID19 group and COVID19 patients hospitalized in intensive care unit and N=16 in healthy controls. TRBV11-2 is the transcript encoding for Vβ21.3.

fig.S4

A Type I & II Interferons



B Other cytokines

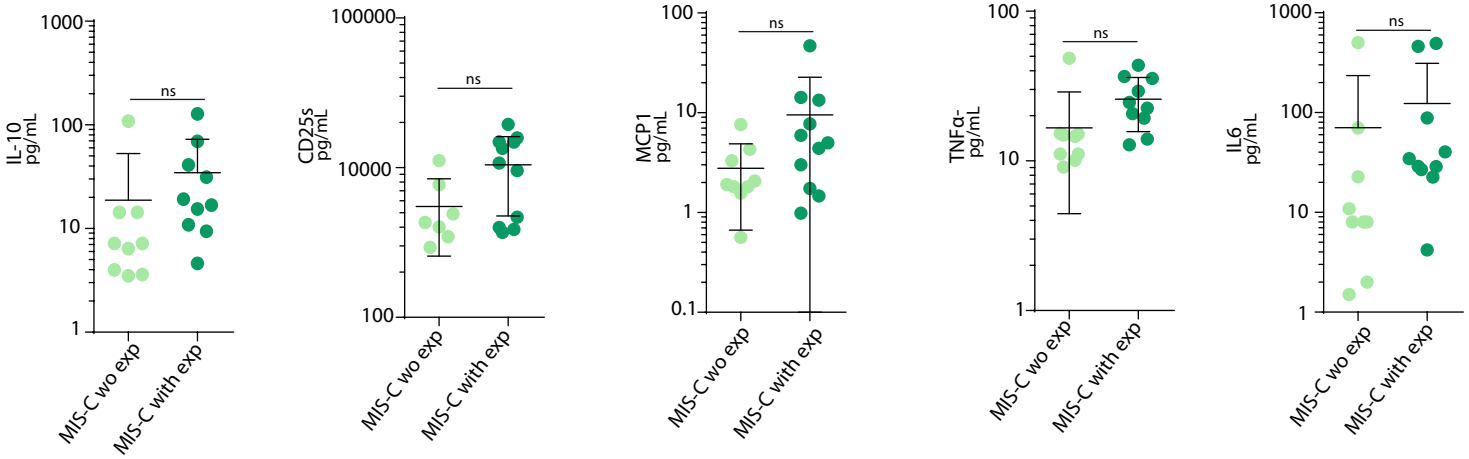


Figure S4: Cytokine assessment in MIS-C

(A-B) Blood ISG score, as described in Fig.2 legend, or serum cytokine levels in MIS-C patients with or without Vβ21.3 expansion. N=6 to 11 per group, as indicated in Table S2; Statistical test: Mann-Withney.

fig.S5

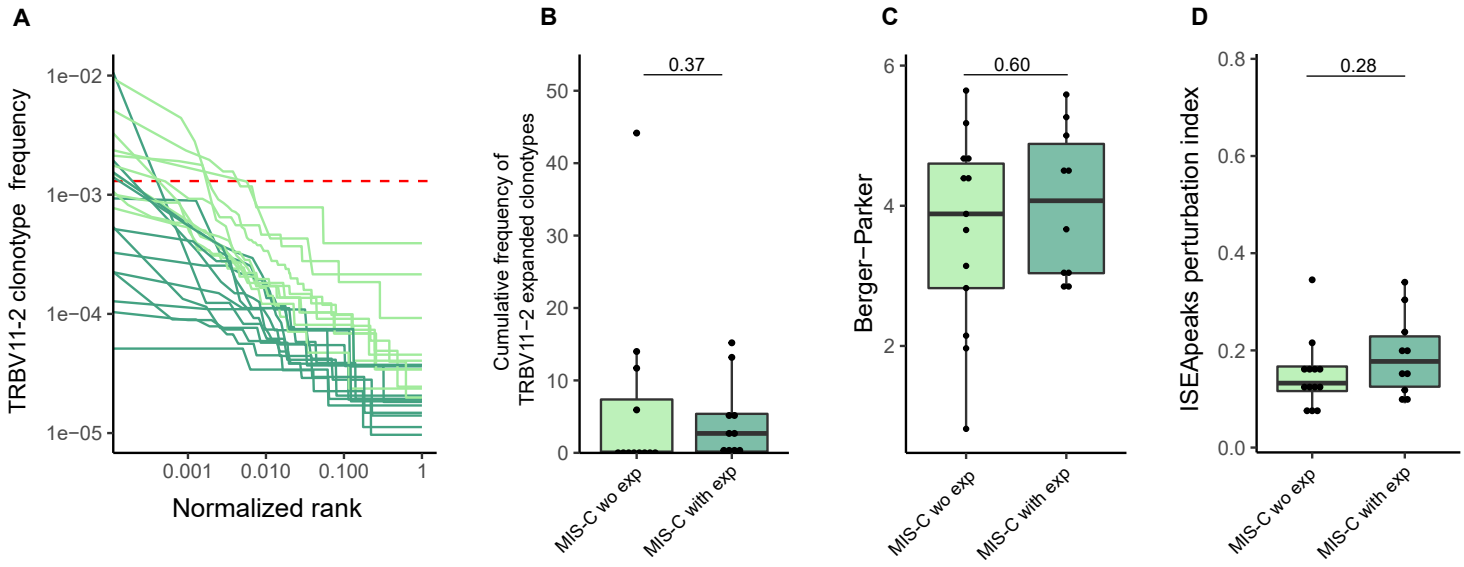

A
B
C
D
TRBV11-2 clonotype frequency
1e−02
1e−03
1e−04
1e−05
0.001
0.010
0.100
1
Normalized rank
Cumulative frequency of TRBV11−2 expanded clonotypes
0.37
0
10
20
30
40
50
MIS-C wo exp
MIS-C with exp
Berger−Parker
0.60
2
4
6
ISEApeaks perturbation index
0.28
0.0
0.2
0.4
0.6
0.8

Figure S5: TRBV11-2 polyclonality assessment

(A) Comparison of TRBV11-2 clonotype distribution between MIS-C patients with TRBV11-2 expansion (light green) and without TRBV11-2 expansion (dark green). TRBV11-2 clonotypes are plotted by increasing ranks as a function of their frequency in the repertoire. The red line indicates the expansion threshold determined as described in the method section. (B) Percentage of expanded clonotypes among TRBV11-2 clonotypes. (C). Berger-Parker index computed on TRBV11-2 clonotypes. (D) Perturbation score of the TRBV11-2 CDR3 length distribution during and after the inflammatory response of MIS-C patients with TRBV11-2 expansions during the acute phase. Statistics were calculated using the Mann-Whitney test.

fig.S6

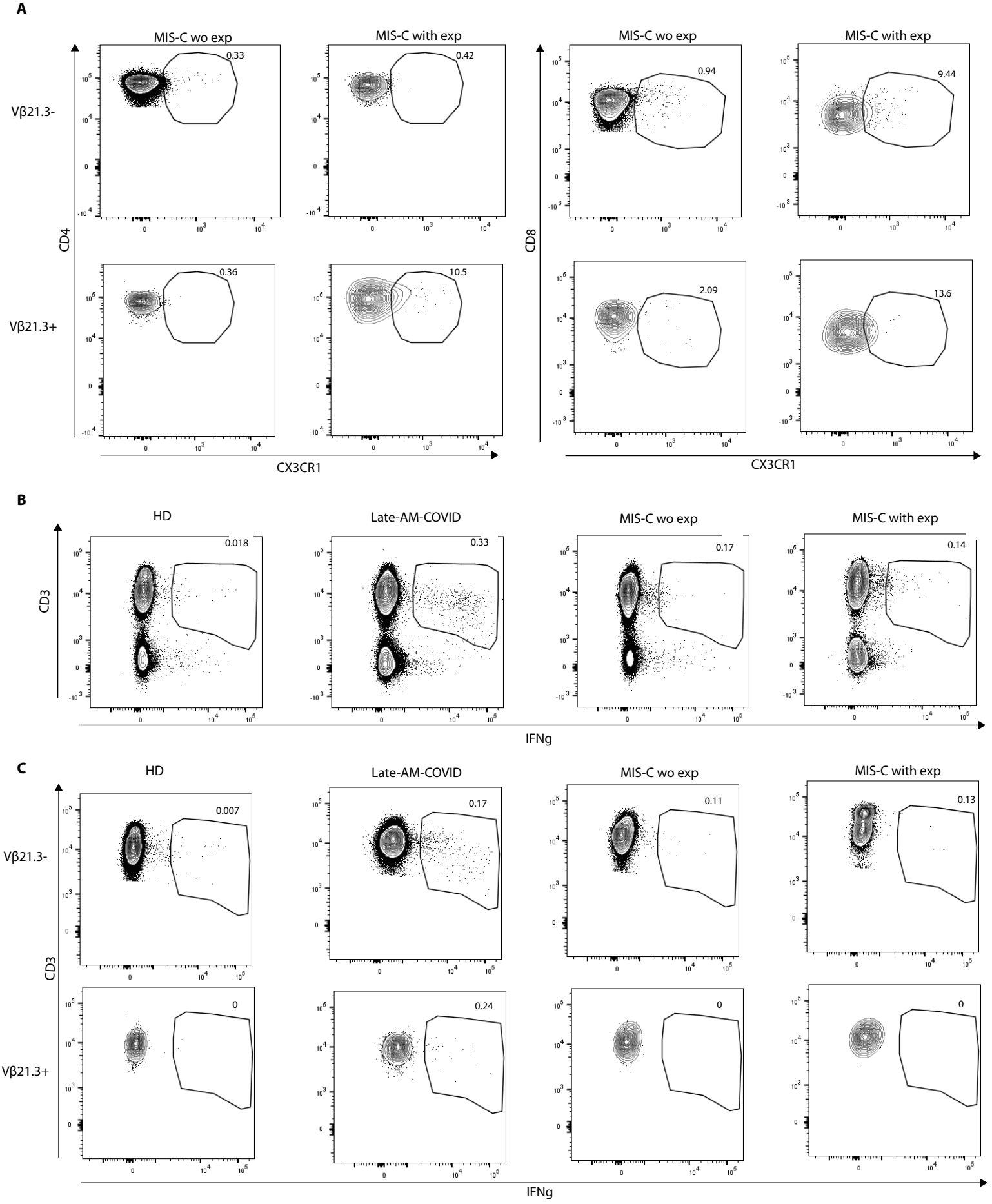

A
MIS-C wo exp
MIS-C with exp
MIS-C wo exp
MIS-C with exp
Vβ21.3-
Vβ21.3+
CD4
CD8
0.33
0.42
0.94
9.44
0.36
10.5
2.09
13.6
CX3CR1
CX3CR1
B
HD
Late-AM-COVID
MIS-C wo exp
MIS-C with exp
CD3
0.018
0.33
0.17
0.14
IFNg
C
HD
Late-AM-COVID
MIS-C wo exp
MIS-C with exp
Vβ21.3-
Vβ21.3+
CD3
0.007
0.17
0.11
0.13
0
0.24
0
0
IFNg

Figure S6: Analysis of T cell activation within Vβ21.3 by flow cytometry

(A) Representative flow cytometry plots of CX3CR1 expression in Vβ21.3- and Vβ21.3+ CD4 or CD8 T cells of one patient in each group as shown in Figure 4E. (B-C) Representative flow cytometry plots of IFNg-expressing T cells following SARS-CoV2 peptide stimulation within CD3+ T cells (B) and within Vβ21.3+ and Vβ21.3- T

fig.S7

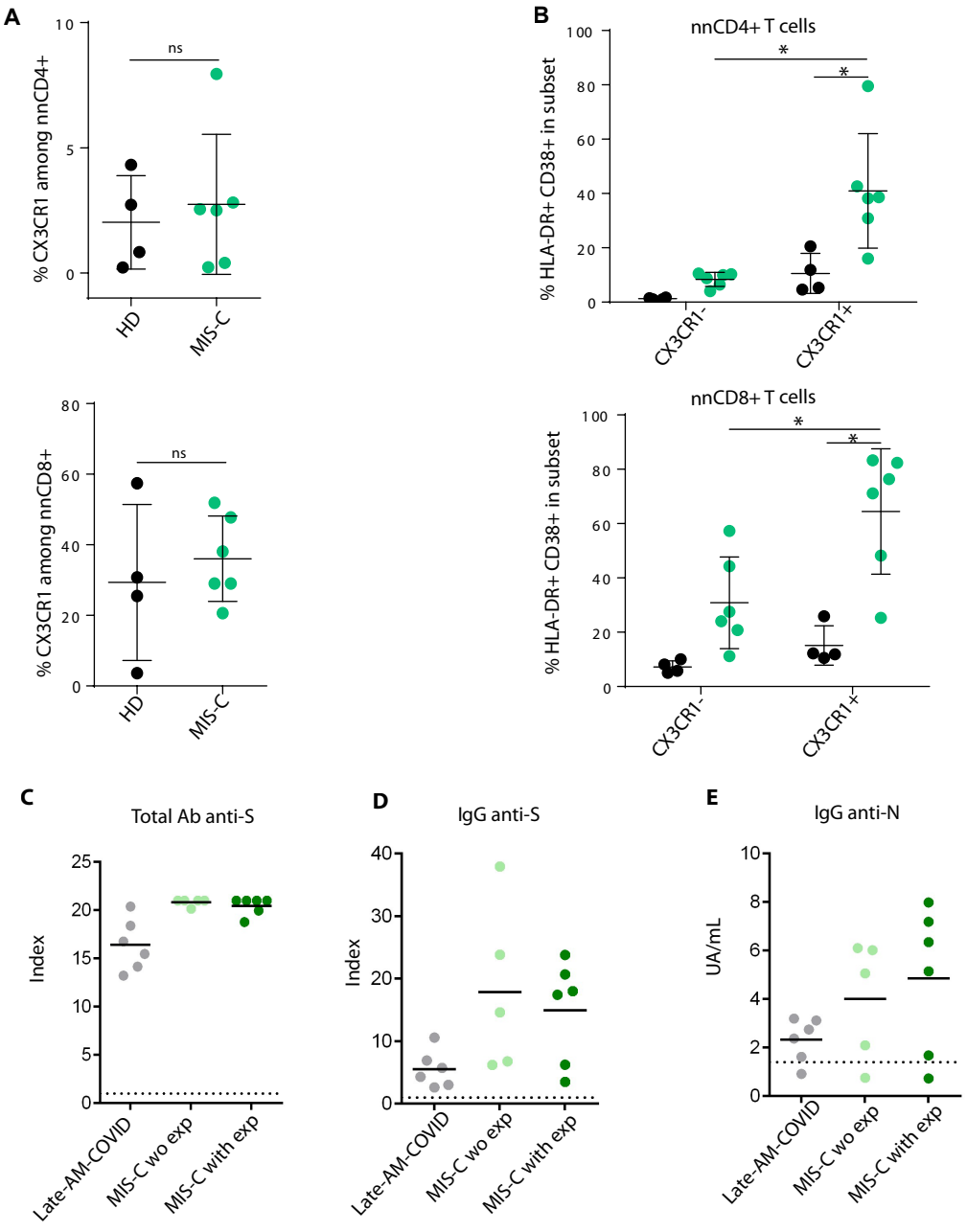

A
% CX3CR1 among nnCD4+
ns
HD
MIS-C
% CX3CR1 among nnCD8+
ns
HD
MIS-C
B
nnCD4+ T cells
% HLA-DR+ CD38+ in subset
*
*
CX3CR1-
CX3CR1+
nnCD8+ T cells
% HLA-DR+ CD38+ in subset
*
*
CX3CR1-
CX3CR1+
C
Total Ab anti-S
Index
Late-AM-COVID
MIS-C wo exp
MIS-C with exp
D
IgG anti-S
Index
Late-AM-COVID
MIS-C wo exp
MIS-C with exp
E
IgG anti-N
UA/mL
Late-AM-COVID
MIS-C wo exp
MIS-C with exp

Figure S7: T cell phenotyping and SARS-CoV2 serology

(A) CX3CR1 expression in non-naive (nn) CD4 or CD8 T cells. (B) HLA-DR and CD38 expression in nn CD4 or CD8 T cells expressing or not CX3CR1 in healthy donors (black) and MIS-C (green). N=4 to 6 per group, as indicated in Table S2; Statistical test: paired or unpaired Mann-Withney. (C-E) SARS-CoV-2 serology (as indicated) of patients with MIS-C or mild COVID-19 that were explored for SARS-Cov 2 T cell activation in Figure 4F-G Dotted lines represent positive threshold recommended by each manufacturer.

Table S1: Clinical description of all MIS-C patients included in the study.

MIS-C patient characteristics: **N=36**	
Age, median [Min-Max]	7.8 [1.2-15.2]
Sex Ratio (Male/Female)	26/10
Main Ethnicity (n, %)	
White	14 (40%)
Afro-Caribbean	20 (54%)
Middle Eastern	2 (6%)
WHO MIS-C criterias* (n, %)	
-fever > 3 days	36 (100%)
-Rash	25 (69%)
-Conjunctivitis	22 (61%)
-Muco-cutaneous inflammation signs (oral, hands or feet)	24 (67%)
-Gastrointestinal symptoms (diarrhoea, vomiting, or abdominal pain)	35 (97%)
-Hypotension/shock	16 (44%)
-Cardiac dysfunction or abnormalities	
elevated Troponin	26 (72%)
elevated NT-pro-BNP	32 (89%)
-Coagulopathy	
elevated D-Dimers	34 (94%)
-Inflammatory markers	
elevated C-reactive protein	36 (100%)
Evidence of COVID-19 (n, %)	36 (100%)
-SARS-Cov-2 PCR	11 (31%)
-SARS-Cov-2 serology	33 (92%)
Vasoactive medications	14 (40%)
Intensive care unit (ICU) admission	31 (86%)

Table S2: Demographic and clinical data of pediatric patients with Kawasaki Disease or Toxic shock syndrome *(S.aureus or S. pyogenes) or COVID-19* and adult patients with mild or severe COVID-19

	pediatric COVID-19	children with Kawasaki Disease	children with Toxic shock syndrome (*S.aureus*)	children with Toxic shock syndrome (*S.pyogenes*)	adult with mild COVID-19	adult severe COVID-19
	N=11	N=16	N=39	N=19	N=21	N=10
Age (y). median [Min-Max]	2.5 [0.1-17.6]	2.9 [0.1-15.8]	14.7 [0.4-18]	4.1 [0.7-18]	42 [29.2-57.3]	60.8 [42.3-78.8]
Sex ratio (M/F)	8/3	7/9	10/29	9/10	3/18	5/5
ICU admission	3 (27%)	3 (18%)	36 (100%) (n=36)	19 (100%)	0 (0%)	10 (100%)
Vasoactive medications	0 (0%)	1 (6%)	19 (65%) (n=29)	15 (83%)	0 (0%)	3 (30%)

Table S3: Patients clinical characteristics

Patient	Age (y)	Sex	ICU admission	Vasoactive medication	TCR Vbeta repertoire analysis by flow cytometry	TCR repertoire sequencing	Cytokines analysis or IFN signature	Treatment (IVIG or Systemic glucocorticoids (SGCs))
MISC-1	4.3	M	Yes	Yes	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-2	12.4	F	Yes	Yes	Yes	Yes	Yes	IVIG 1 dose
MISC-3	11.3	F	Yes	Yes	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-4	2.9	M	No	No	Yes	No	No	IVIG 2 doses + SGCs
MISC-5	1.5	M	No	Yes	Yes	No	Yes	IVIG 1 dose
MISC-6	4.0	F	Yes	Yes	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-7	5.8	M	Yes	Yes	Yes	Yes	Yes	IVIG 1 dose + SGCs
MISC-8	14.3	F	Yes	Yes	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-9	9.8	M	Yes	No	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-10	8.8	M	Yes	No	Yes	Yes	Yes	IVIG 1 dose + SGCs
MISC-11	5.8	F	No	No	Yes	Yes	Yes	SGCs
MISC-12	10.3	F	Yes	Yes	Yes	No	Yes	IVIG 2 doses
MISC-13	5.1	M	Yes	No	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-14	5.0	M	Yes	Yes	Yes	No	No	IVIG 2 doses + SGCs
MISC-15	12.4	M	Yes	No	Yes	Yes	Yes	IVIG 1 dose + SGCs
MISC-16	13.2	M	Yes	Yes	Yes	Yes	Yes	IVIG 2 doses
MISC-17	6.8	M	Yes	No	Yes	No	No	IVIG 2 doses + SGCs
MISC-18	9.4	M	Yes	Yes	No	Yes	Yes	IVIG 2 doses

								+ SGCs
MISC-19	3.8	M	Yes	No	No	Yes	Yes	IVIG 1 dose
MISC-20	7.1	M	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-21	12.4	F	Yes	No	Yes	No	No	IVIG 2 doses + SGCs
MISC-22	1.2	M	Yes	No	No	Yes	Yes	IVIG 2 doses + SGCs
MISC-23	12.6	M	Yes	No	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-24	10.6	M	Yes	No	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-25	15.2	M	Yes	Yes	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-26	10.3	M	Yes	No	Yes	Yes	Yes	IVIG 2 doses + SGCs
MISC-27	10.3	F	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-28	2.3	M	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-29	10.0	M	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-30	4.6	M	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-31	7.5	M	Yes	Yes	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-32	6.8	M	Yes	Yes	Yes	No	Yes	IVIG 1 dose + SGCs
MISC-33	9.6	F	No	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-34	8.8	M	No	No	Yes	No	No	IVIG 2 doses + SGCs
MISC-35	7.7	F	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
MISC-36	4.0	M	Yes	No	Yes	No	Yes	IVIG 2 doses + SGCs
KD-1	4.0	F	No	No	Yes	Yes	Yes	IVIG 1 dose + SGCs
KD-2	8.0	M	Yes	Yes	Yes	No	No	IVIG 1 dose + SGCs
KD-3	2.9	F	No	No	No	Yes	Yes	IVIG 1 dose
KD-4	1.6	F	No	No	No	Yes	Yes	IVIG 1 dose
KD-5	2.5	M	Yes	No	Yes	No	No	-
KD-6	15.8	F	No	No	Yes	No	No	-
KD-7	6.8	F	No	No	Yes	No	No	-
KD-8	0.7	F	No	No	Yes	No	No	-
KD-9	6.2	F	No	No	Yes	No	No	-
KD-10	2.8	F	No	No	Yes	No	No	IVIG 1 dose
KD-11	1.2	F	Yes	No	Yes	No	Yes	IVIG 1 dose
KD-12	2.2	M	No	No	Yes	No	Yes	IVIG 1 dose
KD-13	0.1	M	No	No	Yes	No	Yes	IVIG 1 dose
KD-14	1.3	F	No	No	Yes	No	Yes	IVIG 1 dose
KD-15	5.3	F	No	No	Yes	No	Yes	IVIG 1 dose
KD-16	3.8	M	No	No	Yes	No	Yes	IVIG 1 dose
TSS-Sta-1	11.4	F	Yes	Yes	Yes	No	No	-
TSS-Sta-2	7.3	M	Yes	Yes	Yes	No	No	-
TSS-Sta-3	14.3	F	Yes	Yes	Yes	No	No	-
TSS-Sta-4	15.8	F	Yes	Yes	Yes	No	No	-
TSS-Sta-5	14.2	F	Yes	No	Yes	No	No	-
TSS-Sta-6	18.2	F	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Sta-7	12.7	F	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Sta-8	16.1	F	Yes	Yes	Yes	No	Yes	-
TSS-Sta-9	14.2	F	Yes	Yes	Yes	No	Yes	IVIG 1 dose
TSS-Sta-10	4.5	M	Yes	Yes	Yes	No	Yes	-
TSS-Sta-11	4.8	M	Yes	No	Yes	No	Yes	IVIG 1 dose
TSS-Sta-12	16.0	F	Yes	Yes	Yes	No	No	-
TSS-Sta-13	14.0	F	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Sta-14	17.4	M	Yes	Yes	Yes	No	No	-
TSS-Sta-15	18.0	M	Yes	NA	Yes	No	No	NA
TSS-Sta-16	12.3	F	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Sta-17	15.6	F	Yes	No	Yes	No	No	-
TSS-Sta-18	16.6	F	Yes	No	Yes	No	No	-
TSS-Sta-19	16.7	F	Yes	Yes	Yes	No	No	-

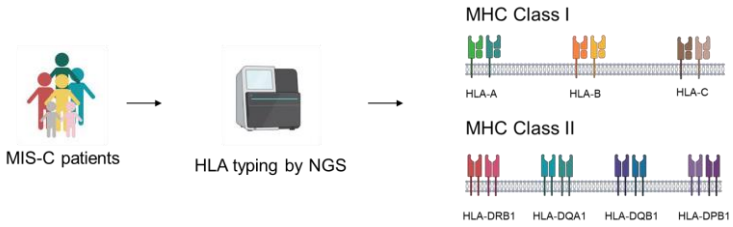
TSS-Sta-20	16.3	F	Yes	NA	Yes	No	No	NA
TSS-Sta-21	11.8	F	Yes	Yes	Yes	No	No	-
TSS-Sta-22	14.7	F	Yes	NA	Yes	No	No	NA
TSS-Sta-23	14.3	F	Yes	No	Yes	No	No	-
TSS-Sta-24	17.3	F	Yes	No	Yes	No	No	-
TSS-Sta-25	16.7	F	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Sta-26	7.8	F	Yes	Yes	Yes	No	No	-
TSS-Sta-27	17.4	F	Yes	NA	Yes	No	No	NA
TSS-Sta-28	17.5	F	Yes	No	Yes	No	No	-
TSS-Sta-29	4.3	M	Yes	NA	Yes	No	No	NA
TSS-Sta-30	4.9	M	Yes	No	Yes	No	No	-
TSS-Sta-31	0.4	M	Yes	Yes	Yes	No	No	-
TSS-Sta-32	12.8	M	Yes	No	Yes	No	No	-
TSS-Sta-33	14.2	M	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Sta-34	12.8	F	Yes	NA	Yes	No	No	NA
TSS-Sta-35	18.2	F	Yes	NA	Yes	No	No	NA
TSS-Sta-36	18.0	F	Yes	No	Yes	No	No	-
TSS-Sta-37	13.5	F	NA	NA	Yes	No	Yes	NA
TSS-Sta-38	17.8	F	NA	NA	Yes	No	Yes	NA
TSS-Sta-39	18.0	F	NA	NA	YEs	No	Yes	NA
TSS-Str-1	3.4	F	Yes	No	Yes	No	No	IVIG 1 dose
TSS-Str-2	0.7	F	Yes	Yes	Yes	No	Yes	-
TSS-Str-3	1.3	F	Yes	Yes	Yes	No	Yes	IVIG 1 dose
TSS-Str-4	3.3	M	Yes	Yes	Yes	No	Yes	IVIG 1 dose
TSS-Str-5	4.1	M	Yes	Yes	Yes	No	Yes	IVIG 1 dose
TSS-Str-6	0.8	F	Yes	Yes	Yes	No	Yes	-
TSS-Str-7	18.0	F	Yes	Yes	Yes	No	No	-
TSS-Str-8	7.6	F	Yes	Yes	Yes	No	No	IVIG 2 doses
TSS-Str-9	12.7	F	Yes	No	Yes	No	No	-
TSS-Str-10	2.0	M	Yes	No	Yes	No	No	-
TSS-Str-11	18.0	M	Yes	Yes	Yes	No	No	-
TSS-Str-12	15.4	F	Yes	Yes	Yes	No	No	-
TSS-Str-13	18.0	M	Yes	No	Yes	No	No	-
TSS-Str-14	0.8	M	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Str-15	5.3	F	Yes	Yes	Yes	No	No	-
TSS-Str-16	13.8	F	Yes	Yes	Yes	No	No	-
TSS-Str-17	0.4	M	Yes	Yes	Yes	No	No	IVIG 1 dose
TSS-Str-18	18.0	F	Yes	Yes	Yes	No	No	-
TSS-Str-19	1.0	M	Yes	Yes	Yes	No	No	IVIG 1 dose
C-COVID1	0.1	M	No	No	Yes	No	Yes	-
C-COVID2	1.3	M	Yes	No	Yes	No	Yes	-
C-COVID3	16.8	F	Yes	No	Yes	No	Yes	SGCs
C-COVID4	6.7	F	Yes	No	Yes	No	Yes	SGCs
C-COVID5	0.5	M	No	No	Yes	No	Yes	-
C-COVID6	2.5	M	No	No	Yes	No	Yes	SGCs
C-COVID7	17.6	M	No	No	Yes	No	Yes	-
C-COVID8	14.8	F	No	No	Yes	No	Yes	SGCs
C-COVID9	0.1	M	No	No	Yes	No	Yes	-
C-COVID10	0.1	M	No	No	Yes	No	Yes	-
C-COVID11	14.2	M	No	No	Yes	No	Yes	-
mild-COVID1	28.7	F	No	No	Yes	No	Yes	-
mild - COVID2	56.7	F	No	No	Yes	No	Yes	-
mild - COVID3	51.3	F	No	No	Yes	No	Yes	-
mild - COVID4	45.5	F	No	No	Yes	No	Yes	-
mild - COVID5	37.3	F	No	No	Yes	No	Yes	-
mild - COVID6	41.1	F	No	No	No	No	Yes	-
mild - COVID7	29.2	M	No	No	No	No	Yes	-
mild - COVID8	31.0	F	No	No	No	No	Yes	-
mild - COVID9	33.9	M	No	No	No	No	Yes	-
mild - COVID10	31.2	F	No	No	No	No	Yes	-
mild - COVID11	57.2	F	No	No	No	No	Yes	-

mild - COVID12	57.3	F	No	No	No	No	Yes	-
mild - COVID13	43.0	F	No	No	No	No	Yes	-
mild - COVID14	35.1	F	No	No	No	No	Yes	-
mild - COVID15	29.3	F	No	No	No	No	Yes	-
mild - COVID16	48.1	F	No	No	No	No	Yes	-
mild - COVID17	31.2	F	No	No	No	No	Yes	-
mild - COVID18	29.4	M	No	No	No	No	Yes	-
mild - COVID19	48.0	F	No	No	No	No	Yes	-
mild - COVID20	48.7	F	No	No	No	No	Yes	-
mild - COVID21	47.4	F	No	No	No	No	Yes	-
sev-COVID1	70.7	M	Yes	No	Yes	No	Yes	SGCs
sev-COVID2	60.2	M	Yes	No	Yes	No	Yes	SGCs
sev-COVID3	55.8	F	Yes	Yes	Yes	No	Yes	SGCs
sev-COVID4	58.1	F	Yes	No	Yes	No	Yes	SGCs
sev-COVID5	78.8	F	Yes	No	Yes	No	Yes	SGCs
sev-COVID6	53.5	M	Yes	No	No	No	Yes	SGCs
sev-COVID7	61.3	F	Yes	No	No	No	Yes	SGCs
sev-COVID8	64.8	M	Yes	Yes	No	No	Yes	SGCs
sev-COVID9	65.4	M	Yes	Yes	No	No	Yes	SGCs
sev-COVID10	42.3	F	Yes	No	No	No	Yes	SGCs

Table S4: Sample distribution as used for each experiment and figure panels

Figure		Number of patients analyzed	
2	**A**	19-30 MIS-C, 3-9 KD, 3 TSS-Sta, 11 ped-COVID, 5-21 mild-COVID, 4-5 sev-COVID, 31 HD	MIS-C 1-3, 5-13, 15-16, 18-20, 22-33, -35; KD -1,3-4, 11-16; TSS-Sta -12, 37-39; ped-COVID 1-11; mild-COVID 1-21; sev-COVID 6-10, 31 HD
	B	29-30 MIS-C, 5 TSS-Str, 8 TSS-Sta, 11 ped-COVID, 10 mild-COVID, 5-10 sev-COVID, 17 HD	MIS-C 1-3, 5-13, 15-16, 18-20, 22-33, -36; KD -1,3-4, 11-16; TSS-Str 2-6; TSS-Sta 8-12, 37-39; ped-COVID 1-11; mild-COVID 1-10; sev-COVID 1-10 ,17 HD
	C	21 MIS-C, 4 KD	MIS-C 1-12, 14-17, 20-24 ; KD 1-4
	D	13 MIS-C, 3 KD	MIS-C 1-3, 6-13, 16-17 ; KD -1,-2, -4
3	**A**	18 MIS-C, 14 KD, 7 ped-COVID, 5 mild-COVID, 4 TSS-Str, 4 TSS-Sta	MIS-C -1, 3-10, 13-15, -17, 20-24; KD 1-2, 5-16; ped-COVID 1-7; mild-COVID 1-5, TSS-Str 3-6, TSS-Sta 8-10, -12
	B	26 MIS-C, 14 KD, 7 ped-COVID, 5 mild-COVID, 5 sev-COVID, 36 TSS-Sta, 19 TSS-Str	MIS-C -1, 3-10, 13-15, 17-24, -26, MIS-C 30-36; KD 1-2, KD 6-16; ped-COVID 1-7; mild-COVID 1-5; sev-COVID 1-5; TSS-Sta 1-36; TSS-Str 1-19
	C-D	20 MIS-C	MIS-C 1-3, -6, 8-9, 12-13, -16, 19-20, 22-26, 30-33
	E	8 MIS-C	MIS-C -5, -7, 12-13-, -22, -24, -26
	F-G	4 MIS-C	MIS-C -3, 8-9, -23
	H	11 MIS-C	MIS-C -1, -3, -7, -9, -15, 20-21, 25-26, 30-31
	I	6 MIS-C, 4 HD	MIS-C -5, -7, -22, -25, 31-32, 4 HD
4	**A-E**	7 MIS-C, 4 HD	MIS-C -5, -7, 22-23, -25, 31-32, 4 HD
	F	14 MIS-C, 7 late-AM-COVID, 6 HD	MIS-C -3, -5, 7-12, -15, -20, -22, -26, -28, -32; late-AM-COVID 1-7, 6 HD
	G	9 MIS-C, 4 late-AM-COVID	MIS-C -3, -5, -7, -9, -20, -22, -26, -28, -32; late-AM-COVID 4-7
S1	**A**	1 MIS-C	MIS-C -36
	B	1 MIS-C	MIS-C -36
	C	1 MIS-C	MIS-C -16
S2	**A**	1 KD	KD-12
	B	2 MIS-C, 1 HD	MIS-C -5, -31
S3	**A**	18 MIS-C	MIS-C -1, 3-10, 13-15, -17, 20-24
	B	21 MIS-C, 14 KD, 7 ped-COVID, 5 mild-COVID, 4 TSS-Str , 5 TSS-Sta	MIS-C 1-15, -17, 21-24; KD 1-2, 5-16; ped-COVID 1-7; mild-COVID 1-5, TSS-Str 3-6; TSS-Sta 8-12
	C	6 MIS-C	MIS-C -3, 8-10, -15- 23
	D	25 MIS-C, 5 Early and late Mild-COVID, 5 Sev-COVID, 1-16 HD	MIS-C 1-12, 15-16, 18-20, 22-26, -30, -32, -36; Early and late COVID 1-5; Sev-COVID 1-5; 16 HD
S4	**A-B**	20 MIS-C	MIS-C 1-3, -5, 8-9, 12-13, -16, 19-20, 22-26, 30-31, -33
S5	**B-D**	18 MIS-C	MIS-C -1, 3-10, 13-15, -17, 20-24
S6	**A**	2 MIS-C	MIS-C 22-23
	B	2 MIS-C, 1 late-AM-COVID, 1HD	MIS-C -5, -9; late-AM-COVID -4
	C	2 MIS-C, 1 late-AM-COVID, 1HD	MIS-C -5, -9; late-AM-COVID -4
S7	**A-B**	6 MIS-C, 4 HD	MIS-C -5, -7, 22-23, -25, 31-32, 4 HD
	C-E	11 MIS-C, 6 late-AM-COVID	MIS-C -5, 7-12, -15, -20, -28, -32; late-AM-COVID 1-3, 5-7

Table S5: HLA sequencing in 13 MIS-C patients



		MHC Class I						MHC Class II							
		HLA-A		HLA-B		HLA-C		HLA-DRB1		HLA-DQA1		HLA-DQB1		HLA-DPB1	
Patient	Vbeta 21.3 expansion	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2	allele 1	allele 2
MISC-1	Yes	02:01	24:02	27:05	51:01	02:02	15:02	11:04	13:01	01:03	05:05	03:01	06:03	04:01	04:02
MISC-3	Yes	01:01	02:01	39:24	51:01	01:02	07:01	15:54	16:01	01:02	01:04	05:02	05:03	02:01	03:01
MISC-8	Yes	02:01	02:01	35:01	44:02	02:10	05:01	04:02	15:03	01:02	03:01	03:02	06:03	01:01	05:01
MISC-10	Yes	02:01	02:01	44:02	57:01	05:01	06:02	07:01	16:01	01:02	02:01	03:03	05:02	04:01	104:01
MISC-15	Yes	03:01	68:01	15:01	51:01	03:03	15:02	03:01	04:01	03:03	05:01	02:01	03:01	03:01	04:01
MISC-16	Yes	23:01	30:02	35:01	39:10	06:02	12:03	03:02	07:01	02:01	04:01	02:02	04:02	01:01	01:01
MISC-23	Yes	03:01	11:01	07:02	44:03	07:02	16:01	07:01	15:01	01:02	02:01	02:02	06:02	02:01	04:01
MISC-24	Yes	01:01	24:02	15:17	57:01	06:02	07:01	04:02	13:02	01:02	03:01	03:02	05:01	04:01	17:01
MISC-2	No	01:01	29:02	14:02	41:01	07:01	08:02	01:02	13:05	01:01	05:05	03:01	05:01	04:01	04:01
MISC-12	No	34:02	74:01	07:02	35:01	04:01	07:02	13:01	15:03	01:02	01:03	06:02	06:08	02:01	18:01
MISC-13	No	03:01	23:01	15:03	18:01	02:10	05:01	07:01	15:03	01:02	02:01	02:02	06:02	01:01	01:01
MISC-18	No	24:02	31:01	51:01	55:01	01:02	15:02	13:01	13:01	01:03	01:03	06:03	06:03	01:01	14:01
MISC-27	No	02:01	23:01	44:02	44:03	04:01	05:01	04:02	07:01	02:01	03:01	02:02	03:02	02:01	17:01

HLA haplotypes in MIS-C patients

Table S6: TRBV11-2 clonotype expansions

	Patients	Cumulative frequency of TRBV11-2 clonotypes	% of TRBV11-2 expanded clonotypes within TRBV11-2	Cumulative frequency of TRBV11-2 expanded clonotypes within TRBV11-2	Cumulative frequency of TRBV11-2 expanded clonotypes within the full repertoire
MIS-C **with** TRBV11-2 expansion	**MISC-9a**	**43.61**	0.50	2.79	1.25
	MISC-3a	**28.35**	0.04	0.00	0.00
	MISC-7	**25.95**	0.05	0.67	0.18
	MISC-26	**19.42**	0.05	0.00	0.00
	MISC-8a	**13.37**	0.18	15.19	2.40
	MISC-6	**7.13**	0.01	0.00	0.00
	MISC-23a	**6.64**	0.01	4.70	0.33
	MISC-13	**5.74**	0.01	2.56	0.15
	MISC-25	**3.97**	1.19	5.61	0.24
	MISC-19	**3.83**	0.01	13.18	0.58
MIS-C **without** TRBV11-2 expansion	**MISC-23b**	3.07	0.02	5.93	0.19
	MISC-15b	2.73	0.07	0.00	0.00
	MISC-10a	2.60	0.00	0.00	0.00
	MISC-3b	2.24	0.06	0.00	0.00
	MISC-12	2.15	0.03	0.00	0.00
	MISC-9b	2.00	0.03	0.00	0.00
	MISC-24	1.84	0.11	0.00	0.00
	MISC-5	1.75	0.01	0.00	0.00
	MISC-22	1.44	0.00	0.00	0.00
	MISC-10b	1.34	0.06	44.16	1.06
	MISC-8b	1.06	0.05	11.68	0.14
	MISC-15a	0.95	0.20	13.99	0.15
	p-value[a]	3.09×10^{-6}	0.722	0.378	0.053
	Significance[b]	***	*ns*	*ns*	*ns*

[a] For each column, values have been compared between the two groups, MIS-C patients with and without TRBV11-2 expansion. The p-values have been obtained using a Mann-Whitney test. [b]The statistical significance of the Mann-Whitney test is shown. *** : p<0,0001; *ns*: non-significant.

Table S7: Raw Data File (excel spreadsheet)

3.3 CHAPTER 5: A DISTINCT T CELL RECEPTOR SIGNATURE ASSOCIATES WITH CARDIAC OUTCOME IN MYOCARDIAL INFARCTION PATIENTS

After showing how TCR can help understand the pathophysiology of an emergent disease, I aimed to determine whether we could identify a minimal TCR feature that could serve as biomarker of disease. This was done in the context of the AIR-MI project in the field of cardio-immunology.

Cardiovascular diseases regroup a large panel of diseases. Myocardial infarction is an ischemic disease that leads to the brutal and massive death of cells from the myocardium. This very inflammatory context is crucial for the establishment of a myocardial injury repair. However, to date, there is no mean to qualitatively assess the ongoing myocardial repair process and predict their outcome.

There are mounting evidences showing that myocardial repair is mediated by T cells (Choo et al., 2017; Epelman et al., 2015; Tang et al., 2019). T cells exert their function upon their activation through their TCR and we hypothesised that, by looking at the collection of TCRs of circulating blood in patients, we could identify signals that can predict the cardiac outcome in patients.

As biopsies repertoire is limited, we used the peripheral blood as a surrogate from the damaged tissues to identify early markers of the myocardial repair. From a cohort of 28 patients who suffered myocardial infarction, my work first consisted in identifying a cardiac signature of TCR that could predict the heart repair outcome, defined as good or poor healing. Patients' classification was defined by the percentage of recuperation of their left ventricular ejection fraction (%δLVEF) between their admission at the 12month follow-up. We also sequenced the peripheral blood of healthy donors from French donation blood centers (EFS) as controls.

The work on this project is summarized as a research article currently under review. In brief, we found that the classical metrics to evaluate repertoires, such as diversity or gene usage, do not provide meaningful result in this context. To tackle this, we developed an innovative approach to identify TCRs that are both correlated to selection pressure and disease outcome. We focused our study on the CDR3, the TCR region in contact with the peptide, and obtained 348 CDR3 forming a discriminant signature between good and poor healers. We then compared

this signature to the healthy donors' repertoire and showed that our signature was significantly enriched in good healers.

A distinct T cell receptor signature associates with cardiac outcome in myocardial infarction patients

Kenz Le Gouge*[1], Diyaa Ashour*[3,4], Margarete Heinrichs**[3,4], Paul Stys**[1], Pierre Barennes**[1], Verena Stangl[5], Lavinia Rech[6], Gerald Hoefler[5], Tobias Gassenmaier[7], Valerie Boivin-Jahns[8], Roland Jahns[9], Ulrich Hofmann[3,4], Dominik Schmitt[3], Anna Frey[3,4], Stefan Störk[3,4], Stefan Frantz[3,4], Peter P. Rainer[6,10,11§], Gustavo Campos Ramos[3,4§], Encarnita Mariotti-Ferrandiz[1,2§]

[1] Sorbonne Université, INSERM, Immunology-Immunopathology-Immunotherapy (i3), Paris, France
[2] Institut Universitaire de France (IUF)
[3] Department of Internal Medicine I, University Hospital Würzburg, Würzburg, Germany
[4] Comprehensive Heart Failure Center, University Hospital Würzburg, Würzburg, Germany
[5] Diagnostic and Research Institute of Pathology, Medical University of Graz
[6] Division of Cardiology, Medical University of Graz, Graz, Austria
[7] Department of Diagnostic and interventional Radiology, University Hospital of Würzburg, Germany
[8] Institute of Pharmacology and Toxicology, University of Würzburg, Würzburg, Germany.
[9] Interdisciplinary Bank of Biomaterials and Data Würzburg (IBDW), University Hospital of Würzburg, Würzburg, Germany.
[10] BioTechMed Graz, Graz, Austria
[11] St. Johann in Tirol General Hospital, St. Johann in Tirol, Austria
* Equally contributed
** Equally contributed
§ Corresponding authors

Peter P. Rainer
CardioScience Lab
Division of Cardiology
Medical University of Graz
A-8036 Graz, Austria
Phone: +43 316 385 12544
Fax: +43 316 385 13733
Email:
peter.rainer@medunigraz.at

Gustavo Ramos
Immunocardiology Lab
University Hospital Würzburg,
Department of Internal Medicine I,
Am Schwarzenberg 15, Haus A15
D-97078 Würzburg, Germany
Phone: +49 931 201 46477
Fax: +49 931 201 46485
Email: Ramos_G@ukw.de

Encarnita Mariotti-Ferrandiz
UMRS959, Immunology-Immunopathology-Immunotherapy lab
Sorbonne Université/INSERM
83 boulevard de l'Hôpital
75013 Paris
France
Phone: +33 1 42 17 74 68
Email:
encarnita.mariotti@sorbonne-universite.fr

Keywords: Myocardial infarction, ejection fraction, T cell, TCR, lymphocytes.

Article type: Rapid communication

List of non-standard abbreviations:

MRI: cardiac magnetic resonance imaging, **ESV**: end-systolic volume, **EDV**: end-diastolic volume **FUP:** follow-up visit, **HF:** Heart failure, **LVEF:** left-ventricular ejection fraction, **MI**: myocardial infarction, **STEMI**: ST-elevation myocardial infarction, **TCR:** T cell receptor, **TRA**: T cell receptor, alpha chain, **TRB**: T cell receptor, beta chain

Main text

Inflammatory processes govern post-myocardial infarction (MI) healing and remodelling. Yet, we cannot predict how acute post-MI immune responses shape long-term cardiac functional outcomes in individual patients. Previous studies showed that $CD4^+$ and $CD8^+$ T cells are key actors of the post-MI repair regulation(1, 2) but did not investigate how they could be used as predictors of the cardiac outcome. Here, we sought to identify T cell receptor (TCR) signatures predicting cardiac functional outcomes in a well-characterized MI patient cohort(3).

TCRs are heterodimeric antigen-specific receptors expressed by T cells, generated through somatic recombination of multiple gene segments in a process that generates a potential repertoire of 10^{19} unique receptors(4). Thus, we hypothesised that analysing TCR repertoires in patients after MI could provide valuable information on the individual immunological status underlying post-MI healing outcomes.

First, we selected acute ST-elevation MI-patients from the ETiCS cohort (**E**tiology, **Ti**tre-Course, and effect of autoimmunity on **S**urvival study(3), Würzburg arm) exhibiting reduced left ventricular ejection fraction (LVEF) assessed by cardiac magnetic resonance (cMRI) on day 4 post-MI (LVEF <50%, 54/150 patients). Amongst those, we further selected patients with complete serial cMRI scans at baseline and 12 months of follow-up (FU) (38/54 patients) (**Figure 1A**). Next, we stratified these patients into “good” versus “poor” healers based on a priori defined cMRI criteria(5)). In brief, patients

showing a ∆LVEF <13% between baseline and follow-up (FUP) were defined as “poor healers” (25/38), whereas those showing greater improvement (∆LVEF >13%) were considered “good healers” (13/38) (**Figure 1B**). In addition to lower ∆LVEF values (P<0.0001), poor healers had significantly greater ∆ end-systolic volumes (ESV) (P<0.0001), whereas ∆ end-diastolic volumes (EDV) did not differ. Age, BMI, infarct size, and routine blood biomarkers were similar in poor and good healers at baseline (**Figure 1B**). After defining groups of MI-patients with diverging healing phenotypes, we extracted RNA of cryopreserved whole blood samples that were collected at hospital admission from all 25 poor and 13 good healers (Nucleospin RNA blood kit, Macherey-Nagel, Düren, Germany). Due to technical limitations inherent to the processing of archived biomaterial, we obtained RNA with sufficient quality from 19 poor and 9 good healers, which were then used for RT-PCR, amplification of TCR alpha and beta chains (TRA and TRB respectively), library preparation and sequencing(6).

To evaluate potential global changes in the TCR repertoires found in blood sampled at baseline, we first measured the diversity of TCRs in each group by computing the Shannon entropy and the gene usage frequency (**Figure 1C-D**), two canonical measures of immune repertoire diversity(7). These basic analyses did not reveal any significant differences between groups. We then hypothesised that distinct TCR repertoire composition may explain healing outcomes. Thus, we devised a refined strategy to identify a set of TCRs that effectively differentiates between cardiac functional outcomes (**Figure 1E**). Accordingly, using statistical modelling from Pogorely *et al*.(8) (**Figure 1E,** panel 1), we computed the probability of each TCR being present in the dataset (Pdata) and identified TCRs that were statistically enriched compared to their expected generation (Pgen). Additionally, we searched for TCRs that were differentially represented between good and bad healers(9) (**Figure 1E,** panel 2). Finally, we selected TCRs that met both criteria: being more represented in one group than the other and having a statistically higher probability of being present in the dataset than expected (**Figure 1E,** panel 3). As a result, we discovered a signature of 348 unique TCRs (**Figure 1F**). Hierarchical clustering based on the presence/absence of this signature allowed a separation of good from poor healers (cluster purity = 0.93). Interestingly, this signature comprising 237 unique TRA and 111 unique TRB was primarily present among the good rather than the poor healers. We then assessed the relevance of the signature by comparing its cumulative frequency in the ETiCS patients versus a group of healthy subjects (n = 27). **Figure H** shows that the cumulative

frequency of the signature is significantly higher in good healers when compared with poor healers and healthy volunteers ($p<0.001$ for both TRA and TRB chains). Moreover, the TCR signature was more enriched in healthy subjects when compared with poor healers ($p<0.001$ for TRA and $p<0.01$ for TRB), suggesting that the clones found in the signature are specifically depleted in MI patients with poor cardiac functional recovery (**Figure 1G**).

In summary, we identified a distinct TCR signature associated with cMRI-assessed post-MI cardiac healing outcomes. Specifically, we found that MI-patients showing a substantial post-MI LVEF improvement were enriched for a set of 348 unique TCRs in the peripheral blood already at baseline (hospitalization). These exploratory findings provide clinical evidence that the individual TCR profile at early stages post-MI contributes to long-term healing outcomes in humans. This supports the notion that modulation of T cell responses may eventually help to improve post-MI remodelling(10) and further confirms that the circulating T cell compartment is impacted by the ongoing cardiac repair. Moreover, the present findings provide a compelling example on the value of immune-based diagnostics in cardiology. Still, our study has some important limitations. First, it was a rather small - though well stratified and characterized - retrospective patient collective. Moreover, since we analysed cryopreserved whole blood samples, we were unable to phenotype T cell subsets that might have accounted for the identified TCR signature. Furthermore, as it is currently not possible to infer which antigens are recognised by the TCRs comprised in this distinct signature, the mechanisms underlying their association with the cardiac functional outcomes remain unknown. Still, our findings underline the potential predictive value of TCR repertoire analyses in MI and might encourage further mechanistic investigations and confirmation in larger patient cohorts.

Author contribution: RJ, VJ, DS, AA, SS recruited the patients for this study and coordinated the clinical study; TG analysed the cMRI images, MH, DA, LV, VS, PPR and GCR retrospectively stratified and selected samples; DA, KLG, PS, PB, and EMF processed the RNA samples, developed bioinformatic tools and analysed the TCR-seq data. DA, MH, KLG, GH, UH, SF, PPR, EMF, GCR made substantial contributions to the conception and design of the present work. KLG, DA, PPR, EMF, and GCR drafted the manuscript. PPR, EMF, and GCR supervised this study. All co-authors critically

revised it, made significant intellectual contribution, and approved the manuscript in its final version to be published.

Study approval
Study was approved by the Ethics committee of the university hospital of Würzburg (approval number 186/07) and all participants gave written informed consent.

Acknowledgements:
This work benefited from equipment and services from the iGenSeq core facility at Institut du Cerveau (ICM, Paris, France) for all the TCR data production.

Data availability statement
The anonymised TCR-seq raw files can be made available upon reasonable request.

Conflict of interest statement
Authors declare no conflict of interest

Funding statement

This work was supported by the European Research Area Network—Cardiovascular Diseases [ERANET-CVD JCT2018, AIR-MI Consortium to GCR (01KL1902), PPR (4168-B) and EMF (ANR-18-ECVD-0001)]. The ETiCS study was supported by the Bundesministerium für Bildung und Forschung (BMBF), grant "Molecular Diagnostics" FKZ 01ES0901 and FKZ 01ES0802 (to V.B.-J. and R.J.). KLG doctoral fellowship was supported by ERANET-CVD JCT2018 (ANR-18-ECVD-0001) and additional support from Sorbonne Université. EMF work was supported by the iReceptorPlus (H2020 Research and Innovation Programme 825821) and SirocCo (ANR-21-CO12-0005-01) grants and the Institut Universitaire de France. GCR is supported by the Interdisciplinary Centre for Clinical Research Würzburg [E-354]. GCR, UH, and SF received funding from the German Research foundation (through the Collaborative Research Centre "Cardio-Immune interfaces" SFB1525, grant number 453989101).

References

1. Forte E, et al. Cross-Priming Dendritic Cells Exacerbate Immunopathology After Ischemic Tissue Damage in the Heart. Circulation. 2021;143(8):821–836.
2. Rieckmann M, et al. Myocardial infarction triggers cardioprotective antigen-specific T helper cell responses. Journal of Clinical Investigation. 2019;129(11):4922–4936.
3. Deubner N, et al. Cardiac beta1-adrenoceptor autoantibodies in human heart disease: rationale and design of the Etiology, Titre-Course, and Survival (ETiCS) Study. Eur J Heart Fail. 2010;12(7):753–762.
4. Bradley P, Thomas PG. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. Annu Rev Immunol. 2019;37(1):547–570.

5. Bulluck H, et al. Defining left ventricular remodeling following acute ST-segment elevation myocardial infarction using cardiovascular magnetic resonance. Journal of Cardiovascular Magnetic Resonance. 2017;19(1):26.
6. Barennes P, et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. Nat Biotechnol. 2021;39(2):236–245.
7. Six A, et al. The Past, Present, and Future of Immune Repertoire Biology – The Rise of Next-Generation Repertoire Analysis. Front Immunol. 2013;4. https://doi.org/10.3389/fimmu.2013.00413.
8. Pogorelyy MV, et al. Method for identification of condition-associated public antigen receptor sequences. eLife. 2018;7:e33050.
9. Emerson RO, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat Genet. 2017;49(5):659–665.
10. Zhao TX, et al. Regulatory T-Cell Response to Low-Dose Interleukin-2 in Ischemic Heart Disease. NEJM Evidence. 2022;1(1):EVIDoa2100009.

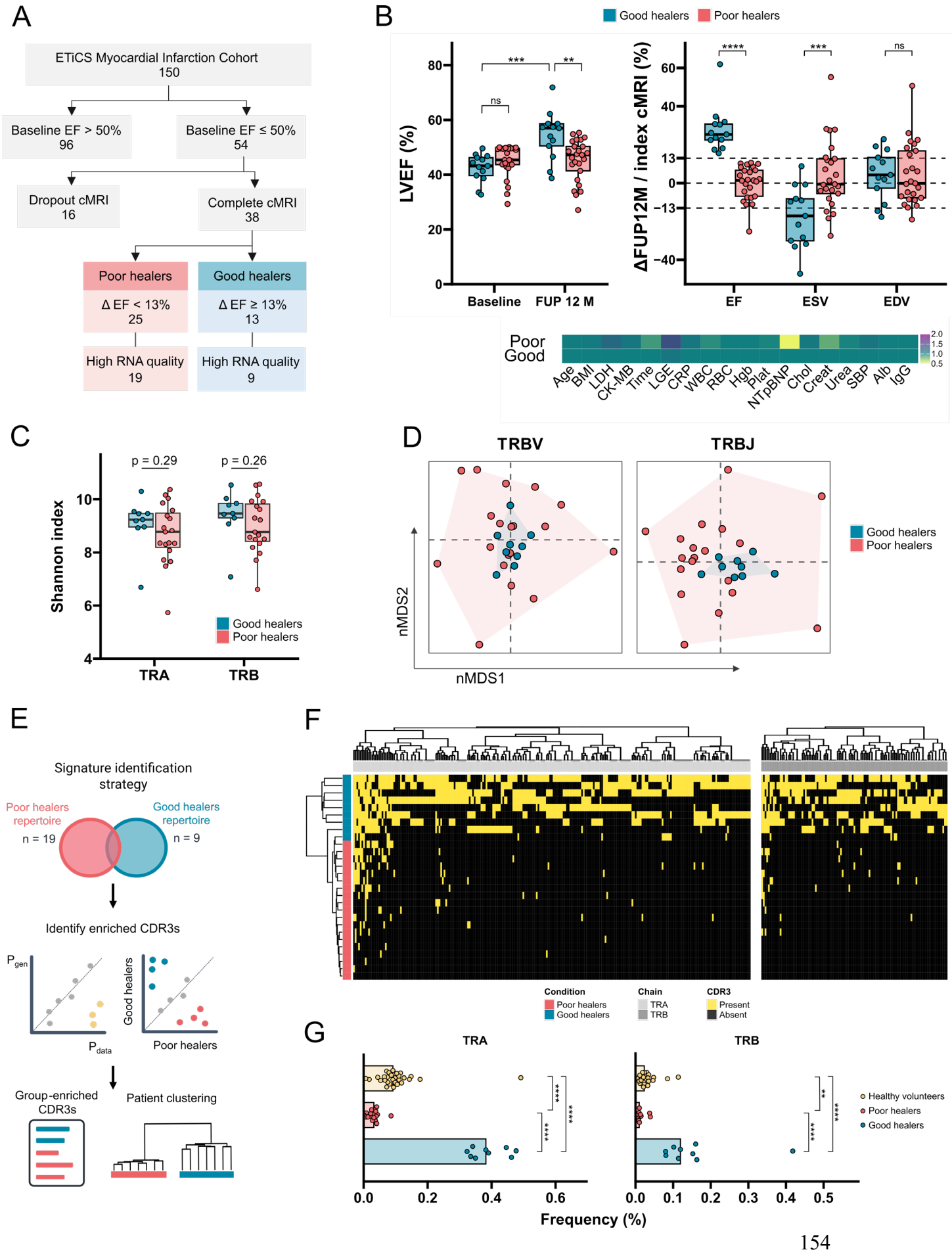

A
ETiCS Myocardial Infarction Cohort
150
Baseline EF > 50%
96
Baseline EF ≤ 50%
54
Dropout cMRI
16
Complete cMRI
38
Poor healers
Δ EF < 13%
25
Good healers
Δ EF ≥ 13%
13
High RNA quality
19
High RNA quality
9
B
Good healers
Poor healers
LVEF (%)
Baseline
FUP 12 M
ΔFUP12M / index cMRI (%)
EF
ESV
EDV
Poor
Good
Age
BMI
LDH
CK-MB
Time
LGE
CRP
WBC
RBC
Hgb
Plat
NTpBNP
Chol
Creat
Urea
SBP
Alb
IgG
C
Shannon index
p = 0.29
p = 0.26
TRA
TRB
Good healers
Poor healers
D
TRBV
TRBJ
nMDS2
nMDS1
Good healers
Poor healers
E
Signature identification strategy
Poor healers repertoire
n = 19
Good healers repertoire
n = 9
Identify enriched CDR3s
Pgen
Pdata
Good healers
Poor healers
Group-enriched CDR3s
Patient clustering
F
Condition
Poor healers
Good healers
Chain
TRA
TRB
CDR3
Present
Absent
G
TRA
TRB
Frequency (%)
Healthy volunteers
Poor healers
Good healers

Figure 1: A distinct T cell receptor signature associates with cardiac functional outcomes in myocardial infarction patients. (A) Patient selection and stratification into good vs poor healers based on cMRI findings. The number of patients is indicated within the boxes (**B**) cMRI findings, including LVEF at baseline and FUP 12 months in good vs poor healers, ΔLVEF (%), ΔESV and ΔEDV (mL), and distribution of confounding factors between groups (BMI: body-mass index; LDH: lactate dehydrogenase; Time: time from pain to intervention; LGE: infarct size assessed by late gadolinium enhancement, CRP: C-reactive protein, WBC: white blood cell count; RBC: red blood cell count, Hgb: hemoglobin; Plat: platelets count; NTproBNP; N-terminal pro-brain natriuretic peptide; Chol: cholesterol; Creat: creatinine; SBP: systolic blood pressure; Alb: albumin. Values are expressed as fold change from the average levels found in the "good healer" group). (**C**) Shannon index based on clone counts from aligned TCR samples for the alpha and beta chain. P-value was computed using Wilcoxon U-test. (**D**) Non-metric dimensional scaling of the TRB variable (TRBV) and joining (TRBJ) gene usage frequencies. (**E**) Strategy used to identify Complementarity-determining region 3 (CDR3, antigen-binding domain) signature. (**F**) Heatmap of the 348 CDR3 signature found in patients. CDR3, by column, were clustered independently by chain. (**G**) Sample's cumulative frequencies (in %) of the CDR3 from the signature. Healthy volunteers are depicted in yellow.

In addition to the results shown in the publication manuscript, we further characterized the signature obtained. Especially, we were interested in characterizing the specificities of the signature, to see if the TCR associated to cardiac outcome were already reported in databases of known TCR specificities. To this end, I have aggregated the TCR specificity information from 3 public databases: Mc-PAS, VDJdb and iEDB (Dhanda et al., 2019; Shugay et al., 2018; Tickotsky et al., 2017). We refined the results by filtering reported TCR with sufficient confidence in the methodology (stimulation with peptide, protein, or direct peptide-MHC binding (tetramer, dextramers). Studies that did not report such TCRs, or typically reported association of TCR with the disease were excluded. CDR3 were categorised as "pathogens" when annotated as specific for bacterial or viral agents, "autoimmune" if derived from self-reactive TCRs, "cancer" if

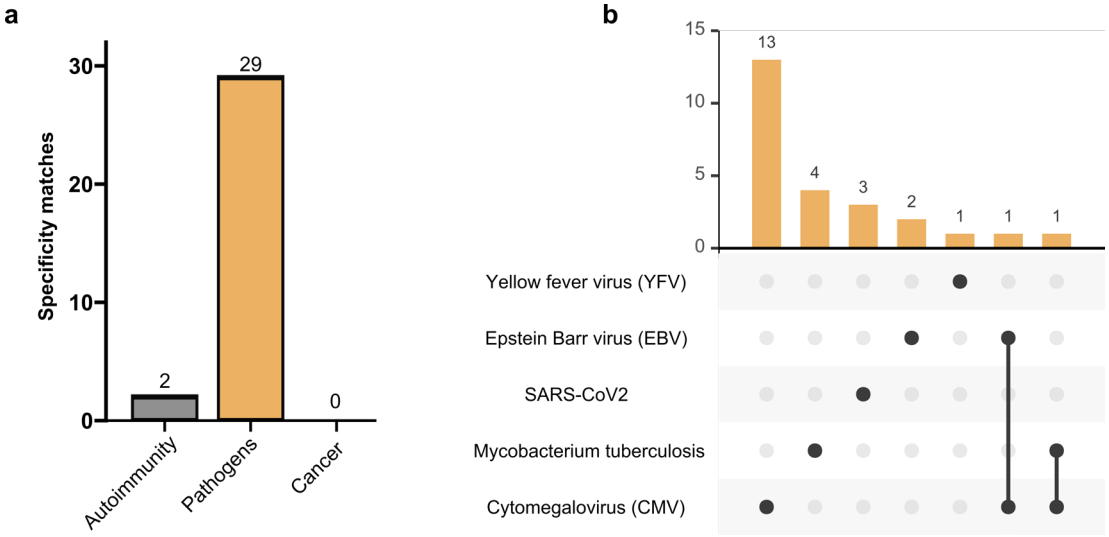


*Figure 14: **ETiCS signature is associated to viral infections**. A: Number of specificity matches from aggregated databases of TCR specificities B: UpSet plot of the viral matches from our ETiCS signature.*

derived from neo-antigens' reactive TCRs. Other categories (Mc-PAS reports "Allergy" for instance) were classified as "others". We then looked for perfect matches between the curated aggregated database and our signature. The results are presented in Figure 19. Out of the 348 CDR3 of our cardiac outcome signature, 31/348 (9%) of CDR3 were also found in the database, of which 29 (94%) were associated to pathogens and only 2 (6%) were reported to bind autoimmune antigens (Figure 19A). We then sought to identify which pathogen specificities were in our signature (Figure 19B). In this UpSet diagram, we can see that more than half (15/29) CDR3s present were reported to bind to

CMV. Second most reported match is Myobacterium tuberculosis with 5/29 CDR3. Of note, Epstein Barr virus (EBV) has been found in 2/29 matches.

These results were also reported by other groups, with some clear links between CVD and CMV infection although no clear mechanisms has been proposed. A meta-analysis of CMV-associated risk to CVD incidence found a significant risk of CMV infection to the contribution of CVD (Wang et al., 2017). A more recent study suggested that active CMV infection could be associated with poorer prognosis in patient by impairing endothelial function (Lebedeva et al., 2020). Supportive evidence of a link between myocardial infarction and CMV infection has been reported with a case of CMV reactivation leading to myocardial infarction in an immunocompetent individual (Yousaf et al., 2021). Atherosclerosis, the most common leading cause of MI, can be initiated by CMV infection. Presence of CMV DNA in arterial walls was associated to further ischemic heart disease (Horváth et al., 2000; Melnick et al., 1983). On the other hand, the link with EBV is still debated. There are reports in the literature of correlation between EBV infection and acute coronary events, but it seems related to overall inflammation rather than autoimmune mechanisms (Binkley et al., 2013). Actively infected patients with EBV can present myocarditis clinical manifestations (Ace and Domb, 2019; Walenta et al., 2006; Watanabe et al., 2019), but these reports are considered unusual and suggest other underlying phenomenon at play to trigger it.

Our results, however, do suggest a link between past encounters with viruses and myocardial repair. Our signature was identified in the peripheral blood, and not in the infarcted tissue. It is possible that all CMV-specific T cells from poor healers were already drawn into the myocardium and present in lower numbers in the circulating blood. There is also evidence that previous viral infections trigger tissue repair through T cell independant mechanisms. For instance H1N1 influenza virus triggers stem cells to undergo rapid proliferation and regenerate the pulmonary tract (Kumar et al., 2011). There is no evidence in the literature of such a mechanism happening for cardiomyocytes.

3.4 CHAPTER 6: IDENTIFYING T CELL ASSOCIATED TO CARDIAC OUTCOME IN OTHER DATASETS

As we have just described, it is possible to identify a CDR3 signature from circulating blood of patients MI. We sought to confirm the relevance of these results in different contexts using infiltrating T cells from internal or external datasets, or a relevant cohort prepared in similar conditions.

3.4.1 Cardiac biopsies

Indeed, we hypothesised that T cells from the peripheral blood could be a good proxy to study what was happening in the tissues, here the heart. Several studies have highlighted how the infiltrating repertoire was perturbed during MI. Murine models of MI demonstrated how $CD4^+$ repertoires of heart were distinct from draining lymph nodes (Rieckmann et al., 2019). They observed T cell expansions and distinct TRBV usage, with high TRBV19. These observations have been evaluated in humans, and suggested a similar low sharing of T cells between circulating and infiltrating compartments (Tang et al., 2019). The high expansions, and private properties of repertoires, prompted us to determine whether the healing signature found in Chapter 5 (p. 146) was linked to infiltrating T cells, or was correlated with the outcome and rather bystander immune response. To study the overlap between circulating and infiltrating T cells, we had access to matched cardiac (n=27) and lungs biopsies (n= 27) from an additional cohort of patients suffering from fatal MI. Tissues were stored in FFPE and multiplex PCR on DNA sequencing was performed using LymphoTrack TRB assay (PGM Invivoscribe). DNA was preferred on these samples as it was considered more stable and more reliable to recover TCR from FFPE samples of low infiltrated tissues such as the cardiac or pulmonary tissues (Pai and Satpathy, 2021). As these data only cover β chain, we limited our analysis to this part of the signature.

The number of CDR3 recovered from lung and cardiac biopsies ranged from 1 to 770 unique sequences, with a high variability between samples (Figure 15A). We also recovered 3 times more CDR3 per sample in lungs than in cardiac tissues (81 ± 100 for cardiac biopsy vs 265±195 CDR3 per lung biopsy). In line with previously published studies, yet not neglecting the low throughput, the cardiac infiltrate CDR3 repertoire was mostly private (Figure 15B) with 96% (2026/2104 unique CDR3 recovered) CDR3s

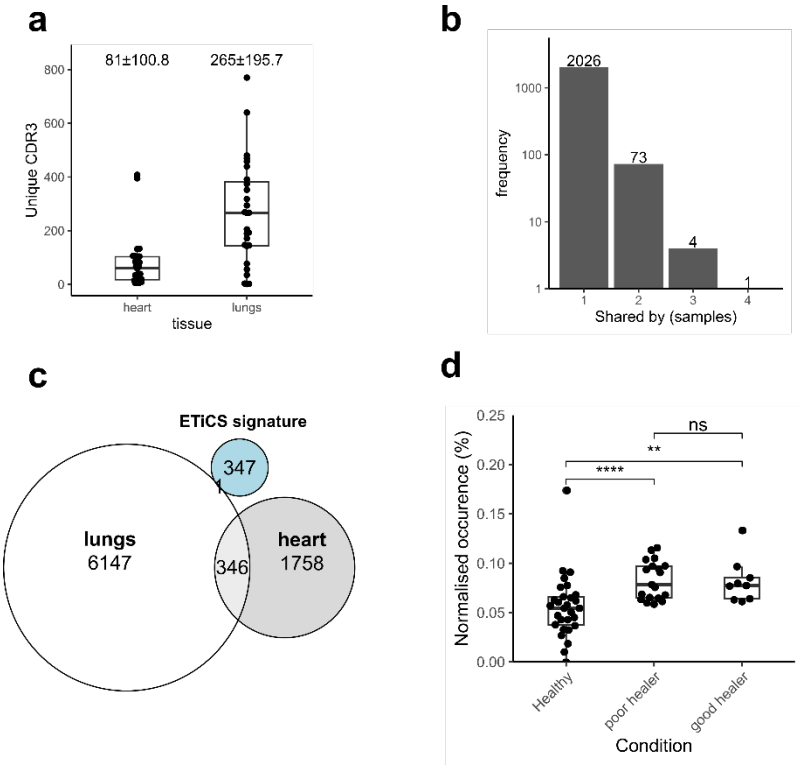


*Figure 15: **Overlap of cardiac biopsies with circulating blood**. **A**: Boxplot of unique CDR3 collected per sample from n=27 cardiac and n=27 patient-matched lung biopsies. Numbers are mean +- standard deviation. **B**: Histogram of the cardiac CDR3 sharing between all cardiac biopsies, logarithmic scale. **C**: Overlap of unique CDR3 from cardiac (heart), pulmonary lungs) and ETiCS signature. **D**: Normalised sharing of infiltrating cardiac CDR3 with repertoire of healthy volunteers, poor healers and good healers. Occurrence is expressed as percentage of the total unique CDR3 in the sample. Wilcoxon U-test, significant symbols: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$*

private to a given individual. There is no evidence for a shared response, as most shared CDR3 were found in 3/27 donors.

We then investigated whether the healing outcome signature found in Chapter 5 (p. 146) was overlapping with the cardiac infiltrate of fatal cardiac MI. To this end, we compared the aggregated CDR3 repertoires of cardiac and pulmonary biopsies with the ETiCS signature. The results displayed in Figure 15C show no overlap between the cardiac infiltrate CDR3s and the circulating T cell signature.

Finally, we sought to look whether the T cell infiltrate was found in the circulating repertoire of patients. I computed the number of unique CDR3 from circulating T cell samples shared with the cardiac infiltrate biopsies, and normalised it by the number of total unique CDR3 in patients. The results are presented in Figure 15D. We found here that both poor healer and good healer share 0.07% of their total CDR3 repertoire with

the cardiac infiltrate, compared to 0.05% to healthy donors. The overlap increase between healthy donors and MI patients is statistically significant ($p<0.01$ for good healers, $p<0.0001$ for bad healers), which could suggest that circulating T cells in MI patients and cardiac infiltrate are not completely distinct compartments.

Altogether, these results suggest that the signature found in Chapter 5 is mainly specific to the circulating T cell compartment. Indeed, the healing signature is not found in cardiac biopsies, but the cardiac infiltrate CDR3s can be modestly, yet significantly, found in the peripheral blood of patients, regardless of the outcome, as compared with healthy donors. We can hypothesise that the TCR involved in the typical myocardial infarction are different to the TCR predicting myocardial outcome.

3.4.2 Circulating blood of Type II diabetes patients after MI

We have demonstrated how our 348 CDR3 signature was specific of the circulating blood and not of the damaged tissues, however we still wanted to confirm its relevance in another control cohort. To this end, we had access to another cohort, from the academic, multicentre, double-blind EMMY trial (NCT03087773) (Tripolt et al., 2020; von Lewinski et al., 2022). EMMY participants are type II diabetes patients with myocardial infarction and evidence of significant myocardial necrosis. Patients were divided in 2 groups at admission, placebo or treated with Empagliflozin, a sodium-glucose co-transport 2 inhibitor (SGLT2i). Empagliflozin are used as a first line of

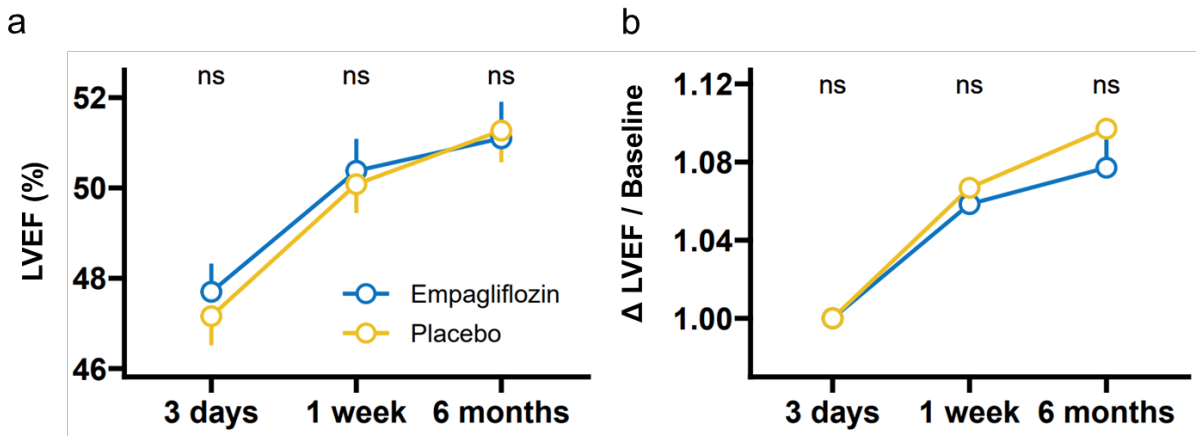


*Figure 16: **Impact of Empagliflozin on echocardiographic parameters between treatments groups**. **a**: LVEF between treated (blue) and placebo (yellow) participants across time-points of visit 1 (3 days), visit 2 (1 week) and visit 4 (6 months). **b**: Percentage of evolution of LVEF between timepoint and baseline (3 days). Results expressed as mean with standard error, Wilcoxon U-test. NS denotes non-significant differences, *: $p<0.05$.*

treatments in patient with chronic heart failure (McDonagh et al., 2021) and acute myocardial infarction (Heidenreich et al., 2022). SGLT2i have been shown to reduce death in patient with or without type II diabetes (Lopaschuk and Verma, 2020). Although the mechanisms are not completely understood, it has been hypothesised that SGLT2i positive effects on cardiovascular health are mediated though metabolic and anti-inflammatory effects (García-Ropero et al., 2019; Ye et al., 2018). Since the original EMMY study on n=476 concluded in 1.5% improvement of LVEF over the course of the study (von Lewinski et al., 2022), we controlled whether it was needed to segregate groups based on treatment.

When comparing the LVEF over the 6 month follow-up of the study, we found no difference between treatment and placebo at all time-points (Figure 16A). We found similar results when comparing the evolution of LVEF between time-points and baseline (Figure 16B). Although not consistent with published results of the clinical trial, these data were collected with only 42% of the initial cohort, thus probably limited to observe the moderate (1.5%) improvement. Given the absence of difference in LVEF between groups or δLVEF and the absence of definitive evidence for immune effect of SGLT2i on T cells, we decided to analyse TCR repertoires without taking into account treatments subgroups.

We had access to 199 blood samples at baseline. We used the same threshold as previously described (Bulluck et al., 2017b) to diagnose patients into poor healers, with δLVEF>13% between baseline and latest timepoint (6-month) and poor healers (δLVEF<13%), regardless of treatment.

Extracted RNA was of exceptionally low quality (Figure 13, p. 65), and resulted in technical difficulties to optimally amplify samples with SMARTer Human TCR a/b Profiling Kit (Takara Bio) as in the ETiCS study. Libraries were sequenced on a MiSeq sequencing platform using a 300bp protocol at the iGenSeq sequencing facility (Institut du Cerveau, Paris). Raw FASTQ files were aligned with MiXCR v3.0.13. To minimise biases due to the undersequenced samples, we

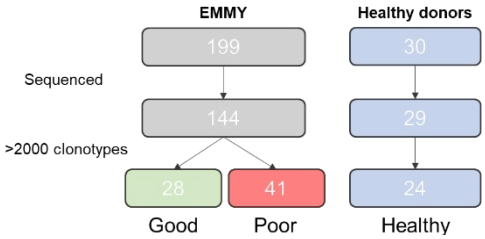


Figure 17: EMMY and healthy volunteers datasets after sequencing and filtering of samples with less than 2000 unique clonotypes.

removed samples with less than 2000 unique sequences (Figure 17). This leaves a final dataset of 93 samples, with n=28 good healers, n=41 poor healers and n=24 healthy volunteers.

This aggregated dataset of 93 samples is highly sparse. When looking at the incidence matrix of "samples x CDR3", more than 98% of the dataset is zero. We found that more than 93% of unique CDR3 are private and only 0.7% of CDR3s are sthared by four or more patients. Similarly, computed Jaccard index between samples based on CDR3 presence/absence shows a very limited sharing between patients (results not shown).

Our previous signature approach was based on similarity between samples. However, this low to non-existent sharing between patients make our previous signature approach not relevant for this specific type of highly sparse data. We decided to orient our analysis between shared features of CDR3s. Indeed, several studies have explored how grouping CDR3 by their specificity rather than matching their exact sequence could provide solid insight and alleviate the low throughput of some datasets (Glanville et al., 2017; Montagne et al., 2020; Musvosvi et al., 2023; Ostmeyer et al., 2020). Based on the solid rational for this type of approach (see section 1.5.1.3), we designed a pipeline of analysis that would not focus on CDR3 but on their shared properties (Figure 18A). This approach of TCR data agglomeration would bring several advantages: apart from highlighting relevant shared properties, this would remove noise from non-shared

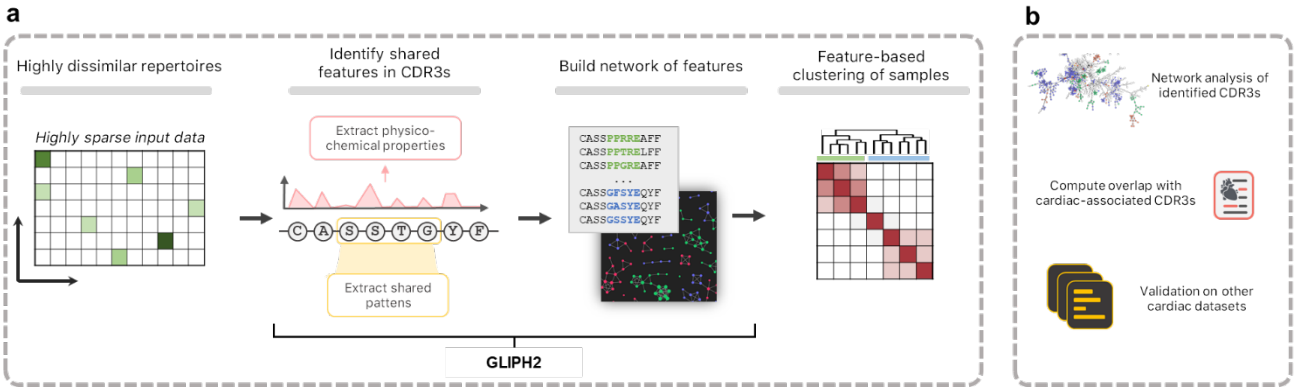


Figure 18: ***Pipeline of identification of healing related CDR3 in the highly sparse EMMY cohort. a****: Starting from highly sparse CDR3 repertoires, we used GLIPH2 tool to find common pattern in CDR3s sharing similar properties. Resulting clusters would then be used for feature based clustering, circumventing the absence of shared unique CDR3.* ***b****: Traditional clustering of CDR3s, by looking at shared CDR3 properties with TCR of Myocardial Interest, or using other datasets of cardiac interest, would perform biological validation of CDR3.*

features and reduce the data complexity by working with a more restricted set of clusters of CDR3s. This analysis is however restricted to the β chain, as GLIPH2 authors have not yet implemented support for α chains.

We used GLIPH2 algorithms on standard internal parameters, with the following protocol: to identify group-specific clusters, each group was compared to the two others. Each exclusive signature was then merged, and CDR3 identified in clusters were pooled. We then filtered out clusters with two or less CDR3s, and selected CDR3s with signing patterns between 3 to 6 amino acid, to prevent long patterns that could cover the whole CDR3. Selecting clusters was based on GLIPH2 significance scoring < 0.01, enrichment of a cluster in any of the group was assessed by a Fisher test p-value <0.01, and an odd ratio > 2.

This strategy selected 237 clusters comprised of 2046 unique CDR3s. Most of the clusters signed the Healthy group (136/237, 57%), while 38 (16%) signed the poor healers, and 63 (27%) were associated with the good healer outcome. These results are similar with the one obtained with the ETiCS cohort, the majority of the signature would be associated to better cardiac repair.

We used this strategy to cluster the patients; results are presented in Figure 19. In this heatmap, each column is a patient and each line is a cluster. If a patient finds in its repertoire at least one CDR3 of the cluster, we assigned the "presence" value to the patient for this cluster, with the assumption that cluster links antigen-specific convergent CDR3s. Patients' and CDR3 clusters were then arranged using ward.D2 hierarchical classification algorithm (Murtagh and Legendre, 2014). We observed four metaclusters of CDR3 signing the patient groups. Metacluster 1 is associated with the poor healer group, metacluster 2 with the good healer, metacluster 3 with the healthy volunteers, while the 4th metacluster is shared by several conditions. To determine the contribution of each CDR3 metacluster to patient classification, I determined each metacluster importance for classification in each of our three classes by building a generalised linear model (GLM). From then, I extracted the coefficients associated with each cluster, (Figure 19, right panel). A positive value indicates that the CDR3 metacluster is associated to the classification of said group. As such, we found that the metaclusters 1,

2 and 3 were positively associated with poor healers, good healers and healthy volunteers respectively

3.4.3 ToCIs

The classification heatmap in 3.4.2 is obtained from clusters of CDR3s. Because we wanted to characterize more deeply the patterns and assess their cardiac relevance, we sought to link CDR3 clusters to their putative specificity. This strategy to link a GLIPH2 cluster to a specificity has been adopted or demonstrated by several other groups (Chiou et al., 2021; Higdon et al., 2021; Huang et al., 2020; Quiniou et al., 2023).

The first step was to identify TCRs of Cardiac Interest. To this end, we identified TCR datasets from the literature with relevant CVD. We looked for i) TCR libraries, ii) T cell enriched 5' single-cell or iii) single-cell TCR libraries of cardiac diseases or

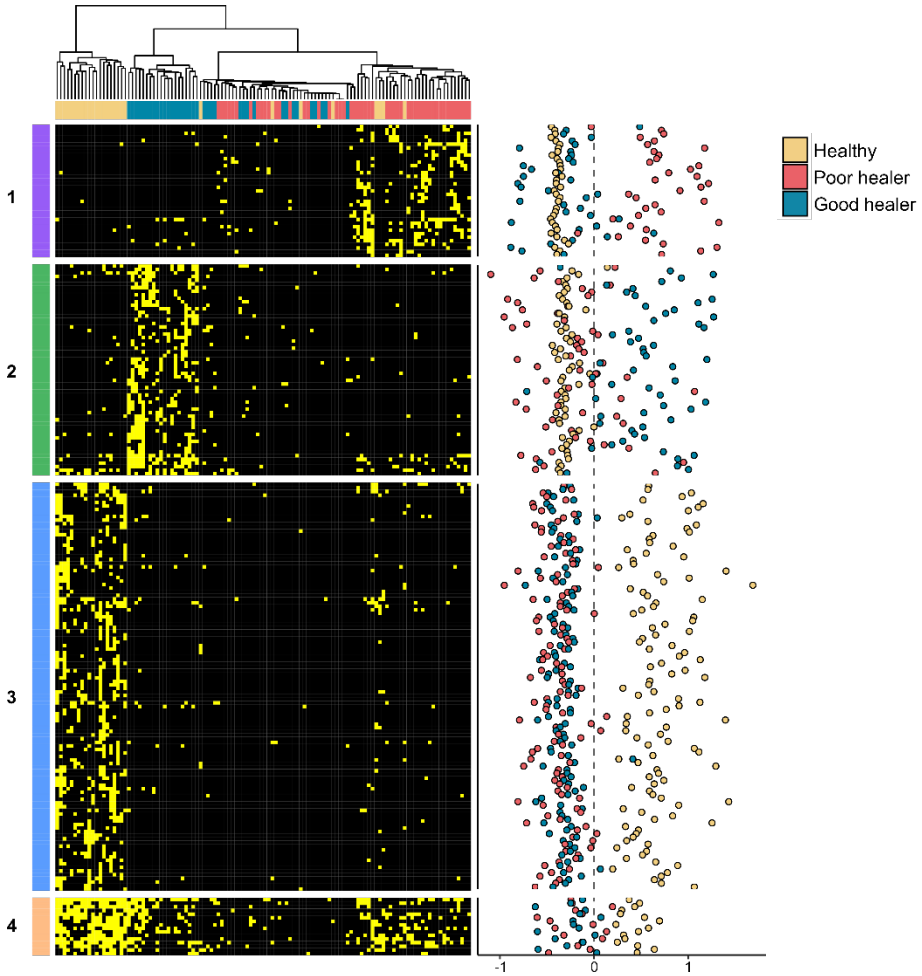


*Figure 19: **Analysis of CDR3 repertoires using GLIPH2 reveals sets of group-specific clusters**. **(Left)** Heatmap of presence absence of clusters identified using GLIPH2. Patient clustering was performed using Ward D2 algorithm. Metaclusters, labelled from 1 to 4, were determined based on the hierarchical clustering. **(Right)** Contribution of each cluster to the clustering observed, using generalised linear modelling. A positive value indicates that the cluster contributes to classify samples in this group.*

healthy controls. Three datasets were selected for this study (Table 2). The massive single-cell project "Cells of the Human heart" (Litviňuková et al., 2020) provided data consisting of healthy hearts from 25 adult donors, 14 female and 11 males. I used the sorted *Immune cells*[2] dataset, from which 10 000 T cells were sequenced. I aligned the data to reconstruct complete TCR sequences from the single-cell RNAseq libraries with TRUST4 (Song et al., 2021). We identified a second dataset used from *ex-vivo* expanded T cells from patients suffering immune checkpoint inhibitor (ICI)-associated myocarditis (Axelrod et al., 2022). Peripheral blood $CD8^+$ T cells were stimulated with α-myosin antigens and activated were sequenced using 5' single cell 10X VDJ. TCRs were reconstructed with TRUST4. Last dataset used is from our own cardiac biopsies. For all these datasets, I focused on the TRB rearrangments, as it was the one used for identifation of the signature in 3.4.2.

Dataset (short name)	Accession code	Condition	Method	Cells	Unique CDR3s
Cells of the Human heart (Atlas)	PRJEB39602	Healthy	5' Single-cell 10X GEX	Infiltrating cardiac T cells	39
T cells specific for α myosin drive immunotherapy related myocarditis (Myocarditis)	SRR21598102	Myocarditis	5' Single-cell 10X VDJ	α-myosin expanded circulating T cells	115
Würzburg biopsies (MI)	NA	MI	TCR VDJ	Infiltrating cardiac T cells	8239

Table 2: Datasets used for the TCR of Myocardial Interest analysis.

[2] https://www.heartcellatlas.org/immune

There is a striking imbalance in the data, with only 115 unique CDR3s recovered in the VDJ enriched datasets. Nonetheless, we computed the overlap between these datasets and the CDR3s from the clusters identified in Figure 20. Clusters were split based on the group they signed. Hence, cluster 1 is called MI_good_healer, cluster 2 is MI_bad_healer, and cluster 3-4 are "Healthy". Results of overlap are shown in Figure 19 in an Euler diagram. As we can see, the myocardial infiltrates from failing hearts (MI, grey) form larger dataset and is the one sharing most, although few, CDR3s with the other datasets. Strikingly, ex vivo expanded T cells from myocarditis dataset (Myocarditis) does not share any CDR3 with either of our circulating TCR signatures, but shares 2 CDR3 with the MI dataset. Healthy and MI_good_healer share CDR3 only with the MI patients (6 and 2 CDR3 respectively), but not with other datasets. CDR3 recovered from *Cells of the Human heart* atlas (Atlas) do not share any CDR3 with our signature but share two CDR3s with MI. Finally, there is no overlap between the poor healing signature and any other datasets.

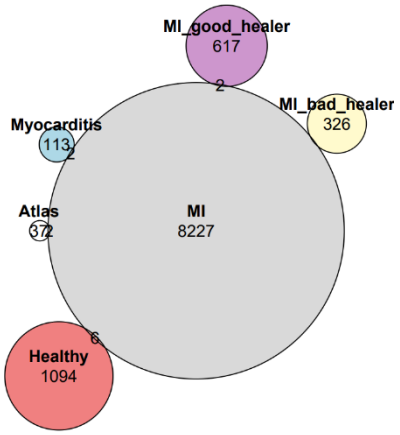


Figure 20: ***Overlap of EMMY dataset CDR3 clusters with the ToCIs datasets using perfect match of CDR3 sequences.***

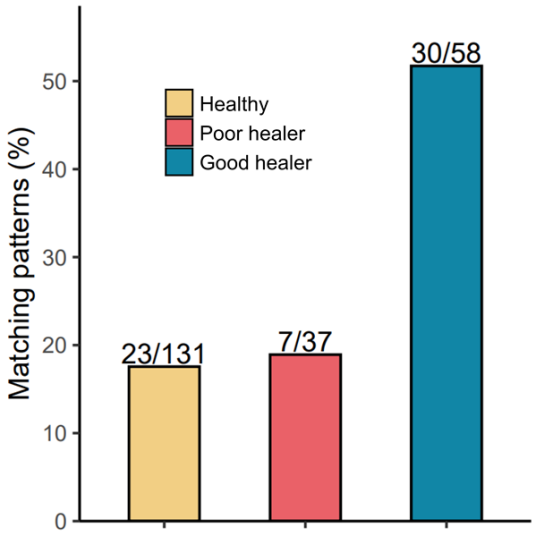


*Figure 21: **Myocarditis CDR3s matching signature patterns**. Number of Myocarditis expanded CDR3s matching amino-acid patterns of our signature. Results are represented as a percentage of total patterns from this group.*

As we have highlighted previously, this analysis is limited due to the constraints of poor recovery of CDR3s in biopsies. To this end, we used the clustering analysis (Figure 18B) and the patterns identified from GLIPH2 (results shown in Figure 19) to find CDR3 matching these patterns and assign them to groups. I first focused on the myocarditis dataset, as it was enriched in relevant CDR3s with known cardiac impact, while MI and Atlas datasets were more representative of recirculating T cells. The results are shown in Figure 21. To our surprise, more than half (30/58) patterns associated to good healing outcome were associated with myosin-expanded TCRs, as compared with 17% (23/131) and 19% (7/37) of healthy volunteers and poor healer groups.

These results show how TCR signatures from cardiac T cell biopsies or circulating T-cells are very private when considering them at the strict sequence level, and show almost no overlap. This does not confirm or infirm the presence of absence of autoimmunity linked to cardiac outcome, but could rather reflect the lack of information due to the low number of sequences recovered from both the EMMY and external datasets. Indeed, even during infarct, the heart is not as highly infiltrated by T cells in numbers comparable to circulating blood or lymph nodes, with 8 cell per mm^2 in immunostaining of $CD3^+$ (Devaux et al., 1997; Ohta-Ogo et al., 2022), hence limiting what can be recovered in small biopsies. To circumvent this, we took advantage of TCR similarities to build networks of CDR3s from our datasets and α-myosin expanded T cells. We found a significant association between α-myosin expanded TCRs and good healer outcome TCR signature compared to healthy volunteers and poor healers, suggesting the relevance of myocardium specific T cells during repair. These results

should however be confirmed in larger cohorts, as we do not know the CD4/CD8 origin of the CDR3s, and therefore the MHC-I or MHC-II restriction associated with the identified CDR3s. Analysis of samples from fresh, sorted blood T cells should provide data of sufficient depth and quality to better characterise the role this signature. High-throughput single cell analysis should in the future allow to accurately tackling this question. These results demonstrate however how analysing TCR as a network of similarities can be a powerful tool to leverage data from sparse matrices.

4 DISCUSSION

4.1 THE NECESSITY OF CONTROLLING THE EXPERIMENTAL LAYOUT

4.1.1 Open data is happy data

The results I presented stressed the need for experimental controls. Computational methods for controlling the data are limited, especially in AIRR-seq, as shown in the 1st part. Results obtained from these methods display large variability, due to the experimental choices (amplification approach, sequencing layout, depth) and software (choice of aligner, genome of reference, definition of a clonotype, downstream normalisation).

These issues are not restricted to the TCR field, but rather influence all fields working with NGS, as have highlighted the recent stimmed controversy over microbial structures. Indeed, in a 2020 landmark study, Poore et al. analysed 18,116 microbiota samples and demonstrated how machine learning could almost perfectly discriminate tumours' types in patients (Poore et al., 2020). The highly cited study (over 600 citations) was however flagged as "entirely wrong". In a long pre-print, Gihawi and colleagues reanalysed results from the study, starting from raw data (Gihawi et al., 2023). They reported "two fundamental flaws" in the methods, namely normalisation and poor contamination removal. They claimed that the original paper used poor normalisation that created artificial tumor specific signatures. The normalisation even introduced signal, by inferring species abundances if they were missing. A second issue with the paper, noted the reply, was the contamination removal. Many reads were aligned to the wrong species, doubling the number of bacterial reads for samples, introducing false signals and ultimately leading to aberrant results. The reply also identified several papers using the same flawed methodology. Original authors provided a quick counter-reply, admitting flaw but stating it did not affect overall results (Sepich-Poore et al., 2023). To date, the community has provided no definitive answer to this controversial study, but called for more oversight of machine learning use in microbiome data (Offord, 2023). Confirmation by new experimental data shall shed light on this matter.

This investigative work could only be done because of i) a team of dedicated experts and ii) a FAIR-compliant open dataset (Wilkinson et al., 2016). The issues raised here are similar to what we can observe in the TCR field. One of the most significant study in circulating cardiac repertoires is from (Tang et al., 2019). I have been trying to replicate their results, without success. They performed TCR-seq on HiSeq-2000, with 60_{R1} bp and 50_{R2} bp paired-end read layout, aligned with BGI software and performed in-house normalisation. Although I requested raw data, they could not provide it. FASTQ were "unavailable", as they did not store it, however they instead sent me their processed data. In the light of Chapter 3 results (p. 68), it appears unlikely to be able to replicate those results. Indeed, the difference in reads length probably affected the alignment. We used MiXCR while they used BGI, and I did not benchmark the differences in alignment performance with the aligners. Moreover, the genome references might have impacted gene and CDR3 determination. Finally, processed data did not feature raw counts, only normalised frequencies, which further complexify the eventual cleaning and quality control of the data. Although there is certainly no malicious intent from the laboratory, this seems like a hasty data management that could be avoided. For these reasons, it appears that integrating the processed data in pipelines will most likely lead to spurious results.

A very interesting initiative from the ecology community investigated how teams with the same datasets managed to find different results for a given scientific question (Gould et al., 2023). A total of 174 teams of analysts from around the world analysed the same dataset of *Cyanistes caeruleus* (blue tit) nesting. Surprisingly, models provided by analysts for the study were not unanimous, but rather lead to heterogeneous, sometimes contradictory conclusions. Although several explanations for the widely different results were provided in the final manuscript, no definitive answer was formulated. Rather, authors defend how analytical choices, which may lead to diverging results, are the path toward the consensus. It is the replicability, and confrontations of data, methods and opinions that pushes the community towards the good answer.

4.1.2 The use of experimental controls

This advocates for the use of technical controls in every bulk TCR sequencing to ensure that, in each study/batch, a set of known TCRs is correctly aligned and counted.

This also ensures that any new protocol, sequencer or aligner will yield the same results. If every TCR sequencing run would add one sample, with known complete TCR sequences and VJ joining, it would ensure the reliability of i) the method used and ii) the experimentation. Such approaches would allow a fair and unbiased benchmark of any new methods developed. This issue has been greatly documented in the survey of the AIRR community, where the use of use of spike-ins could solve most (8/9) common documented errors arising during AIRR-seq workflows (Trück et al., 2021).

Global suppliers have already started to explore this topic. TaKaRaBio adds in each of its TCR kits a "positive control" corresponding to PBMC from healthy volunteers; however, this does not constitute a good control. Since donors differ between batches of kits, it becomes challenging to compare experiments conducted using different batches. This also complicates the process of comparing results across various laboratories, unless the control batch is specified. Moreover, complete sequences of all TCR in this type of positive control is unknown, which is required to assess annotation reliability. I would rather advocate for a set of spike-ins TCR in Jurkat cells for instance, with known, complete sequences, that would cover a wide array of V-J combinations and fixed concentrations, similar to what has been done in the TCRpower paper (Dahal-Koirala et al., 2022). Cellecta currently offers this strategy for human TCR, covering 10 TRBV-TRBJ combinations, one TRAV-TRAJ, one TRGV-TRGJ and one TRDV-TRDJ with 3 variations of each CDR3 for a total of 39 spike-ins (Synthetic Spike-in Controls for Immune Repertoire Profiling, Cellecta). Although most manufacturers do not provide this kind of controls, this is a step towards a better validation of experimental sequencers.

In contrast to openly available raw data, a significant limitation arises when data suppliers, operating in a closed-source manner, do not disclose raw data and instead apply their proprietary or undisclosed methods, such as Adaptive Biotechnologies. This issue of replicability and inability to re-process data from raw reads is problematic (Miyakawa, 2020) and of utmost concern. Indeed, as I have observed, results from TCR annotation drastically vary simply by going from 300bp to 250bp. It also hinders data reuse, a hallmark of FAIR standards (Wilkinson et al., 2016). These issues are still a point of concern. As such, a meta-analysis of 134 studies conducted between 2006 and 2022 found that "only 38.1% had made publicly available raw TCR-seq data in public

repositories" (Huang et al., 2022). To make matters worse, an astonishing 70% of non-public datasets remained unshared even after requesting them from the authors.

4.2 CIRCULATING CARDIAC SIGNATURES IN LIGHT OF PUBLISHED DATA

4.2.1 Challenges of frozen whole blood TCR sequencing

Using circulating T cell as a biomarker of cardiac diseases was a bold project when it was formulated. Indeed, there was scarce literature about the existing role of T cells in myocardial infarction, and the clinical features at our disposal were very limited. The biggest challenge has been about capturing the TCR itself (see 2.2.2.1). Indeed, samples of both the ETiCS cohort and the EMMY cohort were challenging to amplify. Several factors were at play. First, it was whole blood samples, frozen without any preservative agent for up to a decade. During the storage period, EMMY samples were placed for a few weeks at -20°C rather than -80°C, which played a crucial role in the degradation of the samples. As these samples were not originally designed to be used for RNA sequencing, hence the suitable temperature for storage was not taken into consideration.

Additionally, RNA extraction was also a challenging step. The formation of water crystals during freezing caused cell damage, likely leading to the release of RNase in the samples, which, in turn, contributed to the degradation of the material. One of the EMMY cohort recruitment criteria was type II diabetes participants. Blood of these patients was especially rich in fatty deposits, and several samples could not be extracted as they clogged kits filters. Subsequently, samples with high fatty content resulted in lower extraction yields, when successful. This might have affected downstream results, as patients with higher circulating fats could not be sequenced. Several samples were so degraded they could not give any RIN. Fortunately, even with abysmally low RINs, we managed to obtain viable libraries in 34% of samples, proving the relevance of 5'RACE in rescuing information from heavily altered samples.

This poor recovery of samples led to several adjustments from our initial analysis pipeline. In the ETiCS cohort, out of the initial 61 samples sent to us, only 28 were of

sufficient quality for sequencing. The slight imbalance in groups (n=19 *vs* n=9) also limited what could be done in terms of modelling and statistical approaches. Likewise, in the EMMY cohort, only 34% (69/199) of the samples met the quality requirements for analysisHowever, this did not result in an imbalance between groups (28 good healers *vs* 41 poor healers).

4.2.2 Circulating cardiac signatures in light of published data

Murine models experiments published by other groups have demonstrated a TRBV19 bias in infarcted mice in the CD4 compartment (Rieckmann et al., 2019). Although we investigated the use of VJ usage bias in Chapter 4, we could not confirm nor infirm it. Indeed, the cohort was neither designed nor suitable for this usage, as a circulating blood TCR repertoire in humans could not be used to recapitulate results obtained from a sorted population in mice. It also showed differences in diversity in infiltrating CD4 T cells in infarcted vs non-infarcted myocardium. However, we did not had access to cardiac biopsies in the human ETiCS cohort.

Our results, given the low recovery from peripheral blood T cells, demonstrate how even a low input repertoire from circulating blood can be used for prediction. The Emerson *et al.*, study, featuring the most comprehensive, deepest assessment of T cell repertoire dataset published, managed to find T cell signatures in patients infected with CMV in a 666 patient cohort (Emerson et al., 2017). Although their method was simple, it was the largest T cell repertoire study conducted in NGS and showed that, it was possible to identify signatures in blood. It was even more impressive given that past encounter with CMV is a "low signal", meaning it is not an active infection with a large portion of the repertoire dedicated to fighting it. We pushed this even further with a smaller cohort (n=666 *vs* n=28). The major advantage of Emerson study was the information on the HLA of each patient, which allowed distinguishing subtypes of patients sharing different groups of CDR3 of the CMV signature. Individual sharing of the signature was, interestingly, rather low, with the most shared CDR3 only found in 61 patients, highlighting the huge heterogeneity of the immune response, and most patients sharing ~20 CDR3 of the 164 CDR3 complete signature (~12%).

Comparing these results to our own signature obtained in good and poor healers (see Chapter 5, p. 146), we find very similar sharing of CDR3, highlighting how circulating blood signature do not necessarily encompass "universal TCR" shared by all patients. As opposed to the Emerson dataset, we did not have access to a validation cohort to propose a prediction model. Our results do not give a definitive answer to what constitute the healing outcome repertoire, but rather demonstrate how a signal can be identified using adequate methods, and pave the way for new studies to validate our findings. As we showed in the EMMY cohort, if antigen-specific clones mediate the studied disease, one could alleviate the issue by inferring specificities from clones and looking directly at pathology-driven T cells, rather than correlating TCR to disease. This field however, is still blooming and need further refinements.

With only one timepoint from a 12-month differential diagnostic in an aseptic cardiac disease, a limited cohort, low amount of input material, no HLA information, non-sorted cells, the question we asked has certainly been very ambitious. If one of these parameters had been to change, having a second timepoint would have probably been the best option. Indeed, it would have opened many opportunities such as tracking the CDR3s across timepoints. It can be hypothesised that the clones identified are positively correlated to patients' outcome because they fuel the early inflammation, and are not expanded at further timepoints, results supported by the presence of TCR with viral specificities in the signature. The Antigen-specific Lymphocyte Identification by Clustering of Expanded sequences (ALICE) (Pogorelyy et al., 2019) is a dedicated tool to track clonal expansions between timepoints. It was developed in a collaboration with the same group that developed the tool we used to identify the original signature; further confirming its relevance to associate the two.

ETiCS cohort really showed how TCR specificity inference is not difficult, as we have matched CDR3 to reported ones. In the ETiCS cohort, we found that the good healer outcome signature was enriched for TCR binding to viral peptides, especially CMV. These results are surprising in light of the existing literature, as reports of recent viral exposition was a marker associated with an increased risk of CVD (Lebedeva et al., 2020; Wang et al., 2017; Yousaf et al., 2021). However, the challenge lies in the validation of this specificity.Indeed, without HLA or MHC restriction, we can only rely on published,

public data compiled into databases, and ignore the CD4-CD8 antigen restriction of the identified CDR3. Moreover, even if we actually find CDR3 linked to viral antigen, we could not infer a role for them. As T_{REG} and T_{EFF}/CTL have antagonist roles, our signature could very well be driven either by inflammatory T cells, which would make sense in the context of a 3-days' inflammation, or by anti-inflammatory driven with T_{REG}. Finally, this signature has been observed in the circulating blood. We do not have matching biopsies from the heart of the good healer patients to confirm that the infiltrating T cells also bear enriched viral (CMV/EBV) specificities. By having access to matched tissues and circulating blood in a new cohort, it could both confirm the preliminary results obtained here and further explore the specificity of the early cardiac infiltrate associated to patient outcome.

Indeed, by mobilising earlier a pro-inflammatory environment, the T cells can recruit more macrophages, DCs and other parts of the immune system, as well as other T cells. This massive mobilisation allows the rapid clearance of debris and starting the collage scar formation, limiting cardiac chronic insufficiency. On the other hand, T_{REG} would dampen inflammation, reduce initial cardiac damage, and overall switch the inflammatory microenvironment towards a steady-state one faster. Mathematical modelling to infer the CD4 to CD8 ratio from repertoires have been reported, but we could not apply to our data (R. Emerson et al., 2013). Similarly, inferring the the MHC-I or MHC-II restriction from a TCR sequence (and hence, the CD4 or CD8 subtype associated) was attempted, but showed poor performance in a single-chain setting (Carter et al., 2019; Hou et al., 2020). Seeking to identify the factors promoting TCR towards a T_{REG} fate, Lagattuta *et al.* presented a model where they developed the TCR-intrinsic regulatory potential (TiRP) score (Lagattuta et al., 2022). TiRP could provide additional information in screening patients with T_{REG} associated parts of the signature.

4.2.3 HLA is not just another clinical parameter in CVDs

HLA typing would have greatly helped in differentiating our subgroups of patients. For instance, the HLA-DRB1 haplotypes have been identified as a predisposing factor for multiple cardiovascular diseases such as coronary heart disease (Sun et al., 2011), myocardial infarction (Björkbacka et al., 2010; Paakkanen et al., 2012; Sengar et al., 1985), cardiovascular mortality in inflammatory polyarthritis (Sharma et al., 2022),

cardiomyopathies after hepatitis C virus infections (Matsumori, 2005). In a broader extent, other groups have demonstrated how MIS-C was linked to a set of class I HLA (Porritt et al., 2021), while cardiovascular remodelling was associated with arterial blood pressure in several populations (Diamantopoulos et al., 2003; Luque Otero et al., 1983). Novel ongoing work from a german group has identified HLA alleles with different levels of risks in patient with chronic heart failure (Merten et al, *personnal communication*). With over 120 patients and their complete HLA information, they showed how HLA characterisation in failing hearts could be used to stratify patients' prognostic. Specifically, the analysis demonstrated that the presence of HLA-DR2 conferred a protective effect in this patient population, whereas the presence of HLA-DR5 was associated with a 4-fold increased odds ratio of mortality following myocardial infarction (MI). HLA could then be seen as an etiologic factor, rather than a risk factor. Patient bearing those immunogenic HLA might trigger more inflammatory response, ultimately leading to cardiovascular complications. Alternatively, those without the predisposing HLA could have triggered the autoimmunity through other unknown means. This opens some new perspectives on further research.

4.2.4 "Healthy" volunteers, or is it?

Healthy volunteers' recruitment is also a source of variation that can be looked into. Indeed, I used blood sampling from French Blood Centres donations, spanning the 2021-2022 years. We selected donors to match age and sex of the ETiCS cohort to not introduce bias, but we do not have access to similar level of information. Especially, patients could have atherosclerosis, the primary underlying disease process of myocardial infarction (Timmis et al., 2018), which does not exclude you from donation in French Centres. Moreover, our recruitment did not take into account the seasonality of CVD, where we see a higher risk for CVD during the cold

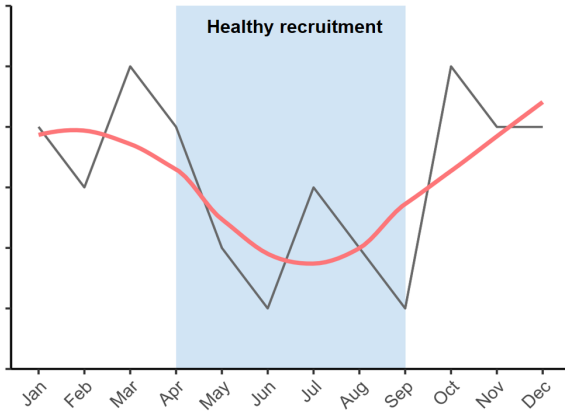


Figure 22: *Myocardial infarction per month in the ETiCS cohort (n = 38). Gray line: ETiCS patients MI inclusion, red line: smoothed approximation. Blue; healthy volunteers recruitment window*

months (reviewed by Stewart et al., 2017), consistent with observations in our cohort (Figure 22). On the other hand, our "healthy volunteers" cohort been constituted between the months of April to September, which would also correlate to the lowest occurrences of MI. Among the seasonal factors involved in increased cardiovascular risks, IL-6 was found to contribute to disease onset (Sartini et al., 2017), results corroborated by other large scale clinical studies (Sattar et al., 2009). The clinical relevance of IL-6 compared to the fattier diet, increased blood pressure and more sedentary lifestyles might be anecdotal, although it may be related to impaired response to winter viruses (Nguyen et al., 2016). The link between myocardial infarction seasonality and T cell repertoire has however not been investigated. SARS-CoV2 pandemic is another major factor that must be taken into account in our "healthy" cohort. Indeed, a large proportion of the french population contracted the virus or was subject to several rounds of vaccination against spike-proteins. At the end of December 2021, it was estimated that 80% of the population has received at least one dose of vaccine (Haugomat, 2022). This brings a substantial bias in circulating repertoire. Even if SARS-CoV2 specific T cell clones pre-exist in healthy population (Shimizu et al., 2021), vaccinated patients share a higher proportion of the virus T signature (Gittelman et al., 2022). Moreover, SARS-CoV can penetrate the myocardium through ACE2 and induce myocarditis (Lindner et al., 2020; Nägele et al., 2022). The risk of CVDs such as MI or stroke has been considered safe in light of vaccination (Botton et al., 2022; Jabagi et al., 2023). We indeed thought about using already published datasets from healthy volunteers datasets from the pre-COVID-19 era. However, they either did not match our experimental protocol (unsorted, poorly stored T cells) or sequencing layout, thus were not comparable to our cohort.

4.3 SPEAKING ABOUT SPECIFICITY

Despite multiple T cell epitopes described in CVDs (see section 1.2.4), few to none TCR of cardiovascular interest (ToCI) have been reported in human databases of TCR binding, at least with sufficient confidence. Screening of cardiac epitopes has been performed in mice (Rieckmann et al., 2019), but this work was not reported in humans. Datasets of T cells with ToCI are scarce and not annotated. My effort to recover TCR from infiltrating cardiac T cells in healthy (Litviňuková et al., 2020) or failing (Tang et

al., 2019) hearts, or *ex-vivo* stimulated $CD8^+$ T cells in autoimmune context (Axelrod et al., 2022) revealed itself limited in both number of T cells recovered and what I could computationally do from it.

This topic goes beyond CVDs epitopes. Inferring specificity of T cells might be an even more challenging task than currently thought. Authors have recently pointed TCR-epitope prediction as the "holy grail of immunology" (Hudson et al., 2023). When I started my thesis in 2019, existing methods to predict specificity were based on sequence clustering (GLIPH2, tcrdist…) with already annotated TCR. Since then, two things changed. First, single-cell studies became much more affordable, with the possibility of coupling dextramers to explore the specificity along with transcriptome (Boutet et al., 2019). This has been greatly popularised with COVID-19 studies, looking for SARS-CoV2 specificities. Second major advance arrived with new deep learning approaches. Previous distance-based methods were somewhat simple, and intelligible, but required a lot of knowledge to bring something. Indeed, models were based on observed data and knowledge of the mechanisms lying behind TCR recognition. On the other hand, deep learning requires data and self-trains itself. This method can be called intuitive, as it does not require prior knowledge of the topic: the model finds by itself the best weight and parameters. This intuitive reasoning can be at the cost of transparency, as this type of black-box models is unable to justify mechanistic basis of its predictions (Yeo and Selvarajoo, 2022). This accelerated the process of developing new prediction tools and allowed to a large panel of new approaches to be published in the last years.

The absence of consensual dataset or requirement to publish TCR specificity prediction methods have led to a prolific publishing, with dozens of papers and preprint in the last 3 years dedicated to this topic. Although competition is always the better alternative, this multiplicity of poorly designed, or biased comparisons[3], is a tedious work. Indeed, authors that compare their method to outdated or irrelevant tools, or use interesting metrics for their benchmark can be deemed unfair. This feeling is shared by other researchers, which decided to take action. A community driven effort to benchmark

[3] Meysman *et al*. (2023) note that "[It is a] commonly accepted phenomenon that a method will always score best when applied by its own authors and on the data set in the paper where it is introduced"

23 methods original and published has been carried on (Meysman et al., 2023) to determine the good practices to observe when developing TCR prediction tools. This work could define the different advantages and weakness of some methods. However, authors noted how this remains largely insufficient, because of i) the multiplicity of tools published in the literature and ii) the lack of accessible code or tunability of tools. Several issues were raised by this community study, notably the lack of benchmark datasets, which become necessary with the current surge of prediction tools, along with the training data issue. These issues, and notably the negative data bias, need to be addressed as it stirs controversy in high impact journals (Dens et al., 2023; Gao et al., 2023b, 2023a).

In a few words, negative data bias is the fundamental flaw that lies in TCR specificity modelling. To build a TCR model, you need both positive and negative labels, and a continuous metric to assess the response (binding strength, interferon production, or any other biological output). The TCR repertoire however, possess this powerful property to be highly cross-reactive, with estimates of a single TCR being able to recognise 10^6 peptides (Mason, 1998). Hence, except in rare demonstration of a large cross-reactivity of T cells (Quiniou et al., 2023; Verhoeven et al., 2008), usually authors that indicate that TCR *X* does bind to epitope *1*, did not test whether or not this TCR could bind to other epitopes. If they did test it, and the TCR did not bind, it would often go unreported as the protocol would focus on binding-positive cells. This happens in cell-sorting studies, where labs would only sequence (or report) tetramer positive cells, or in antigen stimulation protocols where T cells that do not recognise the antigen do not survive or get outnumbered. Moreover, single-chain (α or β) would often be reported as "specific" although their association with their complementary chain might define their specificity. Hong et al showed how pairing of αβ TCR was critical, as screening of a 6 TCRα paired to 6 TCRβ chains would yield a responsive TCR in 5/36 combinations (Hong et al., 2022). These results should be compared to those of Quiniou and colleagues which showed how in some cases, TCRβ was the main contributor to specificity, with 1 β chain paired with 131 different α, was able to bind different viruses (Quiniou et al., 2023), with the α chain fine tuning the binding strength to some epitopes. These observations have been confirmed by multiple TCR-epitope specificity prediction tools

that found better results when having pairing information than just α or β chain (Jokinen et al., 2022). These data demonstrate how we should remain cautious about reported specificity, and experimentally confirm it when possible.

4.4 So, what is next for circulating T cells screening?

4.4.1 The circulating repertoire as biomarker

Using circulating T cells as a biomarker, or a proxy of a tissue-specific response, is however a powerful idea that has been formulated and investigated through cytometry (Blum and Pabst, 2007; Westermann and Pabst, 1990). Indeed, one could draw a parallel with circulating antibodies levels screening in patients to circulating T cells, as they are all immunoglobulins after all. However, as opposed to antibodies, which are circulating proteins, T cells are cells with a complete set of functions and cover much more potential.

Circulating repertoire has been harnessed in multiple contexts to monitor patient health. In advanced lung cancer patient, circulating β repertoire could predict prognosis of patients (Liu et al., 2019). Same results were obtained for renal cell carcinoma (Guo et al., 2020), but the repertoire was not different from healthy volunteers in ovarian cancer (R. O. Emerson et al., 2013) or pancreatic cancer (Bai et al., 2015). T cell diversity in peripheral blood has been used to predict patient's response to personalised treatment with neoantigen therapy and anti-PD1 (Bortone et al., 2021; Poran et al., 2020), anti-CTLA4 (Postow et al., 2015) or anti-neoantigen (Kansy et al., 2018). More recently, circulating CDR3β repertoire was used to detect high grade serous ovarian cancer before conventional diagnostic (X. Yu et al., 2023, *preprint*)

Although cancer studies provides a large spectrum of applications, similar studies were performed in autoimmune context such as type 1 diabetes (Tong et al., 2016), systemic lupus erythematosus (Thapa et al., 2015), multiple sclerosis (Amoriello et al., 2021), or cardiovascular diseases such as acute myocardial infarction or unstable angina (Sudong Liu et al., 2020) or ischemic heart failure (Tang et al., 2019).

All of these studies however have low impact, due to their incapacity to draw a clear link between T cells and either i) function or ii) specificity of T cells. Studies that can do both have a much better impact. This is the case for the lab that designed GLIPH2 and used their knowledge to infer disease progression following *Mycobacterium tuberculosis* infection based on blood sampling, by looking at clusters of T cells associated with previously published Mtb antigens (Musvosvi et al., 2023).

This kind of work, which does not only rely on sequencing data, but can confidently associate TCR with specificity, function, and predict the progression of multifactorial diseases, are probably the high impact papers of the future. However, such studies are costly, as they require a lot of input data to accurately identify sometimes subtle signatures. The recent partnership between Adaptive Biotechnologies and Microsoft to provide this very type of patient diagnostic, with the aim of monitoring patient health through peripheral blood sampling and TCR sequencing, confirms the growing interest of large pharma players to the "niche" field that is T cell repertoire analysis.

4.4.2 Bulk TCR sequencing: has the ship sailed?

Massive sequencing has revolutionised the T cell repertoire field. Nonetheless, this paradigm is quickly changing; multimodal studies, previously requiring consequent amounts of money and input data, has been completely revamped by single-cell technologies. With one experiment, you can now have access simultaneously to transcriptome, epigenetic modifications (Kakaradov et al., 2017), surface molecules expression, proteome (Specht et al., 2021) and epitope binding (Son et al., 2021). From this, you can infer MHC allelic information (Darby et al., 2020), single-nucleotide variants (Schnepp et al., 2019) or T cell activation (Deering et al., 2023). This can be coupled with spatial information of dissected tissues (Liu et al., 2022) and multiple timepoints (Yang et al., 2021). Their obvious limitation for whole blood screening, apart from affordability, is the low number of cells that can be simultaneously sequenced. With only thousands or tens of thousands of cells per experiment that can be recovered, this remains largely inefficient to accurately sample subtle differences in blood, where there is typically 1 x 10^6 T cells per mL. Thus, exploring T cell diversity in single-cell experiments through entropy measures or similarity indices could generate inaccurate results. Repetitive sampling of 10 000 most abundant circulating TCR clonotypes in bulk

sequencing showed that 1 sample captured 75% of top clonotypes from a pool of 10 replicates (Simon et al., 2018). Models of rarefaction have shown that 1 x 10^3 T cells could accurately represent the richness and evenness of a 10^3 bigger repertoire, but uncertainty remains on how representative a 10^3 TCR repertoire could be of circulating blood, which is estimated to contain more than 10^{10} T cells (Clark et al., 1999).

Although bulk TCR does not offer the multidimensional possibilities of single-cell, interesting approaches were employed that went tangent from the classical V-CDR3-J paradigm. Exploring the hypothesis that immunogenicity of certain self-antigens could be encoded in the T cell genes, several studies have demonstrated how some TRBV were poised for autoimmunity. TRBV polymorphisms were identified in immune-related adverse events during checkpoint blockade immunotherapy (Khan et al., 2019; Stephen et al., 2023), diabetes onset (Pierce et al., 2013), and variability of susceptibility to Epstein-Barr virus infection (Gras et al., 2010). Although these polymorphisms were not identified in previous GWAS studies on cardiac infarction, they probably flew under the radar due to the poor annotation of those genes. Indeed, immunoglobulins are highly polymorphic and poorly annotated (Lees et al., 2023; Omer et al., 2022). Despite the massive work of IMGT, we have mentioned a few issues of solely relying on their work (see 1.1.2). To this extent, the AIRR consortium launched the Open Germline Receptor Database (OGRDB), a curated database dedicated to list all known alleles, focused on documenting less studied populations (Collins et al., *preprint*). Although it is unlikely that TCR genes polymorphisms could be the sole driver for myocardial repair, it could be used to further stratify patients or identify outliers in the repair.

4.4.3 A bright future for paired-end bulk sequencing?

As we have covered so far in this introduction, the weakness of TCR diagnostic predictions can be due to unpaired TCRαβ bulk sequencing (see 4.3), paucity of literature about reported specificities of TCRs (see 4.2.2 and 4.3), and the limited throughput of single-cell methods along with their prohibitive price (see 4.4.2).

Single-cell sequencing focusing only on TCR reads could offer a seducing compromise between high throughput and paired αβ chains (DeKosky et al., 2015; McDaniel et al., 2016). Technology is already ready and mature for efficient single cell

encapsulation (Zheng et al., 2017) and unique barcoding of a large number of cells (> 10^7) has already been proposed several years ago (Bhang et al., 2015). Single-cell barcoding for targeted TCR sequencing could also offer efficient counting, as reads would be identified with unique cell identifier and not unique molecular identifiers (although both techniques could be combined).The large number of unique barcodes (>10^{18} for a 30bp-long sequence) allows robust multiplexing without barcode collusion (Smith et al., 2017). Improvement of the current inDrop approach[4] (Klein et al., 2015) allows the fast preparation of 10^6 cells in a couple of hours (Juzenas et al., 2023, *preprint*). The development of these methods has been ported to commercial use for large public. Omniscope OS-T technology (Omniscope) offers a 1 x 10^6 single-cell TCR sequencing, without transcriptome (Nadeu et al., 2022). It uses RNA reverse transcription and multiplex PCR, along with UMI.

Other methods relying on single-cell tracking through creative use of TdT to individually track cell fate at the genomic level as also been recently proposed, although it is limited by the murine model used (Li et al., 2023). For spatial tracking of T cells in biopsies, molecular barcoding with photon emission (Light-seq) has been developed, which could be useful to dissect T cell repertoires distribution in cardiac infiltrates (Kishi et al., 2022)

Large-scale single-cell TCR sequencing would not only bring better understanding, but also more accurate information of T cells. The precise unique cell identifiers would be superior to UMI for quantifying clonality, which provides better assessment of T cell diversity. The pairing of full-length αβ chains along with HLA information will pave new ways to screen epitope banks with unprecedented precision. As previously emphasised, it is imperative not to overlook the preliminary cell-sorting step in this context. The comprehensive characterisation of cells requires the consideration of both their phenotypical and functional attributes. Notably, sorting of cells into main compartments, such as CD4/CD8 and effector/regulator subsets, proves indispensable to prevent the pitfalls of failing to infer specificity effectively. This higher throughput coupled with new TCR-pMHC modelling strategies coming from artificial intelligence

[4] from which the 10X Chromium (10X Genomics) is based on

breakthrough (Bradley, 2023; Jokinen et al., 2022) are likely to be the new disruptive innovation in T cell repertoire analysis.

REFERENCES

1. Abe J, Kotzin BL, Jujo K, Melish ME, Glode MP, Kohsaka T, Leung DY. 1992. Selective expansion of T cells expressing T-cell receptor variable regions V beta 2 and V beta 8 in Kawasaki disease. *Proceedings of the National Academy of Sciences* **89**:4066–4070. doi:10.1073/pnas.89.9.4066

2. Abe J, Kotzin BL, Meissner C, Melish ME, Takahashi M, Fulton D, Romagne F, Malissen B, Leung DY. 1993. Characterization of T cell repertoire changes in acute Kawasaki disease. *Journal of Experimental Medicine* **177**:791–796. doi:10.1084/jem.177.3.791

3. Ace O, Domb S. 2019. Myocarditis as the initial presentation of Epstein-Barr virus infection in a 17-year-old male patient. *Can Fam Physician* **65**:897–899.

4. Acosta-Rodriguez EV, Rivino L, Geginat J, Jarrossay D, Gattorno M, Lanzavecchia A, Sallusto F, Napolitani G. 2007. Surface phenotype and antigenic specificity of human interleukin 17–producing T helper memory cells. *Nat Immunol* **8**:639–646. doi:10.1038/ni1467

5. Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, Bock C, Boehm B, Campo E, Caricasole A, Dahl F, Dermitzakis ET, Enver T, Esteller M, Estivill X, Ferguson-Smith A, Fitzgibbon J, Flicek P, Giehl C, Graf T, Grosveld F, Guigo R, Gut I, Helin K, Jarvius J, Küppers R, Lehrach H, Lengauer T, Lernmark Å, Leslie D, Loeffler M, Macintyre E, Mai A, Martens JH, Minucci S, Ouwehand WH, Pelicci PG, Pendeville H, Porse B, Rakyan V, Reik W, Schrappe M, Schübeler D, Seifert M, Siebert R, Simmons D, Soranzo N, Spicuglia S, Stratton M, Stunnenberg HG, Tanay A, Torrents D, Valencia A, Vellenga E, Vingron M, Walter J, Willcocks S. 2012. BLUEPRINT to decode the epigenetic signature written in blood. *Nat Biotechnol* **30**:224–226. doi:10.1038/nbt.2153

6. Adessi C, Matton G, Ayala G, Turcatti G, Mermod J, Mayer P, Kawashima E. 2000. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic acids research* **28**. doi:10.1093/nar/28.20.e87

7. Akira S, K O, H S. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. *Science (New York, NY)* **238**. doi:10.1126/science.3120312

8. Alcover A, Mariuzza RA, Ermonval M, Acuto O. 1990. Lysine 271 in the transmembrane domain of the T-cell antigen receptor beta chain is necessary for its assembly with the CD3 complex but not for alpha/beta dimerization. *J Biol Chem* **265**:4131–4135.

9. Alt FW, Baltimore D. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions.

Proceedings of the National Academy of Sciences **79**:4118–4122. doi:10.1073/pnas.79.13.4118

10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–410. doi:10.1016/S0022-2836(05)80360-2

11. Amadi-Obi A, Yu C-R, Liu X, Mahdi RM, Clarke GL, Nussenblatt RB, Gery I, Lee YS, Egwuagu CE. 2007. TH17 cells contribute to uveitis and scleritis and are expanded by IL-2 and inhibited by IL-27/STAT1. *Nat Med* **13**:711–718. doi:10.1038/nm1585

12. Amoriello R, Chernigovskaya M, Greiff V, Carnasciali A, Massacesi L, Barilaro A, Repice AM, Biagioli T, Aldinucci A, Muraro PA, Laplaud DA, Lossius A, Ballerini C. 2021. TCR repertoire diversity in Multiple Sclerosis: High-dimensional bioinformatics analysis of sequences from brain, cerebrospinal fluid and peripheral blood. *EBioMedicine* 103429. doi:10.1016/j.ebiom.2021.103429

13. Amoriello R, Greiff V, Aldinucci A, Bonechi E, Carnasciali A, B P, Am R, A M, R S, B M, L M, C B. 2020. The TCR Repertoire Reconstitution in Multiple Sclerosis: Comparing One-Shot and Continuous Immunosuppressive Therapies. *Frontiers in immunology* **11**. doi:10.3389/fimmu.2020.00559

14. Andrews S. 2010. FASTQC. A quality control tool for high throughput sequence data.

15. Antman E, Anbe D, Armstrong P, Bates E, La G, M H, Js H, Hm K, Fg K, Ga L, Cj M, Jp O, Dl P, Ma S, Sc S, Js A, Jl A, Dp F, V F, Rj G, G G, Jl H, Lf H, Sa H, Jacobs A. 2004. ACC/AHA guidelines for the management of patients with ST-elevation myocardial infarction--executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Writing Committee to Revise the 1999 Guidelines for the Management of Patients With Acute Myocardial Infarction). *Circulation* **110**. doi:10.1161/01.CIR.0000134791.68010.FA

16. Anzai A, Mindur JE, Halle L, Sano S, Choi JL, He S, McAlpine CS, Chan CT, Kahles F, Valet C, Fenn AM, Nairz M, Rattik S, Iwamoto Y, Fairweather D, Walsh K, Libby P, Nahrendorf M, Swirski FK. 2019. Self-reactive CD4+ IL-3+ T cells amplify autoimmune inflammation in myocarditis by inciting monocyte chemotaxis. *J Exp Med* **216**:369–383. doi:10.1084/jem.20180722

17. Aplan PD, Lombardi DP, Ginsberg AM, Cossman J, Bertness VL, Kirsch IR. 1990. Disruption of the Human SCL Locus by "Illegitimate" V-(D)-J Recombinase Activity. *Science* **250**:1426–1429. doi:10.1126/science.2255914

18. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. 1999. A Direct Estimate of the Human αβ T Cell Receptor Diversity. *Science* **286**:958–961. doi:10.1126/science.286.5441.958

19. Axelrod ML, Meijers WC, Screever EM, Qin J, Carroll MG, Sun X, Tannous E, Zhang Y, Sugiura A, Taylor BC, Hanna A, Zhang S, Amancherla K, Tai W, Wright JJ, Wei SC, Opalenik SR, Toren AL, Rathmell JC, Ferrell PB, Phillips EJ, Mallal S, Johnson DB, Allison JP, Moslehi JJ, Balko JM. 2022. T cells specific for α-myosin drive immunotherapy-related myocarditis. *Nature* **611**:818–826. doi:10.1038/s41586-022-05432-3

20. Bai X, Zhang Q, Wu S, Zhang X, Wang M, He F, Wei T, Yang J, Lou Y, Cai Z, Liang T. 2015. Characteristics of Tumor Infiltrating Lymphocyte and Circulating Lymphocyte Repertoires in Pancreatic Cancer by the Sequencing of T Cell Receptors. *Sci Rep* **5**:1–9. doi:10.1038/srep13664

21. Bains I, Antia R, Callard R, Yates AJ. 2009. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood* **113**:5480–5487. doi:10.1182/blood-2008-10-184184

22. Baker PE, Gillis S, Smith KA. 1979. Monoclonal cytolytic T-cell lines. *The Journal of experimental medicine* **149**:273–278. doi:10.1084/jem.149.1.273

23. Balakrishnan A, Morris GP. 2016. The highly alloreactive nature of dual TCR T cells. *Curr Opin Organ Transplant* **21**:22–28. doi:10.1097/MOT.0000000000000261

24. Baltcheva I, Veel E, Volman T, Koning D, Brouwer A, Le Boudec J, Tesselaar K, de Boer R, Borghans J. 2012. A generalized mathematical model to estimate T- and B-cell receptor diversities using AmpliCot. *Biophysical journal* **103**. doi:10.1016/j.bpj.2012.07.017

25. Bansal SS, Ismahil MA, Goel M, Patel B, Hamid T, Rokosh G, Prabhu SD. 2017. Activated T Lymphocytes are Essential Drivers of Pathological Remodeling in Ischemic Heart Failure. *Circ Heart Fail* **10**:e003688. doi:10.1161/CIRCHEARTFAILURE.116.003688

26. Barbet G, Nair-Gupta P, Schotsaert M, Yeung ST, Moretti J, Seyffer F, Metreveli G, Gardner T, Choi A, Tortorella D, Tampé R, Khanna KM, García-Sastre A, Blander JM. 2021. TAP dysfunction in dendritic cells enables noncanonical cross-presentation for T cell priming. *Nat Immunol* **22**:497–509. doi:10.1038/s41590-021-00903-7

27. Barennes P, Quiniou V, Shugay M, Egorov ES, Davydov AN, Chudakov DM, Uddin I, Ismail M, Oakes T, Chain B, Eugster A, Kashofer K, Rainer PP, Darko S, Ransier A, Douek DC, Klatzmann D, Mariotti-Ferrandiz E. 2021. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat Biotechnol* **39**:236–245. doi:10.1038/s41587-020-0656-3

28. Bautista JL, Lio C-WJ, Lathrop SK, Forbush K, Liang Y, Luo J, Rudensky AY, Hsieh C-S. 2009. Intraclonal competition limits the fate determination of regulatory T cells in the thymus. *Nat Immunol* **10**:610–617. doi:10.1038/ni.1739

29. Bayat S, Hashemi Nazari SS, Mehrabi Y, Sistanizad M. 2022. Long-term Survival Rate Following Myocardial Infarction and the Effect of Discharge Medications on the Survival Rate. *J Res Health Sci* **22**:e00567. doi:10.34172/jrhs.2022.102

30. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IMJ, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DMD, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E. Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang G-D, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike Andrew C., Pike Alger C., Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers John, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, vandeVondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers Jane, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**:53–59. doi:10.1038/nature07517

31. Bergmann O, Bhardwaj RD, Bernard S, Zdunek S, Barnabé-Heider F, Walsh S, Zupicich J, Alkass K, Buchholz BA, Druid H, Jovinge S, Frisén J. 2009. Evidence for Cardiomyocyte Renewal in Humans. *Science* **324**:98–102. doi:10.1126/science.1164680

32. Bertolino P, Trescol-Biémont M-C, Thomas J, de St Groth BF, Pihlgren M, Marvel J, Rabourdin-Combe C. 1999. Death by neglect as a deletional mechanism of peripheral tolerance. *International Immunology* **11**:1225–1238. doi:10.1093/intimm/11.8.1225

33. Beshnova D, Ye J, Onabolu O, Moon B, Zheng W, Fu Y-X, Brugarolas J, Lea J, Li B. 2020. De novo prediction of cancer-associated T cell receptors for noninvasive cancer detection. *Science translational medicine* **12**. doi:10.1126/scitranslmed.aaz3738

34. Best K, Oakes T, Heather JM, Shawe-Taylor J, Chain B. 2015. Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. *Sci Rep* **5**:14629. doi:10.1038/srep14629

35. Beverly B, Kang SM, Lenardo MJ, Schwartz RH. 1992. Reversal of in vitro T cell clonal anergy by IL-2 stimulation. *Int Immunol* **4**:661–671. doi:10.1093/intimm/4.6.661

36. Bhang HC, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, Singh AP, Kao I, Rakiec D, Shaw P, Balak M, Raza A, Ackley E, Keen N, Schlabach MR, Palmer M, Leary RJ, Chiang DY, Sellers WR, Michor F, Cooke VG, Korn JM, Stegmeier F. 2015. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med* **21**:440–448. doi:10.1038/nm.3841

37. Bhardwaj V, Kumar V, Geysen HM, Sercarz EE. 1993. Degenerate recognition of a dissimilar antigenic peptide by myelin basic protein-reactive T cells. Implications for thymic education and autoimmunity. *The Journal of Immunology* **151**:5000–5010. doi:10.4049/jimmunol.151.9.5000

38. Binkley PF, Cooke GE, Lesinski A, Taylor M, Chen M, Laskowski B, Waldman WJ, Ariza ME, Williams MV, Knight DA, Glaser R. 2013. Evidence for the Role of Epstein Barr Virus Infections in the Pathogenesis of Acute Coronary Events. *PLoS One* **8**:e54008. doi:10.1371/journal.pone.0054008

39. Björkbacka H, Lavant EH, Fredrikson GN, Melander O, Berglund G, Carlson JA, Nilsson J. 2010. Weak associations between human leucocyte antigen genotype and acute myocardial infarction. *Journal of Internal Medicine* **268**:50–58. doi:10.1111/j.1365-2796.2009.02209.x

40. Blackman M, Yagüe J, Kubo R, Gay D, Coleclough C, Palmer E, Kappler J, Marrack P. 1986. The T cell repertoire may be biased in favor of MHC recognition. *Cell* **47**:349–357. doi:10.1016/0092-8674(86)90591-X

41. Blanco-Domínguez R, Fuente H de la, Rodríguez C, Martín-Aguado L, Sánchez-Díaz R, Jiménez-Alejandre R, Rodríguez-Arabaolaza I, Curtabbi A, García-Guimaraes MM, Vera A, Rivero F, Cuesta J, Jiménez-Borreguero LJ, Cecconi A, Duran-Cambra A, Taurón M, Alonso J, Bueno H, Villalba-Orero M, Enríquez JA, Robson SC, Alfonso F, Sánchez-Madrid F, Martínez-González J, Martín P. 2022. CD69 expression on regulatory T cells protects from immune damage after myocardial infarction. *J Clin Invest* **132**. doi:10.1172/JCI152418

42. Blum KS, Pabst R. 2007. Lymphocyte numbers and subsets in the human blood. Do they mirror the situation in all organs? *Immunol Lett* **108**:45–51. doi:10.1016/j.imlet.2006.10.009

43. Bodger MP, Janossy G, Bollum FJ, Burford GD, Hoffbrand AV. 1983. The Ontogeny of Terminal Deoxynucleotidyl Transferase Positive Cells in the Human Fetus. *Blood* **61**:1125–1131. doi:10.1182/blood.V61.6.1125.1125

44. Boehm U, Klamp T, Groot M, Howard JC. 1997. CELLULAR RESPONSES TO INTERFERON-γ. *Annu Rev Immunol* **15**:749–795. doi:10.1146/annurev.immunol.15.1.749

45. 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:10.1038/nmeth.3364

46. Bomberger C, Singh-Jairam M, Rodey G, Guerriero A, Yeager AM, Fleming WH, Holland HK, Waller EK. 1998. Lymphoid Reconstitution After Autologous PBSC Transplantation With FACS-Sorted CD34+ Hematopoietic Progenitors. *Blood* **91**:2588–2600. doi:10.1182/blood.V91.7.2588

47. Bories JC, Demengeot J, Davidson L, Alt FW. 1996. Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility. *Proc Natl Acad Sci U S A* **93**:7871–7876. doi:10.1073/pnas.93.15.7871

48. Bortone DS, Woodcock MG, Parker JS, Vincent BG. 2021. Improved T-cell Receptor Diversity Estimates Associate with Survival and Response to Anti–PD-1 Therapy. *Cancer Immunology Research* **9**:103–112. doi:10.1158/2326-6066.CIR-20-0398

49. Botton J, Jabagi MJ, Bertrand M, Baricault B, Drouin J, Le Vu S, Weill A, Farrington P, Zureik M, Dray-Spira R. 2022. Risk for Myocardial Infarction, Stroke, and Pulmonary Embolism Following COVID-19 Vaccines in Adults Younger Than 75 Years in France. *Ann Intern Med* **175**:1250–1257. doi:10.7326/M22-0988

50. Boutet SC, Walter D, Stubbington MJT, Pfeiffer KA, Lee JY, Taylor SEB, Montesclaros L, Lau JK, Riordan DP, Barrio AM, Brix L, Jacobsen K, Yeung B, Zhao X, Mikkelsen TS. 2019. Scalable and comprehensive characterization of antigen-specific CD8 T cells using multi-omics single cell analysis. *The Journal of Immunology* **202**:131.4. doi:10.4049/jimmunol.202.Supp.131.4

51. Bouvier G, Watrin F, Naspetti M, Verthuy C, Naquet P, Ferrier P. 1996. Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proc Natl Acad Sci U S A* **93**:7877–7881. doi:10.1073/pnas.93.15.7877

52. Boyer SW, Rajendiran S, Beaudin AE, Smith-Berdan S, Muthuswamy PK, Perez-Cunningham J, Martin EW, Cheung C, Tsang H, Landon M, Forsberg EC. 2019. Clonal and Quantitative In Vivo Assessment of Hematopoietic Stem Cell Differentiation Reveals Strong Erythroid Potential of Multipotent Cells. *Stem Cell Reports* **12**:801–815. doi:10.1016/j.stemcr.2019.02.007

53. Bradley P. 2023. Structure-based prediction of T cell receptor:peptide-MHC interactions. *eLife* **12**:e82813. doi:10.7554/eLife.82813

54. Bradley P, Thomas PG. 2019. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. *Annu Rev Immunol* **37**:547–570. doi:10.1146/annurev-immunol-042718-041757

55. Bragado R, Lauzurica P, López D, López de Castro JA. 1990. T cell receptor V beta gene usage in a human alloreactive response. Shared structural features among HLA-B27-specific T cell clones. *Journal of Experimental Medicine* **171**:1189–1204. doi:10.1084/jem.171.4.1189

56. Brennecke P, Reyes A, Pinto S, Rattay K, Nguyen M, Küchler R, Huber W, Kyewski B, Steinmetz LM. 2015. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nat Immunol* **16**:933–941. doi:10.1038/ni.3246

57. Brinkmann V, Geiger T, Alkan S, Heusser CH. 1993. Interferon alpha increases the frequency of interferon gamma-producing human CD4+ T cells. *Journal of Experimental Medicine* **178**:1655–1663. doi:10.1084/jem.178.5.1655

58. Britanova OV, Putintseva EV, Shugay M, Merzlyak EM, Turchaninova MA, Staroverov DB, Bolotin DA, Lukyanov S, Bogdanova EA, Mamedov IZ, Lebedev YB, Chudakov DM. 2014. Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling. *The Journal of Immunology* **192**:2689–2698. doi:10.4049/jimmunol.1302064

59. Brown SD, Raeburn LA, Holt RA. 2015. Profiling tissue-resident T cell repertoires by RNA sequencing. *Genome Med* **7**:125. doi:10.1186/s13073-015-0248-x

60. Bulluck H, Go YY, Crimi G, Ludman AJ, Rosmini S, Abdel-Gadir A, Bhuva AN, Treibel TA, Fontana M, Pica S, Raineri C, Sirker A, Herrey AS, Manisty C, Groves A, Moon JC, Hausenloy DJ. 2017a. Defining left ventricular remodeling following acute ST-segment elevation myocardial infarction using cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance* **19**:26. doi:10.1186/s12968-017-0343-9

61. Bulluck H, Go YY, Crimi G, Ludman AJ, Rosmini S, Abdel-Gadir A, Bhuva AN, Treibel TA, Fontana M, Pica S, Raineri C, Sirker A, Herrey AS, Manisty C, Groves A, Moon JC, Hausenloy DJ. 2017b. Defining left ventricular remodeling following acute ST-segment elevation myocardial infarction using cardiovascular

magnetic resonance. *J Cardiovasc Magn Reson* **19**:26. doi:10.1186/s12968-017-0343-9

62. Burnet M. 1959. Auto-immune disease. I. Modern immunological concepts. *Br Med J* **2**:645–650. doi:10.1136/bmj.2.5153.645

63. Burzyn D, Kuswanto W, Kolodin D, Shadrach JL, Cerletti M, Jang Y, Sefik E, Tan TG, Wagers AJ, Benoist C, Mathis D. 2013. A Special Population of Regulatory T Cells Potentiates Muscle Repair. *Cell* **155**:1282–1295. doi:10.1016/j.cell.2013.10.054

64. Cao Y, Goods BA, Raddassi K, Nepom GT, Kwok WW, Love JC, Hafler DA. 2015. Functional inflammatory profiles distinguish myelin-reactive T cells from patients with multiple sclerosis. *Sci Transl Med* **7**:287ra74. doi:10.1126/scitranslmed.aaa8038

65. Carico Z, K RC, B Z, Y Z, Ms K. 2017. Tcrd Rearrangement Redirects a Processive Tcra Recombination Program to Expand the Tcra Repertoire. *Cell reports* **19**. doi:10.1016/j.celrep.2017.05.045

66. Carlson CM, Endrizzi BT, Wu J, Ding X, Weinreich MA, Walsh ER, Wani MA, Lingrel JB, Hogquist KA, Jameson SC. 2006. Kruppel-like factor 2 regulates thymocyte and T-cell migration. *Nature* **442**:299–302. doi:10.1038/nature04882

67. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung M-W, Parsons JM, Steen MS, LaMadrid-Herrmannsfeldt MA, Williamson DW, Livingston RJ, Wu D, Wood BL, Rieder MJ, Robins H. 2013. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* **4**:2680. doi:10.1038/ncomms3680

68. Carter JA, Preall JB, Grigaityte K, Goldfless SJ, Jeffery E, Briggs AW, Vigneault F, Atwal GS. 2019. Single T Cell Sequencing Demonstrates the Functional Role of αβ TCR Pairing in Cell Lineage and Antigen Specificity. *Frontiers in Immunology* **10**.

69. Casrouge A, Beaudoing E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P. 2000. Size Estimate of the αβ TCR Repertoire of Naive Mouse Splenocytes. *The Journal of Immunology* **164**:5782–5787. doi:10.4049/jimmunol.164.11.5782

70. Chaara W, Gonzalez-Tort A, Florez L-M, Klatzmann D, Mariotti-Ferrandiz E, Six A. 2018. RepSeq Data Representativeness and Robustness Assessment by Shannon Entropy. *Front Immunol* **9**:1038. doi:10.3389/fimmu.2018.01038

71. Chai V, Vassilakos A, Lee Y, Wright JA, Young AH. 2005. Optimization of the PAXgene blood RNA extraction system for gene expression analysis of clinical samples. *J Clin Lab Anal* **19**:182–188. doi:10.1002/jcla.20075

72. Chang M, Jin W, Chang J-H, Xiao Y, Brittain GC, Yu J, Zhou X, Wang Y-H, Cheng X, Li P, Rabinovich BA, Hwu P, Sun S-C. 2011. The ubiquitin ligase Peli1

negatively regulates T cell activation and prevents autoimmunity. *Nat Immunol* **12**:1002–1009. doi:10.1038/ni.2090

73. Chari T, Pachter L. 2023. The specious art of single-cell genomics. *PLOS Computational Biology* **19**:e1011288. doi:10.1371/journal.pcbi.1011288

74. Chaudhry A, Samstein RM, Treuting P, Liang Y, Pils MC, Heinrich J-M, Jack RS, Wunderlich FT, Brüning JC, Müller W, Rudensky AY. 2011. Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. *Immunity* **34**:566–578. doi:10.1016/j.immuni.2011.03.018

75. Chen S-Y, Liu C-J, Zhang Q, Guo A-Y. 2020. An ultra-sensitive T-cell receptor detection method for TCR-Seq and RNA-Seq data. *Bioinformatics* **36**:4255–4262. doi:10.1093/bioinformatics/btaa432

76. Chen W, Jin W, Hardegen N, Lei K-J, Li L, Marinos N, McGrady G, Wahl SM. 2003. Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. *J Exp Med* **198**:1875–1886. doi:10.1084/jem.20030152

77. Cheng X, Liao Y-H, Ge H, Li B, Zhang Jinying, Yuan J, Wang M, Liu Y, Guo Z, Chen J, Zhang Jin, Zhang L. 2005. Th1/Th2 Functional Imbalance After Acute Myocardial Infarction: Coronary Arterial Inflammation or Myocardial Inflammation. *J Clin Immunol* **25**:246–253. doi:10.1007/s10875-005-4088-0

78. Chicz R, Rg U, Ws L, Jc G, Lj S, Da V, Jl S. 1992. Predominant naturally processed peptides bound to HLA-DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* **358**. doi:10.1038/358764a0

79. Chiou S-H, Tseng D, Reuben A, Mallajosyula V, Molina IS, Conley S, Wilhelmy J, McSween AM, Yang X, Nishimiya D, Sinha R, Nabet BY, Wang C, Shrager JB, Berry MF, Backhus L, Lui NS, Wakelee HA, Neal JW, Padda SK, Berry GJ, Delaidelli A, Sorensen PH, Sotillo E, Tran P, Benson JA, Richards R, Labanieh L, Klysz DD, Louis DM, Feldman SA, Diehn M, Weissman IL, Zhang J, Wistuba II, Futreal PA, Heymach JV, Garcia KC, Mackall CL, Davis MM. 2021. Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery. *Immunity* **54**:586-602.e8. doi:10.1016/j.immuni.2021.02.014

80. Chong AJ, Shimamoto A, Hampton CR, Takayama H, Spring DJ, Rothnie CL, Yada M, Pohlman TH, Verrier ED. 2004. Toll-like receptor 4 mediates ischemia/reperfusion injury of the heart. *The Journal of Thoracic and Cardiovascular Surgery* **128**:170–179. doi:10.1016/j.jtcvs.2003.11.036

81. Choo EH, Lee J-H, Park E-H, Park HE, Jung N-C, Kim T-H, Koh Y-S, Kim E, Seung K-B, Park C, Hong K-S, Kang K, Song J-Y, Seo HG, Lim D-S, Chang K. 2017. Infarcted Myocardium-Primed Dendritic Cells Improve Remodeling and Cardiac Function After Myocardial Infarction by Modulating the Regulatory T

Cell and Macrophage Polarization. *Circulation* **135**:1444–1457. doi:10.1161/CIRCULATIONAHA.116.023106

82. Chothia C, Boswell DR, Lesk AM. 1988. The outline structure of the T-cell alpha beta receptor. *The EMBO Journal* **7**:3745–3755. doi:10.1002/j.1460-2075.1988.tb03258.x

83. Chowdhury RR, D'Addabbo J, Huang X, Veizades S, Sasagawa K, Louis DM, Cheng P, Sokol J, Jensen A, Tso A, Shankar V, Wendel BS, Bakerman I, Liang G, Koyano T, Fong R, Nau AN, Ahmad H, Gopakumar J, Wirka R, Lee AS, Boyd J, Woo YJ, Quertermous T, Gulati GS, Jaiswal S, Chien Y-H, Chan CKF, Davis MM, Nguyen PK. 2022. Human Coronary Plaque T Cells Are Clonal and Cross-React to Virus and Self. *Circ Res* **130**:1510–1530. doi:10.1161/CIRCRESAHA.121.320090

84. Chronister WD, Crinklaw A, Mahajan S, Vita R, Koşaloğlu-Yalçın Z, Yan Z, Greenbaum JA, Jessen LE, Nielsen M, Christley S, Cowell LG, Sette A, Peters B. 2021. TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors. *Front Immunol* **12**. doi:10.3389/fimmu.2021.640725

85. Chung J-W, Karau MJ, Greenwood-Quaintance KE, Ballard AD, Tilahun A, Khaleghi SR, David CS, Patel R, Rajagopalan G. 2014. Superantigen profiling of Staphylococcus aureus infective endocarditis isolates. *Diagnostic Microbiology and Infectious Disease* **79**:119–124. doi:10.1016/j.diagmicrobio.2014.03.009

86. Clark DR, de Boer RJ, Wolthers KC, Miedema F. 1999. T Cell Dynamics in HIV-1 Infection In: Dixon FJ, editor. Advances in Immunology. Academic Press. pp. 301–327. doi:10.1016/S0065-2776(08)60789-0

87. Coatnoan N, Berneman A, Chamond N, Minoprio P. 2009. Proline racemases: insights into Trypanosoma cruzi peptides containing D-proline. *Memorias do Instituto Oswaldo Cruz* **104 Suppl 1**. doi:10.1590/s0074-02762009000900039

88. Cochet M, Pannetier C, Regnault A, Darche S, Leclerc C, Kourilsky P. 1992. Molecular detection and in vivo analysis of the specific T cell response to a protein antigen. *European Journal of Immunology* **22**:2639–2647. doi:10.1002/eji.1830221025

89. Collette A, Bagot S, Ferrandiz ME, Cazenave P-A, Six A, Pied S. 2004. A Profound Alteration of Blood TCRB Repertoire Allows Prediction of Cerebral Malaria1. *The Journal of Immunology* **173**:4568–4575. doi:10.4049/jimmunol.173.7.4568

90. Collette A, Six A. 2002. ISEApeaks: an Excel platform for GeneScan and Immunoscope data retrieval, management and analysis. *Bioinformatics (Oxford, England)* **18**. doi:10.1093/bioinformatics/18.2.329

91. Collins AM, Ohlin M, Corcoran M, Heather JM, Ralph D, Law M, Martínez-Barnetche J, Ye J, Richardson E, Gibson WS, Rodriguez OL, Peres A, Yaari G, Watson CT, Lees WD. 2023. AIRR-C Human IG Reference Sets: curated sets of immunoglobulin heavy and light chain germline genes. doi:10.1101/2023.09.01.555348

92. Coutinho A. 2005. The Le Douarin phenomenon: a shift in the paradigm of developmental self-tolerance.

93. Cunha-Neto E, Coelho V, Guilherme L, Fiorelli A, Stolf N, Kalil J. 1996. Autoimmunity in Chagas' disease. Identification of cardiac myosin-B13 Trypanosoma cruzi protein crossreactive T cell clones in heart lesions of a chronic Chagas' cardiomyopathy patient. *J Clin Invest* **98**:1709–1712. doi:10.1172/JCI118969

94. Curotto de Lafaille MA, Kutchukhidze N, Shen S, Ding Y, Yee H, Lafaille JJ. 2008. Adaptive Foxp3+ regulatory T cell-dependent and -independent control of allergic inflammation. *Immunity* **29**:114–126. doi:10.1016/j.immuni.2008.05.010

95. Curtis N, Zheng R, Lamb JR, Levin M. 1995. Evidence for a superantigen mediated process in Kawasaki disease. *Arch Dis Child* **72**:308–311.

96. Dahal-Koirala S, Balaban G, Neumann RS, Scheffer L, Lundin KEA, Greiff V, Sollid LM, Qiao S-W, Sandve GK. 2022. TCRpower: quantifying the detection power of T-cell receptor sequencing with a novel computational pipeline calibrated by spike-in sequences. *Briefings in Bioinformatics* **23**:bbab566. doi:10.1093/bib/bbab566

97. Darby CA, Stubbington MJT, Marks PJ, Martínez Barrio Á, Fiddes IT. 2020. scHLAcount: allele-specific HLA expression from single-cell gene expression data. *Bioinformatics* **36**:3905–3906. doi:10.1093/bioinformatics/btaa264

98. Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**:395–402. doi:10.1038/334395a0

99. Davydov AN, Bolotin DA, Poslavsky SV, Chudakov DM. 2023. Comment on 'rigorous benchmarking of T cell receptor repertoire profiling methods for cancer RNA sequencing.' *Briefings in Bioinformatics* **24**:bbad354. doi:10.1093/bib/bbad354

100. De Boer RJ, Perelson AS. 2013. Quantifying T lymphocyte turnover. *Journal of Theoretical Biology* **327**:45–87. doi:10.1016/j.jtbi.2012.12.025

101. De Neuter N, Bittremieux W, Beirnaert C, Cuypers B, Mrzic A, Moris P, Suls A, Van Tendeloo V, Ogunjimi B, Laukens K, Meysman P. 2018. On the feasibility of mining CD8+ T cell receptor patterns underlying immunogenic peptide recognition. *Immunogenetics* **70**:159–168. doi:10.1007/s00251-017-1023-5

102. Deering RP, Blumenberg L, Li L, Dhanik A, Jeong S, Pourpe S, Song H, Boucher L, Ragunathan S, Li Y, Zhong M, Kuhnert J, Adler C, Hawkins P, Gupta NT, Moore M, Ni M, Hansen J, Wei Y, Thurston G. 2023. Rapid TCR:Epitope Ranker (RAPTER): a primary human T cell reactivity screening assay pairing epitope and TCR at single cell resolution. *Sci Rep* **13**:8452. doi:10.1038/s41598-023-35710-7

103. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, Georgiou G. 2015. In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* **21**:86–91. doi:10.1038/nm.3743

104. Delgobo M, Weiß E, Ashour D, Richter L, Popiolkowski L, Arampatzi P, Stangl V, Arias-Loza P, Mariotti-Ferrandiz E, Rainer PP, Saliba A-E, Ludewig B, Hofmann U, Frantz S, Campos Ramos G. 2023. Myocardial Milieu Favors Local Differentiation of Regulatory T Cells. *Circulation Research* **132**:565–582. doi:10.1161/CIRCRESAHA.122.322183

105. Dengjel J, Schoor O, Fischer R, Reich M, Kraus M, Müller M, Kreymborg K, Altenberend F, Brandenburg J, Kalbacher H, Brock R, Driessen C, Rammensee H-G, Stevanovic S. 2005. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc Natl Acad Sci U S A* **102**:7922–7927. doi:10.1073/pnas.0501190102

106. Dens C, Laukens K, Bittremieux W, Meysman P. 2023. The pitfalls of negative data bias for the T-cell epitope specificity challenge. *Nat Mach Intell* 1–3. doi:10.1038/s42256-023-00727-0

107. Depuydt MAC, Schaftenaar FH, Prange KHM, Boltjes A, Hemme E, Delfos L, de Mol J, de Jong MJM, Bernabé Kleijn MNA, Peeters JAHM, Goncalves L, Wezel A, Smeets HJ, de Borst GJ, Foks AC, Pasterkamp G, de Winther MPJ, Kuiper J, Bot I, Slütter B. 2023. Single-cell T cell receptor sequencing of paired human atherosclerotic plaques and blood reveals autoimmune-like features of expanded effector T cells. *Nat Cardiovasc Res* **2**:112–125. doi:10.1038/s44161-022-00208-4

108. Desponds J, Mora T, Walczak AM. 2016. Fluctuating fitness shapes the clone-size distribution of immune repertoires. *Proceedings of the National Academy of Sciences* **113**:274–279. doi:10.1073/pnas.1512977112

109. Deubner N, Berliner D, Schlipp A, Gelbrich G, Caforio ALP, Felix SB, Fu M, Katus H, Angermann CE, Lohse MJ, Ertl G, Störk S, Jahns R, Etiology, Titre-Course, and Survival-Study Group. 2010a. Cardiac beta1-adrenoceptor autoantibodies in human heart disease: rationale and design of the Etiology, Titre-Course, and Survival (ETiCS) Study. *Eur J Heart Fail* **12**:753–762. doi:10.1093/eurjhf/hfq072

110. Deubner N, Berliner D, Schlipp A, Gelbrich G, Caforio ALP, Felix SB, Fu M, Katus H, Angermann CE, Lohse MJ, Ertl G, Störk S, Jahns R, Group on behalf of the Et-S. 2010b. Cardiac β1-adrenoceptor autoantibodies in human heart disease: rationale and design of the Etiology, Titre-Course, and Survival (ETiCS) Study. *European Journal of Heart Failure* **12**:753–762. doi:10.1093/eurjhf/hfq072

111. Devaux B, Scholz D, Hirche A, Klövekorn WP, Schaper J. 1997. Upregulation of cell adhesion molecules and the presence of low grade inflammation in human chronic heart failure. *Eur Heart J* **18**:470–479. doi:10.1093/oxfordjournals.eurheartj.a015268

112. Dhalla F, Baran-Gale J, Maio S, Chappell L, Holländer GA, Ponting CP. 2020. Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. *The EMBO Journal* **39**:e101828. doi:10.15252/embj.2019101828

113. Dhanda SK, Mahajan S, Paul S, Yan Z, Kim H, Jespersen MC, Jurtz V, Andreatta M, Greenbaum JA, Marcatili P, Sette A, Nielsen M, Peters B. 2019. IEDB-AR: immune epitope database—analysis resource in 2019. *Nucleic Acids Research* **47**:W502–W506. doi:10.1093/nar/gkz452

114. Diamantopoulos EJ, Andreadis EA, Vassilopoulos CV, Vlachonikolis IG, Tarassi KE, Chatzis NA, Tsourous GI, Papasteriades CA. 2003. HLA phenotypes as promoters of cardiovascular remodelling in subjects with arterial hypertension. *J Hum Hypertens* **17**:63–68. doi:10.1038/sj.jhh.1001502

115. 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. doi:10.1093/bioinformatics/bts635

116. Donermeyer DL, Beisel KW, Allen PM, Smith SC. 1995. Myocarditis-inducing epitope of myosin binds constitutively and stably to I-Ak on antigen-presenting cells in the heart. *Journal of Experimental Medicine* **182**:1291–1300. doi:10.1084/jem.182.5.1291

117. Dong D, Zheng L, Lin J, Zhang B, Zhu Y, Li N, Xie S, Wang Y, Gao N, Huang Z. 2019. Structural basis of assembly of the human T cell receptor–CD3 complex. *Nature* **573**:546–552. doi:10.1038/s41586-019-1537-0

118. Dupic T, Marcou Q, Walczak AM, Mora T. 2019. Genesis of the αβ T-cell receptor. *PLoS Computational Biology* **15**. doi:10.1371/journal.pcbi.1006874

119. Early P, Huang H, Davis M, Calame K, Hood L. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* **19**:981–992. doi:10.1016/0092-8674(80)90089-6

120. Egerton M, Scollay R, Shortman K. 1990. Kinetics of mature T-cell development in the thymus. *Proc Natl Acad Sci U S A* **87**:2579–2582. doi:10.1073/pnas.87.7.2579

121. Ehrlich P, Hübener W. 1894. Ueber die Vererbung der Immunität bei Tetanus. *Zeitschr f Hygiene* **18**:51–64. doi:10.1007/BF02216834

122. Elhage R, Gourdy P, Brouchet L, Jawien J, Fouque M-J, Fiévet C, Huc X, Barreira Y, Couloumiers JC, Arnal J-F, Bayard F. 2004. Deleting TCR alpha beta+ or CD4+ T lymphocytes leads to opposite effects on site-specific atherosclerosis in female apolipoprotein E-deficient mice. *Am J Pathol* **165**:2013–2018. doi:10.1016/s0002-9440(10)63252-x

123. Elhanati Y, Murugan A, Callan CG, Mora T, Walczak AM. 2014. Quantifying selection in immune receptor repertoires. *Proceedings of the National Academy of Sciences* **111**:9875–9880. doi:10.1073/pnas.1409572111

124. Ellul P, Rosenzwajg M, Peyre H, Fourcade G, Mariotti-Ferrandiz E, Trebossen V, Klatzmann D, Delorme R. 2021. Regulatory T lymphocytes/Th17 lymphocytes imbalance in autism spectrum disorders: evidence from a meta-analysis. *Molecular Autism* **12**:1–7. doi:10.1186/s13229-021-00472-4

125. Eltahla AA, Rizzetto S, Pirozyan MR, Betz-Stablein BD, Venturi V, Kedzierska K, Lloyd AR, Bull RA, Luciani F. 2016. Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. *Immunol Cell Biol* **94**:604–611. doi:10.1038/icb.2016.16

126. ElTanbouly MA, Noelle RJ. 2021. Rethinking peripheral T cell tolerance: checkpoints across a T cell's journey. *Nat Rev Immunol* **21**:257–267. doi:10.1038/s41577-020-00454-2

127. Emerson R, Sherwood A, Desmarais C, Malhotra S, Phippard D, Robins H. 2013. Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data. *J Immunol Methods* **391**:14–21. doi:10.1016/j.jim.2013.02.002

128. Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, Desmarais C, Klinger M, Carlson CS, Hansen JA, Rieder M, Robins HS. 2017. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet* **49**:659–665. doi:10.1038/ng.3822

129. Emerson RO, Sherwood AM, Rieder MJ, Guenthoer J, Williamson DW, Carlson CS, Drescher CW, Tewari M, Bielas JH, Robins HS. 2013. High-throughput sequencing of T-cell receptors reveals a homogeneous repertoire of tumour-infiltrating lymphocytes in ovarian cancer. *The Journal of Pathology* **231**:433–440. doi:10.1002/path.4260

130. Epelman S, Liu PP, Mann DL. 2015. Role of innate and adaptive immune mechanisms in cardiac injury and repair. *Nature Reviews Immunology* **15**:117–129. doi:10.1038/nri3800

131. Espinoza CR, Feeney AJ. 2007. Chromatin accessibility and epigenetic modifications differ between frequently and infrequently rearranging VH genes. *Molecular Immunology* **44**:2675–2685. doi:10.1016/j.molimm.2006.12.002

132. Essery G, Feldmann M, Lamb JR. 1988. Interleukin-2 can prevent and reverse antigen-induced unresponsiveness in cloned human T lymphocytes. *Immunology* **64**:413–417.

133. Fang H, Yamaguchi R, Liu X, Daigo Y, Yew PY, Tanikawa C, Matsuda K, Imoto S, Miyano S, Nakamura Y. 2014. Quantitative T cell repertoire analysis by deep cDNA sequencing of T cell receptor α and β chains using next-generation sequencing (NGS). *OncoImmunology* **3**:e968467. doi:10.4161/21624011.2014.968467

134. Fanti S, Stephenson E, Rocha-Vieira E, Protonotarios A, Kanoni S, Shahaj E, Longhi MP, Vyas VS, Dyer C, Pontarini E, Asimaki A, Bueno-Beti C, De Gaspari M, Rizzo S, Basso C, Bombardieri M, Coe D, Wang G, Harding D, Gallagher I, Solito E, Elliott P, Heymans S, Sikking M, Savvatis K, Mohiddin SA, Marelli-Berg FM. 2022. Circulating c-Met–Expressing Memory T Cells Define Cardiac Autoimmunity. *Circulation* **146**:1930–1945. doi:10.1161/CIRCULATIONAHA.121.055610

135. Farber DL, Netea MG, Radbruch A, Rajewsky K, Zinkernagel RM. 2016. Immunological memory: lessons from the past and a look to the future. *Nat Rev Immunol* **16**:124–128. doi:10.1038/nri.2016.13

136. Feng D, Bond CJ, Ely LK, Maynard J, Garcia KC. 2007. Structural evidence for a germline-encoded T cell receptor–major histocompatibility complex interaction "codon." *Nat Immunol* **8**:975–983. doi:10.1038/ni1502

137. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, Khan NS, Wong CK, Shamailova R, Hill CA, Wang Z, Remark R, Li JR, Pina C, Faries C, Awad AJ, Moss N, Bjorkegren JLM, Kim-Schulze S, Gnjatic S, Ma'ayan A, Mocco J, Faries P, Merad M, Giannarelli C. 2019. Single-cell immune landscape of human atherosclerotic plaques. *Nat Med* **25**:1576–1588. doi:10.1038/s41591-019-0590-4

138. Feuerer M, Hill JA, Kretschmer K, von Boehmer H, Mathis D, Benoist C. 2010. Genomic definition of multiple ex vivo regulatory T cell subphenotypes. *Proc Natl Acad Sci U S A* **107**:5919–5924. doi:10.1073/pnas.1002006107

139. Fischer DS, Wu Y, Schubert B, Theis FJ. 2020. Predicting antigen specificity of single T cells based on TCR CDR3 regions. *Molecular Systems Biology* **16**:e9416. doi:10.15252/msb.20199416

140. Fontenot JD, Gavin MA, Rudensky AY. 2003. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol* **4**:330–336. doi:10.1038/ni904

141. Forte E, Perkins B, Sintou A, Kalkat HS, Papanikolaou A, Jenkins C, Alsubaie M, Chowdhury RA, Duffy TM, Skelly DA, Branca J, Bellahcene M, Schneider MD, Harding SE, Furtado MB, Ng FS, Hasham MG, Rosenthal N, Sattler S. 2021. Cross-Priming Dendritic Cells Exacerbate Immunopathology After Ischemic Tissue Damage in the Heart. *Circulation* **143**:821–836. doi:10.1161/CIRCULATIONAHA.120.044581

142. Frank SJ, Niklinska BB, Orloff DG, Merćep M, Ashwell JD, Klausner RD. 1990. Structural Mutations of the T Cell Receptor ζ Chain and Its Role in T Cell Activation. *Science* **249**:174–177. doi:10.1126/science.2371564

143. Freitas AA, Rocha B. 2000. Population Biology of Lymphocytes: The Flight for Survival. *Annu Rev Immunol* **18**:83–111. doi:10.1146/annurev.immunol.18.1.83

144. Frohman MA, Dush MK, Martin GR. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci U S A* **85**:8998–9002. doi:10.1073/pnas.85.23.8998

145. Frostegård J, Ulfgren AK, Nyberg P, Hedin U, Swedenborg J, Andersson U, Hansson GK. 1999. Cytokine expression in advanced human atherosclerotic plaques: dominance of pro-inflammatory (Th1) and macrophage-stimulating cytokines. *Atherosclerosis* **145**:33–43. doi:10.1016/s0021-9150(99)00011-8

146. Fu X, Khalil H, Kanisicak O, Boyer JG, Vagnozzi RJ, Maliken BD, Sargent MA, Prasad V, Valiente-Alandi I, Blaxall BC, Molkentin JD. 2018. Specialized fibroblast differentiated states underlie scar formation in the infarcted mouse heart. *J Clin Invest* **128**:2127–2143. doi:10.1172/JCI98215

147. Fukunaga T, Soejima H, Irie A, Sugamura K, Oe Y, Tanaka T, Kojima S, Sakamoto T, Yoshimura M, Nishimura Y, Ogawa H. 2007. Expression of interferon-g and interleukin-4 production in CD4+ T cells in patients with chronic heart failure. *Heart Vessels* **22**:178–183. doi:10.1007/s00380-006-0955-8

148. Gammon G, Sercarz E. 1989. How some T cells escape tolerance induction. *Nature* **342**. doi:10.1038/342183a0

149. Gangi-Peterson L, Sorscher DH, Reynolds JW, Kepler TB, Mitchell BS. 1999. Nucleotide pool imbalance and adenosine deaminase deficiency induce alterations of *N*-region insertions during V(D)J recombination. *J Clin Invest* **103**:833–841. doi:10.1172/JCI4320

150. Gao Yicheng, Gao Yuli, Dong K, Wu S, Liu Q. 2023a. Reply to: The pitfalls of negative data bias for the T-cell epitope specificity challenge. *Nat Mach Intell* 1–3. doi:10.1038/s42256-023-00725-2

151. Gao Yicheng, Gao Yuli, Fan Y, Zhu C, Wei Z, Zhou C, Chuai G, Chen Q, Zhang H, Liu Q. 2023b. Pan-Peptide Meta Learning for T-cell receptor–antigen binding recognition. *Nat Mach Intell* **5**:236–249. doi:10.1038/s42256-023-00619-3

152. Garcia K, Adams JJ, Feng D, Ely LK. 2009. The molecular basis of TCR germline bias for MHC is surprisingly simple. *Nat Immunol* **10**:143–147. doi:10.1038/ni.f.219

153. Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, Teyton L, Wilson IA. 1998. Structural Basis of Plasticity in T Cell Receptor Recognition of a Self Peptide-MHC Antigen. *Science* **279**:1166–1172. doi:10.1126/science.279.5354.1166

154. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, Teyton L, Wilson IA. 1996. An αβ T Cell Receptor Structure at 2.5 Å and Its Orientation in the TCR-MHC Complex. *Science* **274**:209–219. doi:10.1126/science.274.5285.209

155. García-Ropero Á, Santos-Gallego CG, Badimon JJ. 2019. The anti-inflammatory effects of SGLT inhibitors. *Aging (Albany NY)* **11**:5866–5867. doi:10.18632/aging.102175

156. Gauss G, Lieber MR. 1996. Mechanistic Constraints on Diversity in Human V(D)J Recombination. *Molecular and Cellular Biology* **16**:258–269. doi:10.1128/MCB.16.1.258

157. Genolet R, Bobisse S, Chiffelle J, Arnaud M, Petremand R, Queiroz L, Michel A, Reichenbach P, Cesbron J, Auger A, Baumgaertner P, Guillaume P, Schmidt J, Irving M, Kandalaft LE, Speiser DE, Coukos G, Harari A. 2023. TCR sequencing and cloning methods for repertoire analysis and isolation of tumor-reactive TCRs. *Cell Rep Methods* **3**:100459. doi:10.1016/j.crmeth.2023.100459

158. Gibbs RA. 2020. The Human Genome Project changed everything. *Nat Rev Genet* **21**:575–576. doi:10.1038/s41576-020-0275-3

159. Gielis S, Moris P, Bittremieux W, De Neuter N, Ogunjimi B, Laukens K, Meysman P. 2019. Detection of Enriched T Cell Epitope Specificity in Full T Cell Receptor Sequence Repertoires. *Frontiers in Immunology* **10**.

160. Gihawi A, Cooper CS, Brewer DS. 2023. Caution Regarding the Specificities of Pan-Cancer Microbial Structure. doi:10.1101/2023.01.16.523562

161. Gillis S, Baker PE, Ruscetti FW, Smith KA. 1978. Long-term culture of human antigen-specific cytotoxic T-cell lines. *Journal of Experimental Medicine* **148**:1093–1098. doi:10.1084/jem.148.4.1093

162. Gittelman RM, Lavezzo E, Snyder TM, Zahid HJ, Carty CL, Elyanow R, Dalai S, Kirsch I, Baldo L, Manuto L, Franchin E, Vecchio CD, Pacenti M, Boldrin C, Cattai M, Saluzzo F, Padoan A, Plebani M, Simeoni F, Bordini J, Lorè NI, Lazarević D, Cirillo DM, Ghia P, Toppo S, Carlson JM, Robins HS, Crisanti A, Tonon G. 2022. Longitudinal analysis of T cell receptor repertoires reveals shared patterns of antigen-specific response to SARS-CoV-2 infection. *JCI Insight* **7**. doi:10.1172/jci.insight.151849

163. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, Ji X, Han A, Krams SM, Pettus C, Haas N, Arlehamn CSL, Sette A, Boyd SD, Scriba TJ, Martinez OM, Davis MM. 2017. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**:94–98. doi:10.1038/nature22976

164. Glatman Zaretsky A, Taylor JJ, King IL, Marshall FA, Mohrs M, Pearce EJ. 2009. T follicular helper cells differentiate from Th2 cells in response to helminth antigens. *J Exp Med* **206**:991–999. doi:10.1084/jem.20090303

165. Gocayne J, Robinson D, FitzGerald M, Chung F, Kerlavage A, Ku L, J L, Cd W, Cm F, Jc V. 1987. Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors obtained by automated DNA sequence analysis: further evidence for a multigene family. *Proceedings of the National Academy of Sciences of the United States of America* **84**. doi:10.1073/pnas.84.23.8296

166. Goldman AS, Palkowetz KH, Rudloff HE, Brooks EG, Schmalstieg FC. 1992. Repertoire of Vα and Vβ regions of T cell antigen receptors on CD4+ and CD8+ peripheral blood T cells in a novel X-linked combined immunodeficiency disease. *European Journal of Immunology* **22**:1103–1106. doi:10.1002/eji.1830220435

167. Gorochov G, Neumann AU, Kereveur A, Parizot C, Li T, Katlama C, Karmochkine M, Raguin G, Autran B, Debré P. 1998. Perturbation of CD4+ and CD8+ T-cell repertoires during progression to AIDS and regulation of the CD4+ repertoire during antiviral therapy. *Nat Med* **4**:215–221. doi:10.1038/nm0298-215

168. Gould E, Fraser HS, Parker TH, Nakagawa S, Griffith SC, Vesk PA, Fidler F, Hamilton DG, Abbey-Lee RN, Abbott JK, Aguirre LA, Alcaraz C, Aloni I, Altschul D, Arekar K, Atkins JW, Atkinson J, Baker C, Barrett M, Bell K, Bello SK, Beltrán I, Berauer BJ, Bertram MG, Billman PD, Blake CK, Blake S, Bliard L, Bonisoli-Alquati A, Bonnet T, Bordes CNM, Bose APH, Botterill-James T, Boyd MA, Boyle SA, Bradfer-Lawrence T, Bradham J, Brand JA, Brengdahl MI, Bulla M, Bussière L, Camerlenghi E, Campbell SE, Campos LLF, Caravaggi A, Cardoso P, Carroll CJW, Catanach TA, Chen X, Chik HYJ, Choy ES, Christie AP, Chuang A, Chunco AJ, Clark BL, Contina A, Covernton GA, Cox MPC,

Cressman KA, Crouch CD, D'Amelio PB, Sousa AA de, Döbert TF, Dobler R, Dobson AJ, Doherty TS, Drobniak SM, Duffy AG, Duncan AB, Dunn RP, Dunning J, Dutta T, Eberhart-Hertel L, Elmore JA, Elsherif MM, English HM, Ensminger DC, Ernst UR, Ferguson SM, Fernández-Juricic E, Ferreira-Arruda TF-A, Fieberg J, Finch EA, Fiorenza EA, Fisher DN, Fontaine A, Forstmeier W, Fourcade Y, Frank GS, Freund CA, Fuentes-Lillo E, Gandy SL, Gannon DG, García-Cervigón AI, Garretson AC, Ge X, Geary WL, Géron C, Gilles M, Girndt A, Gliksman D, Goldspiel HB, Gomes DGE, Good MK, Goslee SC, Gosnell JS, Grames EM, Gratton P, Grebe NM, Greenler SM, Griffioen M, Griffith DM, Griffith FJ, Grossman JJ, Güncan A, Haesen S, Hagan JG, Hager HA, Harrison ND, Hasnain SS, Havird JC, Heaton A, Herrera-Chaustre ML, Howard TJ, Hsu B-Y, Iannarilli F, Iranzo EC, Iverson ENK, Jimoh SO, Johnson DH, Johnsson M, Jorna J, Jucker T, Jung M, Kačergytė I, Kaltz O, Ke A, Kelly CD, Keogan K, Keppeler FW, Killion AK, Kim D, Kochan DP, Korsten P, Kothari S, Kuppler J, Kusch JM, Lagisz M, Lalla KM, Larkin DJ, Larson CL, Lauck KS, Lauterbur ME, Law A, Léandri-Breton D-J, Lembrechts J, L'Herpiniere K, Lievens EJP, Lima DO de, Lindsay S, Luquet M, Macphie KH, Mair MM, Malm LE, Mammola S, Mandeville CP, Manhart M, Manrique-Garzon LM, Mäntylä E, Marchand P, Marshall BM, Martin CA, Martin DA, Martin JM, Martinig AR, McCallum ES, McCauley M, McNew SM, Meiners SJ, Merkling T, Michelangeli M, Moiron M, Moreira B, Mortensen J, Mos B, Muraina TO, Murphy PW, Nelli L, Niemelä P, Nightingale J, Nilsonne G, Nolazco S, Nooten SS, Novotny JL, Olin AB, Organ CL, Ostevik KL, Palacio FX, Paquet M, Parker DJ, Pascall DJ, Pasquarella VJ, Paterson JH, Payo-Payo A, Pedersen KM, Perez G, Perry KI, Pottier P, Proulx MJ, Proulx R, Pruett JL, Ramananjato V, Randimbiarison FT, Razafindratsima OH, Rennison DJ, Riva F, Riyahi SR, Roast MJ, Rocha FP, Roche DG, Román-Palacios C, Rosenberg MS, Ross J, Rowland FE, Rugemalila D, Russell AL, Ruuskanen S, Saccone P, Sadeh A, Salazar SM, Sales K, Salmón P, Sanchez-Tojar A, Santos LP, Santostefano F, Schilling HT, Schmidt M, Schmoll T, Schneider AC, Schrock AE, Schroeder J, Schtickzelle N, Schultz NL, Scott DA, Shapiro JT, Sharma NS, Shearer CL, Simón D, Sitvarin MI, Skupien FL, Slinn HL, Smith GP, Smith JA, Sollmann R, Whitney KS, Still SM, Stuber EF, Sutton GF, Swallow B, Taff CC, Takola E, Tanentzap AJ, Tarjuelo R, Telford RJ, Thawley CJ, Thierry H, Thomson J, Tidau S, Tompkins EM, Tortorelli CM, Trlica A, Turnell BR, Urban L, Vondel SV de, Wal JEM van der, Eeckhoven JV, Oordt F van, Vanderwel KM, Vanderwel MC, Vanderwolf KJ, Vergara-Florez DC, Verrelli BC, Vieira MV, Villamil N, Vitali V, Vollering J, Walker XJ, Walter JA, Waryszak P, Weaver RJ, Wedegärtner REM, Weller DL, Whelan S, White R, Wolfson DW, Wood A, Yanco SW, Yen JDL, Youngflesh C, Zilio G, Zimmer C, Zimmerman GM, Zitomer RA. 2023. Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology.

169. Govaerts A. 1960. Cellular Antibodies in Kidney Homotransplantation1. *The Journal of Immunology* **85**:516–522. doi:10.4049/jimmunol.85.5.516

170. Gras S, Chen Z, Miles JJ, Liu YC, Bell MJ, Sullivan LC, Kjer-Nielsen L, Brennan RM, Burrows JM, Neller MA, Khanna R, Purcell AW, Brooks AG, McCluskey J, Rossjohn J, Burrows SR. 2010. Allelic polymorphism in the T cell

receptor and its impact on immune responses. *Journal of Experimental Medicine* **207**:1555–1567. doi:10.1084/jem.20100603

171. Gras S, Saulquin X, Reiser J-B, Debeaupuis E, Echasserieau K, Kissenpfennig A, Legoux F, Chouquet A, Le Gorrec M, Machillot P, Neveu B, Thielens N, Malissen B, Bonneville M, Housset D. 2009. Structural Bases for the Affinity-Driven Selection of a Public TCR against a Dominant Human Cytomegalovirus Epitope1. *The Journal of Immunology* **183**:430–437. doi:10.4049/jimmunol.0900556

172. Gray D, Abramson J, Benoist C, Mathis D. 2007. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. *J Exp Med* **204**:2521–2528. doi:10.1084/jem.20070795

173. Greiff V, Bhat P, Cook SC, Menzel U, Kang W, Reddy ST. 2015a. A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. *Genome Med* **7**:1–15. doi:10.1186/s13073-015-0169-8

174. Greiff V, Miho E, Menzel U, Reddy ST. 2015b. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends in Immunology* **36**:738–749. doi:10.1016/j.it.2015.09.006

175. Gross CP, Sepkowitz KA. 1998. The myth of the medical breakthrough: smallpox, vaccination, and Jenner reconsidered. *Int J Infect Dis* **3**:54–60. doi:10.1016/s1201-9712(98)90096-0

176. Guasch-Ferré M, Liu X, Malik VS, Sun Q, Willett WC, Manson JE, Rexrode KM, Li Y, Hu FB, Bhupathiraju SN. 2017. Nut Consumption and Risk of Cardiovascular Disease. *Journal of the American College of Cardiology* **70**:2519–2532. doi:10.1016/j.jacc.2017.09.035

177. Guerau-de-Arellano M, Mathis D, Benoist C. 2008. Transcriptional impact of Aire varies with cell type. *Proc Natl Acad Sci U S A* **105**:14011–14016. doi:10.1073/pnas.0806616105

178. Guo L, Bi X, Li Y, Wen L, Zhang W, Jiang W, Ma J, Feng L, Zhang K, Shou J. 2020. Characteristics, dynamic changes, and prognostic significance of TCR repertoire profiling in patients with renal cell carcinoma. *The Journal of Pathology* **251**:26–37. doi:10.1002/path.5396

179. Hadaschik EN, Wei X, Leiss H, Heckmann B, Niederreiter B, Steiner G, Ulrich W, Enk AH, Smolen JS, Stummvoll GH. 2015. Regulatory T cell-deficient scurfy mice develop systemic autoimmune features resembling lupus-like disease. *Arthritis Research & Therapy* **17**. doi:10.1186/s13075-015-0538-0

180. Hafler DA, Saadeh MG, Kuchroo VK, Milford E, Steinman L. 1996. TCR usage in human and experimental demyelinating disease. *Immunology Today* **17**:152–159. doi:10.1016/0167-5699(96)80611-6

181. Hapke N, Heinrichs M, Ashour D, Vogel E, Hofmann U, Frantz S, Campos Ramos G. 2022. Identification of a novel cardiac epitope triggering T-cell responses in patients with myocardial infarction. *Journal of Molecular and Cellular Cardiology* **173**:25–29. doi:10.1016/j.yjmcc.2022.09.001

182. Harrington LE, Hatton RD, Mangan PR, Turner H, Murphy TL, Murphy KM, Weaver CT. 2005. Interleukin 17–producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat Immunol* **6**:1123–1132. doi:10.1038/ni1254

183. Hassin D, Garber OG, Meiraz A, Schiffenbauer YS, Berke G. 2011. Cytotoxic T lymphocyte perforin and Fas ligand working in concert even when Fas ligand lytic action is still not detectable. *Immunology* **133**:190–196. doi:10.1111/j.1365-2567.2011.03426.x

184. Haugomat T (DREES/OSAM/LABSANTE). 2022. Les taux de personnes vaccinées et non vaccinées contre le Covid-19 en France.

185. Heath WR, Allison J, Hoffmann MW, Schönrich G, Hämmerling G, Arnold B, Miller JF. 1992. Autoimmune diabetes as a consequence of locally produced interleukin-2. *Nature* **359**:547–549. doi:10.1038/359547a0

186. Heath WR, Carbone FR, Bertolino P, Kelly J, Cose S, Miller JF. 1995. Expression of two T cell receptor alpha chains on the surface of normal murine T cells. *Eur J Immunol* **25**:1617–1623. doi:10.1002/eji.1830250622

187. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, Deswal A, Drazner MH, Dunlay SM, Evers LR, Fang JC, Fedson SE, Fonarow GC, Hayek SS, Hernandez AF, Khazanie P, Kittleson MM, Lee CS, Link MS, Milano CA, Nnacheta LC, Sandhu AT, Stevenson LW, Vardeny O, Vest AR, Yancy CW. 2022. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* **145**:e895–e1032. doi:10.1161/CIR.0000000000001063

188. Heidt T, Courties G, Dutta P, Sager HB, Sebas M, Iwamoto Y, Sun Y, Da Silva N, Panizzi P, van der Laan AM, Swirski FK, Weissleder R, Nahrendorf M. 2014. Differential Contribution of Monocytes to Heart Macrophages in Steady-State and After Myocardial Infarction. *Circulation Research* **115**:284–295. doi:10.1161/CIRCRESAHA.115.303567

189. Hempel WM, Stanhope-Baker P, Mathieu N, Huang F, Schlissel MS, Ferrier P. 1998. Enhancer control of V(D)J recombination at the TCRbeta locus: differential effects on DNA cleavage and joining. *Genes Dev* **12**:2305–2317. doi:10.1101/gad.12.15.2305

190. Higdon LE, Schaffert S, Huang H, Montez-Rath ME, Lucia M, Jha A, Saligrama N, Margulies KB, Martinez OM, Davis MM, Khatri P, Maltzman JS. 2021. Evolution of cytomegalovirus-responsive T cell clonality following solid-

organ transplantation. *J Immunol* **207**:2077–2085. doi:10.4049/jimmunol.2100404

191. Hirokawa S, Chure G, Belliveau NM, Lovely GA, Anaya M, Schatz DG, Baltimore D, Phillips R. 2020. Sequence-dependent dynamics of synthetic and endogenous RSSs in V(D)J recombination. *Nucleic Acids Research* **48**:6726–6739. doi:10.1093/nar/gkaa418

192. Hoffmann J, Fiser K, Weaver J, Dimmick I, Loeher M, Pircher H, Martin-Ruiz C, Veerasamy M, Keavney B, von Zglinicki T, Spyridopoulos I. 2012. High-Throughput 13-Parameter Immunophenotyping Identifies Shifts in the Circulating T-Cell Compartment Following Reperfusion in Patients with Acute Myocardial Infarction. *PLoS ONE* **7**:e47155. doi:10.1371/journal.pone.0047155

193. Hofmann U, Beyersdorf N, Weirather J, Podolskaya A, Bauersachs J, Ertl G, Kerkau T, Frantz S. 2012. Activation of CD4+ T lymphocytes improves wound healing and survival after experimental myocardial infarction in mice. *Circulation* **125**:1652–1663. doi:10.1161/CIRCULATIONAHA.111.044164

194. Hofmann U, Frantz S. 2015. Role of Lymphocytes in Myocardial Injury, Healing, and Remodeling After Myocardial Infarction. *Circulation Research* **116**:354–367. doi:10.1161/CIRCRESAHA.116.304072

195. Holst A, Jensen G, Prescott E. 2010. Risk factors for venous thromboembolism: results from the Copenhagen City Heart Study. *Circulation* **121**. doi:10.1161/CIRCULATIONAHA.109.921460

196. Hong C-H, Pyo H-S, Baek I-C, Kim T-G. 2022. Rapid identification of CMV-specific TCRs via reverse TCR cloning system based on bulk TCR repertoire data. *Front Immunol* **13**:1021067. doi:10.3389/fimmu.2022.1021067

197. Horváth R, Cerný J, Benedík J, Hökl J, Jelínková I, Benedík J. 2000. The possible role of human cytomegalovirus (HCMV) in the origin of atherosclerosis. *J Clin Virol* **16**:17–24. doi:10.1016/s1386-6532(99)00064-5

198. Horwitz MS, Bradley LM, Harbertson J, Krahl T, Lee J, Sarvetnick N. 1998. Diabetes induced by Coxsackie virus: initiation by bystander damage and not molecular mimicry. *Nat Med* **4**:781–785. doi:10.1038/nm0798-781

199. Hou X, Chen W, Zhang X, Wang G, Chen J, Zeng P, Fu X, Zhang Q, Liu X, Diao H. 2020. Preselection TCR repertoire predicts CD4+ and CD8+ T-cell differentiation state. *Immunology* **161**:354–363. doi:10.1111/imm.13256

200. Howie B, Sherwood AM, Berkebile AD, Berka J, Emerson RO, Williamson DW, Kirsch I, Vignali M, Rieder MJ, Carlson CS, Robins HS. 2015. High-throughput pairing of T cell receptor α and β sequences. *Sci Transl Med* **7**. doi:10.1126/scitranslmed.aac5624

201. Hsieh C-S, Macatonia SE, Tripp CS, Wolf SF, O'Garra A, Murphy KM. 1993. Development of TH1 CD4+ T Cells Through IL-12 Produced by Listeria-Induced Macrophages. *Science* **260**:547–549. doi:10.1126/science.8097338

202. Hu J, Zhang Y, Zhao L, Frock RL, Du Z, Meyers RM, Meng F, Schatz DG, Alt FW. 2015. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell* **163**:947–959. doi:10.1016/j.cell.2015.10.016

203. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. 2020. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nature Biotechnology* **38**:1194–1202. doi:10.1038/s41587-020-0505-4

204. Huang Y-N, Patel NA, Mehta JH, Ginjala S, Brodin P, Gray CM, Patel YM, Cowell LG, Burkhardt AM, Mangul S. 2022. Data Availability of Open T-Cell Receptor Repertoire Data, a Systematic Assessment. *Frontiers in Systems Biology* **2**.

205. Huang Y-N, Vahed M, Peng K, Alachkar H, Mangul S. 2023. Response to 'comment on rigorous benchmarking of T cell receptor repertoire profiling methods for cancer RNA sequencing' by Davydov A.N.; Bolotin D.A.; Poslavsky S. V. and Chudakov D.M. *Briefings in Bioinformatics* **24**:bbad355. doi:10.1093/bib/bbad355

206. Hudson D, Fernandes RA, Basham M, Ogg G, Koohy H. 2023. Can we predict T cell specificity with digital biology and machine learning? *Nat Rev Immunol* **23**:511–521. doi:10.1038/s41577-023-00835-3

207. Hue S, Ahern P, Buonocore S, Kullberg MC, Cua DJ, McKenzie BS, Powrie F, Maloy KJ. 2006. Interleukin-23 drives innate and T cell–mediated intestinal inflammation. *The Journal of Experimental Medicine* **203**:2473–2483. doi:10.1084/jem.20061099

208. Huseby ES, Crawford F, White J, Kappler J, Marrack P. 2003. Negative selection imparts peptide specificity to the mature T cell repertoire. *Proceedings of the National Academy of Sciences* **100**:11565–11570. doi:10.1073/pnas.1934636100

209. Huter EN, Punkosdy GA, Glass DD, Cheng LI, Ward JM, Shevach EM. 2008. TGF-beta-induced Foxp3+ regulatory T cells rescue scurfy mice. *Eur J Immunol* **38**:1814–1821. doi:10.1002/eji.200838346

210. Hwang ES, Hong J-H, Glimcher LH. 2005. IL-2 production in developing Th1 cells is regulated by heterodimerization of RelA and T-bet and requires T-bet serine residue 508. *J Exp Med* **202**:1289–1300. doi:10.1084/jem.20051044

211. Irving BA, Weiss A. 1991. The cytoplasmic domain of the T cell receptor ζ chain is sufficient to couple to receptor-associated signal transduction pathways. *Cell* **64**:891–901. doi:10.1016/0092-8674(91)90314-O

212. Jabagi M-J, Bertrand M, Botton J, Le Vu S, Weill A, Dray-Spira R, Zureik M. 2023. Stroke, Myocardial Infarction, and Pulmonary Embolism after Bivalent Booster. *N Engl J Med* **388**:1431–1432. doi:10.1056/NEJMc2302134

213. Jaccard P. 1901. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. Rouge.

214. Jackson DA. 1997. Compositional Data in Community Ecology: The Paradigm or Peril of Proportions? *Ecology* **78**:929–940. doi:10.1890/0012-9658(1997)078[0929:CDICET]2.0.CO;2

215. Jenkins MK, Chu HH, McLachlan JB, Moon JJ. 2010. On the Composition of the Preimmune Repertoire of T Cells Specific for Peptide–Major Histocompatibility Complex Ligands. *Annu Rev Immunol* **28**:275–294. doi:10.1146/annurev-immunol-030409-101253

216. Jenkins MK, Schwartz RH. 1987. Antigen presentation by chemically modified splenocytes induces antigen-specific T cell unresponsiveness in vitro and in vivo. *J Exp Med* **165**:302–319. doi:10.1084/jem.165.2.302

217. Jenkinson EJ, Kingston R, Owen JJ. 1990. Newly generated thymocytes are not refractory to deletion when the alpha/beta component of the T cell receptor is engaged by the superantigen staphylococcal enterotoxin B. *Eur J Immunol* **20**:2517–2520. doi:10.1002/eji.1830201125

218. Jokinen E, Dumitrescu A, Huuhtanen J, Gligorijević V, Mustjoki S, Bonneau R, Heinonen M, Lähdesmäki H. 2022. TCRconv: predicting recognition between T cell receptors and epitopes using contextualized motifs. *Bioinformatics* btac788. doi:10.1093/bioinformatics/btac788

219. Jolicoeur C, Hanahan D, Smith KM. 1994. T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. *Proc Natl Acad Sci U S A* **91**:6707–6711. doi:10.1073/pnas.91.14.6707

220. Jonasson L, Holm J, Skalli O, Bondjers G, Hansson G. 1986. Regional accumulations of T cells, macrophages, and smooth muscle cells in the human atherosclerotic plaque. *Arteriosclerosis (Dallas, Tex)* **6**. doi:10.1161/01.atv.6.2.131

221. Josefowicz SZ, Niec RE, Kim HY, Treuting P, Chinen T, Zheng Y, Umetsu DT, Rudensky AY. 2012. Extrathymically generated regulatory T cells control mucosal TH2 inflammation. *Nature* **482**:395–399. doi:10.1038/nature10772

222. Juzenas S, Kiseliovas V, Goda K, Zvirblyte J, Quintinal-Villalonga A, Nainys J, Mazutis L. 2023. inDrops-2: a flexible, versatile and cost-efficient droplet microfluidics approach for high-throughput scRNA-seq of fresh and preserved clinical samples. doi:10.1101/2023.09.26.559493

223. Kägi D, Ledermann B, Bürki K, Seiler P, Odermatt B, Olsen KJ, Podack ER, Zinkernagel RM, Hengartner H. 1994. Cytotoxicity mediated by T cells and natural killer cells is greatly impaired in perforin-deficient mice. *Nature* **369**:31–37. doi:10.1038/369031a0

224. Kakaradov B, Arsenio J, Widjaja CE, He Z, Aigner S, Metz PJ, Yu B, Wehrens EJ, Lopez J, Kim SH, Zuniga EI, Goldrath AW, Chang JT, Yeo GW. 2017. Early transcriptional and epigenetic regulation of CD8+ T cell differentiation revealed by single-cell RNA sequencing. *Nat Immunol* **18**:422–432. doi:10.1038/ni.3688

225. Kansy BA, Shayan G, Jie H-B, Gibson SP, Lei YL, Brandau S, Lang S, Schmitt NC, Ding F, Lin Y, Ferris RL. 2018. T cell receptor richness in peripheral blood increases after cetuximab therapy and correlates with therapeutic response. *Oncoimmunology* **7**:e1494112. doi:10.1080/2162402X.2018.1494112

226. Kappler J, Kubo R, Haskins K, White J, Marrack P. 1983. The mouse T cell receptor: Comparison of MHC-restricted receptors on two T cell hybridomas. *Cell* **34**:727–737. doi:10.1016/0092-8674(83)90529-9

227. Kappler JW, Roehm N, Marrack P. 1987. T cell tolerance by clonal elimination in the thymus. *Cell* **49**:273–280. doi:10.1016/0092-8674(87)90568-X

228. Kappler JW, Staerz U, White J, Marrack PC. 1988. Self-tolerance eliminates T cells specific for Mls-modified products of the major histocompatibility complex. *Nature* **332**:35–40. doi:10.1038/332035a0

229. Keşmir C, Borghans JAM, de Boer RJ. 2000. Diversity of Human αβ T Cell Receptors. *Science* **288**:1135–1135. doi:10.1126/science.288.5469.1135a

230. Khan Z, Hammer C, Guardino E, Chandler GS, Albert ML. 2019. Mechanisms of immune-related adverse events associated with immune checkpoint blockade: using germline genetics to develop a personalized approach. *Genome Med* **11**:1–3. doi:10.1186/s13073-019-0652-8

231. Kim J, Chang D-Y, Lee HW, Lee H, Kim JH, Sung PS, Kim KH, Hong S-H, Kang W, Lee J, Shin SY, Yu HT, You S, Choi YS, Oh I, Lee Dong Ho, Lee Dong Hyeon, Jung MK, Suh K-S, Hwang S, Kim W, Park S-H, Kim HJ, Shin E-C. 2018. Innate-like Cytotoxic Function of Bystander-Activated CD8+ T Cells Is Associated with Liver Injury in Acute Hepatitis A. *Immunity* **48**:161-173.e5. doi:10.1016/j.immuni.2017.11.025

232. Kim J-H, Jin H-O, Park J-A, Chang YH, Hong YJ, Lee JK. 2014. Comparison of three different kits for extraction of high-quality RNA from frozen blood. *Springerplus* **3**:76. doi:10.1186/2193-1801-3-76

233. Kirabo A, Fontana V, de Faria A, R L, Galindo C, Wu J, Bikineyeva A, Dikalov S, Xiao L, Chen W, Saleh M, Trott D, Itani H, Vinh A, Amarnath V,

Amarnath K, Guzik T, Bernstein K, Shen X, Shyr Y, Chen S, Mernaugh R, Laffer C, Elijovich F, Davies S, Moreno H, Madhur M, Roberts J, Harrison D. 2014. DC isoketal-modified proteins activate T cells and promote hypertension. *The Journal of clinical investigation* **124**. doi:10.1172/JCI74084

234. Kishi JY, Liu N, West ER, Sheng K, Jordanides JJ, Serrata M, Cepko CL, Saka SK, Yin P. 2022. Light-Seq: light-directed in situ barcoding of biomolecules in fixed cells and tissues for spatially indexed sequencing. *Nat Methods* **19**:1393–1402. doi:10.1038/s41592-022-01604-1

235. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J. 2012. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* **9**:72–74. doi:10.1038/nmeth.1778

236. Klarin D, Lynch J, Aragam K, Chaffin M, Assimes TL, Huang J, Lee KM, Shao Q, Huffman JE, Natarajan P, Arya S, Small A, Sun YV, Vujkovic M, Freiberg MS, Wang L, Chen J, Saleheen D, Lee JS, Miller DR, Reaven P, Alba PR, Patterson OV, DuVall SL, Boden WE, Beckman JA, Gaziano JM, Concato J, Rader DJ, Cho K, Chang K-M, Wilson PWF, O'Donnell CJ, Kathiresan S, Tsao PS, Damrauer SM. 2019. Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat Med* **25**:1274–1279. doi:10.1038/s41591-019-0492-5

237. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. 2015. Droplet barcoding for single cell transcriptomics applied to embryonic stem cells. *Cell* **161**:1187–1201. doi:10.1016/j.cell.2015.04.044

238. Komatsu N, Mariotti-Ferrandiz M, Wang Y, Malissen B, Waldmann H, S H. 2009. Heterogeneity of natural Foxp3+ T cells: a committed regulatory T-cell lineage and an uncommitted minor population retaining plasticity. *Proceedings of the National Academy of Sciences of the United States of America* **106**. doi:10.1073/pnas.0811556106

239. Kondo M, Weissman IL, Akashi K. 1997. Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow. *Cell* **91**:661–672. doi:10.1016/S0092-8674(00)80453-5

240. Kowalewski DJ, Schuster H, Backert L, Berlin C, Kahn S, Kanz L, Salih HR, Rammensee H-G, Stevanovic S, Stickel JS. 2015. HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL). *Proceedings of the National Academy of Sciences of the United States of America* **112**:E166. doi:10.1073/pnas.1416389112

241. Krebs P, Kurrer MO, Kremer M, De Giuli R, Sonderegger I, Henke A, Maier R, Ludewig B. 2007. Molecular mapping of autoimmune B cell responses

in experimental myocarditis. *Journal of Autoimmunity* **28**:224–233. doi:10.1016/j.jaut.2007.01.003

242. Kretschmer K, Apostolou I, Hawiger D, Khazaie K, Nussenzweig MC, von Boehmer H. 2005. Inducing and expanding regulatory T cell populations by foreign antigen. *Nat Immunol* **6**:1219–1227. doi:10.1038/ni1265

243. Krummel MF. 1995. CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *Journal of Experimental Medicine* **182**:459–465. doi:10.1084/jem.182.2.459

244. Kumar PA, Hu Y, Yamamoto Y, Hoe NB, Wei TS, Mu D, Sun Y, Joo LS, Dagher R, Zielonka EM, Wang DY, Lim B, Chow VT, Crum CP, Xian W, McKeon F. 2011. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. *Cell* **147**:525–538. doi:10.1016/j.cell.2011.10.001

245. Kumar PG, Laloraya M, Wang CY, Ruan QG, Davoodi-Semiromi A, Kao KJ, She JX. 2001. The autoimmune regulator (AIRE) is a DNA-binding protein. *J Biol Chem* **276**:41357–41364. doi:10.1074/jbc.M104898200

246. Kuo C, Shor A, Campbell L, Fukushi H, Patton D, Grayston J. 1993. Demonstration of Chlamydia pneumoniae in atherosclerotic lesions of coronary arteries. *The Journal of infectious diseases* **167**. doi:10.1093/infdis/167.4.841

247. Kurosawa Y, von Boehmer H, Haas W, Sakano H, Trauneker A, Tonegawa S. 1981. Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature* **290**:565–570. doi:10.1038/290565a0

248. Lagattuta KA, Kang JB, Nathan A, Pauken KE, Jonsson AH, Rao DA, Sharpe AH, Ishigaki K, Raychaudhuri S. 2022. Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. *Nat Immunol* **23**:446–457. doi:10.1038/s41590-022-01129-x

249. Langrish CL, Chen Y, Blumenschein WM, Mattson J, Basham B, Sedgwick JD, McClanahan T, Kastelein RA, Cua DJ. 2005. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *Journal of Experimental Medicine* **201**:233–240. doi:10.1084/jem.20041257

250. Larijani M, Cc Y, R G, Ql L, Ge W. 1999. The role of components of recombination signal sequences in immunoglobulin gene segment usage: a V81x model. *Nucleic acids research* **27**. doi:10.1093/nar/27.11.2304

251. Laroumanie F, Douin-Echinard V, Pozzo J, Lairez O, Tortosa F, Vinel C, Delage C, Calise D, Dutaur M, Parini A, Pizzinat N. 2014. CD4+ T Cells Promote the Transition From Hypertrophy to Heart Failure During Chronic Pressure Overload. *Circulation* **129**:2111–2124. doi:10.1161/CIRCULATIONAHA.113.007101

252. Lathrop SK, Huddleston CA, Dullforce PA, Montfort MJ, Weinberg AD, Parker DC. 2004. A Signal through OX40 (CD134) Allows Anergic, Autoreactive T Cells to Acquire Effector Cell Functions1. *The Journal of Immunology* **172**:6735–6743. doi:10.4049/jimmunol.172.11.6735

253. Lathrop SK, Santacruz NA, Pham D, Luo J, Hsieh C-S. 2008. Antigen-specific peripheral shaping of the natural regulatory T cell population. *J Exp Med* **205**:3105–3117. doi:10.1084/jem.20081359

254. Laydon DJ, Bangham CRM, Asquith B. 2015. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. *Phil Trans R Soc B* **370**:20140291. doi:10.1098/rstb.2014.0291

255. Lebedeva A, Maryukhnich E, Grivel J-C, Vasilieva E, Margolis L, Shpektor A. 2020. Productive cytomegalovirus infection is associated with impaired endothelial function in ST-elevation myocardial infarction. *Am J Med* **133**:133–142. doi:10.1016/j.amjmed.2019.06.021

256. Lee AI, Fugmann SD, Cowell LG, Ptaszek LM, Kelsoe G, Schatz DG. 2003. A Functional Analysis of the Spacer of V(D)J Recombination Signal Sequences. *PLoS Biology* **1**. doi:10.1371/journal.pbio.0000001

257. Lees WD, Christley S, Peres A, Kos JT, Corrie B, Ralph D, Breden F, Cowell LG, Yaari G, Corcoran M, Karlsson Hedestam GB, Ohlin M, Collins AM, Watson CT, Busse CE. 2023. AIRR community curation and standardised representation for immunoglobulin and T cell receptor germline sets. *Immunoinformatics (Amst)* **10**:100025. doi:10.1016/j.immuno.2023.100025

258. Lefranc M, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G. 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental and comparative immunology* **27**. doi:10.1016/s0145-305x(02)00039-3

259. Lefranc M-P, Giudicelli V, Ginestoux C, Bodmer J, Müller W, Bontrop R, Lemaitre M, Malik A, Barbié V, Chaume D. 1999. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Research* **27**:209–212. doi:10.1093/nar/27.1.209

260. Lefranc M-P, Lefranc G. 2001. The T cell receptor factsbook, Factsbook series. San Diego, Calif.: Academic Press.

261. Li J, Liang C, Yang KY, Huang X, Han MY, Li X, Chan VW, Chan KS, Liu D, Huang Z-P, Zhou B, Lui KO. 2020. Specific ablation of CD4 $^{+}$ T-cells promotes heart regeneration in juvenile mice. *Theranostics* **10**:8018–8035. doi:10.7150/thno.42943

262. Li J, Xia N, Li D, Wen S, Qian S, Lu Y, Gu M, Tang T, Jiao J, Lv B, Nie S, Hu D, Liao Y, Yang X, Shi G, Cheng X. 2022. Aorta Regulatory T Cells with

a Tissue-Specific Phenotype and Function Promote Tissue Repair through Tff1 in Abdominal Aortic Aneurysms. *Advanced Science* **9**. doi:10.1002/advs.202104338

263. Li L, Bowling S, McGeary SE, Yu Q, Lemke B, Alcedo K, Jia Y, Liu X, Ferreira M, Klein AM, Wang S-W, Camargo FD. 2023. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* S0092867423010401. doi:10.1016/j.cell.2023.09.019

264. Liao Y, Smyth GK, Shi W. 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* **41**:e108. doi:10.1093/nar/gkt214

265. Libby P, Egan D, Skarlatos S. 1997. Roles of Infectious Agents in Atherosclerosis and Restenosis1. *Circulation* **96**:4095–4103. doi:10.1161/01.CIR.96.11.4095

266. Lieber MR. 1991. Site-specific recombination in the immune system. *The FASEB Journal* **5**:2934–2944. doi:10.1096/fasebj.5.14.1752360

267. Lima G, Treviño-Tello F, Atisha-Fregoso Y, Llorente L, Fragoso-Loyo H, Jakez-Ocampo J. 2021. Exhausted T cells in systemic lupus erythematosus patients in long-standing remission. *Clin Exp Immunol* **204**:285–295. doi:10.1111/cei.13577

268. Lindner D, Fitzek A, Bräuninger H, Aleshcheva G, Edler C, Meissner K, Scherschel K, Kirchhof P, Escher F, Schultheiss H-P, Blankenberg S, Püschel K, Westermann D. 2020. Association of Cardiac Infection With SARS-CoV-2 in Confirmed COVID-19 Autopsy Cases. *JAMA Cardiology* **5**:1281–1285. doi:10.1001/jamacardio.2020.3551

269. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, Kanda M, Polanski K, Heinig M, Lee M, Nadelmann ER, Roberts K, Tuck L, Fasouli ES, DeLaughter DM, McDonough B, Wakimoto H, Gorham JM, Samari S, Mahbubani KT, Saeb-Parsy K, Patone G, Boyle JJ, Zhang Hongbo, Zhang Hao, Viveiros A, Oudit GY, Bayraktar OA, Seidman JG, Seidman CE, Noseda M, Hubner N, Teichmann SA. 2020. Cells of the adult human heart. *Nature* **588**:466–472. doi:10.1038/s41586-020-2797-4

270. Liu S, Iorgulescu JB, Li S, Borji M, Barrera-Lopez IA, Shanmugam V, Lyu H, Morriss JW, Garcia ZN, Murray E, Reardon DA, Yoon CH, Braun DA, Livak KJ, Wu CJ, Chen F. 2022. Spatial maps of T cell receptors and transcriptomes reveal distinct immune niches and interactions in the adaptive immune response. *Immunity* **55**:1940-1952.e5. doi:10.1016/j.immuni.2022.09.002

271. Liu Sudong, Zhong Z, Zhong W, Weng R, Liu J, Gu X, Chen Y. 2020. Comprehensive analysis of T-cell receptor repertoire in patients with acute

coronary syndrome by high-throughput sequencing. *BMC Cardiovascular Disorders* **20**:253. doi:10.1186/s12872-020-01538-6

272. Liu Shaojun, Chen J, Shi J, Zhou W, Wang L, Fang W, Zhong Y, Chen X, Chen Y, Sabri A, Liu Shiming. 2020. M1-like macrophage-derived exosomes suppress angiogenesis and exacerbate cardiac dysfunction in a myocardial infarction microenvironment. *Basic Res Cardiol* **115**:22. doi:10.1007/s00395-020-0781-7

273. Liu W, Zhang Q, Chen J, Xiang R, Song H, Shu S, Chen L, Liang L, Zhou J, You L, Wu P, Zhang B, Lu Y, Xia L, Huang L, Yang Y, Liu F, Semple MG, Cowling BJ, Lan K, Sun Z, Yu H, Liu Y. 2020. Detection of Covid-19 in Children in Early January 2020 in Wuhan, China. *New England Journal of Medicine* **382**:1370–1371. doi:10.1056/NEJMc2003717

274. Liu Y-Y, Yang Q-F, Yang J-S, Cao R-B, Liang J-Y, Liu Y-T, Zeng Y-L, Chen S, Xia X-F, Zhang K, Liu L. 2019. Characteristics and prognostic significance of profiling the peripheral blood T-cell receptor repertoire in patients with advanced lung cancer. *International Journal of Cancer* **145**:1423–1431. doi:10.1002/ijc.32145

275. Loh EY, Elliott JF, Cwirla S, Lanier LL, Davis MM. 1989. Polymerase Chain Reaction with Single-Sided Specificity: Analysis of T Cell Receptor δ Chain. *Science* **243**:217–220. doi:10.1126/science.2463672

276. Lombardi V, Speak AO, Kerzerho J, Szely N, Akbari O. 2012. $CD8\alpha^{+}\beta^{-}$ and $CD8\alpha^{+}\beta^{+}$ plasmacytoid dendritic cells induce $Foxp3^{+}$ regulatory T cells and prevent the induction of airway hyper-reactivity. *Mucosal Immunol* **5**:432–443. doi:10.1038/mi.2012.20

277. Lopaschuk GD, Verma S. 2020. Mechanisms of Cardiovascular Benefits of Sodium Glucose Co-Transporter 2 (SGLT2) Inhibitors: A State-of-the-Art Review. *JACC Basic Transl Sci* **5**:632–644. doi:10.1016/j.jacbts.2020.02.004

278. Lu J, Van Laethem F, Bhattacharya A, Craveiro M, Saba I, Chu J, Love NC, Tikhonova A, Radaev S, Sun X, Ko A, Arnon T, Shifrut E, Friedman N, Weng N-P, Singer A, Sun PD. 2019. Molecular constraints on CDR3 for thymic selection of MHC-restricted TCRs from a random pre-selection repertoire. *Nat Commun* **10**:1019. doi:10.1038/s41467-019-08906-7

279. Lu X, Zhang L, Du H, Zhang J, Li YY, Qu J, Zhang W, Wang Y, Bao S, Li Y, Wu C, Liu H, Liu D, Shao J, Peng X, Yang Y, Liu Z, Xiang Y, Zhang F, Silva RM, Pinkerton KE, Shen K, Xiao H, Xu S, Wong GWK, Chinese Pediatric Novel Coronavirus Study Team. 2020. SARS-CoV-2 Infection in Children. *N Engl J Med* **382**:1663–1665. doi:10.1056/NEJMc2005073

280. Lundquist PM, Zhong CF, Zhao P, Tomaney AB, Peluso PS, Dixon J, Bettman B, Lacroix Y, Kwo DP, McCullough E, Maxham M, Hester K, McNitt P, Grey DM, Henriquez C, Foquet M, Turner SW, Zaccarin D. 2008. Parallel

confocal detection of single molecules in real time. *Opt Lett, OL* **33**:1026–1028. doi:10.1364/OL.33.001026

281. Luque Otero M, Martell Claros N, Llorente Pérez L, Fernández Pinilla C, Fernández-Cruz A. 1983. Severe hypertension in the Spanish population. Association with specific HLA antigens. *Hypertension* **5**:V149. doi:10.1161/01.HYP.5.6_Pt_3.V149

282. Lythe G, Callard RE, Hoare RL, Molina-París C. 2016. How many TCR clonotypes does a body maintain? *Journal of Theoretical Biology* **389**:214. doi:10.1016/j.jtbi.2015.10.016

283. Madi A, Poran A, Shifrut E, Reich-Zeliger S, Greenstein E, Zaretsky I, Arnon T, Laethem FV, Singer A, Lu J, Sun PD, Cohen IR, Friedman N. 2017. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife* **6**. doi:10.7554/eLife.22057

284. Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, Chain B, Cohen IR, Friedman N. 2014. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res* **24**:1603–1612. doi:10.1101/gr.170753.113

285. Malchow S, Leventhal DS, Lee V, Nishi S, Socci ND, Savage PA. 2016. Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage. *Immunity* **44**:1102–1113. doi:10.1016/j.immuni.2016.02.009

286. Malchow S, Leventhal DS, Nishi S, Fischer BI, Shen L, Paner GP, Amit AS, Kang C, Geddes JE, Allison JP, Socci ND, Savage PA. 2013. Aire-Dependent Thymic Development of Tumor-Associated Regulatory T Cells. *Science* **339**:1219–1224. doi:10.1126/science.1233913

287. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, Rutten-Jacobs L, Giese A-K, van der Laan SW, Gretarsdottir S, Anderson CD, Chong M, Adams HHH, Ago T, Almgren P, Amouyel P, Ay H, Bartz TM, Benavente OR, Bevan S, Boncoraglio GB, Brown RD, Butterworth AS, Carrera C, Carty CL, Chasman DI, Chen W-M, Cole JW, Correa A, Cotlarciuc I, Cruchaga C, Danesh J, de Bakker PIW, DeStefano AL, den Hoed M, Duan Q, Engelter ST, Falcone GJ, Gottesman RF, Grewal RP, Gudnason V, Gustafsson S, Haessler J, Harris TB, Hassan A, Havulinna AS, Heckbert SR, Holliday EG, Howard G, Hsu F-C, Hyacinth HI, Ikram MA, Ingelsson E, Irvin MR, Jian X, Jiménez-Conde J, Johnson JA, Jukema JW, Kanai M, Keene KL, Kissela BM, Kleindorfer DO, Kooperberg C, Kubo M, Lange LA, Langefeld CD, Langenberg C, Launer LJ, Lee J-M, Lemmens R, Leys D, Lewis CM, Lin W-Y, Lindgren AG, Lorentzen E, Magnusson PK, Maguire J, Manichaikul A, McArdle PF, Meschia JF, Mitchell BD, Mosley TH, Nalls MA, Ninomiya T, O'Donnell MJ, Psaty BM, Pulit SL, Rannikmäe K, Reiner AP, Rexrode KM, Rice K, Rich SS, Ridker PM, Rost NS, Rothwell PM, Rotter JI, Rundek T, Sacco RL, Sakaue S, Sale MM,

Salomaa V, Sapkota BR, Schmidt R, Schmidt CO, Schminke U, Sharma P, Slowik A, Sudlow CLM, Tanislav C, Tatlisumak T, Taylor KD, Thijs VNS, Thorleifsson G, Thorsteinsdottir U, Tiedt S, Trompet S, Tzourio C, van Duijn CM, Walters M, Wareham NJ, Wassertheil-Smoller S, Wilson JG, Wiggins KL, Yang Q, Yusuf S, Bis JC, Pastinen T, Ruusalepp A, Schadt EE, Koplev S, Björkegren JLM, Codoni V, Civelek M, Smith NL, Trégouët DA, Christophersen IE, Roselli C, Lubitz SA, Ellinor PT, Tai ES, Kooner JS, Kato N, He J, van der Harst P, Elliott P, Chambers JC, Takeuchi F, Johnson AD, Sanghera DK, Melander O, Jern C, Strbian D, Fernandez-Cadenas I, Longstreth WT, Rolfs A, Hata J, Woo D, Rosand J, Pare G, Hopewell JC, Saleheen D, Stefansson K, Worrall BB, Kittner SJ, Seshadri S, Fornage M, Markus HS, Howson JMM, Kamatani Y, Debette S, Dichgans M. 2018. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**:524–537. doi:10.1038/s41588-018-0058-3

288. Malissen M, Trucy J, Jouvin-Marche E, Cazenave PA, Scollay R, Malissen B. 1992. Regulation of TCR alpha and beta gene allelic exclusion during T-cell development. *Immunol Today* **13**:315–322. doi:10.1016/0167-5699(92)90044-8

289. Mamula MJ, Lin RH, Janeway CA, Hardin JA. 1992. Breaking T cell tolerance with foreign and self co-immunogens. A study of autoimmune B and T cell epitopes of cytochrome c. *J Immunol* **149**:789–795.

290. Mandric I, Rotman J, Yang HT, Strauli N, Montoya DJ, Van Der Wey W, Ronas JR, Statz B, Yao D, Petrova V, Zelikovsky A, Spreafico R, Shifman S, Zaitlen N, Rossetti M, Ansel KM, Eskin E, Mangul S. 2020. Profiling immunoglobulin repertoires across multiple human tissues using RNA sequencing. *Nat Commun* **11**:3126. doi:10.1038/s41467-020-16857-7

291. Manjula BN, Trus BL, Fischetti VA. 1985. Presence of two distinct regions in the coiled-coil structure of the streptococcal Pep M5 protein: relationship to mammalian coiled-coil proteins and implications to its biological properties. *Proceedings of the National Academy of Sciences* **82**:1064–1068. doi:10.1073/pnas.82.4.1064

292. Manolios N, Bonifacino JS, Klausner RD. 1990. Transmembrane Helical Interactions and the Assembly of the T Cell Receptor Complex. *Science* **249**:274–277. doi:10.1126/science.2142801

293. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology* **14**:e1005944. doi:10.1371/journal.pcbi.1005944

294. Marcou Q, Mora T, Walczak AM. 2018. High-throughput immune repertoire analysis with IGoR. *Nat Commun* **9**:561. doi:10.1038/s41467-018-02832-w

295. Marie-Cardine A, Schraven B. 1999. Coupling the TCR to Downstream Signalling Pathways: The Role of Cytoplasmic and Transmembrane Adaptor Proteins. *Cellular Signalling* **11**:705–712. doi:10.1016/S0898-6568(99)00047-9

296. Mariotti-Ferrandiz E, Pham H-P, Dulauroy S, Gorgette O, Klatzmann D, Cazenave P-A, Pied S, Six A. 2016. A TCRβ Repertoire Signature Can Predict Experimental Cerebral Malaria. *PLOS ONE* **11**:e0147871. doi:10.1371/journal.pone.0147871

297. Marrack P, Kappler J. 1990. The Staphylococcal Enterotoxins and Their Relatives. *Science* **248**:705–711. doi:10.1126/science.2185544

298. Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, Becker A, Bennett R, Berry A, Bhai J, Bhurji SK, Bignell A, Boddu S, Branco Lins PR, Brooks L, Ramaraju SB, Charkhchi M, Cockburn A, Da Rin Fiorretto L, Davidson C, Dodiya K, Donaldson S, El Houdaigui B, El Naboulsi T, Fatima R, Giron CG, Genez T, Ghattaoraya GS, Martinez JG, Guijarro C, Hardy M, Hollis Z, Hourlier T, Hunt T, Kay M, Kaykala V, Le T, Lemos D, Marques-Coelho D, Marugán JC, Merino GA, Mirabueno LP, Mushtaq A, Hossain SN, Ogeh DN, Sakthivel MP, Parker A, Perry M, Piližota I, Prosovetskaia I, Pérez-Silva JG, Salam AIA, Saraiva-Agostinho N, Schuilenburg H, Sheppard D, Sinha S, Sipos B, Stark W, Steed E, Sukumaran R, Sumathipala D, Suner M-M, Surapaneni L, Sutinen K, Szpak M, Tricomi FF, Urbina-Gómez D, Veidenberg A, Walsh TA, Walts B, Wass E, Willhoft N, Allen J, Alvarez-Jarreta J, Chakiachvili M, Flint B, Giorgetti S, Haggerty L, Ilsley GR, Loveland JE, Moore B, Mudge JM, Tate J, Thybert D, Trevanion SJ, Winterbottom A, Frankish A, Hunt SE, Ruffier M, Cunningham F, Dyer S, Finn RD, Howe KL, Harrison PW, Yates AD, Flicek P. 2023. Ensembl 2023. *Nucleic Acids Research* **51**:D933–D941. doi:10.1093/nar/gkac958

299. Mason D. 1998. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today* **19**:395–404. doi:10.1016/S0167-5699(98)01299-7

300. Mastrokolias A, den Dunnen JT, van Ommen GB, 't Hoen PA, van Roon-Mom WM. 2012. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genomics* **13**:1–9. doi:10.1186/1471-2164-13-28

301. Matsumori A. 2005. Hepatitis C Virus Infection and Cardiomyopathies. *Circulation Research* **96**:144–147. doi:10.1161/01.RES.0000156077.54903.67

302. Maxam AM, Gilbert W. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* **65**:499–560. doi:10.1016/s0076-6879(80)65059-9

303. Mayer CT, Floess S, Baru AM, Lahl K, Huehn J, Sparwasser T. 2011. CD8+Foxp3+ T cells share developmental and phenotypic features with classical

CD4+Foxp3+ regulatory T cells but lack potent suppressive activity. *European Journal of Immunology* **41**:716–725. doi:10.1002/eji.201040913

304. Mayer-Blackwell K, Fiore-Gartland A, Thomas PG. 2022. Flexible Distance-Based TCR Analysis in Python with tcrdist3. *Methods in molecular biology (Clifton, NJ)* **2574**:309. doi:10.1007/978-1-0716-2712-9_16

305. McCaughtry TM, Baldwin TA, Wilken MS, Hogquist KA. 2008. Clonal deletion of thymocytes can occur in the cortex with no involvement of the medulla. *J Exp Med* **205**:2575–2584. doi:10.1084/jem.20080866

306. McCaughtry TM, Wilken MS, Hogquist KA. 2007. Thymic emigration revisited. *J Exp Med* **204**:2513–2520. doi:10.1084/jem.20070601

307. McCrindle BW, Rowley AH, Newburger JW, Burns JC, Bolger AF, Gewitz M, Baker AL, Jackson MA, Takahashi M, Shah PB, Kobayashi T, Wu M-H, Saji TT, Pahl E, null null. 2017. Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: A Scientific Statement for Health Professionals From the American Heart Association. *Circulation* **135**:e927–e999. doi:10.1161/CIR.0000000000000484

308. McDaniel JR, DeKosky BJ, Tanno H, Ellington AD, Georgiou G. 2016. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat Protoc* **11**:429–442. doi:10.1038/nprot.2016.024

309. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, Burri H, Butler J, Čelutkienė J, Chioncel O, Cleland JGF, Coats AJS, Crespo-Leiro MG, Farmakis D, Gilard M, Heymans S, Hoes AW, Jaarsma T, Jankowska EA, Lainscak M, Lam CSP, Lyon AR, McMurray JJV, Mebazaa A, Mindham R, Muneretto C, Francesco Piepoli M, Price S, Rosano GMC, Ruschitzka F, Kathrine Skibelund A, ESC Scientific Document Group. 2021. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* **42**:3599–3726. doi:10.1093/eurheartj/ehab368

310. McVean GA, Altshuler (Co-Chair) DM, Durbin (Co-Chair) RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT, McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs (Principal Investigator) RA, Dinh H, Kovar C, Lee S, Lewis L, Muzny D, Reid J, Wang M, Wang (Principal Investigator) J, Fang X, Guo X, Jian M, Jiang H, Jin X, Li G, Li J, Li Y, Li Zhuo, Liu X, Lu Y, Ma X, Su Z, Tai S, Tang M, Wang Bo, Wang G, Wu H, Wu R, Yin Y, Zhang W, Zhao J, Zhao M, Zheng X, Zhou Y, Lander (Principal Investigator) ES, Altshuler DM, Gabriel (Co-Chair) SB, Gupta N, Flicek (Principal Investigator) P, Clarke L, Leinonen R, Smith RE, Zheng-Bradley X, Bentley (Principal Investigator) DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach (Principal Investigator) H, Sudbrak (Project Leader) R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F, Sultan M, Timmermann

B, Yaspo M-L, Sherry (Principal Investigator) ST, McVean (Principal Investigator) GA, Mardis (Co-Principal Investigator) (Co-Chair) ER, Wilson (Co-Principal Investigator) RK, Fulton L, Fulton R, Weinstock GM, Durbin (Principal Investigator) RM, Balasubramaniam S, Burton J, Danecek P, Keane TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt (Principal Investigator) JP, Davies CJ, Gollub J, Webster T, Wong B, Zhan Y, Auton (Principal Investigator) A, Gibbs (Principal Investigator) RA, Yu (Project Leader) F, Bainbridge M, Challis D, Evani US, Lu J, Muzny D, Nagaswamy U, Reid J, Sabo A, Wang Y, Yu J, Wang (Principal Investigator) J, Coin LJM, Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Zhenyu, Lin H, Liu B, Luo R, Qin N, Shao H, Wang Bingqiang, Xie Y, Ye C, Yu C, Zhang F, Zheng H, Zhu H, Marth (Principal Investigator) GT, Garrison EP, Kural D, Lee W-P, Fung Leong W, Ward AN, Wu J, Zhang M, Lee (Principal Investigator) C, Griffin L, Hsieh C-H, Mills RE, Shi X, von Grotthuss M, Zhang C, Daly (Principal Investigator) MJ, DePristo (Project Leader) MA, Altshuler DM, Banks E, Bhatia G, Carneiro MO, del Angel G, Gabriel SB, Genovese G, Gupta N, Handsaker RE, Hartl C, Lander ES, McCarroll SA, Nemesh JC, Poplin RE, Schaffner SF, Shakir K, Yoon (Principal Investigator) SC, Lihm J, Makarov V, Jin (Principal Investigator) H, Kim W, Cheol Kim K, Korbel (Principal Investigator) JO, Rausch T, Flicek (Principal Investigator) P, Beal K, Clarke L, Cunningham F, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zheng-Bradley X, Clark (Principal Investigator) AG, Gottipati S, Keinan A, Rodriguez-Flores JL, Sabeti (Principal Investigator) PC, Grossman SR, Tabrizi S, Tariyal R, Cooper (Principal Investigator) DN, Ball EV, Stenson PD, Bentley (Principal Investigator) DR, Barnes B, Bauer M, Keira Cheetham R, Cox T, Eberle M, Humphray S, Kahn S, Murray L, Peden J, Shaw R, Ye (Principal Investigator) K, Batzer (Principal Investigator) MA, Konkel MK, Walker JA, MacArthur (Principal Investigator) DG, Lek M, Sudbrak (Project Leader), Amstislavskiy VS, Herwig R, Shriver (Principal Investigator) MD, Bustamante (Principal Investigator) CD, Byrnes JK, De La Vega FM, Gravel S, Kenny EE, Kidd JM, Lacroute P, Maples BK, Moreno-Estrada A, Zakharia F, Halperin (Principal Investigator) E, Baran Y, Craig (Principal Investigator) DW, Christoforides A, Homer N, Izatt T, Kurdoglu AA, Sinari SA, Squire K, Sherry (Principal Investigator) ST, Xiao C, Sebat (Principal Investigator) J, Bafna V, Ye K, Burchard (Principal Investigator) EG, Hernandez (Principal Investigator) RD, Gignoux CR, Haussler (Principal Investigator) D, Katzman SJ, James Kent W, Howie B, Ruiz-Linares (Principal Investigator) A, The 1000 Genomes Project Consortium, Corresponding Author, Steering committee, Production group:, Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT and Harvard, European Bioinformatics Institute, Illumina, Max Planck Institute for Molecular Genetics, US National Institutes of Health, University of Oxford, Washington University in St Louis, Wellcome Trust Sanger Institute, Analysis group:, Affymetrix, Albert Einstein College of Medicine, Boston College, Brigham and Women's Hospital, Cold Spring Harbor Laboratory, Dankook University, European Molecular Biology Laboratory, Cornell University, Harvard University, Human Gene Mutation Database, Leiden University Medical Center, Louisiana State University, Massachusetts General Hospital, Pennsylvania State University, Stanford University, Tel-Aviv

University, Translational Genomics Research Institute, University of California SD, University of California SF, University of California SC, University of Chicago, University College London, University of Geneva. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56–65. doi:10.1038/nature11632

311. Mehindate K, Thibodeau J, Dohlsten M, Kalland T, Sékaly RP, Mourad W. 1995. Cross-linking of major histocompatibility complex class II molecules by staphylococcal enterotoxin A superantigen is a requirement for inflammatory cytokine gene expression. *Journal of Experimental Medicine* **182**:1573–1577. doi:10.1084/jem.182.5.1573

312. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ, Johnson R, Segrè AV, Djebali S, Niarchou A, Consortium TGte, Wright FA, Lappalainen T, Calvo M, Getz G, Dermitzakis ET, Ardlie KG, Guigó R. 2015. The human transcriptome across tissues and individuals. *Science* **348**:660–665. doi:10.1126/science.aaa0355

313. Melnick JL, Petrie BL, Dreesman GR, Burek J, McCollum CH, DeBakey ME. 1983. Cytomegalovirus antigen within human arterial smooth muscle cells. *Lancet* **2**:644–647. doi:10.1016/s0140-6736(83)92529-1

314. Meredith M, Zemmour D, Mathis D, Benoist C. 2015. Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat Immunol* **16**:942–949. doi:10.1038/ni.3247

315. Merten M. in preparation. HLA alleles as a potential risk stratifier for chronic HF and T cell activation. *European Journal of Immunology* **53**:290. doi:10.1002/eji.202370300

316. Meuer SC, Acuto O, Hussey RE, Hodgdon JC, Fitzgerald KA, Schlossman SF, Reinherz EL. 1983. Evidence for the T3-associated 90K heterodimer as the T-cell antigen receptor. *Nature* **303**:808–810. doi:10.1038/303808a0

317. Meysman P, Barton J, Bravi B, Cohen-Lavi L, Karnaukhov V, Lilleskov E, Montemurro A, Nielsen M, Mora T, Pereira P, Postovskaya A, Martínez MR, Fernandez-de-Cossio-Diaz J, Vujkovic A, Walczak AM, Weber A, Yin R, Eugster A, Sharma V. 2023. Benchmarking solutions to the T-cell receptor epitope prediction problem: IMMREP22 workshop report. *ImmunoInformatics* **9**:100024. doi:10.1016/j.immuno.2023.100024

318. Mhanna V, Fourcade G, Barennes P, Quiniou V, Pham HP, Ritvo P-G, Brimaud F, Gouritin B, Churlaud G, Six A, Mariotti-Ferrandiz E, Klatzmann D. 2021. Impaired Activated/Memory Regulatory T Cell Clonal Expansion Instigates Diabetes in NOD Mice. *Diabetes* **70**:976–985. doi:10.2337/db20-0896

319. Miller JF. 1961a. Immunological function of the thymus. *Lancet* **2**:748–749. doi:10.1016/s0140-6736(61)90693-6

320. Miller JF. 1961b. Analysis of the Thymus Influence in Leukæmogenesis. *Nature* **191**:248–249. doi:10.1038/191248a0

321. Mingueneau M, Kreslavsky T, Gray D, Heng T, Cruse R, Ericson J, Bendall S, Spitzer MH, Nolan GP, Kobayashi K, von Boehmer H, Mathis D, Benoist C. 2013. The transcriptional landscape of αβ T cell differentiation. *Nat Immunol* **14**:619–632. doi:10.1038/ni.2590

322. Mishra R, Saha P, Datla SR, Mellacheruvu P, Gunasekaran M, Guru SA, Fu X, Chen L, Bolli R, Sharma S, Kaushal S. 2022. Transplanted allogeneic cardiac progenitor cells secrete GDF-15 and stimulate an active immune remodeling process in the ischemic myocardium. *Journal of Translational Medicine* **20**:323. doi:10.1186/s12967-022-03534-0

323. Miyakawa T. 2020. No raw data, no science: another possible source of the reproducibility crisis. *Molecular Brain* **13**:24. doi:10.1186/s13041-020-0552-2

324. Mohrs M, Holscher C, Brombacher F. 2000. Interleukin-4 receptor alpha-deficient BALB/c mice show an unimpaired T helper 2 polarization in response to Leishmania major infection. *Infect Immun* **68**:1773–1780. doi:10.1128/IAI.68.4.1773-1780.2000

325. Montagne JM, Zheng XA, Pinal-Fernandez I, Milisenda JC, Christopher-Stine L, Lloyd TE, Mammen AL, Larman HB. 2020. Ultra-efficient sequencing of T Cell receptor repertoires reveals shared responses in muscle from patients with Myositis. *EBioMedicine* **59**:102972. doi:10.1016/j.ebiom.2020.102972

326. Mora T. 2019. How many different clonotypes do immune repertoires contain? *Current Opinion in Systems Biology*.

327. Mora T, Walczak A, Bialek W, Callan C. 2010. Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences of the United States of America* **107**. doi:10.1073/pnas.1001705107

328. Mora T, Walczak AM. 2016. Quantifying lymphocyte receptor diversity. doi:10.1101/046870

329. Moreews M, Gouge KL, Khaldi-Plassart S, Pescarmona R, Mathieu A-L, Malcus C, Djebali S, Bellomo A, Dauwalder O, Perret M, Villard M, Chopin E, Rouvet I, Vandenesch F, Dupieux C, Pouyau R, Teyssedre S, Guerder M, Louazon T, Moulin-Zinsch A, Duperril M, Patural H, Giovannini-Chami L, Portefaix A, Kassai B, Venet F, Monneret G, Lombard C, Flodrops H, Guillebon J-MD, Bajolle F, Launay V, Bastard P, Zhang S-Y, Dubois V, Thaunat O, Richard J-C, Mezidi M, Allatif O, Saker K, Dreux M, Abel L, Casanova J-L, Marvel J, Trouillet-Assant S, Klatzmann D, Walzer T, Mariotti-Ferrandiz E,

Javouhey E, Belot A. 2021. Polyclonal expansion of TCR Vb 21.3+ CD4+ and CD8+ T cells is a hallmark of multisystem inflammatory syndrome in children. *Science Immunology* **6**. doi:10.1126/sciimmunol.abh1516

330. Morshead KB, Ciccone DN, Taverna SD, Allis CD, Oettinger MA. 2003. Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. *Proc Natl Acad Sci U S A* **100**:11577–11582. doi:10.1073/pnas.1932643100

331. Mosmann TR, Cherwinski H, Bond MW, Giedlin MA, Coffman RL. 1986. Two types of murine helper T cell clone. I. Definition according to profiles of lymphokine activities and secreted proteins. *The Journal of Immunology* **136**:2348–2357. doi:10.4049/jimmunol.136.7.2348

332. Mottet C, Uhlig HH, Powrie F. 2003. Cutting edge: cure of colitis by CD4+CD25+ regulatory T cells. *J Immunol* **170**:3939–3943. doi:10.4049/jimmunol.170.8.3939

333. Muhowski EM, Rogers LM. 2023. Dual TCR-Expressing T Cells in Cancer: How Single-Cell Technologies Enable New Investigation. *Immunohorizons* **7**:299–306. doi:10.4049/immunohorizons.2200062

334. Murphy KM, Heimberger AB, Loh DY. 1990. Induction by antigen of intrathymic apoptosis of CD4+CD8+TCRlo thymocytes in vivo. *Science* **250**:1720–1723. doi:10.1126/science.2125367

335. Murtagh F, Legendre P. 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* **31**:274–295. doi:10.1007/s00357-014-9161-z

336. Murugan A, Mora T, Walczak AM, Callan CG. 2012. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci USA* **109**:16161–16166. doi:10.1073/pnas.1212755109

337. Musvosvi M, Huang H, Wang C, Xia Q, Rozot V, Krishnan A, Acs P, Cheruku A, Obermoser G, Leslie A, Behar SM, Hanekom WA, Bilek N, Fisher M, Kaufmann SHE, Walzl G, Hatherill M, Davis MM, Scriba TJ. 2023. T cell receptor repertoires associated with control and disease progression following Mycobacterium tuberculosis infection. *Nat Med* **29**:258–269. doi:10.1038/s41591-022-02110-9

338. Myers JM, Cooper LT, Kem DC, Stavrakis S, Kosanke SD, Shevach EM, Fairweather D, Stoner JA, Cox CJ, Cunningham MW. 2016. Cardiac myosin-Th17 responses promote heart failure in human myocarditis. *JCI Insight* **1**. doi:10.1172/jci.insight.85851

339. Nadeu F, Royo R, Massoni-Badosa R, Playa-Albinyana H, Garcia-Torre B, Duran-Ferrer M, Dawson KJ, Kulis M, Diaz-Navarro A, Villamor N, Melero JL, Chapaprieta V, Dueso-Barroso A, Delgado J, Moia R, Ruiz-Gil S, Marchese

D, Giró A, Verdaguer-Dot N, Romo M, Clot G, Rozman M, Frigola G, Rivas-Delgado A, Baumann T, Alcoceba M, González M, Climent F, Abrisqueta P, Castellví J, Bosch F, Aymerich M, Enjuanes A, Ruiz-Gaspà S, López-Guillermo A, Jares P, Beà S, Capella-Gutierrez S, Gelpí JL, López-Bigas N, Torrents D, Campbell PJ, Gut I, Rossi D, Gaidano G, Puente XS, Garcia-Roves PM, Colomer D, Heyn H, Maura F, Martín-Subero JI, Campo E. 2022. Detection of early seeding of Richter transformation in chronic lymphocytic leukemia. *Nat Med* **28**:1662–1671. doi:10.1038/s41591-022-01927-8

340. Nägele F, Graber M, Hirsch J, Pölzl L, Sahanic S, Fiegl M, Hau D, Engler C, Lechner S, Stalder AK, Mertz KD, Haslbauer JD, Tzankov A, Grimm M, Tancevski I, Holfeld J, Gollmann-Tepeköylü C. 2022. Correlation between structural heart disease and cardiac SARS-CoV-2 manifestations. *Commun Med* **2**:1–8. doi:10.1038/s43856-022-00204-6

341. Nathan CF, Murray HW, Wiebe ME, Rubin BY. 1983. Identification of interferon-gamma as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. *J Exp Med* **158**:670–689. doi:10.1084/jem.158.3.670

342. Netea MG, Domínguez-Andrés J, Barreiro LB, Chavakis T, Divangahi M, Fuchs E, Joosten LAB, van der Meer JWM, Mhlanga MM, Mulder WJM, Riksen NP, Schlitzer A, Schultze JL, Stabell Benn C, Sun JC, Xavier RJ, Latz E. 2020. Defining trained immunity and its role in health and disease. *Nat Rev Immunol* **20**:375–388. doi:10.1038/s41577-020-0285-6

343. Nevers T, Salvador AM, Grodecki-Pena A, Knapp A, Velázquez F, Aronovitz M, Kapur NK, Karas RH, Blanton RM, Alcaide P. 2015. Left Ventricular T-Cell Recruitment Contributes to the Pathogenesis of Heart Failure. *Circulation: Heart Failure* **8**:776–787. doi:10.1161/CIRCHEARTFAILURE.115.002225

344. Nevers T, Salvador AM, Velazquez F, Ngwenyama N, Carrillo-Salinas FJ, Aronovitz M, Blanton RM, Alcaide P. 2017. Th1 effector T cells selectively orchestrate cardiac fibrosis in nonischemic heart failure. *J Exp Med* **214**:3311–3329. doi:10.1084/jem.20161791

345. Nguyen JL, Yang W, Ito K, Matte TD, Shaman J, Kinney PL. 2016. Seasonal Influenza Infections and Cardiovascular Disease Mortality. *JAMA Cardiol* **1**:274–281. doi:10.1001/jamacardio.2016.0433

346. Ngwenyama N, Kirabo A, Aronovitz M, Velázquez F, Carrillo-Salinas F, Salvador AM, Nevers T, Amarnath V, Tai A, Blanton RM, Harrison DG, Alcaide P. 2021. Isolevuglandin-Modified Cardiac Proteins Drive CD4+ T Cell Activation in the Heart and Promote Cardiac Dysfunction. *Circulation* CIRCULATIONAHA.120.051889. doi:10.1161/CIRCULATIONAHA.120.051889

347. Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang Jie, Hwang S-J, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen L-P, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han B-G, Huang Jianfeng, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki M-L, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon F-R, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim B-J, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M, the CARDIoGRAMplusC4D Consortium. 2015. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**:1121–1130. doi:10.1038/ng.3396

348. Nindl V, Maier R, Ratering D, De Giuli R, Züst R, Thiel V, Scandella E, Di Padova F, Kopf M, Rudin M, Rülicke T, Ludewig B. 2012. Cooperation of Th1 and Th17 cells determines transition from autoimmune myocarditis to dilated cardiomyopathy. *European Journal of Immunology* **42**:2311–2321. doi:10.1002/eji.201142209

349. Nistala K, Adams S, Cambrook H, Ursu S, Olivito B, de Jager W, Evans JG, Cimaz R, Bajaj-Elliott M, Wedderburn LR. 2010. Th17 plasticity in human autoimmune arthritis is driven by the inflammatory environment. *Proceedings of the National Academy of Sciences* **107**:14751–14756. doi:10.1073/pnas.1003852107

350. Nosbaum A, Prevel N, Truong H-A, Mehta P, Ettinger M, Scharschmidt TC, Ali NH, Pauli ML, Abbas AK, Rosenblum MD. 2016. Cutting Edge: Regulatory T Cells Facilitate Cutaneous Wound Healing. *The Journal of Immunology* **196**:2010–2014. doi:10.4049/jimmunol.1502139

351. Nouri N, Kleinstein SH. 2018. A spectral clustering-based method for identifying clones from high-throughput B cell repertoire sequencing data. *Bioinformatics* **34**:i341–i349. doi:10.1093/bioinformatics/bty235

352. Nyren P, Pettersson B, Uhlen M. 1993. Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay. *Analytical Biochemistry* **208**:171–175. doi:10.1006/abio.1993.1024

353. Obata F, Tsunoda M, Ito K, Ito I, Kaneko T, Pawelec G, Kashiwagi N. 1993. A single universal primer for the T-Cell receptor (TCR) variable genes enables enzymatic amplification and direct sequencing of TCRβ cDNA of various T-cell clones. *Human Immunology* **36**:163–167. doi:10.1016/0198-8859(93)90120-P

354. Oettinger MA, Schatz DG, Gorka C, Baltimore D. 1990. RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination. *Science* **248**:1517–1523. doi:10.1126/science.2360047

355. Offord C. 2023. Key study of cancer microbiomes challenged. *Science* **381**:590–591. doi:10.1126/science.adk2103

356. Ohashi PS, Oehen S, Buerki K, Pircher H, Ohashi CT, Odermatt B, Malissen B, Zinkernagel RM, Hengartner H. 1991. Ablation of "tolerance" and induction of diabetes by virus infection in viral antigen transgenic mice. *Cell* **65**:305–317. doi:10.1016/0092-8674(91)90164-t

357. Ohta-Ogo K, Sugano Y, Ogata S, Nakayama T, Komori T, Eguchi K, Dohi K, Yokokawa T, Kanamori H, Nishimura S, Nakamura K, Ikeda Y, Nishimura K, Takemura G, Anzai T, Hiroe M, Hatakeyama K, Ishibashi-Ueda H, Imanaka-Yoshida K. 2022. Myocardial T-Lymphocytes as a Prognostic Risk-Stratifying Marker of Dilated Cardiomyopathy ― Results of the Multicenter Registry to Investigate Inflammatory Cell Infiltration in Dilated Cardiomyopathy in Tissues of Endomyocardial Biopsy (INDICATE Study) ―. *Circulation Journal* **86**:1092–1101. doi:10.1253/circj.CJ-21-0529

358. Okino ST, Kong M, Sarras H, Wang Y. 2016. Evaluation of bias associated with high-multiplex, target-specific pre-amplification. *Biomolecular Detection and Quantification*, Special Issue: Advanced Molecular Diagnostics for Biomarker Discovery – Part II **6**:13–21. doi:10.1016/j.bdq.2015.12.001

359. Oldstone MB, Nerenberg M, Southern P, Price J, Lewicki H. 1991. Virus infection triggers insulin-dependent diabetes mellitus in a transgenic model: role of anti-self (virus) immune response. *Cell* **65**:319–331. doi:10.1016/0092-8674(91)90165-u

360. Olson BJ, Schattgen SA, Thomas PG, Bradley P, Iv FAM. 2022. Comparing T cell receptor repertoires using optimal transport. *PLOS Computational Biology* **18**:e1010681. doi:10.1371/journal.pcbi.1010681

361. Omer A, Peres A, Rodriguez OL, Watson CT, Lees W, Polak P, Collins AM, Yaari G. 2022. T cell receptor beta germline variability is revealed by inference from repertoire data. *Genome Med* **14**:1–19. doi:10.1186/s13073-021-01008-4

362. Onishi Y, Fehervari Z, Yamaguchi T, Sakaguchi S. 2008. Foxp3+ natural regulatory T cells preferentially form aggregates on dendritic cells in vitro and actively inhibit their maturation. *Proceedings of the National Academy of Sciences* **105**:10113–10118. doi:10.1073/pnas.0711106105

363. Onouchi Y, Tamari M, Takahashi A, Tsunoda T, Yashiro M, Nakamura Yoshikazu, Yanagawa H, Wakui K, Fukushima Y, Kawasaki T, Nakamura Yusuke, Hata A. 2007. A genomewide linkage analysis of Kawasaki disease: evidence for linkage to chromosome 12. *J Hum Genet* **52**:179–190. doi:10.1007/s10038-006-0092-3

364. Org T, Rebane A, Kisand K, Laan M, Haljasorg U, Andreson R, Peterson P. 2009. AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. *Human Molecular Genetics* **18**:4699. doi:10.1093/hmg/ddp433

365. Ortmann B, Androlewicz MJ, Cresswell P. 1994. MHC class l/β2-microglobulin complexes associate with TAP transporters before peptide binding. *Nature* **368**:864–867. doi:10.1038/368864a0

366. Ostmeyer J, Christley S, Toby IT, Cowell LG. 2019. Biophysicochemical Motifs in T-cell Receptor Sequences Distinguish Repertoires from Tumor-Infiltrating Lymphocyte and Adjacent Healthy Tissue. *Cancer Res* **79**:1671–1680. doi:10.1158/0008-5472.CAN-18-2292

367. Ostmeyer J, Lucas E, Christley S, Lea J, Monson N, Tiro J, Cowell LG. 2020. Biophysicochemical motifs in T cell receptor sequences as a potential biomarker for high-grade serous ovarian carcinoma. *PLOS ONE* **15**:e0229569. doi:10.1371/journal.pone.0229569

368. Paakkanen R, Lokki M-L, Seppänen M, Tierala I, Nieminen MS, Sinisalo J. 2012. Proinflammatory HLA-DRB1*01-haplotype predisposes to ST-elevation myocardial infarction. *Atherosclerosis* **221**:461–466. doi:10.1016/j.atherosclerosis.2012.01.024

369. Padgett LE, Dinh HQ, Wu R, Gaddis DE, Araujo DJ, Winkels H, Nguyen A, Taylor AM, McNamara CA, Hedrick CC. 2020. Naive CD8+ T Cells Expressing CD95 Increase Human Cardiovascular Disease Severity. *Arteriosclerosis, Thrombosis, and Vascular Biology* **40**:2845–2859. doi:10.1161/ATVBAHA.120.315106

370. Padovan E, Casorati G, Dellabona P, Meyer S, Brockhaus M, Lanzavecchia A. 1993. Expression of Two T Cell Receptor α Chains: Dual Receptor T Cells. *Science* **262**:422–424. doi:10.1126/science.8211163

371. Pai JA, Satpathy AT. 2021. High-throughput and single-cell T cell receptor sequencing technologies. *Nat Methods* **18**:881–892. doi:10.1038/s41592-021-01201-8

372. Pannetier C, Cochet M, Darche S, Casrouge A, Zöller M, Kourilsky P. 1993. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proceedings of the National Academy of Sciences* **90**:4319–4323. doi:10.1073/pnas.90.9.4319

373. Pantaleo G, Demarest JF, Soudeyns H, Graziosi C, Denis F, Adelsberger JW, Borrow P, Saag MS, Shaw GM, Sekalytt RP, Fauci AS. 1994. Major expansion of CD8+ T cells with a predominant Vβ usage during the primary immune response to HIV. *Nature* **370**:463–467. doi:10.1038/370463a0

374. Park H, Li Z, Yang XO, Chang SH, Nurieva R, Wang Y-H, Wang Y, Hood L, Zhu Z, Tian Q, Dong C. 2005. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* **6**:1133–1141. doi:10.1038/ni1261

375. Park J-E, Botting RA, Domínguez Conde C, Popescu D-M, Lavaert M, Kunz DJ, Goh I, Stephenson E, Ragazzini R, Tuck E, Wilbrey-Clark A, Roberts K, Kedlian VR, Ferdinand JR, He X, Webb S, Maunder D, Vandamme N, Mahbubani KT, Polanski K, Mamanova L, Bolt L, Crossland D, de Rita F, Fuller A, Filby A, Reynolds G, Dixon D, Saeb-Parsy K, Lisgo S, Henderson D, Vento-Tormo R, Bayraktar OA, Barker RA, Meyer KB, Saeys Y, Bonfanti P, Behjati S, Clatworthy MR, Taghon T, Haniffa M, Teichmann SA. 2020. A cell atlas of human thymic development defines T cell repertoire formation. *Science* **367**:eaay3224. doi:10.1126/science.aay3224

376. Pasqual N, Gallagher M, Aude-Garcia C, Loiodice M, Thuderoz F, Demongeot J, Ceredig R, Marche PN, Jouvin-Marche E. 2002. Quantitative and Qualitative Changes in V-J α Rearrangements During Mouse Thymocytes Differentiation. *J Exp Med* **196**:1163–1174. doi:10.1084/jem.20021074

377. Paulsson G, Zhou X, Törnquist E, Hansson GK. 2000. Oligoclonal T cell expansions in atherosclerotic lesions of apolipoprotein E-deficient mice. *Arterioscler Thromb Vasc Biol* **20**:10–17. doi:10.1161/01.atv.20.1.10

378. Pawlak M, Ho AW, Kuchroo VK. 2020. Cytokines and transcription factors in the differentiation of CD4+ T helper cell subsets and induction of tissue inflammation and autoimmunity. *Current Opinion in Immunology*, Autoimmunity **67**:57–67. doi:10.1016/j.coi.2020.09.001

379. Peacock T, Heather JM, Ronel T, Chain B. 2021. Decombinator V4: an improved AIRR-C compliant-software package for T-cell receptor sequence annotation? *Bioinformatics* **37**:876–878. doi:10.1093/bioinformatics/btaa758

380. Pegg DE. 2010. The relevance of ice crystal formation for the cryopreservation of tissues and organs. *Cryobiology* **60**:S36-44. doi:10.1016/j.cryobiol.2010.02.003

381. Pender MP, Csurhes PA, Burrows JM, Burrows SR. 2017. Defective T-cell control of Epstein–Barr virus infection in multiple sclerosis. *Clinical & Translational Immunology* **6**:e126. doi:10.1038/cti.2016.87

382. Peng K, Nowicki TS, Campbell K, Vahed M, Peng D, Meng Y, Nagareddy A, Huang Y-N, Karlsberg A, Miller Z, Brito J, Nadel B, Pak VM, Abedalthagafi MS, Burkhardt AM, Alachkar H, Ribas A, Mangul S. 2023. Rigorous benchmarking of T-cell receptor repertoire profiling methods for cancer RNA sequencing. *Briefings in Bioinformatics* **24**:bbad220. doi:10.1093/bib/bbad220

383. Petrie HT, Livak F, Schatz DG, Strasser A, Crispe IN, Shortman K. 1993. Multiple rearrangements in T cell receptor alpha chain genes maximize the production of useful thymocytes. *J Exp Med* **178**:615–622. doi:10.1084/jem.178.2.615

384. Pierce BG, Eberwine R, Noble JA, Habib M, Shulha HP, Weng Z, Blankenhorn EP, Mordes JP. 2013. The Missing Heritability in T1D and Potential New Targets for Prevention. *J Diabetes Res* **2013**:737485. doi:10.1155/2013/737485

385. Pierce BG, Weng Z. 2013. A flexible docking approach for prediction of T cell receptor–peptide–MHC complexes. *Protein Science* **22**:35–46. doi:10.1002/pro.2181

386. Pietra BA, De Inocencio J, Giannini EH, Hirsch R. 1994. TCR V beta family repertoire and T cell activation markers in Kawasaki disease. *J Immunol* **153**:1881–1888.

387. Pinto S, Michel C, Schmidt-Glenewinkel H, Harder N, Rohr K, Wild S, Brors B, Kyewski B. 2013. Overlapping gene coexpression patterns in human medullary thymic epithelial cells generate self-antigen diversity. *Proc Natl Acad Sci U S A* **110**:E3497-3505. doi:10.1073/pnas.1308311110

388. Pogorelyy MV, Minervina AA, Chudakov DM, Mamedov IZ, Lebedev YB, Mora T, Walczak AM. 2018. Method for identification of condition-associated public antigen receptor sequences. *eLife* **7**:e33050. doi:10.7554/eLife.33050

389. 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:10.1371/journal.pbio.3000314

390. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, Kosciolek T, Janssen S, Metcalf J, Song SJ, Kanbar J, Miller-Montgomery S, Heaton R, Mckay R, Patel SP, Swafford AD, Knight R. 2020. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* **579**:567–574. doi:10.1038/s41586-020-2095-1

391. Poran A, Scherer J, Bushway ME, Besada R, Balogh KN, Wanamaker A, Williams RG, Prabhakara J, Ott PA, Hu-Lieskovan S, Khondker ZS, Gaynor RB, Rooney MS, Srinivasan L. 2020. Combined TCR Repertoire Profiles and Blood Cell Phenotypes Predict Melanoma Patient Response to Personalized Neoantigen Therapy plus Anti-PD-1. *Cell Rep Med* **1**:100141. doi:10.1016/j.xcrm.2020.100141

392. Porciello N, Franzese O, D'Ambrosio L, Palermo B, Nisticò P. 2022. T-cell repertoire diversity: friend or foe for protective antitumor response? *J Exp Clin Cancer Res* **41**:356. doi:10.1186/s13046-022-02566-0

393. Porritt RA, Paschold L, Noval Rivas M, Cheng MH, Yonker LM, Chandnani H, Lopez M, Simnica D, Schultheiß C, Santiskulvong C, van Eyk J, McCormick JK, Fasano A, Bahar I, Binder M, Arditi M. 2021. HLA class I-associated expansion of TRBV11-2 T cells in Multisystem Inflammatory Syndrome in Children. *J Clin Invest*. doi:10.1172/JCI146614

394. Postow MA, Manuel M, Wong P, Yuan J, Dong Z, Liu C, Perez S, Tanneau I, Noel M, Courtier A, Pasqual N, Wolchok JD. 2015. Peripheral T cell receptor diversity is associated with clinical outcomes following ipilimumab treatment in metastatic melanoma. *J Immunother Cancer* **3**:23. doi:10.1186/s40425-015-0070-4

395. Potoczna N, Boehncke W-H, Nestle FO, Küenzlen C, Sterry W, Burg G, Dummer R. 1996. T-cell receptor β variable region (Vβ) usage in cutaneous T-cell lymphomas (CTCL) in comparison to normal and eczematous skin. *Journal of Cutaneous Pathology* **23**:298–305. doi:10.1111/j.1600-0560.1996.tb01301.x

396. Powell BR, Buist NRM, Stenzel P. 1982. An X-linked syndrome of diarrhea, polyendocrinopathy, and fatal infection in infancy. *The Journal of Pediatrics* **100**:731–737. doi:10.1016/S0022-3476(82)80573-8

397. Puisieux I, Bain C, Merrouche Y, Malacher P, Kourilsky P, Even J, Favrot M. 1996. Restriction of the T-cell repertoire in tumor-infiltrating lymphocytes from nine patients with renal-cell carcinoma relevance of the CDR3 length analysis for the identification of in situ clonal T-cell expansions. *International Journal of Cancer* **66**:201–208. doi:10.1002/(SICI)1097-0215(19960410)66:2<201::AID-IJC11>3.0.CO;2-F

398. Pummerer CL, Luze K, Grässl G, Bachmaier K, Offner F, Burrell SK, Lenz DM, Zamborelli TJ, Penninger JM, Neu N. 1996. Identification of cardiac myosin peptides capable of inducing autoimmune myocarditis in BALB/c mice. *J Clin Invest* **97**:2057–2062. doi:10.1172/JCI118642

399. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee J-Y, Olshen RA, Weyand CM, Boyd SD, Goronzy JJ. 2014. Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences* **111**:13139–13144. doi:10.1073/pnas.1409155111

400. Quiniou V, Barennes P, Mhanna V, Stys P, Vantomme H, Zhou Z, Martina F, Coatnoan N, Barbie M, Pham H-P, Clémenceau B, Vie H, Shugay M, Six A, Brandao B, Mallone R, Mariotti-Ferrandiz E, Klatzmann D. 2023. Human thymopoiesis produces polyspecific CD8+ α/β T cells responding to multiple viral antigens. *eLife* **12**:e81274. doi:10.7554/eLife.81274

401. Ramos GC, Dalbó S, Leite DP, Goldfeder E, Carvalho CR, Vaz NM, Assreuy J. 2012. The autoimmune nature of post-infarct myocardial healing: oral tolerance to cardiac antigens as a novel strategy to improve cardiac healing. *Autoimmunity* **45**:233–244. doi:10.3109/08916934.2011.647134

402. Ramos GC, van den Berg A, Nunes-Silva V, Weirather J, Peters L, Burkard M, Friedrich M, Pinnecker J, Abeßer M, Heinze KG, Schuh K, Beyersdorf N, Kerkau T, Demengeot J, Frantz S, Hofmann U. 2017. Myocardial aging as a T-cell–mediated phenomenon. *Proceedings of the National Academy of Sciences* **114**:E2420–E2429. doi:10.1073/pnas.1621047114

403. Ramsden DA, Baetz K, Wu GE. 1994. Conservation of sequence in recombination signal sequence spacers. *Nucleic Acids Research* **22**:1785–1796. doi:10.1093/nar/22.10.1785

404. Ramsköld D, Wang ET, Burge CB, Sandberg R. 2009. An Abundance of Ubiquitously Expressed Genes Revealed by Tissue Transcriptome Sequence Data. *PLOS Computational Biology* **5**:e1000598. doi:10.1371/journal.pcbi.1000598

405. Rao M, Wang X, Guo G, Wang L, Chen S, Yin P, Chen K, Chen L, Zhang Z, Chen X, Hu X, Hu S, Song J. 2021. Resolving the intertwining of inflammation and fibrosis in human heart failure at single-cell level. *Basic Res Cardiol* **116**:55. doi:10.1007/s00395-021-00897-1

406. Reichert T, DeBruyère M, Deneys V, Totterman T, Lydyard P, Yuksel F, Chapel H, Jewell D, Van Hove L, Linden J. 1991. Lymphocyte subset reference ranges in adult Caucasians. *Clin Immunol Immunopathol* **60**:190–208. doi:10.1016/0090-1229(91)90063-g

407. Reinherz EL, Kung PC, Goldstein G, Levey RH, Schlossman SF. 1980a. Discrete stages of human intrathymic differentiation: Analysis of normal thymocytes and leukemic lymphoblasts of T-cell lineage. *Proceedings of the National Academy of Sciences* **77**:1588–1592. doi:10.1073/pnas.77.3.1588

408. Reinherz EL, Moretta L, Roper M, Breard J, Mingari M, Cooper M, Schlossman S. 1980b. Human T lymphocyte subpopulations defined by Fc receptors and monoclonal antibodies. A comparison. *The Journal of experimental medicine* **151**. doi:10.1084/jem.151.4.969

409. Rieckmann M, Delgobo M, Gaal C, Büchner L, Steinau P, Reshef D, Gil-Cruz C, Horst EN ter, Kircher M, Reiter T, Heinze KG, Niessen HWM, Krijnen PAJ, van der Laan AM, Piek JJ, Koch C, Wester H-J, Lapa C, Bauer WR,

Ludewig B, Friedman N, Frantz S, Hofmann U, Ramos GC. 2019. Myocardial infarction triggers cardioprotective antigen-specific T helper cell responses. *Journal of Clinical Investigation* **129**:4922–4936. doi:10.1172/JCI123859

410. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, Riddell SR, Warren EH, Carlson CS. 2009. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**:4099–4107. doi:10.1182/blood-2009-04-217604

411. Ronel T, Harries M, Wicks K, Oakes T, Singleton H, Dearman R, Maxwell G, Chain B. 2021. The clonal structure and dynamics of the human T cell response to an organic chemical hapten. *eLife* **10**:e54747. doi:10.7554/eLife.54747

412. Rosenau W, Moon HD. 1961. Lysis of Homologous Cells by Sensitized Lymphocytes in Tissue Culture2. *JNCI: Journal of the National Cancer Institute* **27**:471–483. doi:10.1093/jnci/27.2.471

413. Saito H, Kranz DM, Takagaki Y, Hayday AC, Eisen HN, Tonegawa S. 1984a. Complete primary structure of a heterodimeric T-cell receptor deduced from cDNA sequences. *Nature* **309**:757–762. doi:10.1038/309757a0

414. Saito H, Kranz DM, Takagaki Y, Hayday AC, Eisen HN, Tonegawa S. 1984b. A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes. *Nature* **312**:36–40. doi:10.1038/312036a0

415. Sakaguchi N, Takahashi T, Hata H, Nomura T, Tagami T, Yamazaki S, Sakihama T, Matsutani T, Negishi I, Nakatsuru S, Sakaguchi S. 2003. Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature* **426**:454–460. doi:10.1038/nature02119

416. Sakaguchi S, Sakaguchi N, Asano M, Itoh M, Toda M. 1995. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J Immunol* **155**:1151–1164.

417. Salgado-Pabón W, Breshears L, Spaulding AR, Merriman JA, Stach CS, Horswill AR, Peterson ML, Schlievert PM. 2013. Superantigens Are Critical for Staphylococcus aureus Infective Endocarditis, Sepsis, and Acute Kidney Injury. *mBio* **4**:10.1128/mbio.00494-13. doi:10.1128/mbio.00494-13

418. Saltis M, Criscitiello MF, Ohta Y, Keefe M, Trede NS, Goitsuka R, Flajnik MF. 2008. Evolutionarily conserved and divergent regions of the Autoimmune Regulator (Aire) gene: a comparative analysis. *Immunogenetics* **60**:105–114. doi:10.1007/s00251-007-0268-9

419. Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**:5463. doi:10.1073/pnas.74.12.5463

420. Sansom SN, Shikama-Dorn N, Zhanybekova S, Nusspaumer G, Macaulay IC, Deadman ME, Heger A, Ponting CP, Holländer GA. 2014. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. *Genome Res* **24**:1918–1931. doi:10.1101/gr.171645.113

421. Santos-Zas I, Lemarié J, Zlatanova I, Cachanado M, Seghezzi J-C, Benamer H, Goube P, Vandestienne M, Cohen R, Ezzo M, Duval V, Zhang Y, Su J-B, Bizé A, Sambin L, Bonnin P, Branchereau M, Heymes C, Tanchot C, Vilar J, Delacroix C, Hulot J-S, Cochain C, Bruneval P, Danchin N, Tedgui A, Mallat Z, Simon T, Ghaleb B, Silvestre J-S, Ait-Oufella H. 2021. Cytotoxic CD8 + T cells promote granzyme B-dependent adverse post-ischemic cardiac remodeling. *Nat Commun* **12**:1483. doi:10.1038/s41467-021-21737-9

422. Sartini C, Barry SJ, Whincup PH, Wannamethee SG, Lowe GD, Jefferis BJ, Lennon L, Welsh P, Ford I, Sattar N, Morris RW. 2017. Relationship between outdoor temperature and cardiovascular disease risk factors in older people. *Eur J Prev Cardiol* **24**:349–356. doi:10.1177/2047487316682119

423. Sato K, Suematsu A, Okamoto K, Yamaguchi A, Morishita Y, Kadono Y, Tanaka S, Kodama T, Akira S, Iwakura Y, Cua DJ, Takayanagi H. 2006. Th17 functions as an osteoclastogenic helper T cell subset that links T cell activation and bone destruction. *J Exp Med* **203**:2673–2682. doi:10.1084/jem.20061775

424. Sattar N, Murray HM, Welsh P, Blauw GJ, Buckley BM, Cobbe S, de Craen AJM, Lowe GD, Jukema JW, Macfarlane PW, Murphy MB, Stott DJ, Westendorp RGJ, Shepherd J, Ford I, Packard CJ, Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) Study Group. 2009. Are markers of inflammation more strongly associated with risk for fatal than for nonfatal vascular events? *PLoS Med* **6**:e1000099. doi:10.1371/journal.pmed.1000099

425. Saxena A, Dobaczewski M, Rai V, Haque Z, Chen W, Li N, Frangogiannis NG. 2014. Regulatory T cells are recruited in the infarcted mouse myocardium and may modulate fibroblast phenotype and function. *Am J Physiol Heart Circ Physiol* **307**:H1233-1242. doi:10.1152/ajpheart.00328.2014

426. Schattgen SA, Guion K, Crawford JC, Souquette A, Barrio AM, Stubbington MJT, Thomas PG, Bradley P. 2022. Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nat Biotechnol* **40**:54–63. doi:10.1038/s41587-021-00989-2

427. Schnepp PM, Chen M, Keller ET, Zhou X. 2019. SNV identification from single-cell RNA sequencing data. *Hum Mol Genet* **28**:3569–3583. doi:10.1093/hmg/ddz207

428. Schober K, Voit F, Grassmann S, Müller TR, Eggert J, Jarosch S, Weißbrich B, Hoffmann P, Borkner L, Nio E, Fanchi L, Clouser CR, Radhakrishnan A, Mihatsch L, Lückemeier P, Leube J, Dössinger G, Klein L,

Neuenhahn M, Oduro JD, Cicin-Sain L, Buchholz VR, Busch DH. 2020. Reverse TCR repertoire evolution toward dominant low-affinity clones during chronic CMV infection. *Nat Immunol* **21**:434–441. doi:10.1038/s41590-020-0628-2

429. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**:3. doi:10.1186/1471-2199-7-3

430. Schuldt NJ, Binstadt BA. 2019. Dual TCR T Cells: Identity Crisis or Multitaskers? *The Journal of Immunology* **202**:637–644. doi:10.4049/jimmunol.1800904

431. Schulenburg H, Kurtz J, Moret Y, Siva-Jothy MT. 2009. Introduction. Ecological immunology. *Philos Trans R Soc Lond B Biol Sci* **364**:3–14. doi:10.1098/rstb.2008.0249

432. Schwartz RH. 2003. T Cell Anergy. *Annu Rev Immunol* **21**:305–334. doi:10.1146/annurev.immunol.21.120601.141110

433. Schwarz K, Gh G, L L, U P, Z L, D L, W F, Ra S, Te H-H, S D, Mr L, Cr B. 1996. RAG mutations in human B cell-negative SCID. *Science (New York, NY)* **274**. doi:10.1126/science.274.5284.97

434. Scotto–Lavino E, Du G, Frohman MA. 2006. 5′ end cDNA amplification using classic RACE. *Nat Protoc* **1**:2555–2562. doi:10.1038/nprot.2006.480

435. Sender R, Milo R. 2021. The distribution of cellular turnover in the human body. *Nat Med* **27**:45–48. doi:10.1038/s41591-020-01182-9

436. Sengar DPS, Couture RA, Jindal SL, Catching JD. 1985. Histocompatibility antigens in essential hypertension and myocardial infarction. *Tissue Antigens* **26**:168–171. doi:10.1111/j.1399-0039.1985.tb00954.x

437. Sepich-Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, Kosciolek T, Janssen S, Metcalf J, Song SJ, Kanbar J, Miller-Montgomery S, Heaton R, Mckay R, Patel SP, Swafford AD, Knight R. 2023. Reply to: Caution Regarding the Specificities of Pan-Cancer Microbial Structure. doi:10.1101/2023.02.10.528049

438. Serwold T, F G, J K, R J, N S. 2002. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. *Nature* **419**. doi:10.1038/nature01074

439. Sethna Z, Elhanati Y, Callan C, Walczak A, Mora T. 2019. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics (Oxford, England)* **35**. doi:10.1093/bioinformatics/btz035

440. Sharma S, Plant D, Bowes J, Macgregor A, Verstappen S, Barton A, Viatte S. 2022. HLA-DRB1 haplotypes predict cardiovascular mortality in inflammatory polyarthritis independent of CRP and anti-CCP status. *Arthritis Research & Therapy* **24**:90. doi:10.1186/s13075-022-02775-0

441. Shen M-JR, Boutell JM, Stephens KM, Ronaghi M, Gunderson K, Venkatesan BM, Bowen MS, Vijayan K. 2014. Kinetic exclusion amplification of nucleic acid libraries. US8895249B2.

442. Shen Y, Li R, Tian F, Chen Z, Lu N, Bai Y, Ge Q, Lu Z. 2018. Impact of RNA integrity and blood sample storage conditions on the gene expression analysis. *OTT* **11**:3573–3581. doi:10.2147/OTT.S158868

443. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. 2005. Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome. *Science* **309**:1728–1732. doi:10.1126/science.1117389

444. Sheng J, Wang H, Liu X, Deng Y, Yu Y, Xu P, Shou J, Pan Hong, Li H, Zhou X, Han W, Sun T, Pan Hongming, Fang Y. 2021. Deep Sequencing of T-Cell Receptors for Monitoring Peripheral CD8+ T Cells in Chinese Advanced Non–Small-Cell Lung Cancer Patients Treated With the Anti–PD-L1 Antibody. *Frontiers in Molecular Biosciences* **8**.

445. Shimizu K, Iyoda T, Sanpei A, Nakazato H, Okada M, Ueda S, Kato-Murayama M, Murayama K, Shirouzu M, Harada N, Hidaka M, Fujii S. 2021. Identification of TCR repertoires in functionally competent cytotoxic T cells cross-reactive to SARS-CoV-2. *Commun Biol* **4**:1–13. doi:10.1038/s42003-021-02885-6

446. Shlomchik MJ, Marshak-Rothstein A, Wolfowicz CB, Rothstein TL, Weigert MG. 1987. The role of clonal selection and somatic mutation in autoimmunity. *Nature* **328**:805–811. doi:10.1038/328805a0

447. Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, Komech EA, Sycheva AL, Koneva AE, Egorov ES, Eliseev AV, Van Dyk E, Dash P, Attaf M, Rius C, Ladell K, McLaren JE, Matthews KK, Clemens EB, Douek DC, Luciani F, van Baarle D, Kedzierska K, Kesmir C, Thomas PG, Price DA, Sewell AK, Chudakov DM. 2018. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Research* **46**:D419–D427. doi:10.1093/nar/gkx760

448. Sica A, Mantovani A. 2012. Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest* **122**:787–795. doi:10.1172/JCI59643

449. Sidhom J-W, Larman HB, Pardoll DM, Baras AS. 2021. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun* **12**:1605. doi:10.1038/s41467-021-21879-w

450. Sidorov IA, Romanyukha AA. 1993. Mathematical modeling of T-cell proliferation. *Mathematical Biosciences* **115**:187–232. doi:10.1016/0025-5564(93)90071-H

451. Sim B-C, Zerva L, Greene MI, Gascoigne NRJ. 1996. Control of MHC Restriction by TCR Vα CDR1 and CDR2. *Science* **273**:963–966. doi:10.1126/science.273.5277.963

452. Simon JS, Botero S, Simon SM. 2018. Sequencing the peripheral blood B and T cell repertoire – Quantifying robustness and limitations. *Journal of Immunological Methods* **463**:137–147. doi:10.1016/j.jim.2018.10.003

453. Singleton H, Popple A, Gellatly N, Maxwell G, Williams J, Friedmann PS, Kimber I, Dearman RJ. 2016. Anti-hapten antibodies in response to skin sensitization. *Contact Dermatitis* **74**:197–204. doi:10.1111/cod.12486

454. Sintou A, Mansfield C, Iacob A, Chowdhury RA, Narodden S, Rothery SM, Podovei R, Sanchez-Alonso JL, Ferraro E, Swiatlowska P, Harding SE, Prasad S, Rosenthal N, Gorelik J, Sattler S. 2020. Mediastinal Lymphadenopathy, Class-Switched Auto-Antibodies and Myocardial Immune-Complexes During Heart Failure in Rodents and Humans. *Front Cell Dev Biol* **8**:695. doi:10.3389/fcell.2020.00695

455. Siu G, Clark SP, Yoshikai Y, Malissen M, Yanagi Y, Strauss E, Mak TW, Hood L. 1984. The human T cell antigen receptor is encoded by variable, diversity, and joining gene segments that rearrange to generate a complete V gene. *Cell* **37**:393–401. doi:10.1016/0092-8674(84)90369-6

456. Six A, Mariotti-Ferrandiz ME, Chaara W, Magadan S, Pham H-P, Lefranc M-P, Mora T, Thomas-Vaslin V, Walczak AM, Boudinot P. 2013. The Past, Present, and Future of Immune Repertoire Biology – The Rise of Next-Generation Repertoire Analysis. *Front Immunol* **4**. doi:10.3389/fimmu.2013.00413

457. Smirnova AO, Miroshnichenkova AM, Belyaeva LD, Kelmanson IV, Lebedev YB, Mamedov IZ, Chudakov DM, Komkov AY. 2023. Novel bimodal TRBD1-TRBD2 rearrangements with dual or absent D-region contribute to TRB V-(D)-J combinatorial diversity. *Frontiers in Immunology* **14**.

458. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**:674–679. doi:10.1038/321674a0

459. Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**:491–499. doi:10.1101/gr.209601.116

460. Son ET, Faridi P, Paul-Heng M, Leong ML, English K, Ramarathinam SH, Braun A, Dudek NL, Alexander IE, Lisowski L, Bertolino P, Bowen DG,

Purcell AW, Mifsud NA, Sharland AF. 2021. The self-peptide repertoire plays a critical role in transplant tolerance induction. *J Clin Invest* **131**. doi:10.1172/JCI146771

461. Song L, Cohen D, Ouyang Z, Cao Y, Hu X, Liu XS. 2021. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods* **18**:627–630. doi:10.1038/s41592-021-01142-2

462. Specht H, Emmott E, Petelski AA, Huffman RG, Perlman DH, Serra M, Kharchenko P, Koller A, Slavov N. 2021. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. *Genome Biology* **22**:50. doi:10.1186/s13059-021-02267-5

463. Spinale FG. 2002. Matrix Metalloproteinases. *Circulation Research* **90**:520–530. doi:10.1161/01.RES.0000013290.12884.A3

464. Stephen B, Hajjar J, Sarda S, Duose DY, Conroy JM, Morrison C, Alshawa A, Xu M, Zarifa A, Patel SP, Yuan Y, Kwiatkowski E, Wang L, Ahnert JR, Fu S, Meric-Bernstam F, Lowman GM, Looney T, Naing A. 2023. T-cell receptor beta variable gene polymorphism predicts immune-related adverse events during checkpoint blockade immunotherapy. *J Immunother Cancer* **11**:e007236. doi:10.1136/jitc-2023-007236

465. Stewart S, Keates AK, Redfern A, McMurray JJV. 2017. Seasonal variations in cardiovascular disease. *Nature Reviews Cardiology* **14**:654–664. doi:10.1038/nrcardio.2017.76

466. Stirk ER, Molina-París C, van den Berg HA. 2008. Stochastic niche structure and diversity maintenance in the T cell repertoire. *Journal of Theoretical Biology* **255**:237–249. doi:10.1016/j.jtbi.2008.07.017

467. Struyk L, Hawes GE, Chatila MK, Breedveld FC, Kurnick JT, Elsen PJVD. 1995. T cell receptors in rheumatoid arthritis. *Arthritis & Rheumatism* **38**:577–589. doi:10.1002/art.1780380502

468. Suliga E, Kozieł D, Ciesla E, Rebak D, Głuszek-Osuch M, Naszydłowska E, Głuszek S. 2019. The Consumption of Alcoholic Beverages and the Prevalence of Cardiovascular Diseases in Men and Women: A Cross-Sectional Study. *Nutrients* **11**:1318. doi:10.3390/nu11061318

469. Sun W, Cui Y, Zhen L, Huang L. 2011. Association between HLA-DRB1, HLA-DRQB1 alleles, and CD4+CD28null T cells in a Chinese population with coronary heart disease. *Mol Biol Rep* **38**:1675–1679. doi:10.1007/s11033-010-0279-8

470. Taleb S, Tedgui A, Mallat Z. 2015. IL-17 and Th17 Cells in Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **35**:258–264. doi:10.1161/ATVBAHA.114.303567

471. Tang T-T, Zhu Y-C, Dong N-G, Zhang S, Cai J, Zhang L-X, Han Y, Xia N, Nie S-F, Zhang M, Lv B-J, Jiao J, Yang X-P, Hu Y, Liao Y-H, Cheng X. 2019. Pathologic T-cell response in ischaemic failing hearts elucidated by T-cell receptor sequencing and phenotypic characterization. *Eur Heart J*. doi:10.1093/eurheartj/ehz516

472. Thapa DR, Tonikian R, Sun C, Liu M, Dearth A, Petri M, Pepin F, Emerson RO, Ranger A. 2015. Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing. *Arthritis Res Ther* **17**:1–12. doi:10.1186/s13075-015-0655-9

473. The MHC sequencing consortium. 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**:921–923. doi:10.1038/44853

474. Thomas P, Pang Y, Dong J, Berg AH. 2014. Identification and Characterization of Membrane Androgen Receptors in the ZIP9 Zinc Transporter Subfamily: II. Role of Human ZIP9 in Testosterone-Induced Prostate and Breast Cancer Cell Apoptosis. *Endocrinology* **155**:4250–4265. doi:10.1210/en.2014-1201

475. Thomas-Vaslin V, Altes HK, de Boer RJ, Klatzmann D. 2008. Comprehensive assessment and mathematical modeling of T cell population dynamics and homeostasis. *J Immunol* **180**:2240–2250. doi:10.4049/jimmunol.180.4.2240

476. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**:4673. doi:10.1093/nar/22.22.4673

477. Thompson SD, Pelkonen J, Hurwitz JL. 1990. First T cell receptor alpha gene rearrangements during T cell ontogeny skew to the 5’ region of the J alpha locus. *J Immunol* **145**:2347–2352.

478. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. 2017. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* **33**:2924–2929. doi:10.1093/bioinformatics/btx286

479. Timmis A, Townsend N, Gale C, Grobbee R, Maniadakis N, Flather M, Wilkins E, Wright L, Vos R, Bax J, Blum M, Pinto F, Vardas P, ESC Scientific Document Group, Goda A, Demiraj AF, Weidinger F, Metzler B, Ibrahimov F, Pasquet AA, Claeys M, Thorton Y, Kusljugic Z, Smajic E, Velchev V, Ivanov N, Antoniades L, Agathangelou P, Táborský M, Gerdes C, Viigima M, Juhani PM, Juilliere Y, Cattan S, Aladashvili A, Hamm C, Kuck K-H, Papoutsis K, Bestehorn K, Foussas S, Giannoulidou G, Varounis C, Kallikazaros I, Kiss RG, Czétényi T, Becker D, Gudnason T, Kearney P, McDonald K, Rozenman Y, Ziv B, Bolognese

L, Luciolli P, Boriani G, Berkinbayev S, Rakisheva A, Mirrakhimov E, Erglis A, Jegere S, Marinskis G, Beissel J, Marchal N, Kedev S, Xuereb RG, Tilney T, Felice T, Popovici M, Bax J, Mulder B, Simoons M, Elsendoorn M, Steigen TK, Atar D, Kalarus Z, Tendera M, Cardoso JS, Ribeiro J, Mateus C, Tatu-Chitoiu G, Seferovic P, Beleslin B, Simkova I, Durcikova P, Belicova V, Fras Z, Radelj S, Gonzalez Juanatey JR, Legendre S, Braunschweig F, Kaufmann UP, Rudiger-Sturchler M, Tokgozoglu L, Unver A, Kovalenko V, Nesukay E, Naum A, de Courtelary PT, Martin S, Sebastiao D, Ghislain D, Bardinet I, Logstrup S. 2018. European Society of Cardiology: Cardiovascular Disease Statistics 2017. *European Heart Journal* **39**:508–579. doi:10.1093/eurheartj/ehx628

480. Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* **302**:575–581. doi:10.1038/302575a0

481. Tong Y, Li Z, Zhang H, Xia L, Zhang M, Xu Y, Wang Z, Deem MW, Sun X, He J. 2016. T Cell Repertoire Diversity Is Decreased in Type 1 Diabetes Patients. *Genomics Proteomics Bioinformatics* **14**:338–348. doi:10.1016/j.gpb.2016.10.003

482. Tong Y, Wang J, Zheng T, Zhang X, Xiao X, Zhu X, Lai X, Liu X. 2020. SETE: Sequence-based Ensemble learning approach for TCR Epitope binding prediction. *Comput Biol Chem* **87**:107281. doi:10.1016/j.compbiolchem.2020.107281

483. Trepel F. 1974. Number and distribution of lymphocytes in man. A critical analysis. *Klin Wochenschr* **52**:511–515. doi:10.1007/BF01468720

484. Tripolt NJ, Kolesnik E, Pferschy PN, Verheyen N, Ablasser K, Sailer S, Alber H, Berger R, Kaulfersch C, Leitner K, Lichtenauer M, Mader A, Moertl D, Oulhaj A, Reiter C, Rieder T, Saely CH, Siller-Matula J, Weidinger F, Zechner PM, von Lewinski D, Sourij H. 2020. Impact of EMpagliflozin on cardiac function and biomarkers of heart failure in patients with acute MYocardial infarction—The EMMY trial. *American Heart Journal* **221**:39–47. doi:10.1016/j.ahj.2019.12.004

485. Trück J, Eugster A, Barennes P, Tipton CM, Luning Prak ET, Bagnara D, Soto C, Sherkow JS, Payne AS, Lefranc M-P, Farmer A, Bostick M, Mariotti-Ferrandiz E. 2021. Biological controls for standardization and interpretation of adaptive immune receptor repertoire profiling. *eLife* **10**:e66274. doi:10.7554/eLife.66274

486. Tubo NJ, Pagán AJ, Taylor JJ, Nelson RW, Linehan JL, Ertelt JM, Huseby ES, Way SS, Jenkins MK. 2013. Single Naive CD4+ T Cells from a Diverse Repertoire Produce Different Effector Cell Types during Infection. *Cell* **153**:785–796. doi:10.1016/j.cell.2013.04.007

487. Tuzlak S, Dejean AS, Iannacone M, Quintana FJ, Waisman A, Ginhoux F, Korn T, Becher B. 2021. Repositioning TH cell polarization from single

cytokines to complex help. *Nat Immunol* **22**:1210–1217. doi:10.1038/s41590-021-01009-w

488. Tykocinski L-O, Sinemus A, Rezavandy E, Weiland Y, Baddeley D, Cremer C, Sonntag S, Willecke K, Derbinski J, Kyewski B. 2010. Epigenetic regulation of promiscuous gene expression in thymic medullary epithelial cells. *Proc Natl Acad Sci U S A* **107**:19426–19431. doi:10.1073/pnas.1009265107

489. Unutmaz D, Pileri P, Abrignani S. 1994. Antigen-independent activation of naive and memory resting T cells by a cytokine combination. *J Exp Med* **180**:1159–1164. doi:10.1084/jem.180.3.1159

490. Valkiers S, Houcke MV, Laukens K, Meysman P. 2021. clusTCR: a Python interface for rapid clustering of large sets of CDR3 sequences. *bioRxiv* 2021.02.22.432291. doi:10.1101/2021.02.22.432291

491. Vander Heiden JA, Yaari G, Uduman M, Stern JNH, O'Connor KC, Hafler DA, Vigneault F, Kleinstein SH. 2014. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**:1930–1932. doi:10.1093/bioinformatics/btu138

492. Venturi V, Chin HY, Asher TE, Ladell K, Scheinberg P, Bornstein E, van Bockel D, Kelleher AD, Douek DC, Price DA, Davenport MP. 2008. TCR β-Chain Sharing in Human CD8+ T Cell Responses to Cytomegalovirus and EBV1. *The Journal of Immunology* **181**:7853–7862. doi:10.4049/jimmunol.181.11.7853

493. Verhoeven D, Teijaro JR, Farber DL. 2008. Heterogeneous Memory T Cells in Antiviral Immunity and Immunopathology. *Viral Immunology* **21**:99–114. doi:10.1089/vim.2008.0002

494. von Lewinski D, Kolesnik E, Tripolt NJ, Pferschy PN, Benedikt M, Wallner M, Alber H, Berger R, Lichtenauer M, Saely CH, Moertl D, Auersperg P, Reiter C, Rieder T, Siller-Matula JM, Gager GM, Hasun M, Weidinger F, Pieber TR, Zechner PM, Herrmann M, Zirlik A, Holman RR, Oulhaj A, Sourij H, on behalf of the EMMY Investigators. 2022. Empagliflozin in acute myocardial infarction: the EMMY trial. *European Heart Journal* **43**:4421–4432. doi:10.1093/eurheartj/ehac494

495. Vujović M, Marcatili P, Chain B, Kaplinsky J, Andresen TL. 2023. Signatures of T cell immunity revealed using sequence similarity with TCRDivER algorithm. *Commun Biol* **6**:1–13. doi:10.1038/s42003-023-04702-8

496. Walenta K, Kindermann I, Gärtner B, Kandolph R, Link A, Böhm M. 2006. Dangerous Kisses: Epstein-Barr Virus Myocarditis Mimicking Myocardial Infarction. *The American Journal of Medicine* **119**:e3–e6. doi:10.1016/j.amjmed.2005.11.033

497. Walsh CM, Matloubian M, Liu CC, Ueda R, Kurahara CG, Christensen JL, Huang MT, Young JD, Ahmed R, Clark WR. 1994. Immune function in mice lacking the perforin gene. *Proc Natl Acad Sci U S A* **91**:10854–10858.

498. Wang F, Huang C-Y, Kanagawa O. 1998. Rapid deletion of rearranged T cell antigen receptor (TCR) Vα-Jα segment by secondary rearrangement in the thymus: Role of continuous rearrangement of TCR α chain gene and positive selection in the T cell repertoire formation. *Proceedings of the National Academy of Sciences* **95**:11834–11839. doi:10.1073/pnas.95.20.11834

499. Wang H, Peng G, Bai J, He B, Huang K, Hu X, Liu D. 2017. Cytomegalovirus Infection and Relative Risk of Cardiovascular Disease (Ischemic Heart Disease, Stroke, and Cardiovascular Death): A Meta-Analysis of Prospective Studies Up to 2016. *J Am Heart Assoc* **6**:e005025. doi:10.1161/JAHA.116.005025

500. Wang S, Sontag ED, Lauffenburger DA. 2023. What cannot be seen correctly in 2D visualizations of single-cell 'omics data? *cels* **14**:723–731. doi:10.1016/j.cels.2023.07.002

501. Wang Z, Xie L, Ding G, Song S, Chen L, Li G, Xia M, Han D, Zheng Y, Liu J, Xiao T, Zhang H, Huang Y, Li Y, Huang M. 2021. Single-cell RNA sequencing of peripheral blood mononuclear cells from acute Kawasaki disease patients. *Nat Commun* **12**:5444. doi:10.1038/s41467-021-25771-5

502. Watanabe M, Panetta GL, Piccirillo F, Spoto S, Myers J, Serino FM, Costantino S, Di Sciascio G. 2019. Acute Epstein-Barr related myocarditis: An unusual but life-threatening disease in an immunocompetent patient. *J Cardiol Cases* **21**:137–140. doi:10.1016/j.jccase.2019.12.001

503. Weber CR, Akbar R, Yermanos A, Pavlović M, Snapkov I, Sandve GK, Reddy ST, Greiff V. 2020. immuneSIM: tunable multi-feature simulation of B- and T-cell receptor repertoires for immunoinformatics benchmarking. *Bioinformatics* **36**:3594–3596. doi:10.1093/bioinformatics/btaa158

504. Wei J, Zanker D, Carluccio ARD, Smelkinson MG, Takeda K, Seedhom MO, Dersh D, Gibbs JS, Yang N, Jadhav A, Chen W, Yewdell JW. 2017. Varied Role of Ubiquitylation in Generating MHC Class I Peptide Ligands. *Journal of immunology (Baltimore, Md : 1950)* **198**:3835. doi:10.4049/jimmunol.1602122

505. Weirather J, Hofmann UDW, Beyersdorf N, Ramos GC, Vogel B, Frey A, Ertl G, Kerkau T, Frantz S. 2014. Foxp3+ CD4+ T cells improve healing after myocardial infarction by modulating monocyte/macrophage differentiation. *Circ Res* **115**:55–67. doi:10.1161/CIRCRESAHA.115.303895

506. Westermann J, Pabst R. 1992. Distribution of lymphocyte subsets and natural killer cells in the human body. *The Clinical investigator* **70**. doi:10.1007/BF00184787

507. Westermann J, Pabst R. 1990. Lymphocyte subsets in the blood: a diagnostic window on the lymphoid system? *Immunol Today* **11**:406–410. doi:10.1016/0167-5699(90)90160-b

508. White J, Herman A, Pullen AM, Kubo R, Kappler JW, Marrack P. 1989. The Vβ-specific superantigen staphylococcal enterotoxin B: Stimulation of mature T cells and clonal deletion in neonatal mice. *Cell* **56**:27–35. doi:10.1016/0092-8674(89)90980-X

509. Whittaker RH. 1972. Evolution and Measurement of Species Diversity. *Taxon* **21**:213–251. doi:10.2307/1218190

510. Wiedeman AE, Muir VS, Rosasco MG, DeBerg HA, Presnell S, Haas B, Dufort MJ, Speake C, Greenbaum CJ, Serti E, Nepom GT, Blahnik G, Kus AM, James EA, Linsley PS, Long SA. 2020. Autoreactive CD8+ T cell exhaustion distinguishes subjects with slow type 1 diabetes progression. *J Clin Invest* **130**:480–490. doi:10.1172/JCI126595

511. Wilkins E, Wilson L, Wickramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, Burns R, Rayner M, Townsend N. 2017. European Cardiovascular Disease Statistics 2017. Brussels: European Heart Network.

512. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**:160018. doi:10.1038/sdata.2016.18

513. Wolf D, Gerhardt T, Winkels H, Michel NA, Pramod AB, Ghosheh Y, Brunel S, Buscher K, Miller J, McArdle S, Baas L, Kobiyama K, Vassallo M, Ehinger E, Dileepan T, Ali A, Schell M, Mikulski Z, Sidler D, Kimura T, Sheng X, Horstmann H, Hansen S, Mitre LS, Stachon P, Hilgendorf I, Gaddis DE, Hedrick C, Benedict CA, Peters B, Zirlik A, Sette A, Ley K. 2020. Pathogenic Autoimmunity in Atherosclerosis Evolves From Initially Protective Apolipoprotein B_{100} –Reactive $CD4^+$ T-Regulatory Cells. *Circulation* **142**:1279–1293. doi:10.1161/CIRCULATIONAHA.119.042863

514. Won T, Kalinoski HM, Wood MK, Hughes DM, Jaime CM, Delgado P, Talor MV, Lasrado N, Reddy J, Čiháková D. 2022. Cardiac myosin-specific autoimmune T cells contribute to immune-checkpoint-inhibitor-associated myocarditis. *Cell Reports* **41**:111611. doi:10.1016/j.celrep.2022.111611

515. Wong WF, Kohu K, Nakamura A, Ebina M, Kikuchi T, Tazawa R, Tanaka K, Kon S, Funaki T, Sugahara-Tobinai A, Looi CY, Endo S, Funayama R, Kurokawa M, Habu S, Ishii N, Fukumoto M, Nakata K, Takai T, Satake M. 2012. Runx1 deficiency in CD4+ T cells causes fatal autoimmune inflammatory lung disease due to spontaneous hyperactivation of cells. *J Immunol* **188**:5408–5420. doi:10.4049/jimmunol.1102991

516. Wu LC, Tuot DS, Lyons DS, Garcia KC, Davis MM. 2002. Two-step binding mechanism for T-cell receptor recognition of peptide–MHC. *Nature* **418**:552–556. doi:10.1038/nature00920

517. Wulf MG, Maguire S, Humbert P, Dai N, Bei Y, Nichols NM, Corrêa IR, Guan S. 2019. Non-templated addition and template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-occur and compete with each other. *Journal of Biological Chemistry* **294**:18220–18231. doi:10.1074/jbc.RA119.010676

518. Xia N, Lu Y, Gu M, Li N, Liu M, Jiao J, Zhu Z, Li J, Li D, Tang T, Lv B, Nie S, Zhang M, Liao M, Liao Y, Yang X, Cheng X. 2020. A Unique Population of Regulatory T Cells in Heart Potentiates Cardiac Protection From Myocardial Infarction. *Circulation* **142**:1956–1973. doi:10.1161/CIRCULATIONAHA.120.046789

519. Xie X, Shi X, Liu M. 2018. The Roles of Genetic Factors in Kawasaki Disease: A Systematic Review and Meta-analysis of Genetic Association Studies. *Pediatr Cardiol* **39**:207–225. doi:10.1007/s00246-017-1760-0

520. Xu Y, Qian X, Zhang X, Lai X, Liu Y, Wang J. 2022. DeepLION: Deep Multi-Instance Learning Improves the Prediction of Cancer-Associated T Cell Receptors for Accurate Cancer Detection. *Front Genet* **13**:860510. doi:10.3389/fgene.2022.860510

521. Yamagata H, Kobayashi A, Tsunedomi R, Seki T, Kobayashi M, Hagiwara K, Chen C, Uchida S, Okada G, Fuchikami M, Kamishikiryo T, Iga J, Numata S, Kinoshita M, Kato TA, Hashimoto R, Nagano H, Okamoto Y, Ueno S, Ohmori T, Nakagawa S. 2021. Optimized protocol for the extraction of RNA and DNA from frozen whole blood sample stored in a single EDTA tube. *Sci Rep* **11**:17075. doi:10.1038/s41598-021-96567-2

522. Yan X, Anzai A, Katsumata Y, Matsuhashi T, Ito K, Endo J, Yamamoto T, Takeshima A, Shinmura K, Shen W, Fukuda K, Sano M. 2013. Temporal dynamics of cardiac immune cell accumulation following acute myocardial infarction. *Journal of Molecular and Cellular Cardiology* **62**:24–35. doi:10.1016/j.yjmcc.2013.04.023

523. Yan X, Shichita T, Katsumata Y, Matsuhashi T, Ito H, Ito K, Anzai A, Endo J, Tamura Y, Kimura K, Fujita J, Shinmura K, Shen W, Yoshimura A, Fukuda K, Sano M. 2012. Deleterious Effect of the IL-23/IL-17A Axis and γδT

Cells on Left Ventricular Remodeling After Myocardial Infarction. *J Am Heart Assoc* **1**:e004408. doi:10.1161/JAHA.112.004408

524. Yang H-Q, Wang Y-S, Zhai K, Tong Z-H. 2021. Single-Cell TCR Sequencing Reveals the Dynamics of T Cell Repertoire Profiling During Pneumocystis Infection. *Frontiers in Microbiology* **12**.

525. Yang Z, Day Y-J, Toufektsian M-C, Xu Y, Ramos SI, Marshall MA, French BA, Linden J. 2006. Myocardial infarct-sparing effect of adenosine A2A receptor activation is due to its action on CD4+ T lymphocytes. *Circulation* **114**:2056–2064. doi:10.1161/CIRCULATIONAHA.106.649244

526. Yassai MB, Naumov YN, Naumova EN, Gorski J. 2009. A clonotype nomenclature for T cell receptors. *Immunogenetics* **61**:493–502. doi:10.1007/s00251-009-0383-x

527. Ye Y, Jia X, Bajaj M, Birnbaum Y. 2018. Dapagliflozin Attenuates Na+/H+ Exchanger-1 in Cardiofibroblasts via AMPK Activation. *Cardiovasc Drugs Ther* **32**:553–558. doi:10.1007/s10557-018-6837-3

528. Yen D, Cheung J, Scheerens H, Poulet F, McClanahan T, Mckenzie B, Kleinschek MA, Owyang A, Mattson J, Blumenschein W, Murphy E, Sathe M, Cua DJ, Kastelein RA, Rennick D. 2006. IL-23 is essential for T cell–mediated colitis and promotes inflammation via IL-17 and IL-6. *J Clin Invest* **116**:1310–1316. doi:10.1172/JCI21404

529. Yeo HC, Selvarajoo K. 2022. Machine learning alternative to systems biology should not solely depend on data. *Briefings in Bioinformatics* **23**:bbac436. doi:10.1093/bib/bbac436

530. Yokota R, Kaminaga Y, Kobayashi TJ. 2017. Quantification of Inter-Sample Differences in T-Cell Receptor Repertoires Using Sequence-Based Information. *Front Immunol* **8**:1500. doi:10.3389/fimmu.2017.01500

531. Yousaf Z, Albaz N, Abdelmajid AA, Sabobeh T, Elzouki A. 2021. Reactivation cytomegalovirus leading to acute myocardial infarction—A first reported case in an immunocompetent patient. *Clin Case Rep* **9**:1958–1963. doi:10.1002/ccr3.3914

532. Yu X, Ye J, Hathaway CA, Tworoger S, Lea J, Li B. 2023. Quantifiable TCR repertoire changes in pre-diagnostic blood specimens among high-grade ovarian cancer patients. doi:10.1101/2023.10.12.562056

533. Zacchigna S, Martinelli V, Moimas S, Colliva A, Anzini M, Nordio A, Costa A, Rehman M, Vodret S, Pierro C, Colussi G, Zentilin L, Gutierrez MI, Dirkx E, Long C, Sinagra G, Klatzmann D, Giacca M. 2018. Paracrine effect of regulatory T cells promotes cardiomyocyte proliferation during pregnancy and after myocardial infarction. *Nat Commun* **9**:2432. doi:10.1038/s41467-018-04908-z

534. Zehn D, Bevan M. 2006. T cells with low avidity for a tissue-restricted antigen routinely evade central and peripheral tolerance and cause autoimmunity. *Immunity* **25**. doi:10.1016/j.immuni.2006.06.009

535. Zerrahn J, Held W, Raulet DH. 1997. The MHC Reactivity of the T Cell Repertoire Prior to Positive and Negative Selection. *Cell* **88**:627–636. doi:10.1016/S0092-8674(00)81905-4

536. Zhang H, Liu L, Zhang J, Chen J, Ye J, Shukla S, Qiao J, Zhan X, Chen H, Wu CJ, Fu Y-X, Li B. 2020. Investigation of Antigen-Specific T-Cell Receptor Clusters in Human Cancers. *Clinical Cancer Research* **26**:1359–1371. doi:10.1158/1078-0432.CCR-19-3249

537. Zhang H, Zhan X, Li B. 2021. GIANA allows computationally-efficient TCR clustering and multi-disease repertoire classification by isometric transformation. *Nat Commun* **12**:4699. doi:10.1038/s41467-021-25006-7

538. Zhang J, Wang Y, Yu H, Chen G, Wang L, Liu F, Yuan J, Ni Q, Xia X, Wan Y. 2021. Mapping the spatial distribution of T cells in repertoire dimension. *Molecular Immunology* **138**:161–171. doi:10.1016/j.molimm.2021.08.009

539. Zhang S, Zhang X, Wang K, Xu X, Li M, Zhang J, Zhang Y, Hao J, Sun X, Chen Y, Liu X, Chang Y, Jin R, Wu H, Ge Q. 2018. Newly Generated CD4+ T Cells Acquire Metabolic Quiescence after Thymic Egress. *J Immunol* **200**:1064–1077. doi:10.4049/jimmunol.1700721

540. Zhang Z, Xiong D, Wang X, Liu H, Wang T. 2021. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nature Methods* **18**:92–99. doi:10.1038/s41592-020-01020-3

541. Zhao TX, Sriranjan RS, Lu Y, Hubsch A, Kaloyirou F, Vamvaka E, Helmy J, Kostapanos M, Klatzmann D, Tedgui A, Rudd JHF, Hoole SP, Bond SP, Mallat Z, Cheriyan J. 2020. Low dose interleukin-2 in patients with stable ischaemic heart disease and acute coronary syndrome (LILACS). *European Heart Journal* **41**:ehaa946.1735. doi:10.1093/ehjci/ehaa946.1735

542. Zhao TX, Sriranjan RS, Tuong ZK, Lu Y, Sage AP, Nus M, Hubsch A, Kaloyirou F, Vamvaka E, Helmy J, Kostapanos M, Jalaludeen N, Klatzmann D, Tedgui A, Rudd JHF, Horton SJ, Huntly BJP, Hoole SP, Bond SP, Clatworthy MR, Cheriyan J, Mallat Z. 2022. Regulatory T-Cell Response to Low-Dose Interleukin-2 in Ischemic Heart Disease. *NEJM Evidence* **1**:EVIDoa2100009. doi:10.1056/EVIDoa2100009

543. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R, Wong A, Ness KD, Beppu LW, Deeg HJ, McFarland C, Loeb KR, Valente WJ, Ericson NG, Stevens EA, Radich JP, Mikkelsen TS, Hindson BJ, Bielas JH. 2017. Massively parallel digital

transcriptional profiling of single cells. *Nat Commun* **8**:14049. doi:10.1038/ncomms14049

544. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**:892–897. doi:10.2144/01304pf02

545. Zinkernagel R, Doherty P. 1974. Immunological surveillance against altered self components by sensitised T lymphocytes in lymphocytic choriomeningitis. *Nature* **251**. doi:10.1038/251547a0

546. Zinkernagel RM, Doherty PC. 1974. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**:701–702. doi:10.1038/248701a0

FIGURES AND TABLES

FIGURES

TABLES

ABSTRACT

Modelling the T-cell repertoires of circulating T-cells and its application in cardiovascular diseases

Cardiovascular diseases (CVD) are the leading cause of mortality in Europe, surpassing various cancers, and are a major current public health concern. The immune system is involved in the aetiology of these diseases. Among the many components of the immune system, T lymphocytes play a predominant role in the development, progression, or tissue repair of CVD. T lymphocytes carry out their function following their activation, which is determined by their ability to specifically recognise proteins or antigens, whether they are exogenous, such as those from pathogens or allergens, or endogenous, such as those from physiological cell renewal or tumour cells. This recognition is mediated by a specific receptor called the TCR, which stands for T-cell receptor. Unlike most genes, the TCR is generated through a random somatic recombination process involving around a hundred genes belonging to three major families: V for Variable, D for Diversity, and J for Junction. This unique mechanism, particular to genes encoding antigen-specific receptors, takes place in the thymus during the development and differentiation of T lymphocytes and results in the production of several billion different TCRs in an individual. This diversity forms the TCR repertoire and provides each individual with the ability to recognize any antigen, whether exogenous or endogenous, and initiate an immune response. Therefore, the TCR repertoire is shaped throughout an individual's life depending on variable and successive exposures to the antigenic environment they encounter.

In the context of my thesis, my goal was to assess the connection between the composition of the TCR repertoire and its dynamics with the development and progression of cardiovascular diseases.

First, I evaluated the reliability of sequencing technologies that allow for a detailed analysis of the TCR repertoire's composition through the quantification of sequences from thousands of TCRs. During this work, I developed a quality control tool that enabled me to identify a contamination phenomenon related to the sequencing platform utilizing the exclusion amplification technology and devised an algorithm for data detection and decontamination.

Furthermore, I characterized the TCR repertoire in different cardiovascular disease (MCV) situations. Through a national collaboration, I characterized the circulating TCR repertoire in paediatric cases of multisystem inflammatory syndrome following COVID-19. This study was fundamental in confirming the remote superantigen effect of a Sars-CoV2 infection and demonstrating the relevance of TCR sequencing data in addition to clinical observations for diagnostic assistance. In an international collaboration, we had access to blood samples from patients who had experienced a myocardial infarction. The aim of my work was to identify, from the TCRs in these patients' blood, a set of TCRs capable of predicting cardiac repair in these patients. These efforts demonstrated the feasibility of a method to identify a TCR signature distinguishing good from poor repairers.

Finally, in a second cohort of myocardial infarction patients, I explored the relevance of network analysis of TCRs for low-coverage sequencing data. These studies showed that the signatures associated with cardiac repair prognosis were not linked to TCRs known for cardiac specificity. Collectively, these results have helped to better define the possibilities that massive TCR sequencing can offer in the context of cardiovascular diseases.

Keywords : Cardiovascular diseases (CVD), T-cell receptor, T cell repertoire, myocardial infarction, Sars-CoV-2

Modélisation des répertoires des lymphocytes T circulants et son application aux maladies cardiovasculaires

Les maladies cardio-vasculaires MCV représentent la première cause de mortalité en Europe, devant les différents cancers, et sont un enjeu actuel majeur de santé publique. Le système immunitaire est impliqué dans l'étiologie de ces maladies. Parmi les nombreux acteurs du système immunitaire, les lymphocytes T jouent un rôle prépondérant dans l'établissement de ces maladies, leur progression ou la réparation des tissus endommagés. Les lymphocytes T exercent leur fonction suite à leur activation déterminée par leur capacité à reconnaitre de manière spécifique des protéines, ou antigènes, soit exogènes, issues par exemple de pathogènes, allergènes, ou endogènes issues par exemple du renouvellement physiologique cellulaire ou de cellules tumorales. Cette reconnaissance est opérée par un récepteur particulier, appelé le TCR pour T-cell receptor. Contrairement à la majorité des gènes, le TCR est généré par un processus de recombinaison somatique aléatoire entre une centaine de gènes appartenant à trois grandes familles : V, pour Variable, D pour Diversité et J pour Jonction. Ce mécanisme, unique aux gènes codant pour les récepteurs spécifiques d'antigènes, a lieu dans le thymus, au cours du développement et de la différenciation des lymphocytes T et conduit à la production de plusieurs milliards de TCR différents chez un individu. Cette diversité forme le répertoire TCR et confère à chaque individu la capacité de reconnaitre tout antigène, exogène ou endogène, et de mettre en œuvre une réponse immunitaire. Ainsi, le répertoire TCR va être façonné tout au long de la vie d'un individu au gré des expositions variables et/ou successives à l'environnement antigénique auquel il se confronte.

Dans le cadre de ma thèse, j'avais pour objectif d'évaluer quel pouvait être le lien entre la composition du répertoire TCR et sa dynamique avec le développement et la progression s maladies cardio-vasculaires.

D'une part, j'ai évalué la fiabilité des technologies de séquençage permettant l'analyse fine de la composition du répertoire TCR au travers de la quantification des séquences de plusieurs milliers de TCR. Au cours de ces travaux, j'ai développé un outil de contrôle qualité qui m'a permis d'identifier un phénomène de contamination dépendant de la gamme de séquenceur utilisant la technologie d'amplification par exclusion, et de mettre au point un algorithme de détection et de décontamination de nos données.

D'autre part, j'ai caractérisé le répertoire TCR dans différentes situations de MCV. Au travers d'une collaboration nationale, j'ai caractérisé le répertoire TCR circulant de cas pédiatriques du syndrome d'inflammation multiple survenus suite à la COVID-19. Cette étude a été fondamentale pour confirmer l'effet superantigène à distance d'une infection au Sars-CoV2, et de démontrer la pertinence des données de séquençage TCR en complément d'observations cliniques pour l'aide au diagnostic. Dans le cadre d'une collaboration internationale, nous avons eu accès à des prélèvements sanguins de patients ayant eu un infarctus du myocarde. L'objectif de mes travaux était d'identifier, à partir des TCR du sang de ces patients, un ensemble de TCR capables de prédire la réparation cardiaque chez ces patients. Ces travaux ont permis de démontrer la faisabilité d'une méthode d'identification d'une signature de TCR distinguant les bons des mauvais réparateurs.

Enfin, dans une seconde cohorte de patients ayant fait un infarctus du myocarde, j'ai exploré la pertinence de l'analyse en réseau des TCR pour l'analyse de données de séquençage à faible couverture. Ces travaux ont permis de montrer que les signatures associées au pronostic de réparation cardiaques n'étaient pas liées à des TCR connus pour des spécificités cardiaques.

Ensemble, ces résultats ont permis de mieux délimiter les possibilités que peuvent apporter le séquençage massif de TCR dans le contexte des maladies cardiovasculaires.

Mot-clés : Maladies cardiovasculaires (MCV), Répertoire TCR, infarctus du myocarde, Sars-CoV-2