Université Paris Cité

École doctorale 561 : Hématologie, Oncogénèse et Biothérapies

Laboratoire i3 : Immunologie-Immunopathologie-Immunotherapie

Signatures TCR des réponses immunes dans les pathologies auto-immunes et les infections

Par Pierre Barennes

Thèse de doctorat en Biothérapies et biotechnologies

Présentée et soutenue publiquement le 26 septembre 2022

Co-Dirigée par

Pr. David Klatzmann et Dr. Encarnita Mariotti-Ferrandiz

Devant un jury composé de :

Mme Aleksandra Walczak, PhD-HDR, Rapportrice de thèse - Ecole Normale Supérieure

M. Laurent Chiche, DR, Rapporteur de thèse - Hôpital Européen

M. Benjamin Terrier, PU-PH, Examinateur de thèse - Université Paris Cité

M. Jérémie Sellam, PU-PH, Examinateur de thèse - Sorbonne Université

Mme Encarnita Mariotti-Ferrandiz, MCU-HDR, Directrice de thèse - Sorbonne Université

Pr. David Klatzmann, PU-PH, Directeur de thèse - Sorbonne Université

Except where otherwise noted, this is work licensed under
https://creativecommons.org/licenses/by-nc-nd/3.0/fr/

Remerciements

Tout d'abord, je souhaite remercier mon directeur et ma directrice de thèse David et Encarnita, pour leur encadrement.

David, je vous remercie de m'avoir accueilli dans votre laboratoire, de m'avoir donné ma chance et de m'avoir financé pendant ces années de thèse. Votre méthode de travail libérale m'a permis de prendre confiance en moi et vous avez su, je l'espère, en tirer le meilleur parti. Grâce à votre disponibilité, votre vision de la recherche et vos quelques "coups de gueule", j'ai acquis une solide expérience dans le monde de la recherche.

Encarnita, je te remercie de m'avoir accueilli dans ton équipe. Merci de m'avoir accompagné tout au long de cette aventure. Le temps que tu consacres à tes étudiants et à tes projets m'a été d'une aide précieuse. Je ne sais pas de combien d'heures sont faites tes journées, mais encore merci pour ta détermination, ton soutien, ton écoute et tes nombreux conseils tout au long de ma thèse.

Je tiens également à remercier les membres de mon jury de thèse pour leur disponibilité et le temps qu'ils ont consacré à l'évaluation de mon travail. En espérant vous avoir donné satisfaction.

Merci également aux patients qui intègrent des essais cliniques, pour cette démarche altruiste, pour le bien de la science et la santé des futurs malades.

Merci également aux membres du laboratoire :

Les membres du bureau nommés non moins prétentieusement HIF ou l'imprimerie. À tous les quatre (Valentin, Vanessa, Federica et Paul), merci pour ces conversations scientifiques enrichissantes, mais également pour votre amitié, élément essentiel à mes yeux.

Par ordre de rencontre, Valentin, mon ancien encadrant de M2. Merci pour tes conseils, pour ces réflexions, mais aussi pour ces très bons moments passés en dehors

du laboratoire. J'espère que nos discussions autour de bonnes bouteilles de vin se poursuivront !

Vanessa, mon ancienne stagiaire M2, si bien formée que tu es PhD avant moi ! Quelle belle expérience, merci pour ton aide (notamment en anglais) et les nombreux fous rires que nous avons eus. À bientôt pour une balade à vélo !

Federica, une bien belle rencontre. Merci pour ton aide en statistiques et nos discussions en dehors du laboratoire, mais également pour tes nombreux conseils. Je n'oublie pas non plus ton soutien lors de notre vol pour la Corse !

Enfin, Paul, mon ancien stagiaire de M2, je te remercie pour ton aide dans tous les projets que nous avons réalisés ensemble. Tu feras sans aucun doute une excellente thèse. Rendez-vous à la fête de l'été avec les Limougeauds !

Je tiens à remercier les autres membres de l'équipe « TriPoD » (Faustine, Gwladys, Hélène, Leslie, Kenz, Nicolas), l'équipe clinique (Michelle R., Michèle B-S, Roberta, Claire), les chefs (Adrien, Bertrand, Stéphanie, Guillaume, Nicolas) ainsi que tous les autres membres du laboratoire. Vous avez toutes et tous contribué(e)s de différentes manières à la réalisation de ces projets.

Merci également à Fabien, Toto, Nico, Raph pour ces beaux moments de distractions autour d'un baby-foot, d'un verre, d'un repas, d'un match, d'un tennis et autres… sans oublier Alexandra, Sophie, James et Pax, encore merci pour votre amitié.

Je remercie également les anciens membres de Husson-Mourier, Wahiba, Nicolas, Phuong et Claude pour leur accueil lors de mon arrivée en M2.

Je remercie chaleureusement mes beaux-parents Pascal et Véronique, ainsi que mon beau-frère Romain, ma belle-sœur Thea et leur enfant Marceau. Vos « elle est longue cette thèse » ont sans aucun doute contribué à ma motivation.

Je remercie tendrement mes parents François et Josette, ma sœur Lucie, mon beau-frère Jean-Rémi et leurs enfants Arthur, Raphaël et Marius, mon frère Matthieu ainsi que mes grands-parents pour leur amour, confiance et soutien indéfectibles et également pour leurs sacrifices afin de mener à bien mon parcours étudiant. Trouvez ici ma profonde reconnaissance et mon amour.

Enfin mes remerciements vont à Morgane, tes encouragements m'ont accompagné tout au long de ces années, merci pour ton écoute et ta compréhension. Ta simple présence m'a été d'une aide précieuse. Trouve ici l'expression de mon amour.

Table des matières

Abréviations *8*
Table des figures *10*
Table des tableaux *11*

Chapitre I. Introduction ***13***
I.1. Les lymphocytes T, soldats du système immunitaire pour le meilleur et pour le pire **17**
I.1.1. Les sous-ensembles lymphocytaires 17
I.1.1.1. Les lymphocytes T helpers ou auxiliaires 17
I.1.1.2. Les lymphocytes T cytotoxiques 19
I.1.1.3. Les lymphocytes T régulateurs 19
I.1.2. Les lymphocytes T dans les maladies auto-immunes 23
I.1.2.1. Le diabète de type 1 25
I.1.2.2. La polyarthrite rhumatoïde 27
I.1.2.3. Le lupus érythémateux systémique 28
I.1.3. Les promesses de l'interleukine-2 à faible dose 30
I.2. Développement, sélections et vie des lymphocytes T **31**
I.2.1. Du thymocyte aux lymphocytes T spécialisés 32
I.2.1.1. La sélection β 33
I.2.1.2. La sélection positive et négative 34
I.2.1.3. L'engagement vers une lignée cellulaire 34
I.2.2. La reconnaissance de l'antigène par le TCR 35
I.2.2.1. La structure du CMH et reconnaissance de l'antigène 36
I.2.2.2. Transmission du signal 37
I.2.3. En périphérie : activation, mémoire immunitaire 38
I.2.3.1. Activation des lymphocytes T 38
I.2.3.2. Le façonnement de l'histoire immunitaire passée et future d'un individu 39
I.3. Le Répertoire des récepteurs des lymphocytes T **40**
I.3.1. Le TCR, un objet biologique complexe 41
I.3.1.1. La structure du TCRαβ 41
I.3.1.2. Organisation génomique des gènes codant les chaines α et β du TCR 42
I.3.1.3. La recombinaison VDJ 43
I.3.1.4. Diversité jonctionnelle 45
I.3.1.5. Diversité d'association des chaines 45
I.3.2. Une grande diversité hautement spécifique, un antagonisme ? 46
I.3.2.1. La réactivité croisée des cellules T 48
I.3.2.2. Un potentiel biais à l'origine des MAI 50
I.4. Le répertoire TCR, un omic source de promesse **51**
I.4.1. De l'échantillon à la banque d'amplicons TCR 52
I.4.2. Les Axes d'analyses du répertoire TCR 56

I.4.2.1. L'Alignement des séquences TCR 57
I.4.2.2. La diversité 57
I.4.2.3. L'usage des gènes du TCR 58
I.4.2.4. La convergence des répertoires TCR 58
I.4.2.5. L'architecture des répertoires TCR 59
I.4.2.6. Modèle génératif de recombinaison VDJ 60
I.4.2.7. La spécificité 61
I.4.3. Une source de biomarqueurs 62
I.4.4. Nécessité de standardisation 63

Chapitre II. Expérimentations et objectifs 67
II.1. Cohortes de patients **70**
II.1.1. Transimmunom : un essai observationnel 70
II.1.2. Lupil-2 : un essai clinique longitudinal 71
II.1.3. Données publiques : Sars-Cov-2 72
II.2. Du prélèvement à l'ARN des cellules **73**
II.3. Création des banques TCR et séquençage de nouvelle génération **74**
II.4. Analyses des données **75**
II.5. Objectifs de thèse **77**

Chapitre III. Résultats 83
III.1. Article 1 - « Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases » **85**
III.2. Article 2 - « Autoimmune disease-specific TCR repertoires in peripheral blood T cells » **125**
III.3. Résultats complémentaires **157**
III.3.1. Le répertoire TCR, un biomarqueur de la réponse au traitement 157
III.3.2. Le répertoire TCR comme biomarqueur pronostique face à la Covid-19 162

Chapitre IV. Discussion et perspectives 167
IV.1. L'importance de la méthode de production **169**
IV.2. Les différents types de réponse immunitaire **173**
IV.3. Discriminer les sous-types cellulaires pour mieux observer **176**
IV.4. De la séquence à la spécificité TCR des cellules T **179**
IV.5. Les thérapies spécifiques d'antigènes **180**
IV.6. Conclusion générale **183**

Références bibliographiques 187
Annexes 223

Abréviations

5'RACE : *rapid amplification of cDNA-ends*

ADN : acide désoxyribonucléique

ADNc : ADN complémentaire

ADNg : ADN génomique

AIRE : *autoimmune regulator*

AIRR : *adaptive immune recepetor repertoire*

AIRR-seq : *adaptive immune receptor repertoire sequencing*

ARN : acide ribonucléique

ARNm : ARN messagers

BCR : *B cell recpetor* ou récepteur des cellules B

CD : *cluster* de différenciation

$CD4^+$ Teffs : LT effecteurs $CD4^+$

$CD4^+$ Tregs : LT régulateurs $CD4^+$

$CD8^+$ Teffs : LT effecteurs $CD8^+$

CMH : complexe majeur d'histocompatibilité

CMV : cytomégalovirus

CPA : cellule présentatrice d'antigène

CPL : circulating pathogenic-like lymphocyte

CSH : cellules souches hématopoïétiques

CTLA-4 : *cytotoxic T-lymphocyte-associated protein 4*

DN : Double négatif

DP : Double positif

DT1 : Diabète de type 1

E.Coli : Escherichia coli

EBV : virus Epstein-Barr

FoxP3 : *Forkhead box protein 3*

HLA: *human leukocyte antigen*

iaTreg : inflammation-associated Treg

IL-2 : Interleukine-2

$IL\text{-}2_{fd}$: IL-2 à faible dose

IPEX : Immunodérégulation de polyendocrinopathie et d'entéropathie lié au chromosome X

LAG-3 : *Lymphocyte Activation Gene-3*

LB : lymphocyte B

LES : Lupus érythémateux systémique

LT : lymphocyte T

MAI : Maladie auto-immune

mPCR : PCR multiplexe

NGS : séquençage de nouvelle génération

NOD : *Non Obese Diabetic*

PBMC : cellules mononucléées du sang périphérique

PCR : polymerase chain reaction

Pgen : Probabilité de génération

Ppost : Probabilité post-sélection

PR : Polyarthrite rhumatoïde

RAG1 : *Recombination Activating Gene 1*

RAG2 : *Recombination Activating Gene 2*

RIN : *RNA Integrity Number*

RSS : Séquences Signal de Recombinaison

SC : Single cell

SEP : Sclérose en plaque

SI : Système immunitaire

SMAP : *supramolecular attack particles*

SP : Simple positif

sPLSDA : *sparse Partial Least Squares Discriminant Analysis*

TCR : T cell receptor ou récepteur des cellules T

TCR-seq : TCR sequencing

Tdt : Terminal deoxynucleotidyl transferase

TGF-β : *Transforming Growth Factor Beta*

Th : LT helper

TNF-α : *tumor necrosis factors alpha*

TRA : *T cell receptor alpha*

TRB : *T cell receptor beta*

UMI : *Unique Molecular Identifier*

VIH : Virus de l'immunodéficience humaine

VS : Volontaire sain

Table des figures

Figure 1 : Mécanismes de suppression des Tregs. *21*
Figure 2 : Processus de développement des cellules T dans le thymus. *33*
Figure 3 : Structure et interaction du complexe CD3-TCR avec le complexe peptide-CMH. *37*
Figure 4 : Activation des cellules T. *38*
Figure 5 : Modification du ratio entre les LT naïfs et mémoires suite à l'involution thymique. *40*
Figure 6 : Structure générale du TCR. *42*
Figure 7 : Organisation et mécanisme du réarrangement V(D)J des gènes du TCR. *43*
Figure 8 : La diversité du répertoire TCR en chiffre. *47*
Figure 9 : Illustration des mécanismes impliqués dans la réactivité croisée des TCR à l'égard des molécules peptides-CMH. *49*
Figure 10 : Principes des protocoles de biologie moléculaire. *54*
Figure 11 : Axes d'analyses du répertoire TCR. *56*
Figure 12 : Présentation des maladies incluses dans l'essai Transimmunom *70*
Figure 13 : Description de l'essai clinique LUPIL-2 ..... *71*
Figure 14 : Protocole de tri cellulaire magnétique. *73*
Figure 15 : Schéma du protocole expérimental de création des librairies TCR. *75*
Figure 16 : Schéma explicatif du rééchantillonnage "Leave-One-Out". *77*
Figure 17 : Modulation de la diversité des répertoires TCR Teff et Treg sous IL-2fd. *158*
Figure 18 : Évaluation de la valeur prédictive des répertoires TCR des Tregs face à la réponse au traitement chez les patients traités par IL-2fd. *159*
Figure 19 : La signature TCR prédit l'issue du COVID-19 ..... *163*
Figure 20 : Illustration d'une cellule aléatoire et d'une cellule à motif. *170*
Figure 21 : Impact de l'ajout d'UMI sur la méthode RACE-3. ..... *172*

Table des tableaux

Tableau 1 : Les découvertes de ces dernières années sur le répertoire des TCR dans le diabète de type 1. ..... *26*
Tableau 2 : Les découvertes de ces dernières années sur le répertoire des TCR dans la polyarthrite rhumatoïde. *28*
Tableau 3 : Les découvertes de ces dernières années sur le répertoire des TCR dans le lupus érythémateux systémique. *29*
Tableau 4 : Agents pathogènes associés aux maladies auto-immunes les plus courantes *50*
Tableau 5 : Bilan des échantillons analysés. *80*

Chapitre I. INTRODUCTION

L'espèce humaine, de par son système immunitaire (SI), possède la compétence de se défendre contre de nombreux agents pathogènes tels que les bactéries, les parasites et les virus. La clé de ces défenses immunitaires réside dans la capacité à discriminer le « soi » du « non-soi », c'est-à-dire de reconnaitre une cible comme intruse et possiblement pathogène (Janeway, 1992).

L'immunologie, discipline s'intéressant à l'étude du SI, est incarnée à l'échelle animale par une combinaison de processus et mécanismes de défense mise en œuvre à travers un ensemble de cellules et molécules au sein d'une machinerie hautement régulée et contrôlée. Ce système est dit robuste, « *s'il est capable de résister, de continuer à fonctionner correctement en dépit d'aléas du monde extérieur et de dysfonctionnement interne* » (Pr. Philippe Kourilsky). Cette définition décrit à mon sens relativement bien ce qu'est le SI.

Mécanisme hérité dès la naissance, il s'avère être autonome et doué d'une grande plasticité, il se module au gré des années suite aux divers contacts avec des éléments environnementaux étrangers au corps façonnant à chacun sa propre histoire immunitaire (Alpert et al., 2019). On distingue deux mécanismes de défense établis, l'un supposé non spécifique, agissant comme première ligne de défense, appelé l'immunité innée, l'autre à contrario hautement spécifique, plus tardif mais durable grâce à un mécanisme de mémoire immunitaire, c'est l'immunité adaptative (Delves and Roitt, 2000).

L'ensemble des animaux jusqu'à certaines plantes possèdent, à des degrés variables de complexité, un système dit de l'immunité innée. Découvert en 1980 par Christiane Nüsslein-Volhard (Nüsslein-Volhard and Wieschaus, 1980) et Jules Hoffmann (Lemaitre et al., 1996), respectivement Prix Nobel de physiologie et de médecine en 1995 et 2011, cette première ligne de défense interne contre les pathogènes est sous le contrôle de nombreux types cellulaires, tels que les monocytes, les macrophages neutrophiles et les cellules dendritiques. À elle seule, l'immunité innée permet d'endiguer la majorité des pathogènes extérieurs (Delves and Roitt, 2000).

Lorsque celle-ci est insuffisante, la réponse immunitaire adaptative se met alors en place. Ce mécanisme découle d'une suite de nombreuses découvertes au cours du 20ème siècle notamment grâce aux travaux de Jacques Miller et Graham Mitchell (Miller and Mitchell, 1968), ainsi que Ralph Steinman (Steinman and Cohn, 1973) co-

lauréat du prix Nobel de physiologie et de médecine en 2011. Ce système entraîne deux types de réponses spécialisées, étroitement liées, permettant la destruction des pathogènes, il s'agit de la réponse humorale et de la réponse cellulaire. La réponse humorale, qui vise à détruire les pathogènes extracellulaires, met en jeu les lymphocytes B (LB), globules blancs qui une fois activés, deviennent alors des plasmocytes qui vont sécréter des anticorps spécifiques. La réponse cellulaire, en amont de la précédente, vise à détruire les agents pathogènes intracellulaires, à l'aide des lymphocytes T (LT), globules blancs qui, une fois activés, auront un rôle soit (i) cytotoxique ; soit (ii) auxiliaire mais aussi ; de (iii) contrôle de ces réponses hautement spécifiques grâce aux lymphocytes T régulateurs. Et c'est au travers de ces trois rôles que ces LT, soutenues par d'autres populations cellulaires et signaux, participent à l'homéostasie de ce SI et lui confèrent sa dite robustesse.

Cette homéostasie est guidée par (i) les fonctions variées et complémentaires des sous-populations de LT ; (ii) l'équilibre entre celles-ci et enfin par (iii) la diversité de reconnaissance antigénique dont elles sont dotées au travers des récepteurs spécifiques de l'antigène. Cette notion de diversité de spécificité a été le fil conducteur qui a animé mes années de thèse.

I.1. Les lymphocytes T, soldats du système immunitaire pour le meilleur et pour le pire

Les LT sont les acteurs centraux de la réponse immunitaire adaptative, dont les principales sous populations sont : les LT effecteurs exprimant le *cluster* de différenciation 4 ou $CD4^+$ ($CD4^+$ Teffs), les LT effecteurs $CD8^+$ ($CD8^+$ Teffs) et les LT régulateurs CD4+ ($CD4^+$ Tregs). Leurs rôles respectifs sont bien distincts, je les détaillerai dans les prochaines sections. De façon générale, il est admis que les Teffs ($CD4^+$ et $CD8^+$) contribuent à l'activation de la réponse immunitaire alors que les $CD4^+$ Tregs sont davantage impliqués dans le contrôle voire l'achèvement de cette même réaction. La protection efficace de l'organisme est assurée par un grand nombre de clones cellulaires T différents généré et capable de déclencher une réponse immunitaire spécifique d'antigènes (fragments moléculaires endogènes ou exogènes). Ce mécanisme de défense est fondé sur le recrutement, la sélection et l'expansion de ses clones T antigènes spécifiques, qui une fois engagés, vont se consacrer à l'élimination des agents pathogènes tels que les virus, les bactéries mais aussi les cellules cancéreuses (Zhang and Bevan, 2011). Spécificité portée par le récepteur que les LT expriment à leur surface, le TCR (*T-cell Receptor*). Néanmoins ce système robuste et complexe peut montrer parfois des failles, c'est le cas, par exemple lorsque celui-ci ne fait plus la distinction entre le « soi » et le « non-soi » ou le « soi-modifié », conduisant les cellules reconnaissant des antigènes du soi d'un individu (donc propres à l'organisme) à s'attaquer aux tissus de ce même individu ; on parle alors de rupture de la tolérance au soi, mécanisme à l'origine des maladies auto-immunes (MAI). À la différence des MAI, le développement des tumeurs et la dissémination des métastases sont corrélés entre autres à un excès de tolérance, médié par les lymphocytes $CD4^+$ Tregs.

I.1.1. Les sous-ensembles lymphocytaires

I.1.1.1. Les lymphocytes T helpers ou auxiliaires

Les LT auxiliaires ou *T helper cells* (Th), constituent la sous-population de $CD4^+$ Teffs impliqués dans l'orchestration de la réponse immunitaire en activant d'autres cellules tels que les macrophages, les $CD8^+$ Teffs ou encore les LB. Les Th se distinguent en plusieurs sous-familles dont la différenciation se fait suite à la reconnaissance de

l'antigène et les signaux d'activation produits par les cellules présentatrices d'antigène. Les cellules Th naïves sont activées, subissent une différenciation et sécrètent des cytokines participant à la promotion et la polarisation de la réponse immunitaire. On distingue aujourd'hui 3 populations Th différentes : (i) les cellules Th1 jouent un rôle important dans la protection de l'hôte contre les agents pathogènes intracellulaires. Leur fonction est de promouvoir l'immunité à médiation cellulaire (Mosmann and Coffman, 1989) se caractérisant par une activité cytolytique via des cytokines pro-inflammatoires permettant l'activation des macrophages et la prolifération des $CD8^+$ Teffs. Néanmoins, une activité pro-inflammatoire excessive peut endommager les tissus et provoquer des maladies inflammatoires indésirables, une auto-réactivité, ainsi que des troubles auto-immuns tels que le diabète de type 1 (Wang et al., 1997) et la polyarthrite rhumatoïde (Leung et al., 2000) ; (ii) les cellules Th2 sont davantage associées à la réponse humorale (Mosmann and Coffman, 1989) et jouent un rôle important dans la résistance aux agents pathogènes extracellulaires, notamment via l'activation des LB permettant la production d'anticorps (ou plasmocytes). Une réponse Th2 aberrante conduit souvent à des maladies inflammatoires chroniques des voies respiratoires, telles que l'asthme et l'allergie (Robinson et al., 1992) et (iii) les cellules Th17 découvertes plus récemment ont des propriétés pro-inflammatoires permettant le contrôle des bactéries extracellulaires. Toutefois, elles sont aussi impliquées dans l'induction et la propagation de l'auto-immunité telle que la sclérose en plaques et la polyarthrite rhumatoïde (Langrish et al., 2005; Noack and Miossec, 2014).

Lors de l'activation des Th naïfs, bien que l'environnement cytokinique détermine leurs différenciations (Zhu et al., 2010), le TCR exprimé à leur surface pourrait également jouer un rôle important dans la polarisation de la réponse immunitaire (Baumgartner et al., 2012). Nicholas van Panhuys (van Panhuys, 2016) a proposé que l'intensité du signal TCR contrôle de manière dominante la polarisation des cellules Th en régulant l'expression des récepteurs de cytokines importants pour la polarisation vers les sous-ensembles Th. Lorsque des cellules $CD4^+$ Teffs naïves sont soumises à une forte stimulation du TCR, la différenciation des cellules Th1 est favorisée, à l'inverse, de faibles signaux TCR permettraient la différenciation des cellules Th2 (Brogdon et al., 2002). Bien que ce modèle ne soit, pour l'heure, toujours qu'hypothétique il permettrait d'expliquer comment la reconnaissance d'auto-antigènes entraîne une MAI. Au cours du développement de l'auto-immunité, une forte

signalisation TCR peut se produire en raison de la grande disponibilité d'auto-antigènes dans l'organisme et ainsi contribuer au développement d'un phénomène auto-immun. Ces cellules auto-réactives présentes en périphérie et ayant une grande affinité pour le soi vont s'activer et présenter un profil Th1 et/ou Th17.

I.1.1.2. Les lymphocytes T cytotoxiques

Les lymphocytes T cytotoxiques constituent la population de cellules $CD8^+$ Teffs responsables de la protection contre les agents pathogènes intracellulaires, notamment les virus et les bactéries (Harty et al., 2000), ainsi que la surveillance des tumeurs (Yee et al., 2002). Les $CD8^+$ Teffs émigrent du thymus en tant que cellules naïves et sont activées en cellules tueuses suite à la reconnaissance d'un antigène de façon très spécifique dans les organes lymphoïdes secondaires. Plus précisément, elles disposent de trois mécanismes principaux pour détruire les cellules infectées ou malignes. Le premier est la production de perforines et de granzymes, permettant de perforer la membrane et d'induire l'apoptose de la cellule cible (Froelich et al., 1996; Lopez et al., 2013). Le deuxième mécanisme consiste à activer le processus d'auto-apoptose des cellules. Un contact direct entre les récepteurs de mort Fas, à la surface de la cellule cible, et Fas-Ligan, à la surface des $CD8^+$ Teffs activés, entraine l'activation de la cascade des caspases et provoque l'apoptose de la cellule cible (Tamzalit et al., 2019). Les $CD8^+$ Teffs pouvant exprimer les deux récepteurs, ce mécanisme a également pour rôle d'éliminer les cellules effectrices du SI pendant la phase de contraction à la fin d'une réponse immunitaire. Enfin, le troisième et dernier mécanisme, mis en avant récemment, montre que les $CD8^+$ Teffs sécrètent des *supramolecular attack particles* (SMAP) contenant des molécules précédemment décrites, à savoir des perforines et des granzymes protégées par une enveloppe, faisant de ces entités des particules tueuses extracellulaires autonomes (Bálint et al., 2020).

I.1.1.3. Les lymphocytes T régulateurs

Les lymphocytes T régulateurs (Tregs) constituent une sous-population unique de cellules T, exprimant majoritairement le corécepteur CD4 (bien que certains Tregs expriment le CD8) essentielle pour maintenir l'homéostasie immunitaire et prévenir les désordres auto-immuns, les allergies ainsi que le rejet de greffe (Sakaguchi et al., 2020). Cette sous-population cellulaire se caractérise notamment par l'expression du récepteur membranaire de haute-affinité à l'interleukine-2, le CD25 (Sakaguchi et al., 1995) ainsi que l'expression intracellulaire de FoxP3 (*Forkhead box protein 3*), facteur

de transcription reconnu comme le régulateur principal des fonctions suppressives des Tregs et contrôlant l'expression d'un large éventail de gènes, y compris les cytokines et les molécules de surface (Fontenot et al., 2003, p. 3; Hori et al., 2003; Yagi et al., 2004). Il existe des Tregs provenant de différentes origines. Les Tregs thymiques (tTregs), initialement nommés Tregs naturels, sont générés au niveau du thymus alors que les Tregs périphériques (pTregs) ou induits (iTregs) proviennent de LT $CD4^+$ naïfs qui, en présence de TGF-β (*Transforming Growth Factor Beta*) accompagné d'une stimulation du TCR expriment FoxP3 et acquièrent des fonctions régulatrices (Chen et al., 2003; Kretschmer et al., 2005). La perturbation de la fonction des Tregs engendre une pléthore de pathologies auto-immunes et inflammatoires, illustrant le rôle indispensable des Tregs dans le maintien de la tolérance immunologique aux auto-antigènes et dans la suppression des réponses excessives nuisibles à l'hôte (Kim et al., 2007). Elles exercent cette régulation via divers mécanismes qui peuvent être regroupés en quatre "modes d'action" de base (ici vous sera présenté un aperçu des mécanismes fondamentaux permettant aux Tregs d'exercer leurs fonctions suppressives) :

(i) Suppression par des cytokines inhibitrices (**Figure 1a**) : ce mécanisme fait intervenir 3 cytokines, l'IL-10, qui permet d'inhiber la fonction de présentation des antigènes des cellules présentatrices d'antigènes (CPA ; Mittal and Roche, 2015) ainsi que la réponse pathogène des Th17 (Chaudhry et al., 2011), le TGF-β qui, exprimé sous forme membranaire par Tregs exerce une immunosuppression via une interaction cellule-cellule (Nakamura et al., 2001) et produit sous forme soluble par les Tregs conduit à l'induction des LT $CD4^+$ naïves en iTregs et au maintien de l'expression de FoxP3 (Li et al., 2006), enfin l'IL-35, exprimée essentiellement par les Tregs, est nécessaire pour une activité suppressive maximale (Collison et al., 2007).

(ii) Suppression par cytolyse (**Figure 1b**), via un mécanisme direct de production de granzymes et de perforines (Grossman et al., 2004), qui auraient des effets cytolytiques, conduisant à l'apoptose des cellules T effectrices ($CD4^+$ et $CD8^+$), des LB (Zhao et al., 2006) ainsi que des CPA (Boissonnas et al., 2010).

(iii) Suppression par perturbation métabolique (**Figure 1c**), via divers mécanismes touchant principalement la disponibilité d'une cytokine, l'interleukine-2 (IL-2) essentielle à la prolifération et au fonctionnement des LT, dont nous parlerons

dans la section I.1.3. Cette perturbation se fait grâce à une consommation préférentielle de l'IL-2 par les Tregs (Pandiyan et al., 2007) mais aussi par une diminution de la production d'IL-2 des Teffs (Bopp et al., 2007).

(iv) Suppression par modulation de l'activité des CPA (**Figure 1d**). Le CTLA-4 (*cytotoxic T-lymphocyte-associated protein 4)*, exprimé fortement par les Tregs, est un récepteur inhibiteur qui interagit avec le CD80/CD86 des CPA induisant l'élimination de ces molécules de costimulation nécessaires à l'activation des LT (Qureshi et al., 2011) et l'expression d'indoleamine 2,3-dioxygenase, molécule régulatrice ayant une action suppressive indirecte sur les Teffs (Mellor and Munn, 2004; Oderup et al., 2006). De plus, il semblerait que LAG-3 (*Lymphocyte Activation Gene-3*) exprimé à la surface des Tregs, bloquerait la maturation des DC via son interaction avec le Complexe Majeur d'Histocompatibilité de classe 2 ou CMH-II (section I.2.2.1 ; Huang et al., 2004).

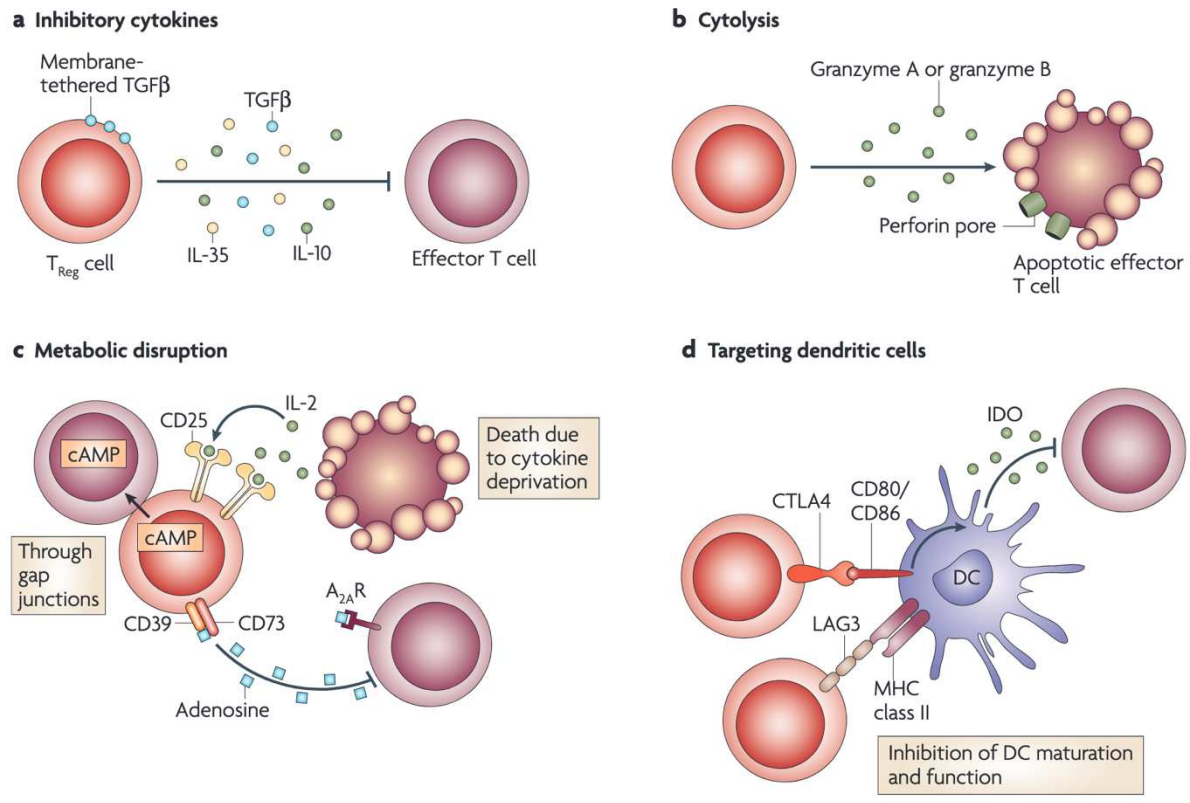


Figure 1 : Mécanismes de suppression des Tregs.
(Issu de Vignali D., Collison L. and Workman C., How regulatory T cells work. Nat Rev Immunol, 2008)

Un temps classé comme naturellement anergique, dû à l'absence de prolifération via une stimulation de leurs TCR in vivo (Takahashi et al., 1998; Thornton and Shevach, 1998), certains des mécanismes de suppression présentés ci-dessus sont dépendants d'une reconnaissance spécifique d'antigènes, et donc d'activation par le TCR. Ceci a été bien illustré par Seddon et al. qui ont montré que les $CD4^+$ Tregs permettant de prévenir l'auto-immunité ne sont générés qu'en présence de l'auto-antigène en question. En effet, le transfert adoptif de $CD4^+$ Tregs provenant d'un rat donneur ayant subi *in utero* une thyroïdectomie, n'étaient plus en mesure de prévenir de l'apparition d'une inflammation auto-immune chronique de la thyroïde (thyroïdite) chez un rat receveur. En revanche ces $CD4^+$ Tregs demeuraient capables de prévenir des atteintes auto-immunes touchant d'autres organes, suggérant que les $CD4^+$ Tregs ont besoin de rencontrer leurs auto-antigènes spécifiques pour survivre et être fonctionnels (Seddon and Mason, 1999). Walker et al. ont quant à eux montré que des $CD4^+$ Tregs exprimant un TCR transgénique prolifèrent et s'accumulent localement en réponse à un antigène tissulaire exprimé par voie transgénique, alors que leurs homologues $CD4^+$ Teff sont appauvris dans ces sites (Walker et al., 2003). Des études ultérieures ont également montré que (i) le blocage des signaux par l'intermédiaire du TCR entraine une distribution anormale des Tregs au travers des organes lymphoïdes ainsi qu'une absence de prolifération (Kim et al., 2009) ; (ii) via des expériences d'ablation des TCR à la surface de Tregs matures, l'expression de ce dernier est critique pour le maintien et la différenciation des cellules compromettant ainsi leurs capacités fonctionnelles (Vahl et al., 2014) et (iii) qu'après un transfert adoptif, les Tregs spécifiques des tissus se divisent et se développent préférentiellement dans les ganglions lymphatiques drainant leurs auto-antigènes cibles (Fisson et al., 2003). Enfin de nombreuses études sur les greffes ont montré un recrutement de Tregs spécifiques, des Tregs alloréactifs, au niveau du site d'expression de l'alloantigène (Adeegbe et al., 2010, 2006; Joffre et al., 2004; Nagahama et al., 2007). Finalement, il a été montré dans le cadre du diabète de type 1, que les Tregs spécifiques avaient une capacité de suppression supérieure aux Treg polyclonéaux permettant une rémission chez l'ensemble des souris (Cabello-Kindelan et al., 2020; Tang et al., 2004a). Ensemble ces données illustrent le lien fort entre les fonctions suppressives des Tregs et leurs spécificités. Nous allons donc maintenant aborder plus en détail le rôle des $CD4^+$ Tregs dans la survenue des MAI, en se focalisant sur trois pathologies majeures :

le diabète de type 1 (DT1), la polyarthrite rhumatoïde (PR) et le lupus érythémateux systémique (LES). Ces trois pathologies se distinguant par les tissus touchés et par leur dynamique.

I.1.2. Les lymphocytes T dans les maladies auto-immunes

Environ 4 % de la population des pays industrialisés est touchée par une MAI, et leur fréquence semble toujours en augmentation. Ces maladies chroniques sont graves et invalidantes, la grande majorité ont un développement relativement lent, rendant le diagnostic tardif et ainsi les traitements inadaptés renvoyant une guérison définitive compliquée. Certaines de ces maladies sont spécifiques d'organe, comme le diabète de type 1 et la polyarthrite rhumatoïde, tandis que d'autres sont systémiques, comme le lupus érythémateux systémique.

Ainsi, les MAI regroupent l'ensemble des maladies dans lesquelles le SI se retourne contre l'hôte et perçoit le « soi » comme un élément étranger, potentiellement pathogène qu'il doit éliminer. Malgré l'implication des LT dans la pathogenèse des MAI, leurs origines sont complexes et résultent d'association de plusieurs facteurs, qu'ils soient, (i) environnementaux, tel que le régime alimentaire (Jörg et al., 2016; Manzel et al., 2014) ; (ii) infectieux, dont nous reparlerons par la suite (Bach, 2005; Bjornevik et al., 2022) ; (iii) hormonaux avec un déséquilibre de genre dans l'incidence des MAI (Cutolo et al., 2006; Rubtsov et al., 2010) mais aussi (iv) génétiques avec des prédispositions de certaines populations, notamment des associations avec certains haplotypes du CMH ou *Human Leucocyte Antigen* (HLA) de classe I (HLA-I) et/ou de classe II (HLA-II) (van Lummel et al., 2012). En effet, les molécules du HLA interagissant directement avec le TCR représentent un biais majeur dans l'incidence des MAI (Gough and Simmonds, 2007). Il est admis que la majorité des LT auto-réactifs sont éliminés lors de la sélection et de la différenciation thymique. Cependant, ce processus de sélection négative, aussi appelé tolérance centrale, n'est pas infaillible (Legoux et al., 2015) et certaines cellules T effectrices potentiellement auto-réactives émergent tout de même du thymus pour circuler en périphérie. Ces cellules peuvent être générées et activées par des mécanismes directement liés à la présentation de peptides via le HLA comme par exemple (i) l'instabilité complexe HLA-peptide, qui au niveau du thymus serait responsable d'un faible signal assurant la sélection positive

(section I.2.1.2) mais qui n'atteindrait pas le seuil requis pour la sélection négative (section I.2.1.2) ; (ii) une présentation altérée des peptides du soi par certains allèles HLA conduisant à une sélection incomplète de Tregs, dans ce cas-là l'apparition de la MAI serait associée à un défaut de protection des Tregs spécifiques des antigènes du soi.

Ces biais mécanistiques ainsi que ceux illustrés en section I.3.2 peuvent modifier le répertoire TCR des LT ou changer la façon dont l'antigène est reconnu en périphérie.

Malgré ces potentiels biais, le SI est complexe et robuste. En effet, il existe un mécanisme connu sous le nom de tolérance périphérique. Dans un SI à l'homéostasie, les lymphocytes T auto-réactifs sont contrôlés par les Tregs et ne déclenchent donc pas de MAI (Sakaguchi et al., 1995). Ce contrôle par les Tregs est bien illustré par plusieurs observations : (i) un déficit génétique en Tregs entraine de multiples MAI spécifiques d'organes. En effet, les patients atteints du syndrome d'immunodérégulation de polyendocrinopathie et d'entéropathie lié au chromosome X (IPEX) ont hérité d'un gène *foxp3* muté, ce qui entraîne une quasi-absence de cellules Tregs. En l'absence de ces cellules Tregs en périphérie, les cellules T auto-réactives qui ont échappé à la tolérance centrale dans le thymus ne sont pas contrôlées, ce qui entraîne une maladie auto-immune systémique (Moraes-Vasconcelos et al., 2008) ; (ii) dans un modèle murin d'auto-immunité multi-organes provoqué par thymectomie, il est possible de maintenir la tolérance périphérique grâce à l'injection de cellules Tregs, maintenant ainsi les LT auto-réactifs potentiellement pathogènes dans un état anergique (Asano et al., 1996) ; (iii) un défaut quantitatif et/ou qualitatif en Tregs est décrit dans de nombreuses MAI (Bonelli et al., 2008; Boyer et al., 2004; Kukreja et al., 2002; Lindley et al., 2005; Noack and Miossec, 2014) et enfin (iv) une déplétion efficace des Tregs chez les individus sains, à tout moment de la vie, induit une réponse auto-immune et inflammatoire conduisant à une auto-immunité multi-organes (Kim et al., 2007). Cet ensemble d'observations a conduit au concept de balance d'immuno-régulation entre ces deux sous-populations de lymphocytes qui serait déséquilibrée dans le cadre des MAI.

Une connaissance plus détaillée des mécanismes conduisant à une auto-immunité destructrice est fortement nécessaire, afin de développer de meilleures méthodes pour prévenir et traiter cet important groupe de maladies. Les LT auto-

réactifs et le déséquilibre entre les $CD4^+$ Teffs et les $CD4^+$ Tregs étant potentiellement impliqués d'une part dans les stades précoces de l'induction des MAI mais également dans le maintien à long terme et leur progression, ils représentent de bons candidats de *monitoring* du SI des patients. Ainsi, la compréhension du répertoire des LT au site de l'inflammation comme dans le sang périphérique des patients pourrait permettre d'identifier des biomarqueurs de la maladie et contribuer au développement de thérapies pour prévenir l'apparition de la maladie. (Pour les sections I.1.2.1, I.1.2.2 et I.1.2.3, certaines notions spécifiques de l'étude du répertoire TCR vont être abordées, si nécessaire voir section I.3).

I.1.2.1. Le diabète de type 1

Le DT1 est l'une des MAI chronique les plus courantes. Elle se caractérise par la destruction des cellules productrices d'insuline du pancréas, les cellules β des îlots de Langerhans. Cette hypo-insulinémie a pour conséquence une augmentation des niveaux de glucose dans le sang traitée via une dépendance à vie à l'insuline exogène. Bien qu'elle soit généralement diagnostiquée pendant l'enfance, la maladie peut rester indétectable durant un certain temps, entraînant une altération définitive des tissus pancréatiques.

L'implication des LT dans le DT1 est maintenant relativement bien établie. En effet, plusieurs études appuient cette hypothèse de travail notamment (i) la présence d'infiltrat de lymphocytes au niveau des ilots de Langerhans des patients souffrant de DT1 (Bottazzo et al., 1985; Rodriguez-Calvo et al., 2014) ; (ii) le ralentissement de la progression de la maladie grâce à des immunosuppresseurs spécifiques des cellules T (Bougneres et al., 1988) et enfin (iii) des études chez la souris NOD (*Non Obese Diabetic*, souche de souris génétiquement modifiée permettant l'apparition d'un diabète insulinodépendant) ont démontré qu'un transfert adoptif de LT $CD4^+$ spécifiques de protéines des îlots pancréatiques était suffisant pour initier et accélérer le processus pathologique chez de jeunes souris NOD (Haskins and McDuffie, 1990). Ce phénomène a également été observé chez l'homme suite à un transfert adoptif de moelle osseuse entre deux membres d'une fratrie dont le donneur était diabétique (Lampeter et al., 1993), alors que de nombreuses autres transplantations de ce type n'ont pas donné lieu à ce phénomène avec de la moelle osseuse appauvrie en LT.

Pour ce qui est des connaissances actuelles du répertoire TCR dans le cadre du DT1, il a été montré que la diversité du répertoire des $CD4^+$ et des $CD8^+$ du sang était diminuée dû à l'apparition d'expansions majeures (Tong et al., 2016). En revanche pour ce qui est des cellules Tregs, nous avons observé chez la souris NOD une diminution des expansions des cellules Tregs activées mémoires pouvant expliquer l'apparition de la maladie (voir Annexe 1, Mhanna et al., 2021). Il est également important de noter que chez l'homme, plusieurs équipes ont identifié un partage plus important de certaines expansions clonales au niveau des îlots de Langerhans et du pancréas entre les patients DT1 en comparaison aux donneurs sains ; et celles-ci étaient détectables dans le sang des mêmes individus, indiquant une recirculation entre les tissus par le sang périphérique appuyant la possibilité de suivre les clones spécifiques de la maladie (Codina-Busqueta et al., 2011; Seay et al., 2016). Enfin, Gomez-Tourino et al. ont remarqué des anomalies dans la distribution de longueur des régions CDR3 (*Complementary Determining Region 3*) du TCR chez les patients DT1, en particulier des CDR3 plus courts et plus fréquemment partagés (Gomez-Tourino et al., 2017). En conclusion, Nakayama et Michels soulignent la nécessité d'identifier des biomarqueurs TCR qui définissent l'activité de la maladie via l'expression TCR des cellules T. Ils mettent en garde la difficulté d'identifier des TCR fiables impliqués dans la pathologie et soulignent que celle-ci passera notamment par l'intégration d'un grand nombre de données ainsi que par l'utilisation du *Big-Data* (Nakayama and Michels,

			Main discovery					
Author (year)	Methods	Samples	TCR repertoire diversity	Related to disease activity	Biased V-usage	Islet-specific T cell clones	Preferential TCR genes or structures	References
Gomez-Tourino et al. (2017)	Sequencing	$CD4^+$ T Cells in peripheral blood	Reduced	Unclear	Unclear	Unclear	Unclear	97]
Cerosaletti et al. (2017)	HTS	$CD4^+$ T Cells in peripheral blood	Unclear	Yes	Yes	$CD4^+$ memory T cells	TRBV5.1 et al.	[98]
Seay et al. (2016)	HTS	$CD4^+$/ $CD8^+$ T cells in PBMC and tissues	Reduced	Unclear	Yes	$CD4^+$ T cell but not $CD8^+$ T cell	Unclear	[99]
Marrero et al. (2016)	HTS	$CD4^+$ T cells in the pancreatic lymph nodes	Reduced	Unclear	Yes	Unclear	TRBV13.2; Vβ8.2	[100]
Tong et al. (2016)	HTS	$CD4^+$ T cells/$CD8^+$ T cells in PBMC	Reduced	Unclear	Yes	Unclear	TRBV15 et al.	[101]
Eugster et al. (2015)	HTS	$CD4^+$ T cells in blood	Unclear	Unclear	Yes	GAD65 autoantigen-specific human $CD4^+$ T cells	TRBV5.1	[102]
Marrero et al. (2013)	HTS	$CD4^+$ T cells in islets	Unclear	Unclear	Yes	$CD4^+CD44^{high}$ memory T cells	TRBV1, TRBV13.3, TRBV19; Vβ2, Vβ6, Vβ8.1	[66]
Diz et al. (2012)	Flow cytometry	$CD4^+$ /$CD8^+$ T Cells in islet	Reduced	Unclear	Yes	$Vβ12^+$effector/ memory $CD4^+$T cells	Vβ12	[103]

HTS, high throughput sequencing; PBMC, peripheral blood mononuclear cells; TRBV, TCR β chain V region genes; Vβ, TCR β chain V region structures.

Tableau 1 : Les découvertes de ces dernières années sur le répertoire des TCR dans le diabète de type 1.
(Issu de Song R., Jia X., Zhao J., Du P. and Zhang J.A., T cell receptor revision and immune repertoire changes in autoimmune diseases. International Reviews of Immunology, 2021)

2021). Les principaux résultats de ces dernières années sont résumés dans le **Tableau 1.**

I.1.2.2. La polyarthrite rhumatoïde

La PR est également une MAI chronique caractérisée par une inflammation et une destruction de la membrane synoviale de l'articulation entraînant la destruction de l'os et du cartilage. Le diagnostic demeure compliqué, étant établi grâce à l'association de différents facteurs : un examen médical mettant en évidence une polyarthrite, une prise de sang caractéristique d'un syndrome inflammatoire combinée à la détection d'auto-anticorps, ainsi qu'un examen d'imagerie médicale. La PR se caractérise également par des cycles de poussée inflammatoire et de rémission. Les traitements actuels, essentiellement des corticostéroides et des immunothérapies ciblant différentes cytokines inflammatoires (anticorprs anti-TNF, anticorps anti-IL6), consistent néanmoins principalement à traiter les symptômes, contrôler les poussées et la destruction tissulaire. Ces traitements sont également proposés à vie, car non curatifs, ce qui nécessite un suivi régulier et implique une atteinte de la qualité de vie des patients sur le long terme.

L'implication des cellules T dans la pathologie est bien établie grâce à des observations similaires à la section précédente, (i) la présence d'infiltrat de lymphocytes au niveau de la synovie articulaire (Ley et al., 2007; Nourshargh and Alon, 2014; Van Boxel and Paget, 1975) ; (ii) chez le rat, la déplétion des LT $CD4^+$ avant l'apparition de la maladie provoque une arthrite légère et une résistance relative à l'induction de la pathologie (Van den Broek et al., 1992) et (iii) enfin, Pearson et Wood ont démontré que le transfert de cellules de rate ou de ganglions lymphatiques de rats arthritiques à des receveurs sains pouvait induire la maladie (Pearson and Wood, 1964). De façon plus précise, il a été montré que les LT $CD4^+$ avaient un rôle majeur dans l'induction et le maintien de la maladie (Kadowaki et al., 1994; Wong et al., 2006).

Concernant les connaissances du répertoire TCR dans le cadre de la PR, il est tout d'abord intéressant d'observer un nombre de Tregs trois fois supérieur chez les souris ayant développé une PR en comparaison avec des souris contrôles, indiquant un défaut qualitatif potentiellement de la spécificité des Tregs (Rankin et al., 2008). Les infiltrats des lésions synoviales comprennent majoritairement des cellules T $CD4^+$, caractérisées par un répertoire TCR oligo-clonal et une distribution réduite des gènes TRBV (Waase et al., 1996). D'autres études ont notamment mis en évidence un

Author (year)	Methods	Samples	Main discovery					
			TCR repertoire diversity	Related to disease activity	Biased V-usage	Synovium\ SF-specific T cell clones	Preferential TCR genes	References
Jiang et al. (2020)	HTS	$CD4^+$ T cells in blood	Unclear	Yes	Unclear	Yes	Unclear	[81]
Liu et al. (2019)	HTS	T cells in PBMC	Higher TCR heterogeneity	Yes	Yes	Yes	Unclear	[75]
Sakurai et al. (2018)	HTS	Naive and memory $CD4^+$ T cells in blood	Reduced in memory $CD4^+$ T cells	Yes	Yes	Unclear	Unclear	[82]
Spreafico et al. (2016)	HTS	$CD4^+$ T cells in blood and synovium	Reduced	Yes	Yes	Yes	Unclear	[6]
Zhou et al. (2014)	PCR	T cells in blood and SF	Unclear	Unclear	Yes	Yes	Unclear	[83]
Klarenbeek et al. (2012)	HTS	T cells in blood and synovium	Reduced	Yes	Yes	Yes	TRBV6.5, TRBV5.6 and TRBV29.1 et al.	[84]

HTS, high throughput sequencing; SF, Synovial fluid; TRBV, TCR β chain V region genes; Vβ, TCR β chain V region structures.

Tableau 2 : Les découvertes de ces dernières années sur le répertoire des TCR dans la polyarthrite rhumatoïde.
(Issu de Song R., Jia X., Zhao J., Du P. and Zhang J.A., T cell receptor revision and immune repertoire changes in autoimmune diseases. International Reviews of Immunology, 2021)

enrichissement significatif de cellules T exprimant le gènes TRBV25 chez les patients PR, résultat confirmé via des essais de stimulation avec du collagène (Di Sante et al., 2015; Zhou et al., 2014). Jiang et al. ont montré que les patients atteints de PR partagent des caractéristiques communes, comme des expansions et contractions clonales de certaines familles de gènes TRBV et notamment, une surexpression du gène TRBV3-1 dans les cellules $CD4^+$ retrouvées au niveau du sang et du liquide synovial (Jiang et al., 2020). Enfin, il a été observé un partage important entre les clones expandus provenant de différentes articulations inflammées (Klarenbeek et al., 2012). Finalement, une étude récente identifie des clones enrichis dans les patients PR en comparaison à des volontaires sains (Liu et al., 2019). Ensemble, ces biais d'usage de gènes et ces partages de clones détectés aussi bien au niveau du sang périphérique que dans différents tissus inflammés suggèrent que les LT auto-réactifs recirculeraient, appuyant ainsi l'utilisation du répertoire TCR comme source de biomarqueurs au niveau du sang de patients atteints de PR. Les principaux résultats de ces dernières années sont résumés dans le **Tableau 2**.

I.1.2.3. Le lupus érythémateux systémique

Le LES est une MAI chronique systémique, dont les manifestations cliniques varient fortement d'une personne à une autre en fonction des organes touchés (les reins, le cœur, les poumons, les cellules du sang, ...) rendant le diagnostic long et difficile à établir. Le LES est caractérisé par la production d'auto-anticorps dirigés contre de nombreux auto-antigènes, notamment ceux dirigés envers les acides

nucléiques, constituant ainsi la principale caractéristique utilisée pour le diagnostic. Là encore la maladie reste à l'heure actuelle incurable et nécessite un traitement et un suivi adapté à chaque patient en fonction de la gravité de ses symptômes et de leur évolution.

Ici l'implication et le rôle central que jouent les lymphocytes T sont démontrés suite à l'étude des tissus affectés, obtenus via des biopsies ou autopsies, montrant des infiltrats importants de LT au sein de ces organes (D'Agati et al., 1986 ; Simpson et al., 2010 ; Tsokos, 2011). L'implication est aussi soutenue par le fait que la maladie peut être induite par transfert adoptif de cellules T dans certains modèles murins (Quddus et al., 1993; Richardson et al., 2012) ainsi que par la corrélation entre l'activité de la maladie et la proportion de cellules CD4+ dans le sang (Choi et al., 2015). Finalement, de nombreuses observations soutiennent l'utilité de l'étude du répertoire TCR dans le cadre du LES. Ainsi, Costa et al ont mis en évidence chez des patients traités par infusion intraveineuse d'immunoglubulines polyclonales des expansions oligo-clonales des LT chez les patients LES, une restauration de la diversité du répertoire TCR et une augmentation de fonctionnalités des Tregs corrélée à une amélioration

			Main discovery				
Author (year)	Method	Samples	TCR repertoire diversity	Related to disease activity	Biased V-usage	Preferential TCR genes or structures	Reference
Ye et al. (2020)	HTS	T cells in PBMC	Reduced	No	Unclear	TRBV/TRBJ genes	[74]
Liu et al. (2019)	HTS	T cells in PBMC	Higher TCR heterogeneity	Yes	Yes	Unclear	[75]
Jakez-Ocampo et al. (2018)	Flow Cytometry	CD3+ lymphocytes	Unclear	Unclear	Unclear	Vβ5.2, Vβ11, Vβ16, Vβ3, Vβ4, Vβ7.1, Vβ17, Vβ9 in $CD3^+$ lymphocytes	[76]
Yu et al. (2017)	HTS	T cells in PBMC	Changed after high-dose glucocorticoid therapy	Yes	Yes	TRBV2, TRBV19, TRBV28 et al.	[73]
Li et al. (2015)	HTS	T cells in Thymus and spleen	Reduced	Yes	Yes	Thymus: TRBV26, TRBV17 et al Spleen:TRBV13.1, TRBV26 et al	[71]
Thapa et al. (2015)	HTS	T cells in PBMC	Reduced	No	Yes	Unclear	[72]
Sui et al. (2015)	HTS	T cells in PBMC	Comparable	Unclear	Yes	TRBV10.2, TRBV23.1 et al	[79]
Alexander et al. (2013)	Flow cytometry	$CD4^+$ T cells in blood	Reduced in $Foxp3^+$ $Helios^+$ Tregs	Yes	Yes	Unclear	[69]
Winchester et al. (2012)	Flow cytometry and PCR	$CD4^+$\$CD8^+$ T cells in blood and renal	Reduced	Yes	Yes	TRBV7 in $CD8^+$ T cells	[68]
Costa et al. (2013)	PCR	T cells in PBMC	Reduced after intravenous IgG therapy	Unclear	Yes	Vβ1, Vβ2, Vβ11 and Vβ14	[70]
Tzifi et al. (2013)	Flow cytometry	$CD4^+$ T cells in blood	Unclear	Unclear	Yes	Vβ16 in $CD4^+$T cells	[78]

HTS, high throughput sequencing; PBMC, peripheral blood mononuclear cells; TRBV, TCR β chain V region genes; TRBJ, TCR β chain joining genes; Vβ, TCR β chain V region structures.

Tableau 3 : Les découvertes de ces dernières années sur le répertoire des TCR dans le lupus érythémateux systémique.
(Issu de Song R., Jia X., Zhao J., Du P. and Zhang J.A., T cell receptor revision and immune repertoire changes in autoimmune diseases. International Reviews of Immunology, 2021)

clinique (Costa et al., 2013). D'autres études ont identifié des variations d'usages de gènes TRBV au niveau du sang de patients LES en comparaison avec des contrôles sains ainsi que des TCR privés aux patients LES (Alexander et al., 2013; Sui et al., 2015). Finalement, une étude récente a identifié des clones enrichis dans les patients LES en comparaison à des volontaires sains (Liu et al., 2019). Les principaux résultats de ces dernières années sont résumés dans le **Tableau 3**.

I.1.3. Les promesses de l'interleukine-2 à faible dose

L'IL-2 est une cytokine auto- et paracrine immuno-modulatrice découverte en 1976 par Morgan et al. (Morgan et al., 1976). Initialement utilisée dans le cadre des cancers, cette cytokine entraine, à forte dose, une prolifération des LT effecteurs $CD4^+$ et $CD8^+$ permettant d'éliminer les cellules cancéreuses (Rosenberg et al., 1985). Cependant, ce traitement à forte dose a entrainé des effets indésirables majeurs, comme des hémorragies sévères (Boyman et al., 2006), compromettant le prognostic vital des patients traités et ayant conduit à la suspension de ce traitement par les autorités de sécurité des médicaments. C'est en 1993 que Sadlack et al. ont montré qu'un déficit en IL-2 chez les souris entrainait non pas une immunodéficience mais bien une lympho-prolifération associée à des atteintes auto-immunes (Sadlack et al., 1993). C'est alors qu'a émergé la double fonction de l'IL-2 dans l'immuno-modulation des LT. En effet, cette cytokine sécrétée en grande partie par la LT $CD4^+$ est indispensable à la différenciation, la survie et la fonction suppressive des Tregs (Fontenot et al., 2003; Malek and Castro, 2010; Yu et al., 2009). De manière dose-dépendante, la cytokine permet de cibler la population effectrice (forte dose) ou la population régulatrice (faible dose) des LT, ceci étant rendu possible grâce à la structure différente du récepteur à l'IL-2 exprimé à la surface de ces LT. Le récepteur à l'IL-2 est un hétérotrimère composé d'une chaine α, une chaine β et une chaine γ. Tandis que les chaines β et γ, exprimées de manière constitutive assurent la fixation de l'IL-2 avec une faible affinité ainsi que la transmission du signal, la chaine α, aussi appelée CD25 confère au récepteur la capacité de lier l'IL-2 avec une haute affinité (Spolski et al., 2018). Bien que l'expression du CD25 soit augmentée de façon transitoire lors de l'activation des cellules T, elle est exprimée plus fortement à la

surface des cellules régulatrices les rendant plus sensibles à l'IL-2 (Sakaguchi et al., 1995).

Ainsi l'administration d'IL-2 à faible dose (IL-2_{fd}) permet de rétablir l'équilibre de la balance d'immuno-régulation Teff/Treg dans le cadre des MAI. Ceci a notamment été prouvé dans le cadre d'essais cliniques dans différentes indications thérapeutiques, pour lesquels le laboratoire i^3 est pionnier, et ayant montré que le traitement IL-2_{fd} entraine une augmentation de la proportion des Tregs dans le sang, associé à un bénéfice clinique (He et al., 2020, 2016; Humrich and Riemekasten, 2019; Rosenzwajg et al., 2019, 2015a; Saadoun et al., 2011; von Spee-Mayer et al., 2016). En effet, rétablir la balance anti-inflammatoire dans une pathologie auto-immune est une stratégie offrant la possibilité de réduire les effets indésirables infectieux souvent observés avec les immunosuppresseurs non spécifiques.

Néanmoins, de nombreuses questions restent encore en suspens concernant l'effet de l'IL-2_{fd}, car si son bénéfice clinique n'est plus à prouver, qu'en est-il des cellules Tregs impliquées dans l'amélioration clinique ? L'IL-2 n'étant pas spécifique de certains clones T, peut-elle activer préférentiellement une composition particulière du répertoire TCR. C'est ce qu'ont étudié Mhanna et al. dans le cadre du DT1 chez la souris NOD (Annexe 1). Dans cette étude, nous avons montré que, l'IL-2_{fd} favorisait l'expansion de Tregs enrichies en TCR spécifiques du DT1 (Mhanna et al., 2021). Aussi, comme les Tregs spécifiques de l'antigène semblent se développer préférentiellement sous traitement par l'IL-2_{fd}, les analyses du répertoire des TCR ont un fort potentiel pour fournir des biomarqueurs de l'efficacité de l'IL-2_{fd}.

I.2. Développement, sélections et vie des lymphocytes T

Les lymphocytes T, soldats du SI adaptatif, initient leurs développements au niveau de la moelle osseuse. C'est ici que nous retrouvons les cellules souches hématopoïétiques (CSH), cellules multipotentes précurseurs notamment des lymphocytes B et T. Une partie de ces CSH vont ensuite migrer par voie sanguine au niveau du thymus et permettre l'engagement de ces précurseurs dans la lignée T, évocatrice de l'organe en question. Suite aux différentes étapes de maturation, les LT matures et fonctionnels vont émerger du thymus et exercer leurs fonctions au travers

de l'ensemble des organes et tissus formant le système lymphoïde (l'ensemble des organes et tissus où résident les cellules du SI).

Au cours de l'évolution, des familles de gènes spécialisés et des mécanismes moléculaires ont été sélectionnés ayant la propriété de générer une population de LT exprimant une diversité suffisamment vaste permettant de répondre à n'importe quel élément extérieur, capable d'interagir avec le CMH du Soi, mais sans répondre aux protéines du Soi.

I.2.1. Du thymocyte aux lymphocytes T spécialisés

Le thymus est l'organe du système lymphoïde primaire responsable de la différenciation et de la sélection des cellules dépendantes du TCR qu'elles expriment. Cette glande, située au niveau de la partie supérieure du thorax, est formée de deux lobes unis, eux-mêmes constitués de multiples lobules, chacun ayant une zone périphérique : le cortex et une zone centrale : la médulla. Dès l'arrivée des CSH dans le thymus, ces précurseurs lymphoïdes que l'on dénomme thymocytes, sont des LT immatures qui vont se différencier progressivement en acquérant divers marqueurs membranaires jusqu'à devenir des LT matures : c'est ce qu'on appelle la thymopoïèse (**Figure 2**). Les principaux stades de développement des LT sont les suivants, le stade double négatif (DN) dans lequel les cellules présentent un phénotype $CD4^-CD8^-$ qui se subdivise lui-même en 4 étapes. Les thymocytes DN1 représentent les cellules qui viennent d'entrer dans le thymus, s'en suit le stade DN2 au cours duquel elles initient le réarrangement somatique de la chaine β du TCR (section I.3.1.1). Au stade DN3, les cellules expriment un pré-TCR et sont soumises à la β-sélection (section I.2.1.1). Celles qui passent cette sélection prolifèrent et commencent à réarranger la chaine α de leurs TCR, c'est le stade DN4. La maturation se poursuit par le stade double positif (DP) dans lequel les cellules présentent un phénotype $CD4^+CD8^+$, exprimant un TCR potentiellement fonctionnel. C'est à ce stade qu'ont lieu les étapes de sélections positive et négative (section I.2.1.2) leurs permettant de passer au dernier stade de maturation, le stade simple positif (SP) en perdant l'expression du CD4 ou du CD8. A l'issue de leurs maturations, les LT $CD4^+$ et $CD8^+$ dits naïfs, car n'ayant jamais rencontré le/les antigène(s) complémentaire(s) de leurs TCR, migrent par voie sanguine dans les

organes lymphoïdes secondaires. L'ensemble de ces étapes constituent un processus stricte auquel seulement 3 à 5% des cellules survivent (Merkenschlager et al., 1997).

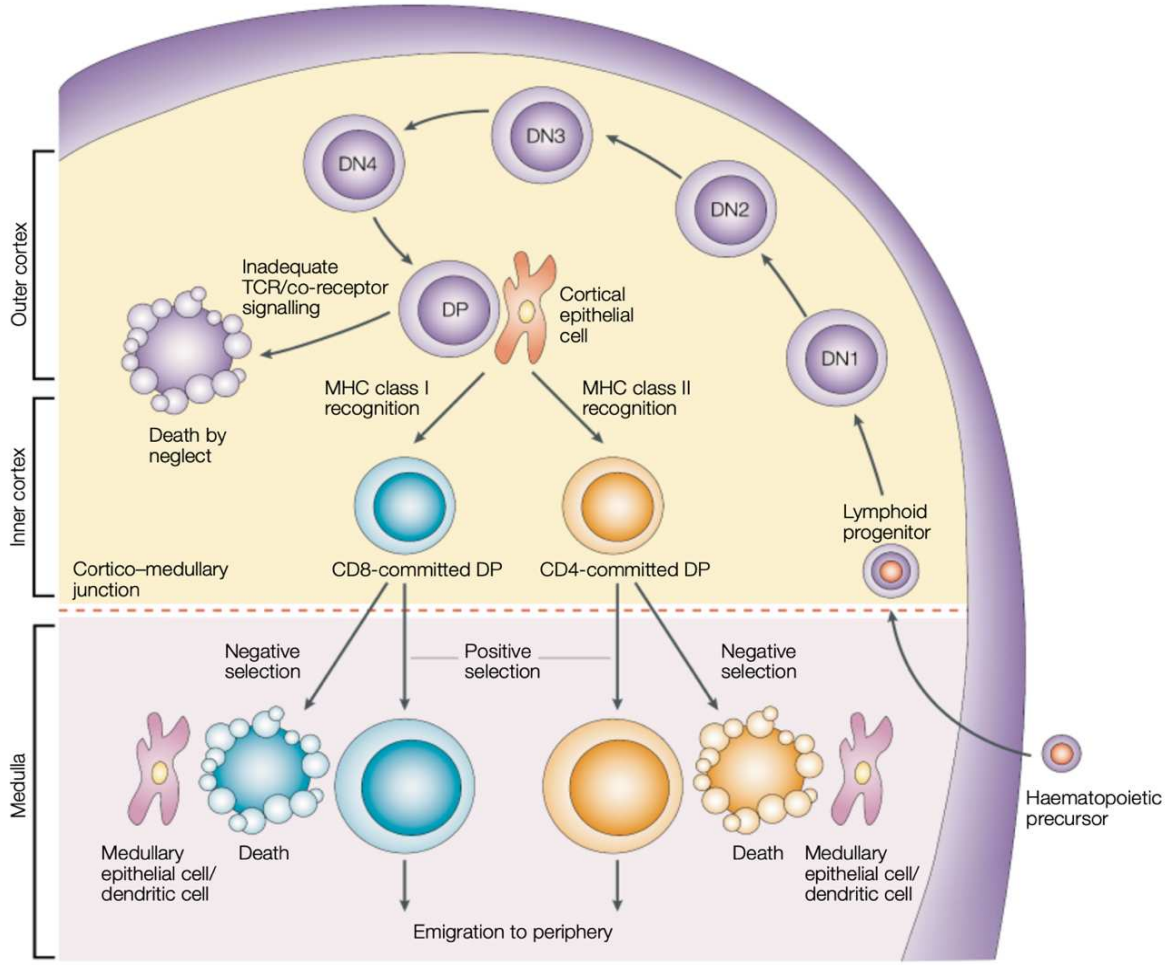


Figure 2 : Processus de développement des cellules T dans le thymus.
Issu de Germain R., T-cell development and the CD4–CD8 lineage decision. Nat Rev Immunol, 2002)

I.2.1.1. <u>La sélection β</u>

La sélection β initie les étapes de sélection des LT et se déroule dans les derniers stades DN de la maturation. Suite au réarrangement somatique (section I.3.1) de la chaîne β du TCR, les thymocytes expriment un TCR inachevé appelé pré-TCR correspondant à l'association d'une chaîne β avec une chaîne pré-TCRα (chaine α incomplète). La signalisation ne semble pas dépendre d'une liaison quelconque avec le pré-TCR. Chen et al sont même allés plus loin en montrant que ces précurseurs T pouvaient passer au stade de différenciation suivant (DP) sans ce pré-TCR. Néanmoins, un assemblage du complexe correct est en faveur de l'expression des gènes de survies et de proliférations (Chen et al., 2021). Au bilan, la signalisation du pré-TCR

permet d'inhiber le réarrangement du second locus de la chaine β, par un phénomène d'exclusion allélique (Brady et al., 2010), suivie d'une phase de prolifération des thymocytes, puis le réarrangement de la chaine α débute, permettant d'augmenter considérablement la diversité des récepteurs en générant des clones de cellules ayant le même réarrangement β mais une chaine α différente. Une fois le réarrangement complet des chaines du TCR achevé, les cellules subissent deux nouvelles étapes de sélection.

I.2.1.2. La sélection positive et négative

Ces lymphocytes vont subir dans le thymus deux autres sélections. La première d'entre elles est la sélection positive et a lieu au niveau du cortex thymique. Au cours de celle-ci les cellules épithéliales thymiques vont présenter un peptide du « soi » associé à une molécule de CMH à chaque lymphocyte T. Seuls les LT dont le TCR est capable de reconnaitre avec suffisamment d'affinité le complexe peptide-CMH vont recevoir le signal de survie et vont être conservés. Les cellules qui ne reçoivent pas ce signal meurent par apoptose et sont éliminées. Cette sélection permet de conserver seulement les lymphocytes T capables de reconnaitre des antigènes dans un contexte restreint au CMH du Soi, c'est la restriction au CMH. Les cellules qui ont survécu à cette étape de sélection positive vont subir une sélection négative. La sélection négative a lieu plus profondément dans le cortex et la médulla. Cette fois-ci, les cellules T vont être mises au contact d'un thymopeptidome unique composé d'une large sélection d'antigènes tissulaires spécifiques d'organes par l'intermédiaire du facteur de transcription AIRE (*Autoimmune regulator* ; Anderson et al., 2002), exprimé exclusivement par les CPA thymiques. Les LT dont le TCR interagit trop fortement avec ces antigènes vont mourir également par apoptose. Cette sélection permet d'éliminer les LT potentiellement auto-réactifs susceptibles de s'attaquer aux cellules de l'organisme, c'est la tolérance du soi (Klein et al., 2009).

I.2.1.3. L'engagement vers une lignée cellulaire

Les thymocytes DP précédemment décrits, sont sélectionnés sur la base de l'affinité de leurs TCR mais doivent aussi s'orienter vers une lignée cellulaire CD4 ou CD8, permettant de finaliser leurs maturations en LT SP matures. Plusieurs modèles peuvent expliquer les modifications d'expressions géniques des DP les conduisant à perdre l'expression de l'un des deux marqueurs, CD4 ou CD8. L'engagement dans la lignée CD4 ou CD8 se fait de façon concomitante à la sélection positive au cours de

laquelle les cellules DP vont être orientées en fonction de la durée de l'interaction TCR-CMH-peptide du soi. Un signal long sans discontinuité entrainera des signaux permettant l'engagement dans la lignée CD4 alors que des signaux courts et discontinus favoriseront l'engagement vers le phénotype CD8, c'est ce qu'on appelle le modèle de la cinétique de signalisation (Singer, 2002). Parmi les cellules s'orientant dans la lignée CD4, 5-10% d'entre elles vont se différencier en Tregs. Cette différenciation est conditionnée par l'affinité avec laquelle le TCR exprimé à la surface des CD4 va reconnaitre les antigènes du soi présentés par le CMH (Hassler et al., 2019). En effet, Moran et al ont démontré qu'un signal TCR de forte intensité, en lien avec l'affinité vis-à-vis de l'antigène favorisait l'orientation vers le phénotype Tregs (Moran et al., 2011). Cependant l'observation de TCR partagé entre les Tregs et les Teffs, indique que le signal TCR n'explique pas à lui seul l'orientation dans la lignée cellulaire (Hsieh et al., 2006). Il apparait que les cytokines ; l'antigène lui-même ; l'affinité (force d'interaction entre une molécule TCR et le complexe peptide-CMH) ; l'avidité (la somme des forces d'interaction) et la durée d'interaction façonnent l'intensité du signal TCR qui entrainerait soit une délétion clonale de ces cellules auto-réactives, soit une augmentation de l'expression du CD25. Celui-ci favoriserait l'affinité pour l'IL-2 permettant une expression stable de FoxP3 et ainsi l'orientation vers la lignée régulatrice (Bautista et al., 2009). Finalement, l'ensemble du processus de maturation des thymocytes (recombinaison et sélection) semble être aléatoire, néanmoins, un partage faible, mais plus important a été observé entre des jumeaux ce qui pourrait suggérer un rôle des composants génétiques dans la sélection du répertoire du TCR ou indiquer des expositions environnementales communes (Tanno et al., 2020).

I.2.2. La reconnaissance de l'antigène par le TCR

Comme expliqué précédemment, les cellules T ne sont pas à même de reconnaitre les antigènes seules. Elles nécessitent l'aide des CPA (Zinkernagel and Doherty, 1974). Les CPA ont pour fonction de présenter les antigènes peptidiques associés aux molécules du CMH. Ces peptides sont issus de la dégradation des protéines par la CPA en fragments de 8 à 15 acides aminés. Par un mécanisme d'apprêtement, les complexes peptide-CMH seront formés selon deux voies différentes, dépendantes de la classe de CMH et de l'origine des antigènes. La voie intracellulaire dégrade les antigènes intracellulaires et les apprête sur le CMH de classe

I, alors que la voie dite extracellulaire dégrade les antigènes d'origine extracellulaire et les apprête aux molécules de classe II. Ce complexe CMH est un élément essentiel au maintien physiologique de l'organisme, en permettant aux LT de « scanner » l'ensemble des cellules de l'organisme via le TCR et ainsi identifier les éléments à détruire.

I.2.2.1. La structure du CMH et reconnaissance de l'antigène

Les molécules du CMH sont elles aussi des hétérodimères glycoprotéiques ayant une grande diversité. Le CMH-I, présent à la surface de l'ensemble des cellules nucléées de l'organisme, est composé de 2 chaines, une chaine α qui comporte 3 sous-unités α1, α2, α3 et une chaine β composée uniquement de la sous-unité β2. Les domaines α1 et α2 interagissent pour former une poche peptidique dans laquelle un peptide de 8 à 10 acides aminés va être présenté (Germain, 1994). Le domaine α3 sert de liaison avec la molécule membranaire CD8 des LT cytotoxiques (**Figure 3**, gauche). Le CMH-I est donc exclusivement reconnu par les cellules exprimant le CD8 et a pour principal rôle la présentation de peptides endogènes ou viraux, généralement d'origine intracellulaire. Le CMH-II, présent à la surface des cellules dendritiques, des macrophages et des lymphocytes B, constitue le groupe des CPA professionnelles (double expression CMH-I et CMH-II). Il est composé aussi de 2 chaines, une chaine α qui comporte 2 sous-unités α1 et α2 et une chaine β composée également de 2 sous-unités, β1 et β2. Les domaines α1 et β1 forment la poche peptidique pouvant présenter des peptides plus longs, de 13 à 18 acides aminés (Germain, 1994). Le domaine β2 sert de liaison avec la moléculaire membranaire CD4 des LT auxiliaires (**Figure 3,** droite). Le CMH-II est donc exclusivement reconnu par les cellules exprimant le CD4 et a pour principal rôle la présentation de peptides exogènes. L'ensemble des gènes qui codent pour le CMH forme le complexe HLA. Situés sur le chromosome 6, ces gènes hautement polymorphes forment la carte d'identité génétique de chaque individu. Les gènes HLA de classe I comprennent HLA-A, HLA-B et HLA-C, tandis que ceux du HLA de classe II sont HLA-DR, HLA-DP et HLA-DQ. Chaque individu hérite d'une combinaison de six allèles de classe I et de six allèles de classe II. Chaque variante du CMH exprimant une protéine possède des préférences uniques pour les peptides qui se lient et sont ensuite présentés pour activer les cellules T (Archbold et al., 2009). Finalement, parmi l'ensemble des structures de ces deux complexes, les boucles CDR1 et CDR2 du TCR interagissent principalement avec les domaines formant la poche

peptidique et permettent de stabiliser la région CDR3, responsables de la liaison à la molécule reconnue ou l'épitope (Borg et al., 2005; Garcia et al., 1999). Dès que le complexe TCR-peptide/CMH établit un contact stable, le signal d'activation est déclenché via la synapse immunitaire.

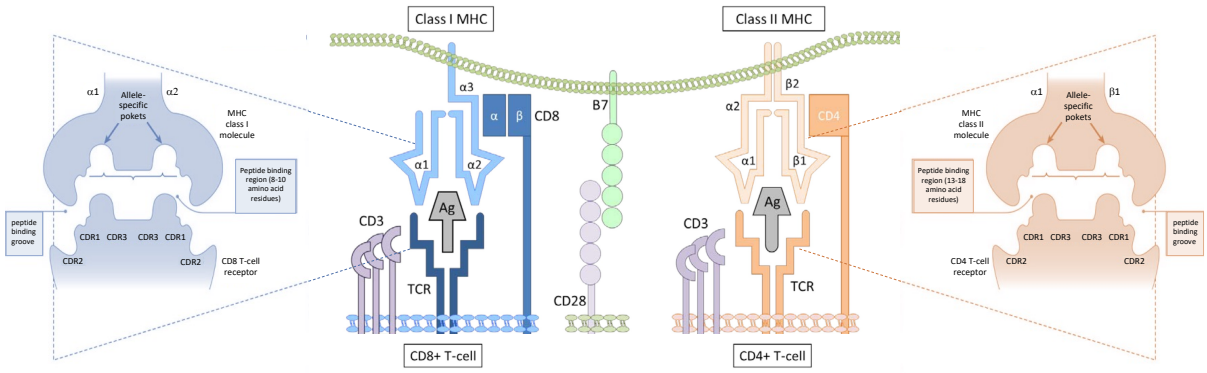


Figure 3 : Structure et interaction du complexe CD3-TCR avec le complexe peptide-CMH. *(Adapté de la thèse de Thomas Poiret "Diversity and focus of CMV Specific T-cell responses in Patients post-HSCT and with solid Tumor")*

I.2.2.2. Transmission du signal

Comme décrit plus tôt, la partie cytoplasmique du TCR ne permet pas la transmission du signal intracellulaire d'activation lors de la reconnaissance du complexe peptide-CMH. Celle-ci passe par la molécule CD3, présente à la surface de l'ensemble des LT. C'est un complexe protéique formé par 4 chaînes transmembranaires s'associant aux TCR qui permet de mettre en œuvre toute la machinerie de signalisation. D'autre part, le CD3 assure également la migration/expression du TCR à la surface des LT, ainsi un déficit d'une des chaines du CD3 entraine un déficit immunitaire (Basile et al., 2004).

Suite à l'association entre un TCR et un complexe peptide-CMH complémentaire, les marqueurs CD4 ou CD8 impliqués dans la reconnaissance du CMH se lient aux sous-unités invariantes des molécules de CMH, respectivement la sous-unité β2 du CMH-II et la sous-unité α3 du CMH-I (**Figure 3,** centre) pour le CD4 et le CD8. Cette interaction stabilise le complexe ce qui accroît l'affinité du TCR (Holler and Kranz, 2003), et initie l'activation des nombreuses voies de signalisations responsables de la migration, la prolifération, et la mise en œuvre des mécanismes d'actions (production de cytokines, perforines, …) des LT. Les cellules immunitaires

porteuses d'un TCR spécifique de l'antigène seront activées et, selon leur phénotype, pourront exercer leurs rôles effecteurs ou régulateurs.

I.2.3. En périphérie : activation, mémoire immunitaire

Nous allons voir plus précisément le contrôle de l'activation de ces LT, leurs évolutions au cours de la vie au travers de la mémoire immunitaire, de la dégénérescence thymique, source de complications liées à l'âge.

I.2.3.1. Activation des lymphocytes T

En périphérie les LT dits naïfs sont en perpétuel mouvement, via la circulation sanguine, entre les organes lymphoïdes secondaires, lieu de stockage et d'activation. Les organes lymphoïdes secondaires sont constitués de structures bien délimitées comme la rate et les ganglions lymphatiques, et de structures plus diffuses comme les tissus lymphoïdes associés aux muqueuses. C'est dans ces organes que l'ensemble des cellules T du SI vont opérer. Elles vont « scanner », via la reconnaissance des milliers de peptides différents provenant de pathogène mais également de protéines et molécules provenant de l'environnement, de l'alimentation et du soi, mais ne s'activent pas forcément (Banchereau and Steinman, 1998). En effet, comme tous systèmes dits robustes, l'activation d'une cellule T nécessite que 3 signaux soient émis dans un ordre défini en vue d'une division cellulaire (**Figure 4**). L'étape initiale implique une interaction entre le TCR et le complexe peptide-CMH et constitue le 1er signal. Plus l'affinité de liaison entre le paratope du TCR (résidus en contact) et l'épitope du peptide immuno-dominant présent dans le sillon du CMH est importante, plus la liaison entre

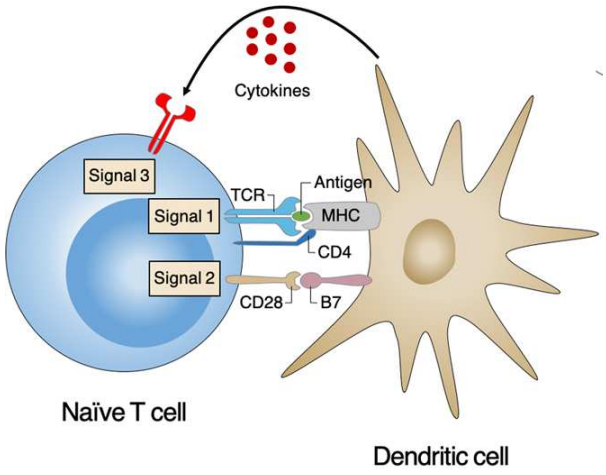


Figure 4 : Activation des cellules T.
(Issu de Lee H.G., Cho M.Z. and Choi J.M., Bystander CD4+ T cells: crossroads between innate and adaptive immunity. Exp Mol Med, 2020)

les deux cellules sera longue et de forte intensité. C'est ce signal prolongé et de forte affinité qui permet l'activation du lymphocyte. Cette liaison est renforcée par les deux marqueurs de différenciation précédemment décrit (CD4 ou CD8), qui, de par leurs parties intracytoplasmiques permettent la transduction du signal lors de la reconnaissance de leurs CMH respectifs. La deuxième étape d'activation des LT est émise par l'interaction entre le CD28 de la cellule T avec une molécule de costimulation, CD80 ou CD86, exprimée par les CPA et constitue le 2ème signal. L'ensemble nécessite également un environnement cytokinique propice permettant à la cellule T d'être activée, c'est le 3ème et dernier signal. Ce clone va se multiplier à l'identique par prolifération et engendrer une expansion clonale. Une partie des cellules divisées seront des cellules activées et effectrices de la réponse immunitaire, elles vont par la suite subir une étape de contraction, caractérisée par la mort d'une grande proportion d'entre elles. Une autre partie va constituer un pool de cellules dites mémoires, spécifiques du pathogène, ayant une durée de demi-vie plus longue (Wherry and Ahmed, 2004). La spécificité des réponses immunitaires adaptatives est étroitement liée à la présence de ces cellules mémoires, ayant l'expérience de la rencontre de l'antigène. En absence de l'un des signaux la cellule restera dans un état naïf. Ainsi, la dynamique de sélection/activation propre à chaque individu va constituer au travers des cellules naïves et mémoire l'histoire immunitaire passée et future d'un individu.

I.2.3.2. Le façonnement de l'histoire immunitaire passée et future d'un individu

Au début de la vie, le thymus est en pleine possession de ses capacités de génération de cellules T fonctionnelles. Il permet à un être relativement vierge de toutes stimulations extérieures de se façonner une protection efficace via un ensemble de cellules naïves relativement large. Ses capacités de génération vont diminuer avec l'âge dû au phénomène d'involution thymique (Gui et al., 2012), représentant le vieillissement de l'organe, ce qui va engendrer une contraction du compartiment des cellules T naïves nouvellement générées (**Figure 5**). Cette diminution est compensée par un ensemble de cellules mémoires acquises au cours de la vie. Ainsi la diversité du répertoire TCR commence à diminuer vers de l'âge de 40 ans (Britanova et al., 2014), entrainant vers l'âge de 70 ans une diminution importante de l'ensemble de cellules naïves différentes (Qi et al., 2014). Ainsi le répertoire immunitaire d'un individu est un objet biologique intéressant à analyser, car il est un reflet de son l'histoire immunitaire passée et possède un pouvoir informatif à la vue des réponses immunitaires futures.

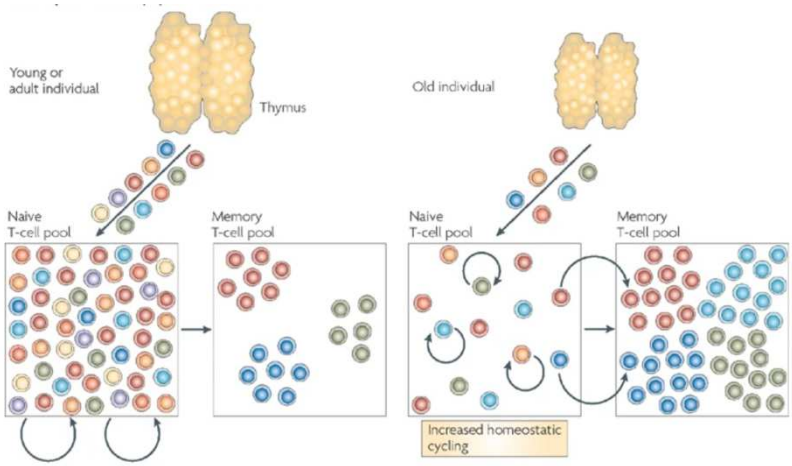


Figure 5 : Modification du ratio entre les LT naïfs et mémoires suite à l'involution thymique.
(Issu de Nikolich-Žugich J., Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections. Nat Rev Immunol, 2008)

I.3. Le Répertoire des récepteurs des lymphocytes T

Le répertoire est constitué de l'ensemble des TCR d'un individu à un moment donné. Chaque étape de différenciation, d'activation, d'expansion et de mort cellulaire va contribuer à la modification du répertoire TCR par la modification des clones cellulaires qui vont le composer. C'est un objet biologique dynamique et complexe qui incarne le statut du SI.

De manière plus concrète, le concept de répertoire immunitaire a été imaginé pour décrire la diversité des lymphocytes impliqués dans le SI d'un individu (Boudinot et al., 2008). Étant donné que chaque lymphocyte exprime un récepteur spécifique d'antigène, le répertoire lymphocytaire disponible à un instant donné conditionne donc le répertoire des antigènes reconnus. Le répertoire lymphocytaire englobe à la fois l'ensemble du potentiel de la diversité permise par les ressources génétiques de l'espèce et de l'ensemble des récepteurs disponibles exprimés dans un tissu. L'étude du répertoire immunitaire est un outil très prometteur pour décrire les populations lymphocytaires et leurs récepteurs lors du développement du SI ou dans le cas de pathologies telles que les maladies auto-immunes, les infections, les cancers ou les traitements thérapeutiques mais aussi dans le cadre de la vaccination. L'analyse de sa diversité et de sa composition devrait permettre de mieux comprendre la dynamique des populations de lymphocytes en physiologie et en pathologie, ainsi que de découvrir de nouveaux biomarqueurs et de nouvelles cibles thérapeutiques. Avec l'évolution des

méthodes de séquençage, cet objet biologique s'étudie maintenant par séquençage de nouvelle génération qui donne un aperçu du répertoire immunitaire à la fois profond et détaillé.

I.3.1. Le TCR, un objet biologique complexe

L'ensemble des LT d'un individu peut ainsi être considéré comme une collection de lymphocytes presque tous différents de par leur TCR, ayant chacun une capacité propre à reconnaître des antigènes donnés. Cette diversité a pour but d'avoir la capacité de reconnaitre n'importe quel antigène. Cependant il existe deux paradoxes entre la diversité quasi « infinie » du répertoire immunitaire. D'une part, (i) le nombre de gènes dans le génome humain (≈ 20000 - 25000 gènes, International Human Genome Sequencing Consortium, 2004) et d'autre part, (ii) le déséquilibre avec le nombre de cellules T présent dans un individu, qui sans limite de nombre de cellules pourrait représenter jusqu'à 500kg (Mason, 1998). Dans cette partie (i) nous décrirons la structure du TCR ; (ii) nous verrons que ce récepteur n'est pas entièrement codé dans le génome, mais génère cette diversité à travers 3 mécanismes et enfin (iii), nous détaillerons les mécanismes de reconnaissance du TCR lui permettant de ne pas avoir à atteindre une diversité potentielle inatteignable.

I.3.1.1. La structure du TCRαβ

Les récepteurs lymphocytaires sont des hétérodimères transmembranaires glycoprotéiques constitués dans 95% des cas par l'appariement d'une chaine TCRα et d'une chaine TCRβ, reliées entre elles par un pont disulfure (Chien et al., 1984). Chaque chaîne TCR est composée (**Figure 6a** ; de haut en bas) d'une courte queue cytoplasmique qui ne permet pas le transfert du signal d'activation (section I.2.2.2), d'un domaine membranaire permettant l'ancrage dans la cellule, d'un domaine constant extracellulaire reliant le TCR aux complexes de transduction du signal et d'un domaine variable, impliqué dans la reconnaissance de l'antigène. Ce dernier formé de six boucles en épingle à cheveux hypervariables, appelées *Complementary Determining Region* (CDR; Chothia et al., 1988) est le produit de réarrangements somatiques entre les multiples gènes Variable (V), Diversité (D, sur la chaine β uniquement) et Jonction (J) (Davis and Bjorkman, 1988) qui le composent. Les boucles CDR1 et CDR2 codées par les gènes V de chacune des chaines se situent à la périphérie

des boucles CDR3 régions hypervariables du TCR qui lui confère sa spécificité antigénique (**Figure 6b**).

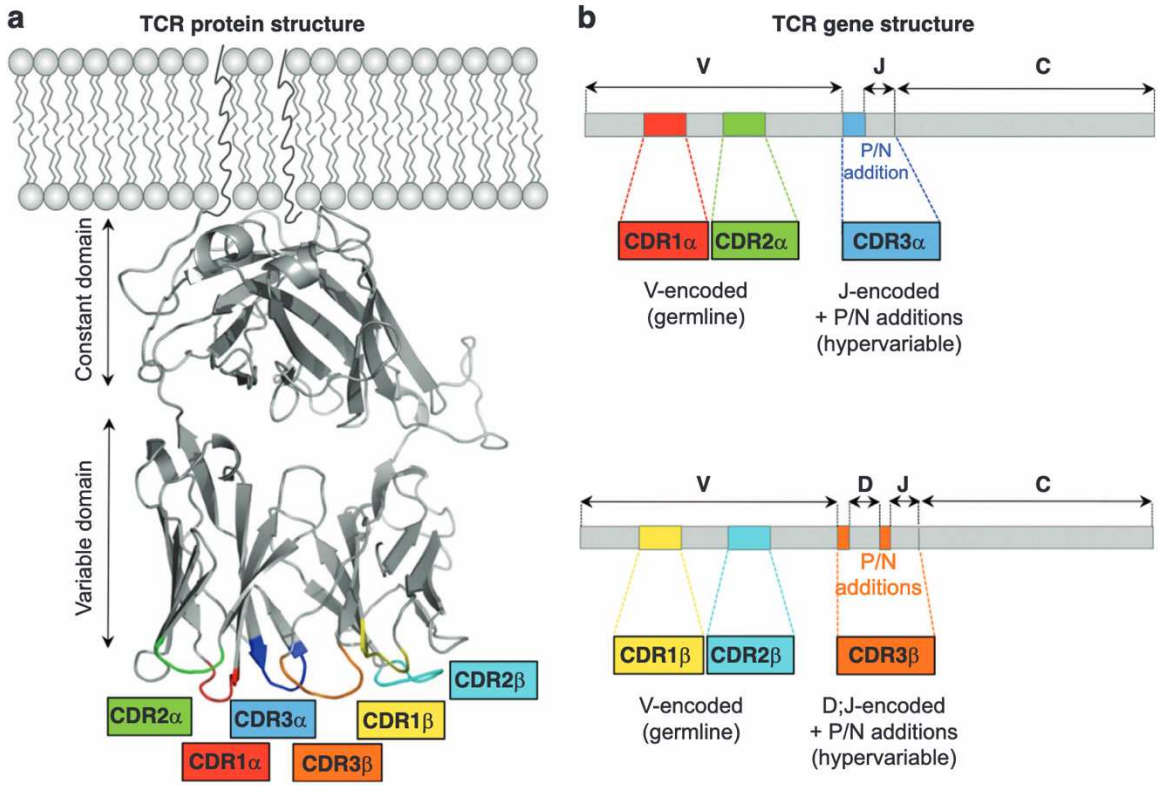


Figure 6 : Structure générale du TCR.
(a) Structure de la protéine du TCR. Structure tridimensionnelle des chaines α et β du TCR. Les boucles en « épingles à cheveux » des CDR du TCRαβ. (b) Structure des gènes des chaines α et β du TCR. Les régions CDR1 et CDR2 sont entièrement codés dans les gènes V de la lignée germinale, tandis que CDR3 se trouve à la jonction entre les segments V et J réarrangés (TCRα) et les segments V, D et J (TCRβ). (Adapté de Attaf M., Legut M., Coleand D. K., Sewell A. K., The T cell antigen receptor: the Swiss army knife of the immune system. CEI, 2015)

I.3.1.2. Organisation génomique des gènes codant les chaines α et β du TCR

La première source de diversité est issue de la diversité combinatoire entre les différents gènes codants les chaines du TCR. L'ensemble des gènes TRAV, TRAJ et TRAC codants pour la chaîne α du TCR (TRA), se situent sur le chromosome 14 et s'étendent sur près de 1000kb. On dénombre, selon ImMunoGeneTics (IMGT, http://www.imgt.org), référence mondiale en immunogénétique, créée par Marie-Paule Lefranc, 54 gènes TRAV dont 44-47 sont fonctionnels, 61 gènes TRAJ dont 50 sont fonctionnels et 1 gène TRAC fonctionnel. Concernant les gènes TRBV, TRBD, TRBJ et TRBC codants pour la chaine β du TCR (TRB), ils sont situés sur le

chromosome 7 et s'étendent sur 620 kb. On dénombre, 64-67 gènes TRBV dont 40-48 fonctionnels, 2 gènes TRBD fonctionnels, 14 gènes TRBJ dont 12-13 fonctionnels et 2 gènes TRBC fonctionnels (Parra et al., 2008; Scaviner and Lefranc, 2000).

I.3.1.3. La recombinaison VDJ

Les gènes présentés précédemment sont assemblés au cours du développement des lymphocytes, par un mécanisme spécifique appelé recombinaison V(D)J. La recombinaison V-J du locus α se déroule en une étape qui consiste à réarranger un des nombreux segments TRAV à un des nombreux segments TRAJ. Les séquences intermédiaires sont épissées, produisant un transcrit de la chaine α du TCR dans lequel les segments V, J et C sont directement adjacents. La recombinaison VDJ du locus β est ordonnée en deux étapes. Dans un premier temps la recombinaison D-J permet de juxtaposer TRBD1 à l'un des six segments TRBJ1 ou TRBD2 à l'un des sept segments TRBJ2. Dans un deuxième temps se produit la recombinaison V-DJ qui amène ensuite la jonction DJ réarrangée à l'un des nombreux segments TRBV (**Figure 7a**). Les

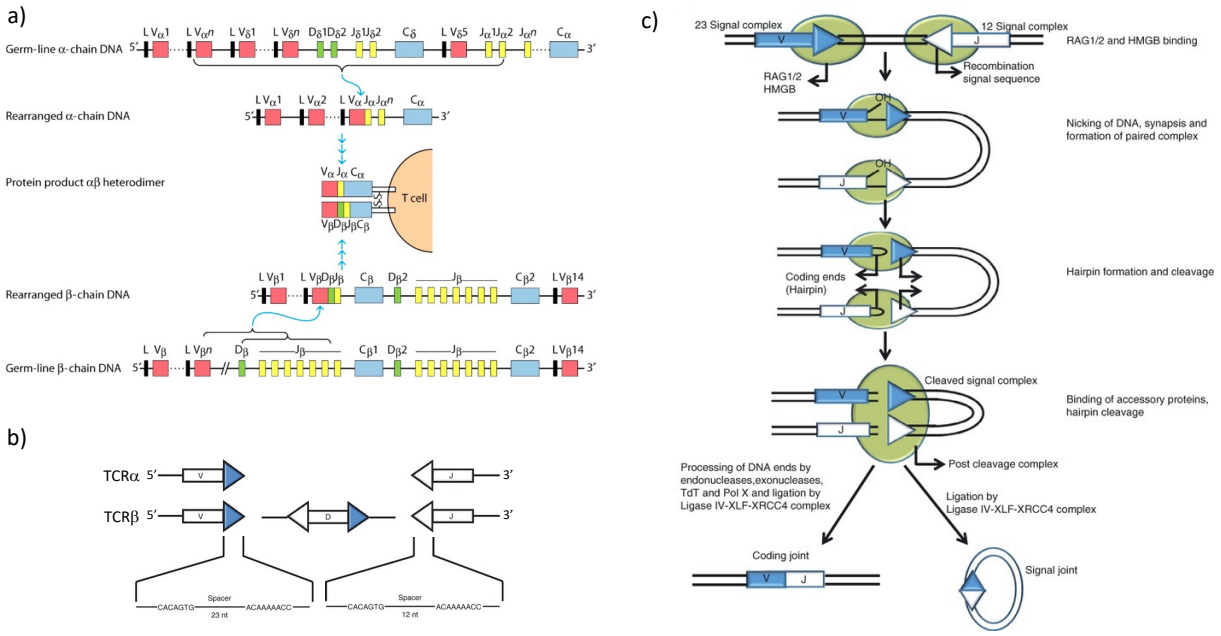


Figure 7 : Organisation et mécanisme du réarrangement V(D)J des gènes du TCR.
(a) Organisation génomique et Recombinaison V(D)J des locus TCRα (haut) et TCRβ (bas). (b) Les Séquences Signal de Recombinaison (RSS) bordant les segments géniques V, D et J. La RSS avec l'espaceur de 23pb (triangle bleu) borné de part et d'autre par la séquence nonamère (CACAGTC) et la séquence heptamère (ACAAAAACC) et la RSS avec l'espaceur de 12pb (triangle blanc). (c) Mécanisme de recombinaison des gènes V(D)J par les enzymes RAG1 et RAG2. (Adapté de Nishana M. and Raghavan S.C, Role of recombination activating genes in the generation of antigen receptor diversity and beyond. Immunology, 2012)

séquences intermédiaires sont ensuite épissées, générant un transcrit de la chaine β du TCR dans lequel les segments V, D, J et C sont directement adjacents.

Cette recombinaison fait intervenir de courtes séquences cibles jouxtant les segments codants appelées Séquences Signal de Recombinaison (RSS). Elles sont composées d'un heptamère (ACAAAAACC) et d'un nonamère (CACAGTC), de structure conservée, séparés par une séquence espaceur aléatoire de 12 (RSS-12) ou 23 (RSS-23) nucléotides (Sakano et al., 1979; Tonegawa, 1983). Les segments V présentent à leur extrémité 3' une RSS-12, les segments J à leur extrémité 5' une RSS-23 et les segments D à leurs extrémités 5' et 3' respectivement des RSS-23 et RSS-12 (**Figure 7b**). La recombinaison V(D)J étant limitée aux segments flanqués de RSS différentes garantit l'inclusion d'un membre de la famille de chaque gène. L'ensemble de ce mécanisme est dépendant d'un complexe enzymatique constitué notamment de enzymes *Recombination Activating Gene 1* (RAG1 ; D. G. Schatz et al., 1989) et *Recombination Activating Gene 2* (RAG2 ; M. A. Oettinger et al., 1990). Les gènes codants pour ces deux enzymes sont exprimés uniquement au cours du développement des LT dans le thymus, reconnaissent et se lient aux RSS, entrainant la formation d'une synapse entre deux RSS (12 et 23) et introduisant un clivage simple brin concentré au niveau des RSS (van Gent et al., 1995). Les extrémités OH attaquent respectivement le brin opposé et forment deux extrémités codantes scellées en épingle à cheveux et deux extrémités signales (**Figure 7c**). Ce complexe protéique est capital pour la recombinaison V(D)J des cellules T mais aussi des cellules B. En effet, les mutations des gènes RAG entraînent une absence totale de réarrangement (Mombaerts et al., 1992; Shinkai et al., 1992) et sont à l'origine de déficits immunitaires combinés sévères (Corneo et al., 2001).

Finalement, bien que IMGT aient initié le référencement des différents gènes qui composent le TCR, cette base de données ne référence pas la diversité allélique. La recherche d'allèles non documentés ne constitue pas un axe majeur dans l'étude du répertoire TCR. Cependant récemment, l'équipe de Gur Yaari, focalisé principalement sur l'étude du répertoire des cellules B a adapté sa stratégie pour l'inférence d'allèles non documentés, de génotypes et d'haplotypes aux cellules T. Ils ont ainsi identifié 39 nouveaux allèles pour la chaine β du TCR et annoncent que de nombreux autres allèles non documentés restent à découvrir. De plus, un déséquilibre d'association entre deux

allèles a été constaté, permettant d'émettre l'hypothèse que la recombinaison somatique des gènes du TCR ne semble pas totalement aléatoire (Omer et al., 2022).

I.3.1.4. Diversité jonctionnelle

Une autre source de diversité est la diversité jonctionnelle. Elle consiste en la délétion de nucléotides aux extrémités des gènes codants lors de la ligature des segments Vα-Jα et Vβ-Dβ-Jβ sous l'action de diverses enzymes et complexes protéiques au cours du réarrangement. Suite à l'action des protéines RAG-1 et RAG-2, le complexe protéique Artemis, ayant une activité endonucléase, clive les extrémités en épingles à cheveux, ajoute une série de nucléotides palindromiques (nucléotides P) (Lafaille et al., 1989), suite à la jonction imprécise entre les segments géniques (Schlissel, 1998), conduisant à la fusion entre les gènes codants. Ensuite il y aura addition et/ou délétion de nucléotides N, de façon aléatoire, par plusieurs enzymes et notamment la *terminal deoxynucleotidyl-transferase* (TdT), une enzyme exprimée exclusivement au cours de la maturation des lymphocytes dans le thymus (Boulé et al., 2001; Martinez-Valdez and Cohen, 1988). Pour finir, la ligature entre les deux segments codants est assurée par l'ADN ligase IV (**Figure 7c**). Ces mécanismes permettent l'apparition dans les chaînes du TCR d'acides aminés qui ne sont pas codés par la lignée germinale. Ainsi, pour des mêmes gènes V, D et J recombinés, des jonctions différentes en séquence et en taille vont être produites. Toutes ces altérations du génome lui-même font que statistiquement chaque lymphocyte T est un lymphocyte T unique. La séquence exacte de la jonction est aussi nommée la région CDR3, région hypervariable qui confère au TCR sa spécificité antigénique. Il a notamment été montré que, en l'absence de TdT, la longueur de la région du CDR3 était raccourcie en moyenne de 3 paires de bases, ce qui avait pour conséquence une réduction importante de la diversité du répertoire TCR ainsi qu'une perturbation de l'usage de certains gènes V (Cabaniols et al., 2001).

I.3.1.5. Diversité d'association des chaines

Enfin, la dernière source de diversité est la diversité d'association entre les deux chaines recombinées que nous venons de décrire précédemment. En effet, la recombinaison de chacune des deux chaines qui composent le TCR étant indépendante, une même chaine TRB pourra être associée avec plusieurs chaines TRA différentes. Il ne semble pas y avoir de phénomène d'association particulier qui impose une contrainte d'appariement au sein du complexe TCRαβ (Shcherbinin et al., 2020),

ce qui d'un point de vue évolutif permet au SI de s'armer des cellules les plus diverses possibles afin d'engager une réponse immunitaire efficace contre tout pathogène, y compris des pathogènes jamais rencontrés.

Ainsi, les récepteurs exprimés par un même lymphocyte T sont dans la majorité des cas parfaitement identiques. Néanmoins, l'exclusion allélique moins soutenue au niveau de la chaine α du TCR peut provoquer le réarrangement somatique sur les deux allèles du chromosome, ainsi un même lymphocyte peut produire deux transcrits pour allèle du locus α et donc exprimer plusieurs TCR différents à sa surface. La présence de ces transcrits α et/ou β secondaires, bien que minoritaire, peut également être considérée comme participant à la diversité d'une population lymphocytaire mais est également mis en cause dans l'induction des MAI (Elliott and Altmann, 1995; Padovan et al., 1993).

I.3.2. Une grande diversité hautement spécifique, un antagonisme ?

Bien que la diversité réelle du répertoire TCR d'un individu reste encore incertaine, à travers l'étude de la machinerie de recombinaison et l'étude par séquençage de nouvelle génération, les estimations de son ordre de grandeur varient depuis 20 ans, passant de $2.5x10^7$ en 1999 (Arstila et al., 1999) à 10^8 en 2014 (Qi et al., 2014). Il semble donc impossible qu'un répertoire TCR puisse être suffisamment divers pour faire face aux $\sim10^{12}$ challenges que les épitopes des CMH vont lui proposer (Sewell, 2012). En effet, comme illustré sur la **Figure 8**, il y a un déséquilibre entre l'ensemble des valeurs exposées. Le répertoire TCR théorique a été estimé à $>10^{61}$ (Mora and Walczak, 2018), permettant d'avoir potentiellement les armes pour répondre à une large gamme de défis immunologique (à titre de comparaison, le nombre d'étoiles estimé dans l'univers est de 6×10^{22} (Manojlović, 2015)). Néanmoins à cela s'ajoute les étapes de sélections (section I.2.1), sans quoi ces nombreux clones auraient sans aucun doute un rôle délétère pour l'hôte. En périphérie, à la suite de défis immunologiques et accompagnée de la production thymique, la véritable diversité du répertoire est estimée à 10^8 (Qi et al., 2014). C'est ici qu'apparait un déséquilibre entre nombre d'antigènes potentiels, la diversité estimée du répertoire TCR et le nombre total de LT présent dans un individu (3.10^{11} ; Jenkins et al., 2010). Mason fut le premier

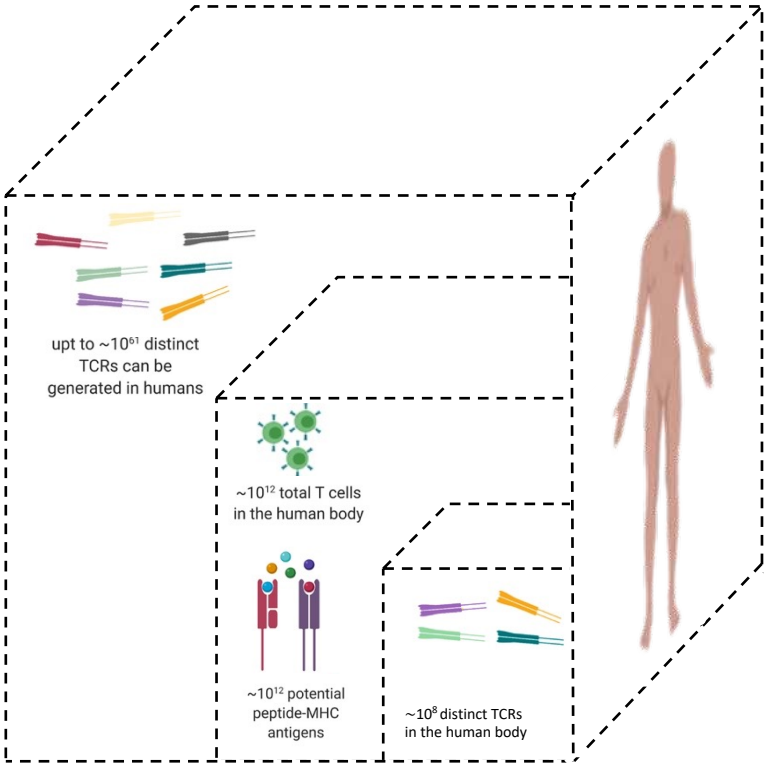


Figure 8 : La diversité du répertoire TCR en chiffre.

à remettre en question la théorie de la sélection clonale (Jerne, 1955 ; théorie selon laquelle un LT individuel serait spécifique d'un seul antigène) par la simple arithmétique qu'une immunité efficace nécessiterait plus de 500kg de cellules T. Par ailleurs, un cas concret d'un patient atteint d'une immunodéficience combinée sévère liée au chromosome X atypique (Stephan et al., 1996), maladie génétique entrainant une lymphopénie avec une absence de LT, est venue renforcer cette remise en question. En effet, ce patient atypique possédait un nombre normal de cellules T mais un répertoire TCR extrêmement restreint, estimé à 1000 TCR différents, lui conférant néanmoins une protection face à diverses infections (Bousso et al., 2000). Ainsi nous pouvons nous demander par quel stratagème un tel répertoire TCR a-t-il pu contrôler les nombreuses infections. Comment le répertoire TCR reconnaît-il une telle diversité potentielle d'antigènes ?

En 1995, Michael G. McHeyzer-Williams et Mark M. Davis ont montré que des clonotypes différents pouvaient répondre au même peptide, indiquant que la recombinaison du TCR est capable de générer plusieurs solutions au problème de la reconnaissance (McHeyzer-Williams and Davis, 1995). La réactivité croisée des cellules T est également bien illustrée par la présence de cellules T $CD4^+$ spécifiques

des épitopes du Virus de l'Immunodéficience Humaine (VIH), du Cytomégalovirus (CMV) et du virus de l'herpès simplex identifiées chez des volontaires sains qui n'avaient jamais été infectés par ces virus (Su et al., 2013).

I.3.2.1. La réactivité croisée des cellules T

La réactivité croisée des cellules T peut s'expliquer par divers mécanismes qui peuvent être intrinsèques à la structure du TCR, à son interaction avec le complexe peptide-CMH ou bien au peptide lui-même, dont voici quelques exemples.

Comme expliqué précédemment, la reconnaissance entre le TCR et le complexe peptide-CMH se fait notamment grâce à des interactions entre les régions CDR1 et CDR2 des chaines αβ du TCR et les régions conservées de la molécule de CMH, tandis que les régions CDR3αβ entrent en contact avec le peptide selon une géométrie de liaison diagonale (Ding et al., 1998). Néanmoins, les interactions entre ces deux complexes ainsi que les complexes eux-mêmes possèdent une certaine flexibilité (Borbulevych et al., 2009), permettant (i) au site de liaison (principalement la région CDR3) de certains TCR de s'adapter à différents complexes peptide-CMH et ceci par des changements de conformation (**Figure 9a** ; Reiser et al., 2003), mais aussi, (ii) d'adopter une orientation atypique permettant une interaction avec différents peptides ayant des similarités structurelles (**Figure 9b** gauche) et finalement (iii) au site de liaison peptidique dans le sillon du CMH de présenter des zones différentes d'un même peptide au TCR (**Figure 9b** droite ; Armstrong et al., 2008). Ainsi les acides aminés qui entrent en contact avec le TCR peuvent changer radicalement la nature de la réponse immunitaire qui en découle. À noter que ce dernier mécanisme peut aussi être observé non pas dans le cadre de la flexibilité mais simplement dans le cadre de la présentation alternative d'un même peptide par des HLA différents, qui aurait les mêmes conséquences (Stadinski et al., 2010).

Un autre mécanisme, source de réactivité croisée est cette fois ci lié au peptide lui-même et à la reconnaissance de certains TCR. Le mimétisme moléculaire est considéré comme une importante source de réactivité croisée, il s'agit d'une proximité structurelle d'une petite zone entre des peptides, positionnés de manière identique et ne présentant que des différences mineures (**Figure 9c**). Ainsi le TCR tolère de multiples substitutions d'acides aminés à d'autres positions dans le peptide (Fujinami et al., 1983).

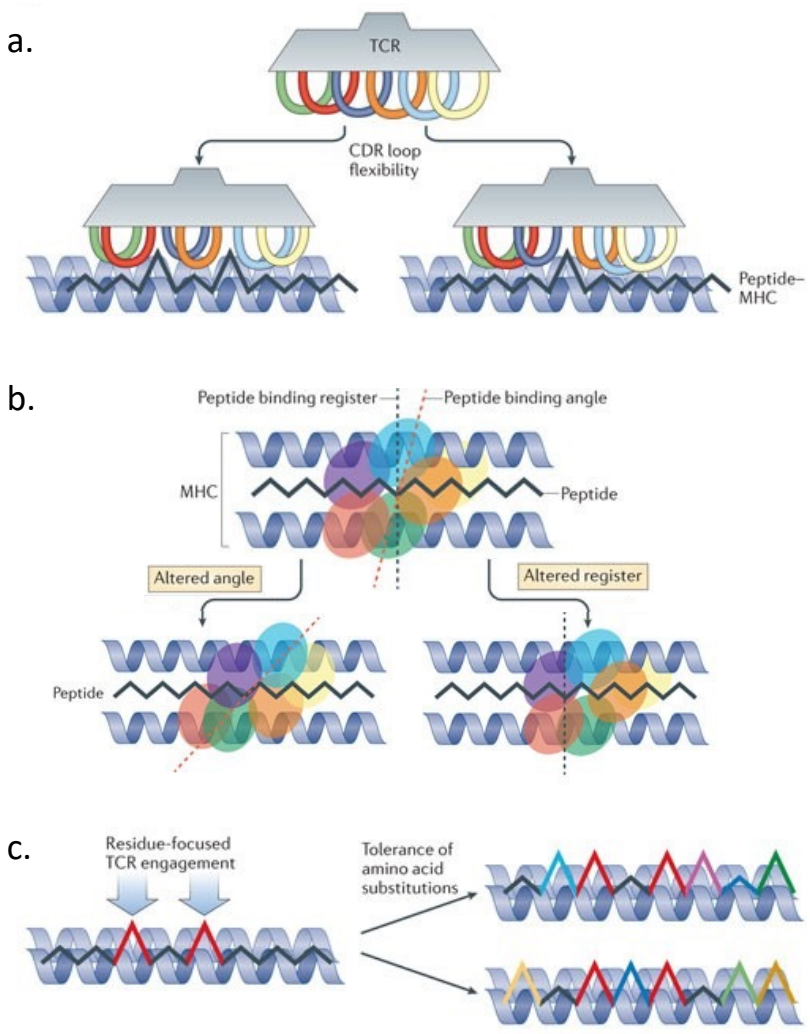


Figure 9 : Illustration des mécanismes impliqués dans la réactivité croisée des TCR à l'égard des molécules peptides-CMH.
(Adapté de Sewell A., Why must T cells be cross-reactive ?. Nat Rev Immunol, 2012)

Finalement, en lien avec l'exclusion allélique précédemment décrite (section I.3.1.5), un phénomène non pas considéré comme de la réaction croisée mais pouvant avoir aussi un impact sur l'étendue du panel de reconnaissance d'un LT est la double expression de la chaine α. En effet on considère qu'environ 10 à 25% (Corthay et al., 2001; Schuldt and Binstadt, 2019) des LT expriment à leurs surface 2 TCR différents composés d'une même chaine β mais de deux chaines α différentes (Padovan et al., 1993). Ce phénomène représente un danger auto-immun en permettant à des TCR auto-réactifs d'échapper éventuellement à la sélection thymique sur la base de leur premier TCR mais être activés en périphérie grâce à leur deuxième TCR (Elliott and Altmann, 1995; Schuldt et al., 2017).

I.3.2.2. <u>Un potentiel biais à l'origine des MAI</u>

A travers les mécanismes de réactivité croisée présentés en section I.1.2 et section I.3.2.1 nous estimons qu'un TCR unique peut reconnaitre jusqu'à 10^6 complexes peptide-CMH différents (Mason, 1998; Wooldridge et al., 2012). Il s'agit d'une caractéristique essentielle pour que le répertoire des cellules T puisse faire face à tout défi immunitaire. Néanmoins, ce haut degré de réactivité croisée peut être responsable de l'échappement des cellules potentiellement autoréactives du thymus mais aussi de leur activation en périphérie. Sur la base d'observations cliniques et épidémiologiques, l'incidence des MAI est étroitement liée à celle des infections (Bach, 2002), suggérant un mécanisme d'activation des LT autoréactifs par les antigènes de certains pathogènes. De nombreuses études récentes mettent en évidence ces liens étroits entre les MAI et certaines infections.

Une récente étude a suivi pendant plus de 20 ans 10 millions de militaires parmi lesquels 801 ont développés une sclérose en plaque (SEP ; maladie auto-immune touchant le système nerveux central). Ces militaires ayant des visites médicales régulières, ils ont pu effectuer un suivi infectiologique notamment du virus Epstein-Barr (EBV). Parmi les militaires ayant développé une SEP, 800/801 (99,8%) se sont avérés être séropositifs contre « seulement » 57% des sujets contrôles (n'ayant pas

AD	PATHOGEN	MECHANISMS	OBSERVATIONS
SLE	CMV	Molecular mimicry	SLE diagnostic after CMV acute infection
			Asociation with high levels of antibodies against CMV and Anti Sm autoantibodies.
	EBV	Molecular mimicry, bystander activation, clearence deficiency	Higher prevalence in patients (99%) than in healty controls (70%).
			Higher viral load in blood cells from patients than in controls.
			B and T cell reactivity to Sm related to antibodies against EBNA and HLA-DR3.
	T. gondii	Molecular mimicry	Higher antibody titers to T. *gondii* in patients than in healty controls.
APS	EBV	Molecular mimicry	Higher prevalence in patients than in healty controls.
	C. tetani	Molecular mimicry	Relationship between tetanus toxoid and the presence of anti β2-GPI antibodies.
T1D	CMV		Relationship of presence of antibodies to islet cells and CMV infection
			Ealry infection during childhood cuases elevated risk of DT1
	Saccharomyces cerevisiae		Higher prevalence of ANCA autoantibodies in patients than in controls
AR	*E. coli*, *K. pneumoniae P. miriabilis*	Molecular mimicry	Relationship of presence of IgM to infection and rheumatoid factor
	P. miriabilis	Molecular mimicry	Recurrent *P. miriabilis* infection in patients with RA.
	EBV	Molecular mimicry	Higher prevalence of infection in patients than in healthy controls.
	Endogenous retrovirus	Molecular mimicry	Evidence of induction of autantibodies
MS	EBV	Molecular mimicry	Higher frequency of EBV infection before MS onset in patients (95%).
	Acinetobacter sp P. aeruginosa	Molecular mimicry	Presence of autoantibodies that recognize autoantigens in patients but not in healthy controls.
SS	Coxsackie virus	Non especific activation	Possible presence of the virus in the cells of salivary glands in SS patients but not in helathy controls.
	EBV		Presence of previous infection or IgM to the virus before disease onset.

Tableau 4 : Agents pathogènes associés aux maladies auto-immunes les plus courantes.
AD : Maladie auto-immune, SLE : lupus érythémateux systémique, APS : syndrome des antiphospholipides, T1D : diabète de type 1, RA : polyarthrite rhumatoïde, MS : sclérose en plaques, SS : syndrome de Sjögren, CMV : cytomégalovirus, EBV : virus d'Epstein-Barr.
Adapté de Arango M.T., Shoenfeld Y., Cervera R. et Anaya J.M., Infection and autoimmune diseases. Autoimmunity: From Bench to Bedside, Chapter 19, 2013.

développé une SEP). Ils ont ainsi prouvé que le risque de développer une SEP était multiplié par 32 après une infection par l'EBV (Bjornevik et al., 2022). D'autres liens entre certaines MAI et des pathogènes sont suspectés, sans être pour autant prouvés ou appuyés comme celui étayé dans l'étude précédemment décrite. Un aperçu non-exhaustif est présenté dans le **Tableau 4**.

En ce sens, nous avons mené une étude au laboratoire (Annexe 2) portant sur le répertoire TCR des LT $CD8^+$ dans le thymus, correspondant en majorité aux cellules naïves sélectionnées sur la base du peptidome thymique. Nous avons confirmé les hypothèses suggérées par Paul G. Thomas et Jeremy Chase Crawford selon lesquelles ces cellules $CD8^+$ après sélection étaient enrichies en clones publics, composées de TCR ayant des spécificités multiples pour les pathogènes communs, indiquant des biais dans la recombinaison des TCR. Finalement ces résultats laissent suggérer que l'évolution a sélectionné un processus de recombinaison des TCR ayant une forte probabilité de générer des TCR spécifiques aux agents pathogènes courants. Ce mécanisme viendrait renforcer l'idée que la nécessité de protéger un individu, notamment en début de vie contre les agents pathogènes est bien plus importante que de limiter l'auto-réactivité (Thomas and Crawford, 2019). Ce mécanisme permettrait d'expliquer comment un groupe limité de TCR peut fournir la large couverture antigénique nécessaire à une immunité efficace (Bousso et al., 2000).

I.4. Le répertoire TCR, un omic source de promesse

Force est de constater que l'information pouvant jaillir de l'analyse du répertoire TCR permettra de mieux comprendre les mécanismes du SI adaptatif ainsi que son potentiel thérapeutique et de biomarqueurs. Historiquement les études du répertoire immunitaire se faisaient via des analyses par cytométrie en flux, donnant un bref aperçu de la diversité (Ciupe et al., 2013). Plus tard, des analyses par spectratypage de la région CDR3 pour chaque gène TRV (principalement de la chaine β, car trop nombreux sur la chaine α), permettaient la détection d'expansions clonales (Cappuccilli et al., 2017; Matsumoto et al., 2006). Les avancées sur le séquençage de nouvelle génération (NGS) offrent aujourd'hui une technologie alliant informations quantitatives et qualitatives, deux éléments essentiels pour l'étude de la diversité et de la spécificité du répertoire immunitaire. En nette augmentation, l'étude du répertoire

TCR par séquençage (TCR-seq) est considérée comme une source prometteuse de biomarqueurs pour l'ensemble des maladies à composantes immunitaires (Six et al., 2013). En effet cette méthode fournit de multiples angles d'analyses, telles que l'analyse de la diversité, l'identification de nouveaux allèles TRV, l'usage de ces gènes, la fréquence de clones, ainsi que l'identification de signatures TCR. Parallèlement à l'augmentation de la granulosité de ces données, de nouveaux outils d'analyses bio-informatiques, ont été développés. Dans cette section, je présenterai brièvement, car détaillés en section III.1, les aspects de biologie moléculaire permettant un accès robuste à la diversité du TCR, ainsi que l'effort apporté pour la standardisation de cet axe d'analyse, puis je donnerai un aperçu de la variété des approches d'analyses statistiques et de modélisations disponibles, les outils adaptés en fonction du type de réponse immunitaire.

I.4.1. De l'échantillon à la banque d'amplicons TCR

Au vu des nombreuses et différentes populations de LT émergentes du thymus, de leurs rôles et de leurs proportions distinctes, de nombreux choix s'imposent lors de la conception d'une étude du répertoire TCR. De manière générale, il est essentiel de définir la question à laquelle nous souhaitons répondre. S'il s'agit d'une recherche de biomarqueurs à visée de diagnostic, il sera plus intéressant d'avoir un minimum de manipulation et un accès simple à l'échantillon. Si au contraire, nous sommes à la recherche d'un biomarqueur à visée thérapeutique, il sera plus intéressant de connaitre son impact dans la pathologie en se focalisant au niveau du site de la réponse immunitaire. Sur la base des connaissances acquises au laboratoire, j'ai choisi d'exposer les différentes étapes de la stratégie expérimentale en énonçant les options disponibles. Concernant mon projet, le choix effectué pour ces étapes sera justifié en section II.2 et II.3.

La première décision concerne l'origine de l'échantillon. Le choix du tissu/fluide est dépendant du caractère invasif du prélèvement. Dans certains cas et notamment dans le cadre des cancers, des biopsies peuvent être réalisées fournissant une capture de l'environnement où se déroule la réponse immunitaire. Toutefois, pour la grande majorité des études cliniques le sang est utilisé car il s'agit d'une source biologique idéale pour étudier les maladies à grande échelle dans des populations de patients

hétérogènes. Accessible et relativement abondant, le prélèvement sanguin est issu d'un geste médical non invasif qui ne nécessite pas nécessairement l'hospitalisation du patient. Bien qu'une faible proportion de cellules T spécifiques d'une réponse immunitaire soit détectable en périphérie (Christophersen et al., 2014; Roep and Peakman, 2012), plusieurs études tendent à montrer que le sang est tout de même représentatif de l'état physiopathologique d'un patient (Liu et al., 2019; Rossetti et al., 2017; Spreafico et al., 2016). Afin de maximiser l'identification de la réponse immunitaire d'intérêt, il convient de caractériser les populations cellulaires impliquées dans la physiopathologie. Leurs phénotypes et proportions sont des éléments essentiels à prendre en compte. Un tri cellulaire d'une population majoritaire ($CD4^+$ Teffs représentent 90% des cellules T $CD4^+$ du sang) n'est pas essentiel, l'information recherchée pourra être détectée malgré les « contaminants ». Si un tri semble avantageux, il conviendra de mettre en place une stratégie de tri efficace, combinant rapidité et efficacité (tri magnétique, rapide mais avec une pureté moindre mais adapté lorsque les échantillons sont nombreux sur de courtes périodes) et tri cellulaire Aria, plus fastidieux mais avec une pureté généralement >95%.

L'étape d'extraction des acides nucléiques est la première étape de biologie moléculaire pour laquelle le choix du matériel génétique de départ est à prendre en compte, à savoir analyser l'ADN génomique (ADNg) ou les ARN messagers (ARNm) des cellules. Brièvement, l'ADN est un élément biologique stable offrant une meilleure quantification des clones TCR uniques (1 molécule/cellule). À l'inverse, bien que plus sensible à la dégradation et à la quantification des clones, l'ARN contenant de multiples transcrits (10 à 100 fois supérieur à celui de l'ADNg) est le reflet du niveau d'expression de chaque cellule offrant une plus grande « flexibilité » (notamment pour des échantillons de faible quantité) sur l'utilisation de certains kits qui exigent une quantité et une concentration minimales d'ARN ou d'ADNg. Finalement, l'ARN permet de se focaliser principalement sur les chaines du TCR exprimées et donc fonctionnelles, réduisant ainsi le bruit de fond des données NGS. Une fois l'ARN ou l'ADN extrait une étape de « normalisation » des quantités de matériel est nécessaire (Mamedov et al., 2013). Pour cela un dosage, et dans le cadre de l'ARN, une qualification de l'intégrité du matériel (RIN) doit être réalisés à l'aide de diverses méthodes. Ces outils basés principalement sur la spectrophotométrie ou l'électrophorèse permettent soit un simple dosage comme le Qubit (Invitrogen®) et le Nanodrop (Thermo Fisher®) soit un dosage associé à une qualification des RIN. Il est essentiel de déterminer l'outil à

utiliser, considéré comme le plus adapté notamment à la quantité de cellules et de ne pas changer au sein d'un même projet. Nous avons observé une grande variabilité entre ces différentes méthodes.

L'étude du répertoire TCR nécessite une étape d'amplification, dépendante du matériel génétique sélectionné. Le choix du kit d'amplification de l'ADN ou de l'ARNm des TCR des cellules est également un choix majeur. De nombreux kits sont maintenant disponibles sur le marché, en prestation de service ou à travers des collaborations, offrant chacune des avantages et des inconvénients. Cette partie faisant l'objet d'un article présenté en section 0, Article 1, j'ai choisi de vous présenter les principes de biologie moléculaire majoritairement utilisés (**Figure 10**). La première approche est l'amplification par PCR (*polymerase chain reaction*) multiplex (mPCR). Le principe se base sur une combinaison d'amorces complémentaires des gènes TRJ ou TRC du TCR avec un mélange d'amorces complémentaires des allèles TRV connus. Cette méthode peut être utilisée aussi bien avec de l'ADNg que de l'ARNm. Le biais concernant la mPCR résulte de la complexité de la réaction. La combinaison d'un grand nombre d'amorces différentes est source de biais d'amplifications dû à la « compétition » entre elles. La deuxième méthode la plus utilisée est l'amplification rapide des extrémités 5′ de l'ADN complémentaires (5'RACE), uniquement utilisable à partir d'ARN. Cette

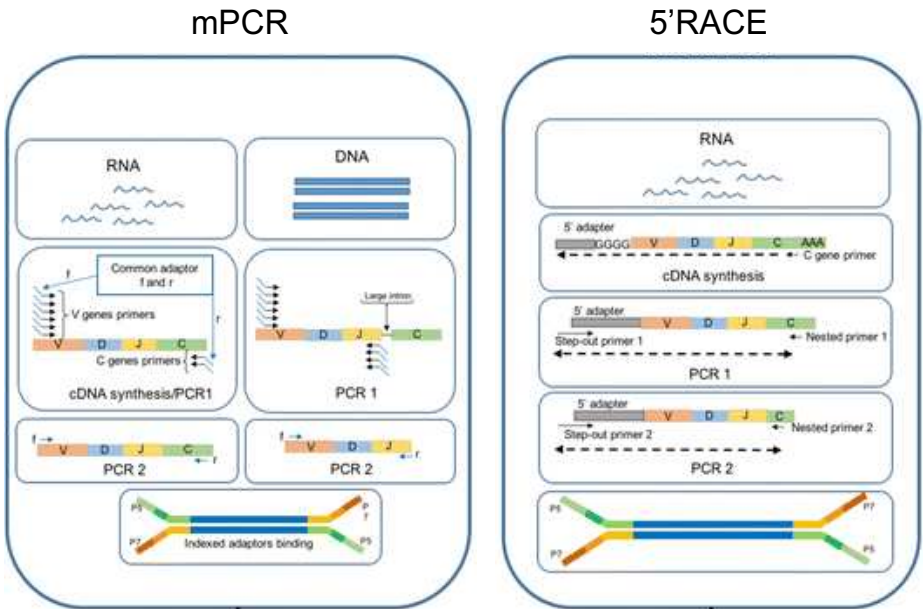


Figure 10 : Principes des protocoles de biologie moléculaire.
(Adapté de Rosati E., Dowds C.M., Liaskou E. et al. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol, 2017)

méthode, s'appuie sur l'activité de la rétro-transcriptase (enzyme permettant de transcrire l'information génétique de l'ARN en ADN complémentaire (ADNc)) qui incorpore une séquence adaptatrice à l'extrémité 5′ de l'ADNc. Cet adaptateur est utilisé comme région cible pour les amplifications suivantes en PCR nichées et permet, en combinaison avec des amorces complémentaires des régions TRC, une amplification ciblée, spécifique des transcrits TCR de façon indépendante des gènes TRV. A la suite de l'une de ces deux méthodes, les amplicons peuvent être ligaturés aux codes-barres uniques permettant leurs identifications et aux adaptateurs de séquençage.

Un grand nombre de plateformes de séquençage différentes sont disponibles sur le marché, car même s'il s'agit d'un même fabricant, les séquenceurs, options (longueurs et nombre des lectures) et chimies utilisées varient et peuvent ne pas convenir pour le séquençage de banques de TCR. Bien que dépendantes du kit et/ou de la stratégie moléculaire utilisée, les banques TCR de part la taille élevée des amplicons, la diversité faible (notamment les amplicons 5'RACE, peu divers dû à la même séquence adaptatrice à l'extrémité 5′) en font des amplicons difficiles à séquencer. Il faut au final trouver un équilibre entre (i) la taille des amplicons, la longueur de lecture et le déséquilibre entre les sens de lectures (exemple : 2x300 pb ou 1x600 pb, TCR ≈ 500 paires de bases ; (Woodsworth et al., 2013)), selon si l'on souhaite une couverture totale ou partielle de l'amplicon ; (ii) le nombre de cellules, la quantité d'ARN ou d'ADN, le nombre d'échantillons par piste de séquençage, lui-même dépendant du nombre d'identifiants fournis par le kit utilisé, la présence d'identificateurs moléculaires uniques (UMI : séquences de 8 à 12 nucléotides aléatoires incorporés, utilisés pour générer une région unique pour chaque molécule d'ADNc permettant de réduire considérablement les erreurs de séquençage), la diversité, le pourcentage de PhiX (génome de phage servant de contrôle tout en fournissant la diversité nécessaire à un séquençage de qualité), la profondeur de séquençage, c'est-à-dire le nombre de lecture par échantillon, pour obtenir une couverture suffisante des répertoires tout en gérant efficacement le taux d'erreur, le prix et la complexité des expériences.

En conclusion, chaque étape de laboratoire que nous venons de voir affecte les observations obtenues, c'est pourquoi il est essentiel de s'assurer que le traitement de tous les échantillons est aussi uniforme que possible.

I.4.2. Les Axes d'analyses du répertoire TCR

Suite à cette première étape de *design* expérimental il est aussi important d'adapter l'analyse au regard de la nature de nos échantillons.

De plus, l'une des principales difficultés dans l'étude du répertoire TCR est de savoir dans quelle mesure le résultat d'analyse obtenu reflète de manière significative l'objet biologique étudié, c'est pourquoi les axes d'analyse sont en constante comparaisons et améliorations. Amélioration nécessaire, car l'analyse de données répertoire immunitaire représente une quantité importante d'informations comme vous le verrez sur le **Tableau 5**, c'est pourquoi des équipes se concentrent sur l'optimisation de l'efficacité des calculs utilisés dans l'analyse du répertoire des cellules immunitaires (Rognes et al., 2022)

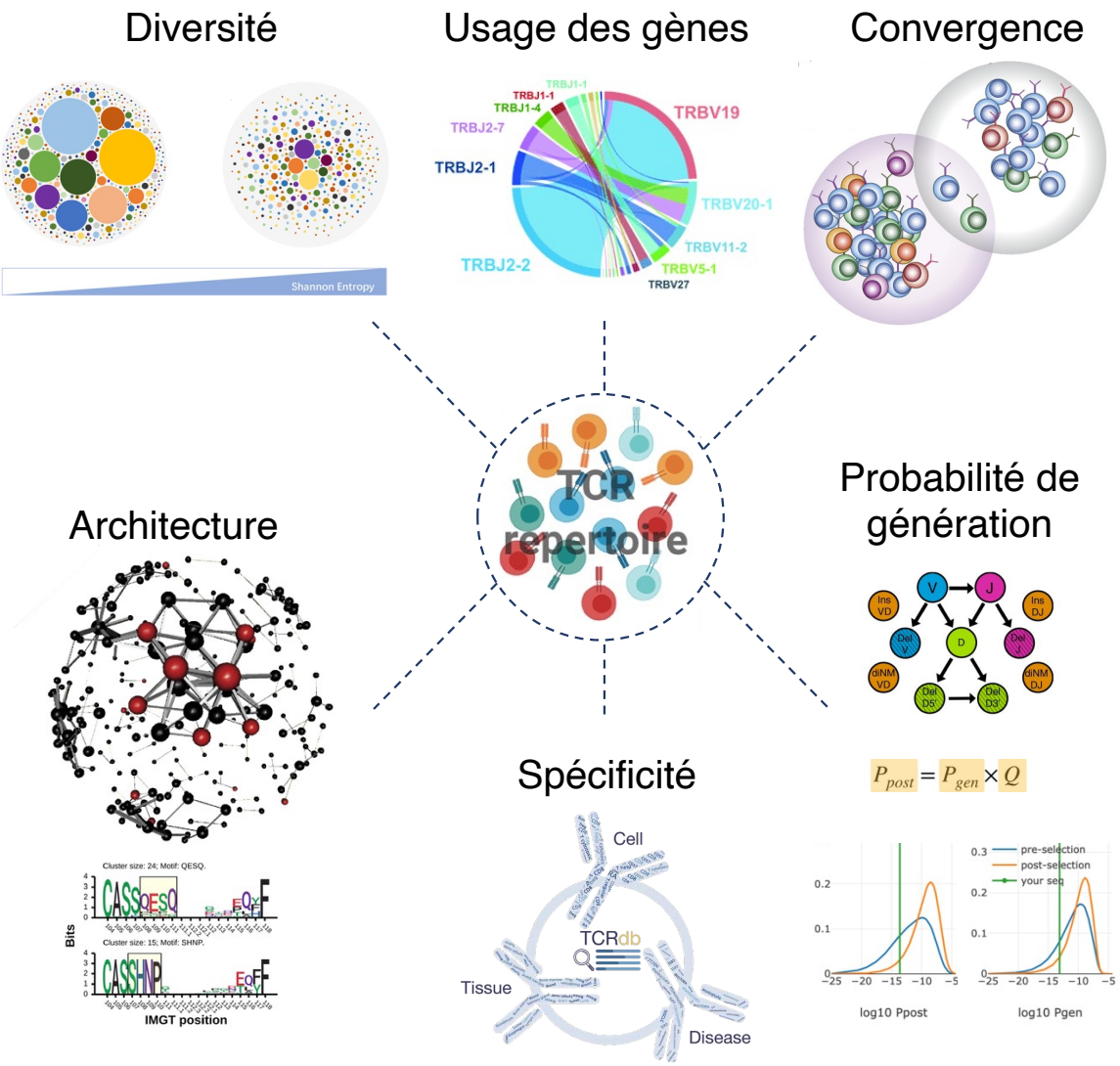


Figure 11 : Axes d'analyses du répertoire TCR.

I.4.2.1. L'Alignement des séquences TCR

L'alignement des séquences représente l'étape à laquelle les lectures brutes issues du séquenceur vont être alignées avec un « génome » de référence. Il existe plus d'une dizaine d'outils d'alignement différents ayant chacun leurs propres caractéristiques et ayant fait l'objet de nombreuses analyses comparatives (Afzal et al., 2019; Heather et al., 2017; Zhang et al., 2020). Ces logiciels varient de par (i) leurs implémentations, le langage (Java, Python, C^{++}, …) ; (ii) la méthode d'alignement (Blast, K-mers, Alignement Global, …) ; (iii) leurs options (correction des erreurs PCR, parallélisation, …) et (iv) les informations extraites. Toutefois, pour l'ensemble, l'objectif principal est la reconstruction et l'annotation des séquences d'amplicons en clonotypes, c'est-à-dire l'identification de l'allèle du gène TRV et TRJ ainsi que la région CDR3, après quoi les clonotypes identiques, caractérisés par une combinaison unique entre un TRV, un TRJ et un CDR3 en acides aminés sont regroupés conduisant ainsi au dénombrement de leurs occurrences. Au travers de ces étapes, un contrôle qualité des séquences est effectué permettant de supprimer les lectures dont la fiabilité du nucléotide détecté n'est pas assurée. Finalement nous obtenons un fichier contenant les informations sur les clonotypes qui composent le répertoire de chaque échantillon analysé.

La suite de cette partie sera organisée selon la **Figure 11** qui offre un aperçu des principaux types et niveaux d'analyses. L'analyse du répertoire TCR se structure en 6 grands axes que je vais présenter et illustrer à l'aide d'articles qui m'ont inspiré au cours de ma thèse.

I.4.2.2. La diversité

Nous avons beaucoup parlé de la diversité du répertoire TCR, en effet, son étude constitue le premier axe d'analyse communément abordé. Considéré comme un écosystème, la diversité du répertoire est analysée grâce à des outils développés pour l'écologie. Il existe un grand nombre d'indices différents portant sur divers aspects de la diversité α (diversité locale d'un répertoire), dont les plus courants sont (i) la richesse, qui correspond au nombre d'espèces uniques ; (ii) l'indice de Shannon, qui prend en compte la richesse et l'abondance relative des « espèces » au sein d'un répertoire (Shannon, 1948) et (iii) l'indice de Simpson, un indice de dominance des « espèces », évaluant la diversité des expansions clonales (Simpson, 1949). Cette diversité peut être analysée à l'échelle des gènes TRV et TRJ ainsi qu'à l'échelle des

clonotypes ou des régions CDR3. Le suivi de la diversité chez les patients peut révéler une dynamique clonale en corrélation avec la réponse au traitement ou d'autres caractéristiques cliniquement pertinentes. De plus, en associant les informations clonales aux phénotypes cellulaires, il est possible d'identifier des types cellulaires pertinents pour la maladie, qui peuvent être utiles pour le pronostique de la maladie face à un traitement et le développement de stratégies d'immunothérapie ciblées.

En examinant l'expression de la chaîne β du TCR par NGS dans le sang de patients traités avec un anticorps anti-CTLA4, Postow et al. ont mis en évidence des informations prédictives du répertoire TCR à travers une analyse de la diversité des combinaisons TRBV-TRBJ. Les patients ayant eu un bénéfice clinique présentaient, avant traitement, une plus grande richesse et une meilleure équitabilité de l'utilisation des gènes TRBV-TRBJ que les patients sans bénéfice clinique (Postow et al., 2015). Plus récemment Bortone et al. ont montré que la diversité de patients atteints par un mélanome est associée à la survie et à la réponse à un traitement, fournissant une preuve de concept de l'utilisation de la diversité du répertoire TCR comme biomarqueurs de prognostique pour cette indication.

I.4.2.3. L'usage des gènes du TCR

L'usage des différents gènes composants les chaines du TCR est une analyse comparative d'expression génique. Elle peut être le reflet d'une réponse face à un super-antigène viral ayant la capacité d'activer en ciblant la région V du TCR, l'ensemble des clones de cette famille de gène (Moreews et al., 2021). Elle peut être aussi utilisée comme élément de distinction entre deux groupes de MAI (Liu et al., 2019), néanmoins, cette différence peut s'expliquer par le biais de prédisposition HLA connus.

I.4.2.4. La convergence des répertoires TCR

La convergence des répertoires correspond à l'identification de clonotypes ou CDR3 partagés. L'hypothèse étant qu'une réponse immunitaire vis-à-vis d'antigènes communs peut être source d'une augmentation du partage entre individus. Cette hypothèse dépend du type de réponse immunitaire, si nous sommes dans le cadre d'une réponse engageant de fortes expansions clonales avec un nombre de patients/échantillons suffisamment important alors cette approche est idéale pour identifier des cellules impliquées dans la pathologie.

C'est ce qu'ont notamment montré Emerson et al. en 2017 dans le cadre de biomarqueurs de diagnostic face à un pathogène. Ils ont identifié des signatures TRB d'infection au CMV à partir de 666 patients dont le statut sérologique face au CMV était connu. Cette signature TCR enrichie en séquences spécifiques du CMV leur a permis de classer avec une excellente fiabilité les patients en fonction de leurs statuts sérologiques CMV et ceci à partir de répertoires TCR de cellules mononucléées du sang périphérique (PBMC) d'une cohorte de validation composée de 120 sujets différents (Emerson et al., 2017). Ces résultats ont été confirmés par De Neuter et al. sur une nouvelle cohorte à partir de LT $CD4^+$ mémoires, apportant la preuve que les signatures TCR spécifiques d'antigènes viraux sont présentent dans la mémoire immunitaire de LT $CD4^+$ (De Neuter et al., 2019).

Cette approche a également été mis en place pour distinguer des patients atteints de LES et de PR face à des volontaires sains (VS). Liu et al. ont pu identifier dans une grande cohorte (877 LES, 206 PR et 439 VS) des clones TRB partagés et uniques à partir des répertoires TCRβ de PBMC des patients. Ces clones se sont révélés avoir un excellent pouvoir de classification des 3 groupes (Liu et al., 2019).

Finalement, avec des approches moins sophistiquées que celles utilisées dans les articles précèdents, Looney et al. ont évalué l'utilisation du TCR-seq comme biomarqueur prédictif de la réponse à une monothérapie anti-CTLA-4 dans un ensemble de 22 sujets atteints de divers cancers. Ils ont constaté que la clonalité (en lien avec la diversité vue précédemment) combinée à la convergence des TCR permet de prédire la réponse au traitement avec une fiabilité, plus importante que chaque métrique prise indépendamment.

I.4.2.5. L'architecture des répertoires TCR

L'ensemble des relations de similarité de clones représente l'architecture de similarité des séquences TCR. Le CDR3, région hypervariable, impliquée dans la spécificité antigénique du TCR, fait l'objet de nombreuses recherches d'homologie de séquences. Le principe de cette stratégie repose sur le regroupement de CDR3 répondant aux mêmes antigènes sur la base de similarité de séquences. Elle s'avère d'une grande utilité lorsque les approches de recherche d'homologie exacte de séquences (section I.4.2.4) ne sont pas fructueuses dû à des biais d'échantillonnages, faible partage ou de quantité de données. De nombreuses stratégies d'enrichissement de groupes de CDR3 ont été créées, se basant sur de simples distances des séquences

en acides aminées (Madi et al., 2017; Pogorelyy and Shugay, 2019; Ritvo et al., 2018), en passant par la recherche de motifs et/ou k-mers (courtes séquences de même longueurs) (Glanville et al., 2017), jusqu'à des algorithmes plus complexes tenant compte des propriétés physico-chimiques des CDR3 (Beshnova et al., 2020; Chronister et al., 2021; Dash et al., 2017; Valkiers et al., 2021; Xu et al., 2022). Ces regroupements de clones potentiellement impliqués dans des réponses immunitaires communes, sont des outils puissants comme en atteste les résultats de Ye et al. qui ont identifié chez des nourrissons atteints d'atrésie des voies biliaires des motifs au niveau de la région CDR3 sous-représentés chez les nourrissons malades, offrant potentiellement de nouvelles cibles pour un traitement clinique (Ye et al., 2021). Wang et al. ont utilisé l'analyse des motifs chez des patients positifs à l'EBV et atteints de carcinome nasopharyngé, traités avec des LT $CD8^+$ autologues spécifiques de l'EBV. Ils ont identifiés des motifs associés à une réponse favorable à l'immunothérapie LT $CD8^+$ contre le carcinome nasopharyngé EBV^+ (Wang et al., 2021). Finalement, Chiou et al. ont identifié un ensemble de clones ayant infiltrés la tumeur pulmonaire, comme étant associés à la reconnaissance des antigènes de l'EBV et d'Escherichia coli (E.Coli). La réactivité croisée de ces TCR pour des peptides dérivés du cancer et ceux d'EBV et d'E.Coli a ensuite été validée expérimentalement, montrant la fiabilité de ces approches (Chiou et al., 2021). Enfin, ces approches peuvent également être intéressantes dans le cadre de cohortes de patients plus limitées mais aussi en cas de réponse immunitaire privée, en analysant la similarité des séquences au sein d'un même répertoire.

I.4.2.6. Modèle génératif de recombinaison VDJ

Le modèle de recombinaison VDJ des chaines du TCR a été étudié comme un modèle mathématique par l'équipe d'Aleksandra Walczak et de Thierry Mora (Murugan et al., 2012). Ainsi, à partir de séquences TCR non fonctionnelles (non contraintes à la sélection thymique), ils ont établi un outil permettant d'estimer la probabilité de génération d'une séquence donnée, (Pgen; Marcou et al., 2018). Cette probabilité est calculée en utilisant des probabilités cumulatives basées sur un modèle qui implique l'usage de chacun des gènes TRV, TRD et TRJ, ainsi que les probabilités de délétions et d'insertions de nucléotides aléatoires au niveau de la région hypervariable. En effet, bien que la recombinaison VDJ soit considérée comme stochastique, nous pouvons observer un large panel de distribution des probabilités de génération. Ainsi cette approche, permettant de modéliser un répertoire TCR sain, peut être utilisée comme mesure contrôle d'un clone et ainsi différencier les clonotypes

partagés dû à la recombinaison de ceux partagés dû à une réponse antigénique commune (Komech et al., 2018; Pogorelyy et al., 2018; Rosati et al., 2020). Plus récemment, la même équipe a créé un nouveau modèle prenant en compte un facteur de sélection (Q) permettant à partir de la Pgen d'estimer la probabilité post-sélection (Ppost ; Isacchini et al., 2020; Sethna et al., 2020). Ce nouvel outil offre la possibilité de comparer les facteurs de sélections (Q) d'une même séquence en fonction d'une maladie, d'un organe ou sous l'impact d'un traitement, d'une vaccination.

I.4.2.7. La spécificité

Comme défini précédemment, la spécificité du TCR reste une thématique complexe à appréhender. Un effort important a été fait depuis quelques années pour proposer des bases de données contenant des séquences TCR avec des spécificités connues, majoritairement, virales, auto-immunes, tumorales (Bagaev et al., 2020; Shugay et al., 2018; Tickotsky et al., 2017; Vita et al., 2019). Néanmoins pour les séquences TCR recensées, la/les spécificité(s) associée(s) n'a(ont) pas été démontrée(s). Néanmoins, c'est une source d'information essentielle pour caractériser, comprendre et interpréter les analyses.

La spécificité peut donc être attribuée à un TCR s'il est déjà répertorié dans des bases de données, mais à ma connaissance nous ne recensons pour le moment que peu de spécificité. C'est pourquoi plusieurs outils d'inférence de la spécificité des TCR ont été publiés (Dash et al., 2017; Gielis et al., 2019; Jokinen et al., 2021; Jurtz et al., 2018; Sidhom et al., 2021; Springer et al., 2020). Ces outils sont basés sur la théorie selon laquelle les TCR qui reconnaissent le même peptide partagent des motifs et des séquences spécifiques au niveau du CDR3. Pour cela, les auteurs ont proposé différentes mesures de similarité ainsi que différentes approches computationnelles. Bien qu'il y ait eu des progrès dans cet aspect important de la spécificité, un réel manque d'évaluation de ces outils de façon indépendante est nécessaire pour guider au mieux les immunologistes.

Néanmoins, ceux-ci ont permis à Gittelman et al. d' identifier des signatures TCR issues de plus de 76 patients atteints par le SARS-CoV-2 et ont constaté que 97,4% des cas présentaient un niveau plus élevé de TCR spécifique de SARS-CoV-2 en comparaison à des patients négatifs au SARS-CoV-2. Ils ont en parallèle identifié 40 infections « suspectes », sans test PCR de confirmation, mais ces sujets étaient principalement composés de personnes ayant eux des symptômes similaires à ceux

provoqués par une infections par le SARS-CoV-2 ou cas contact avec des membres de leurs foyers. Ces résultats nous montrent que les LT constituent une mesure biologique fiable et persistante établissant une infection passée par le SRAS-CoV-2 symptomatique ou non (Gittelman et al., 2020).

I.4.3. Une source de biomarqueurs

Les LT et donc les TCR ont un rôle majeur dans la réponse immunitaire, son évolution et sa résolution, ils sont donc de bons candidats pour surveiller l'état du SI des individus. En effet, les maladies chroniques telles que les maladies auto-immunes et les cancers ont dans leur grande majorité un développement relativement lent. Par exemple, dans le cas de maladies auto-immunes comme le DT1, les premiers symptômes de la maladie peuvent survenir seulement des années après le début de la maladie ou la découverte d'auto-anticorps dans le sérum. L'identification du répertoire TCR potentiellement délétère ou associé à la maladie pourrait aider à prévenir le développement de ces pathologies. Les LT sont également impliqués dans la réponse face aux agents infectieux. Les infections représentent un large spectre de maladies pouvant être bénignes comme le rhume ou l'angine mais aussi graves comme le sida, la tuberculose et plus récemment la COVID-19. Si la contagiosité et la mortalité peut varier d'une pathologie à l'autre, elle varie également d'un individu à l'autre. Ceci en fonction de multiples causes dont son SI. Son implication dans la défense de l'organisme face aux agents pathogènes reste à élucider pour bon nombre d'infections d'un point de vue de la spécificité des cellules impliquées. La recherche a permis d'éradiquer certaines d'entre elles grâce au développement de vaccins permettant une activation spécifique de la réponse humorale, l'accent ayant été mis sur le rôle des anticorps. Néanmoins, de nombreuses données suggèrent que la contribution des LT est essentielle pour assurer une protection (i) précoce, la réponse cellulaire initie la réponse humorale ; (ii) large, les cibles des anticorps sont limitées aux protéines de surface, tandis que les épitopes des LT proviennent à la fois de protéines structurelles et de surface et (iii) durable (Ahmed and Akondy, 2011; Guo et al., 2022; Hellerstein, 2020; Kalimuddin et al., 2021). Il ne fait donc aucun doute que l'étude du répertoire TCR dans le cadre des maladies infectieuses peut aider la communauté scientifique à résoudre certains défis comme celui que nous vivons actuellement. À travers l'ensemble de ces exemples, le répertoire TCR représente un biomarqueur de demain.

Un biomarqueur peut se définir comme un paramètre biologique quantifiable lié à une condition physiologique ou pathologique, permettant une prise de décision. Les biomarqueurs servent un large éventail d'objectifs dans le développement des médicaments, les essais cliniques, les stratégies thérapeutiques et contribueront à la mise en œuvre de la médecine personnalisée. Le principe de l'utilisation d'un « biomarqueur » est de rechercher la signature de l'impact biologique (actuel ou passé). De façon plus précise concernant le TCR-seq, comme décrit précédemment, son étude nous apporte divers paramètres biologiques quantifiables tel que la diversité globale, l'identification d'expansions clonales, l'identification et le suivi de clonotypes particuliers. Ensemble ces découvertes pourront dans un second temps être utilisées à en bioingénierie à travers le transfert adoptif de cellules T avec des TCR génétiquement modifiés spécifiques d'antigènes.

En parallèle de cela, si l'analyse du répertoire TCR par séquençage de nouvelle génération vise à être un outil clinique, il est essentiel d'en assurer sa standardisation, en termes de vocabulaire, de production, contrôle qualité et d'analyse.

I.4.4. Nécessité de standardisation

Au vu de l'émergence et de l'intérêt grandissant de l'étude du répertoire immunitaire, une communauté internationale, « The AIRR community », s'est créée en 2014 sous l'impulsion de Jamie Scott, Tom Kepler, et Felix Breden (https://www.antibodysociety.org/the-airr-community/). Son but étant de promouvoir des normes et des recommandations en vue d'une meilleure standardisation de l'ensemble des études portant sur l'étude des répertoires des récepteurs immunitaires adaptatifs (ou *Adaptive immune receptor repertoire, AIRR*), incluant le récepteur des cellules B (BCR) par séquençage massif (*AIRR-sequencing*, AIRR-seq). Ce groupe, dont je suis membre, est structuré en plusieurs axes visant à (i) une clarification du vocabulaire utilisé par la communauté scientifique à travers une meilleure définition des populations cellulaires accompagnée d'une standardisation des informations complémentaires pour le partage des données dans des fichiers, des schémas et des nomenclatures normalisés ; (ii) la validation des gènes qui composent l'AIRR en donnant accès à un formulaire de dépôt pour l'identification et la soumission de nouveaux allèles avec un processus de soumission efficace tout en validant les bases

de données de références germinales ; (iii) la standardisation des méthodes de production par des recommandations afin de produire puis de partager des données dont la qualité et la fiabilité peuvent être évaluées. Pour cela, les objectifs sont d'identifier et de collecter des calibreurs et/ou contrôles appropriés, de les promouvoir auprès des chercheurs, commerciaux et de tout utilisateur d'AIRR-seq ; (iv) contrôler les outils d'analyses afin d'encourager les bonnes pratiques de création pour assurer une utilisation correcte des logiciels d'analyse en définissant des statistiques adaptées et validées pour caractériser les données (Olson et al., 2019) et proposer des mesures statistiques d'évaluation de la crédibilité du répertoire ; (v) assurer le partages de données AIRR-seq à travers, en lien avec le premier groupe de travail, la standardisation des fichiers de données brutes mis à disposition de la communauté scientifique, la création d'une base de données regroupant celles disponibles. D'autres groupes travaillent en étroite collaboration avec plusieurs autres équipes à travers une thématique, notamment autour des aspects juridiques et éthiques des données issues de l'AIRR-seq mais également un groupe en charge d'explorer, d'évaluer la possibilité d'appliquer la technologie AIRR-seq comme outil de diagnostic en clinique et finalement d'étendre son utilisation. Ensemble, cette communauté a pour objectif d'évaluer, de décider et d'imposer un certain nombre d'exigences afin de garantir une plus grande transparence et crédibilité des résultats AIRR-seq.

Chapitre II. EXPERIMENTATIONS ET OBJECTIFS

Dans cette section, je vais présenter les essais cliniques coordonnés par David Klatzmann dont sont issus les échantillons sur lesquels mes travaux ont été réalisés. Je détaillerai également les méthodes que nous avons choisi d'appliquer pour mon projet. Mais avant cela, voyons comment nous avons guidé ces décisions.

Dans un contexte de recherche de biomarqueurs dans le cadre des MAI chez l'homme, nous avons eu accès uniquement à des prélèvements sanguins, prélèvements non invasifs qui se prêtent parfaitement à l'objectif de la recherche de biomarqueurs. Il est important de tirer l'information de ce fluide, car comme décrit précédemment nous savons que dans le sang circule une information portée par les LT qui caractérise un état pathologique.

Au sujet des populations à étudier, les cellules $CD4^+$ Tregs jouant un rôle majeur dans le cadre des MAI, ne représentent qu'une faible proportion des LT $CD4^+$ (5-10% de $CD4^+$ Tregs et 90-95% de $CD4^+$ Teffs). Il nous a paru essentiel, afin d'attribuer une interprétation biologique aux résultats, de distinguer ces deux sous-populations cellulaires. Dans cette optique, les prélèvements sanguins, provenant du monde entier pour certains essais, ont été traités par tri magnétique dès réception sur des prélèvements non-congelés. Le choix du tri magnétique par rapport à un tri plus efficace en termes de pureté a été décidé pour des raisons logistiques, étant donné la réception au fil de l'eau des prélèvements.

Concernant le choix du matériel génétique à étudier, il découle lui-même de la comparaison des méthodes d'amplifications du matériel génétique du TCR (section III.1). Nous avons ciblé une méthodologie basée sur l'ARN, matériel qui nous semblait également évident de par l'état d'activation des cellules permettant de détecter les transcrits TCR des cellules certes rares mais impliquées dans la réponse immunitaire d'intérêt. De plus, de par la grande diversité des répertoires TCR issus de sous-types cellulaires provenant de prélèvements d'une partie du sang, il était important de sélectionner une méthode nous permettant une capture optimale de cette diversité.

Finalement, l'outil d'alignement sélectionné a été décidé à l'aide de multiples analyses comparatives (Afzal et al., 2019; Heather et al., 2017; Zhang et al., 2020). MiXCR créé en 2015 par Bolotin est le plus populaire d'entre eux (Bolotin et al., 2015), faisant de celui-ci une référence en perpétuelle amélioration.

En raison de la complexité et de la diversité des répertoires TCR individuels, l'identification des signatures TCR associées à un antigène est un défi qui fait l'objet de cette thèse.

II.1. Cohortes de patients

Dans le cadre de ma thèse, j'ai pu étudier le répertoire TCR dans différents contextes immuno-pathologiques au travers d'études translationnelles menées au laboratoire (sections II.1.1 et II.1.2) ou rendues publiques (II.1.3). L'ensemble des jeux de données sont décrits dans les sections suivantes et résumés dans le **Tableau 5**.

II.1.1. Transimmunom : un essai observationnel

L'ensemble des maladies systémiques inflammatoires/auto-immunes (MAI) forme un vaste groupe de maladies complexes pour lesquelles la/les causes/processus physiopathologiques restent mal compris. En 2006, McGonagle et McDermott offrent une nouvelle classification en proposant l'idée qu'il ne s'agirait pas de deux catégories de maladies distinctes mais plutôt d'un continuum autoinflammatoire-autoimmun (McGonagle and McDermott, 2006). Cet ensemble allant des maladies inflammatoires pures aux maladies auto-immunes pures englobe un large panel de maladies inflammatoires avec certaines composantes auto-immunes, et vice versa. En

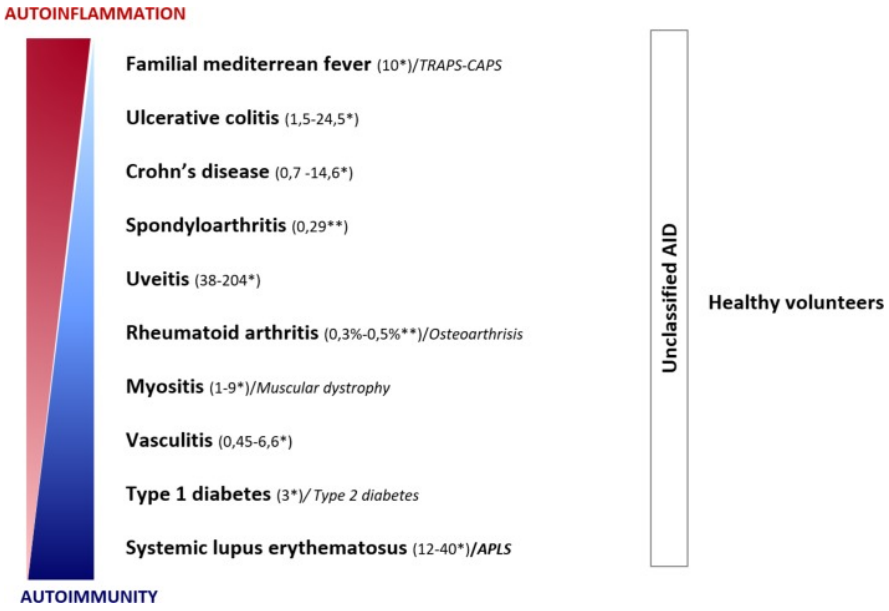


Figure 12 : Présentation des maladies incluses dans l'essai Transimmunom

combinant les sciences fondamentales, cliniques et le traitement de données massives, l'essai observationnel Transimmunom (NCT02466217) a pour objectif de revisiter la nosologie de plusieurs MAI (Lorenzon et al., 2018). Les investigations biologiques appliquées à cet essai ont consisté en un immuno-monitoring large, incluant l'immuno-protéomique sérique, la transcriptomique, l'étude du microbiote et du répertoire TCR.

Les MAI ciblées dans l'étude sont présentées **Figure 12**, illustrant ce continuum autoinflammatoire-autoimmun. Dans le cadre de ce travail, nous nous sommes focalisés sur deux d'entre elles : le diabète de type 1 (n = 52) et la polyarthrite rhumatoïde (n = 68), pour lesquelles nous avons étudié les répertoires TCR des $CD4^+$ Tregs et des $CD4^+$ Teffs du sang, en comparaison à ceux de volontaires sains (n = 54). Ces deux MAI forment un groupe représentatif pour l'étude du TCR, ayant une composante auto-immune certifiée, spécifique d'organe ou de tissus avec une composante auto-inflammatoire variable.

II.1.2. Lupil-2 : un essai clinique longitudinal

LUPIL-2 (NCT02955615) est un essai clinique de phase IIb, multicentrique, randomisé, en double aveugle visant à évaluer l'efficacité et l'innocuité de l'IL-2 à faible dose ($IL\text{-}2_{fd}$; 1,5 Millions d'unités internationales) chez 100 patients atteints de lupus

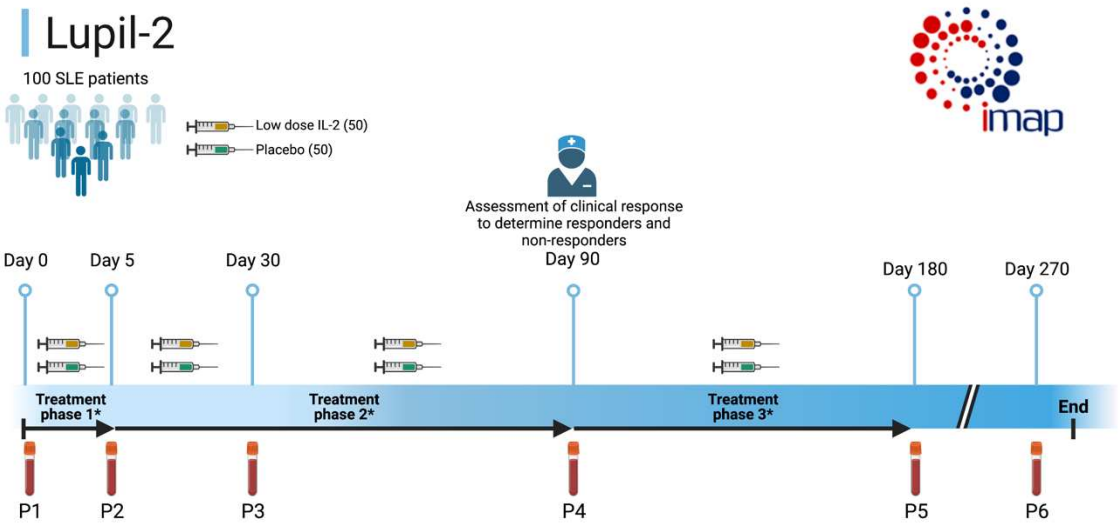


Figure 13 : Description de l'essai clinique LUPIL-2.
**Traitement phase 1 : Tous les patients reçoivent 1 injection par jour. Traitement phase 2 : Tous les patients reçoivent 1 injection par semaine. Traitement phase 3 : Seuls les patients répondeurs au traitement reçoivent 1 injection par semaine.*

érythémateux systémique modéré à sévère (score d'activité de la maladie : SLEDAI>6). Une moitié des patients (50) ont reçu le traitement à IL-2_{fd} et l'autre moitié (50) le placebo. Le traitement a été administré sur une durée minimale de 3 mois et maximale de 6 mois en fonction de la réponse face au traitement ; cette dernière ayant été évaluée au 3ème mois. Les patients sont considérés répondeurs si le score SLEDAI diminue d'au moins 4 points, en l'absence de détérioration lors de l'évaluation globale du médecin (https://clinicaltrials.gov/ct2/show/NCT02955615). Le suivi longitudinal nous a donné accès à 5 points de prélèvement (**Figure 13**) : début du traitement (Prélèvement 1, P1), jour 5 du traitement où la proportion de $CD4^+$ Tregs est au maximum (P2), un mois après le début du traitement (P3), mi-traitement (3ème mois, P4), en fin de traitement (6ème mois, P5) et 3 mois après la fin du traitement (9ème mois, P6). Pour chacun de ces points, un prélèvement sanguin a été effectué pour l'étude des répertoires TCR des $CD4^+$ Tregs et des $CD4^+$ Teffs. Dans le cadre de ce projet, nous proposons ici de suivre la dynamique du répertoire TCR de ces populations sous traitement à l'IL-2_{fd}.

Sur les 100 patients inclus dans l'essai clinique, 2 ont été exclus pour une sortie de protocole. Un patient a refusé les études ancillaires et 8 patients provenants du Mexique n'ont pu être traités suite à des problèmes logistiques. Sur les 89 patients, la qualité des prélèvements (Nombre de cellules>2000 ainsi que la pureté des cellules $CD4^+$ Tregs : FoxP3>80%) nous a conduit à exclure les prélèvements de la première visite de 28 patients. Sur les 61 patients dont les prélèvements de la première visite ont été validés, l'ensemble des échantillons ont été traités selon le protocole décrit en section II.3. La cohorte finalement étudiée ne présente pas de déséquilibres entre le nombre de patients traités par un placebo (n = 29) de celui des patients traités par IL2-$_{fd}$ (n = 32).

II.1.3. Données publiques : Sars-Cov-2

Le Sars-Cov-2 ou la COVID-19 est un nouveau coronavirus, apparu en Chine fin décembre 2019. Cette infection émergente est encore en 2022 un problème majeur de santé mondiale, car bien que des vaccins aient été créés, la pandémie perdure et provoque toujours chez certaines personnes un syndrome respiratoire aigu, responsable à l'heure actuelle de plus de 6,3 millions de décès. La recherche mondiale

s'est alors rapidement engagée dans un effort commun afin de mieux comprendre cette infection et une quantité importante de données a été collectée et mise à disposition de toute équipe. J'ai pu récupérer un ensemble de données public de séquences TCRβ provenant du sang périphérique de sujets infectés par le virus Sars-Cov-2 et dont la progression de la maladie est connue. Composé de 35 patients décédés suite à l'infection et de 38 patients ayant récupéré, nous avons souhaité évaluer le pouvoir prédictif de l'évolution de la maladie chez ces patients.

II.2. Du prélèvement à l'ARN des cellules

L'ensemble des protocoles expérimentaux ont été validés en amont, nous permettant de traiter les échantillons humains provenant de ces différents essais selon le même protocole, rendant ainsi possible une analyse transversale. Ainsi, à partir du sang (30ml) des participants aux différentes études, les $CD4^+$ Tregs et les $CD4^+$ Teffs ont été triés avec le kit d'enrichissement magnétique EasySep Human $CD4^+CD127^{low}CD25^+$ regulatory T cells (Stemcell®) (**Figure 14,** gauche). Brièvement, les PBMC sont isolées par gradient de densité, consistant à déposer le sang sur une solution de polysaccharide ou ficoll puis à centrifuger selon les recommandations du fournisseur. L'anneau de cellules mononucléées contenant entre autres les Tregs et les

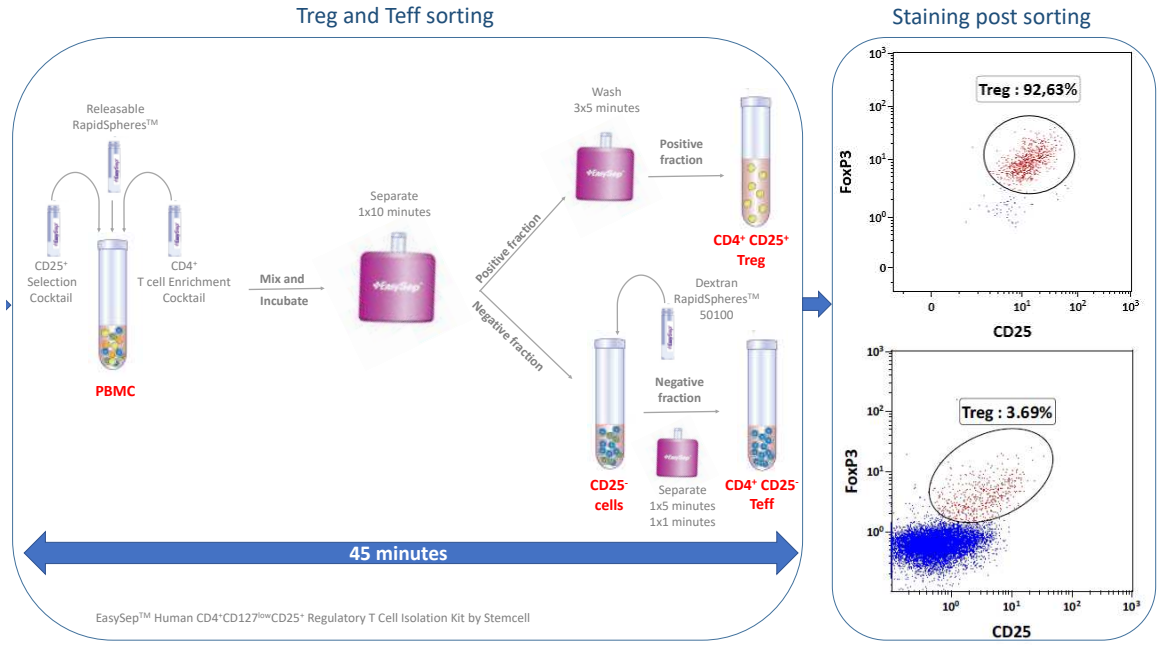


Figure 14 : Protocole de tri cellulaire magnétique.

Teffs est récupéré puis traité avec un cocktail d'anticorps permettant de cibler toutes les cellules non $CD4^+$ (CD8, CD14, CD16, CD19, CD20, CD36, CD56, CD66b, CD123, TCRγ/δ, glycophorin A), et un cocktail anti-CD25 couplés à des billes magnétiques. Les cellules $CD25^+$ sont isolées par sélection positive à l'aide de billes magnétiques Releasable RapidSpheres™. La fraction positive contient les cellules $CD4^+$, $CD25^+$ ($CD4^+$ Tregs) hautement purifiées et la fraction négative l'ensemble des cellules $CD25^-$, dont les cellules $CD4^+$, $CD25^-$ ($CD4^+$ Teffs). Un deuxième tri magnétique à l'aide de billes Dextran RapidSpheres™ venant cibler l'ensemble des cellules $CD4^-$ va permettre d'isoler les $CD4^+$ Teffs dans la fraction négative. Suite au tri, le compte des cellules est effectué puis la suspension cellulaire mise en tampon de lyse (RNAqueous, Ambion®). En parallèle, nous vérifions la pureté cellulaire par cytomètre en flux, sur la base du $CD25^+$ et du FoxP3. La technique permet de récupérer en moyenne 1.10^5 $CD4^+$ Tregs et 1.10^6 $CD4^+$ Teffs avec des puretés de 80-99% (**Figure 14**, droite). Les échantillons pour lesquels une pureté >80% ne serait pas respectée ne sont pas validés.

II.3. Création des banques TCR et séquençage de nouvelle génération

L'ARN des cellules triées a été extrait à l'aide du micro kit RNAqueous (Ambion®), un traitement DNAse a été effectué puis la qualité ainsi que la quantification de l'ARN ont été évaluées au Bioanalyser (Agilent®). Les ARN totaux ont été rétro-transcrits pour la création des banques par la méthode de 5'RACE ciblant des transcrits TCRα/β (SMARTer Human TCR a/b Profiling Kit version 1, Takara®) que j'ai validée (Barennes et al., 2021), standardisée et automatisée au laboratoire. Les résultats de cette étude comparative sont présentés dans la section résultats.

Cette méthode intègre une double identification moléculaire par échantillon nous permettant de créer 96 banques TCR ayant une combinaison d'indexage unique. Les banques d'ADN ont été transmises à une plateforme de séquençage (LIGAN, Lille, France) qui réalisait les mélanges équimolaires pour le séquençage (**Figure 15**). Dans un objectif d'optimisation, et après avoir validé le nombre de séquences nécessaires à 1.10^6 par échantillons, nous avons fait séquencer nos banques sur deux lignes de

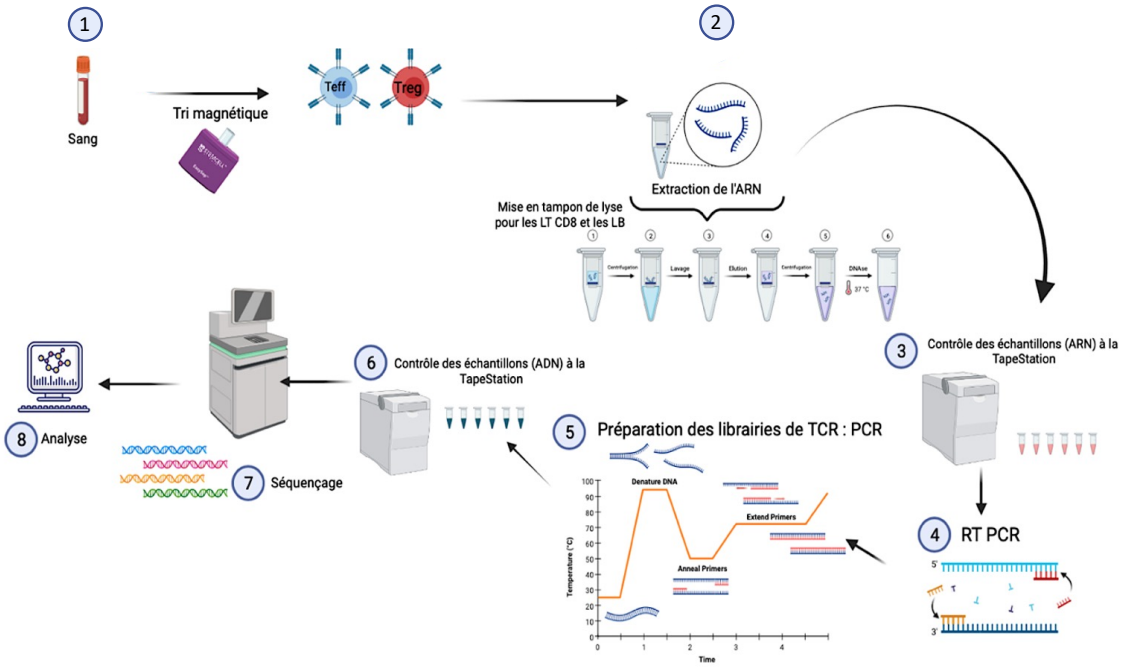


Figure 15 : Schéma du protocole expérimental de création des librairies TCR.
(1) A partir du sang frais, tri magnétique afin de récupérer les Teffs et Tregs. (2a) extraction de l'ARN des LT triés. (3) Contrôle des échantillons (ARN) à la Tapestation. (4) RT-PCR puis (5) préparation des librairies. (6) Contrôle de l'ADN à la TapeStation. (7) Les échantillons validés sont envoyés au séquençage puis (8) analysés par des outils bioinformatiques.

séquençage soit sur séquenceur HiSeq V2500 soit sur séquenceur NovaSeq 6000 (Illumina®). Les deux lignes de séquençage étant séparées, cette configuration me permettait le séquençage simultané de 2x96 banques, soit 192 échantillons. Après démultiplexage réalisé par la plateforme de séquençage à l'aide du logiciel Illumina (HiSeq Analysis Software ou bcl2fastq2 Conversion Software), la qualité des séquences de chaque échantillon a été évaluée avec le package « FastQC » (Andrews, 2015). L'alignement des séquences et l'annotation des séquences TCR sont effectuées avec un outil publié et communément utilisé, MiXCR (Bolotin et al., 2015) fournissant une correction d'erreurs et l'occurrence de chaque TCR unique (clonotype) ainsi que les scores de qualité d'alignement et de séquence.

II.4. Analyses des données

Pour analyser la diversité des répertoires TCR, l'indice de Shannon correspond à un indice qui prend en compte la richesse et la régularité d'occurrence des clonotypes dans l'échantillon. Cet indice est représentatif de la diversité globale d'un répertoire :

plus sa valeur est élevée, plus la distribution des clonotypes est diverse. C'est également l'un des indices les moins affectés par le biais d'échantillonnage, qui est inévitable dans l'étude du répertoire TCR (Bortone et al., 2021). Autrement l'entropie de Renyi offre une courbe représentative de nombreux indices de diversité à travers un facteur α, mettant l'importance de l'analyse sur les catégories de clonotypes de plus en plus expandus au sein du répertoire. Ce paramètre α va de la richesse spécifique (α=0), qui correspond au nombre de clonotypes indépendamment de leur abondance, à la dominance clonale (α→+∞), correspondant à la fréquence du clonotype le plus prédominant. Pour α = 1, on calcule l'indice de diversité de Shannon.

En ce qui concerne la mesure de similarité ou convergence des répertoires, l'indice de Jaccard (Jaccard, 1912), couramment utilisé en immunologie, est défini comme la taille de l'intersection divisée par la taille de l'union des ensembles d'échantillons. Cet indice ne prend en compte que la présence et l'absence d'espèces, sans tenir compte de l'occurrence, et varie de 0 (dissimilarité parfaite) à 1 (similarité parfaite).

Nos analyses d'identifications de variables discriminantes sont exclusivement basées exclusivement sur la région CDR3 des TCRs. En effet, c'est cette région qui interagit avec le peptide antigénique (Danska et al., 1990). De même, ceci nous permet d'effectuer des analyses croisées sur des sujets indépendamment de leur génotype HLA. De telles analyses centrées sur les CDR3 sont courantes dans la littérature et se sont avérées efficaces (Glanville et al., 2017). Nous avons utilisé pour la première fois dans l'étude du répertoire TCR la sPLSDA (*sparse Partial Least Squares Discriminant Analysis*) comme méthode supervisée de sélection de variables biologiquement pertinentes (Lê Cao et al., 2011). Cette stratégie est, comme son nom l'indique, adaptée à des données dispersées (faiblement partagées) telles que le répertoire TCR (Elhanati et al., 2018). Elle vise à identifier, parmi un très grand nombre de variables, celles qui ont la plus grande importance de classification de groupes de type qualitatif (pathologies, réponse au traitement, ...).

Nous combinons l'identification de variables discriminantes avec une stratégie de classification basée sur une méthode de rééchantillonnage "Leave-one-Out" (LOO ; **Figure 16**). Chaque itération représente un modèle dans lequel un échantillon est considéré comme un échantillon d'essai (Test Data) et tous les autres échantillons comme des données d'entraînement (Training Data). Pour chaque modèle, le sPLSDA

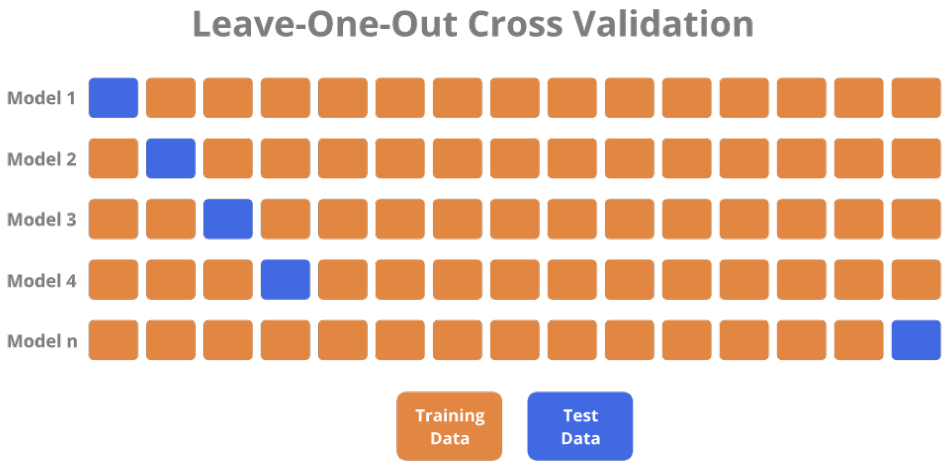


Figure 16 : Schéma explicatif du rééchantillonnage "Leave-One-Out".

est appliqué sur les données d'entraînement, puis les variables sélectionnées sont utilisées pour classer l'échantillon d'essai selon les groupes qualitatifs. Pour toutes ces classifications nous utilisons l'algorithme d'apprentissage « glmnet », qui est une méthode de classification qui stabilise les modèles de régression linéaire grâce à une régularisation de type "lasso" qui limite les problèmes d'instabilité des prédictions par la pénalisation de certaines variables considérées comme moins importantes ou nulles. Cette technique de régularisation est également utilisée dans la méthode sPLSDA.

II.5. Objectifs de thèse

L'étude du répertoire TCR nécessite une étape d'amplification. Afin de détecter le maximum de la diversité des cellules T d'un individu nous avons commencé ce projet par la validation de la méthode la plus appropriée à nos échantillons. Nous avons ensuite évalué plusieurs stratégies d'identification de variables discriminantes et identifié un algorithme adapté aux données TCR que nous avons appliqué dans différents contextes exposés dans cette section, à travers les 4 objectifs qui ont organisé mes années de thèse. Les données analysées dans le cadre de mon projet de thèse sont résumées dans le **Tableau 5**, qui contient des informations sur le nombre de patients par type de cellule, le nombre de séquences et de clonotypes uniques.

Axe 1 : validation méthodologique

Conscient des contraintes imposées par le cadre des essais observationnels et/ou cliniques ainsi que de la nouveauté du séquençage massif du TCR lors de l'initiation de ce projet, il nous a paru essentiel de débuter par la validation du protocole expérimental moléculaire permettant de capturer la diversité du répertoire TCR avec la meilleure fiabilité, sensibilité et reproductibilité, en tenant compte des contraintes imposées par nos échantillons, en particulier concernant la cellularité. Cette validation s'est déroulée dans le cadre d'une collaboration internationale, regroupant les principales méthodes académiques et industrielles créées à l'heure du design expérimental du projet. L'ensemble de cette étude a fait l'objet de mon premier article présenté en section III.1.

Axe 2 : Identification de biomarqueurs prédictifs de l'état physiopathologique

Deuxièmement, dans le cadre de l'essai observationnel Transimmunom (section II.1.1) nous avons cherché à mettre en place une méthode d'identification de réponse immunitaire $CD4^+$ Teffs et/ou $CD4^+$ Tregs entre les MAI et ainsi, mieux comprendre ces différentes maladies, identifier des groupes/catégories de clones impliqués, spécifiques ou transversaux aux maladies, et à découvrir des biomarqueurs et des cibles thérapeutiques potentielles. Ces analyses font l'objet d'un article en cours de soumission et sont présentées en section III.2.

Axe 3 : Caractérisation du pouvoir prédictif dans la réponse à un traitement immunodulateur

Dans le cadre de l'essai clinique LUPIL-2 (section II.1.2) j'ai évalué l'impact de l'$IL\text{-}2_{fd}$ sur la réponse immunitaire des $CD4^+$ Teffs et des $CD4^+$ Tregs chez des patients atteints de LES afin de (i) caractériser la dynamique du répertoire TCR des sous-types cellulaires sous traitement $IL\text{-}2_{fd}$; (ii) déterminer si ces sous-populations de $CD4^+$ partagent des cibles antigéniques similaires et (iii) distinguer les signatures TCR spécifiques de la maladie et/ou du traitement, c'est à dire identifier des biomarqueurs de l'efficacité de l'IL-2 pour la conception de futurs outils de diagnostics/thérapies et de suivi de l'état physiopathologique des patients. Ces analyses sont en cours et sont présentées en section III.3.1.

Axe 4 : Application dans le pronostic de patients infectés par la COVID-19

Ce projet ayant eu lieu pendant la période de pandémie que nous avons vécu, il m'a semblé intéressant d'étudier la réponse immunitaire dans le contexte du Sars-Cov-2, afin de tester l'association du répertoire TCR dans le pronostic de réponse immunitaire. Pour cela, j'ai tiré parti de données publiques et j'ai étudié le répertoire TCRβ des lymphocytes T sanguins chez les patients présentant une maladie sévère et chez ceux ayant « vaincu » l'infection, afin d'évaluer rétrospectivement l'évolution de la maladie et ainsi orienter la prise en charge. Ces résultats sont présentés en section III.3.2.

Project	Disease	other category	Cell subset	Number of sample	Number of total TCR sequences		Number of unique clonotypes	
					TRA	TRB	TRA	TRB
Transimmunom	HV		CD4 Teff	54	19 957 156	9 177 647	1 446 015	1 698 416
			CD4 Treg	40	15 209 413	8 553 179	520 411	633 216
	RA		CD4 Teff	68	29 656 503	14 056 153	2 070 290	2 247 557
			CD4 Treg	49	18 173 574	11 597 655	738 028	1 032 477
	T1D		CD4 Teff	52	20 397 495	10 129 715	1 186 514	1 541 443
			CD4 Treg	50	15 184 744	7 936 162	682 894	812 502
	Total			**313**	**118 578 885**	**61 450 511**	**4 631 095**	**7 213 448**
Public dataset from Snyder et al. 2020	CO19	death	T cell	35	na	4 091 211	na	3 742 740
		recovery	T cell	38	na	2 149 042	na	2 017 430
	Total			**73**	**na**	**6 240 253**	**na**	**5 582 178**
LUPIL-2	SLE	IL-2 Responder visit 1	CD4 Teff	23	24 075 717	12 300 561	951 745	1 605 178
			CD4 Treg	24	17 560 789	19 449 165	355 148	632 480
		IL-2 Responder visit 5	CD4 Teff	23	17 336 172	11 725 008	642 149	395 412
			CD4 Treg	20	14 051 850	10 341 328	1 018 256	520 878
		IL-2 No Responder visit 1	CD4 Teff	8	8 555 616	5 475 878	481 557	753 107
			CD4 Treg	8	5 533 152	5 944 873	80 160	115 453
		IL-2 No Responder visit 5	CD4 Teff	8	5 206 347	3 631 658	302 803	515 972
			CD4 Treg	9	4 969 066	3 870 910	93 407	149 091
		Placebo Responder visit 1	CD4 Teff	17	12 245 326	4 967 178	560 781	695 193
			CD4 Treg	17	10 199 730	4 902 800	175 995	227 363
		Placebo Responder visit 5	CD4 Teff	12	6 605 496	2 900 832	267 476	344 712
			CD4 Treg	12	6 568 328	4 166 741	98 983	119 684
		Placebo No Responder visit 1	CD4 Teff	12	15 787 808	11 749 643	581 100	1 115 407
			CD4 Treg	12	7 967 217	8 879 597	140 563	180 114
		Placebo No Responder visit 5	CD4 Teff	6	4 646 140	2 774 118	233 298	354 709
			CD4 Treg	6	2 997 760	3 247 910	40 211	54 100
	Total			**217**	**164 306 514**	**116 328 200**	**3 138 931**	**6 913 009**

Tableau 5 : Bilan des échantillons analysés.
Le nombre d'échantillons peut varier entre CD4 Teff et CD4 Treg en raison de la validation des cellules triées.

Chapitre III. RESULTATS

III.1. Article 1 - « Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases »

L'étude du répertoire TCR voit, parallèlement à l'amélioration des méthodes de séquençage de nouvelles générations, un engouement de la communauté scientifique. En effet cet axe d'analyse nous offre une clé dans la compréhension du mécanisme de la réponse immunitaire adaptative qui, à terme, pourrait permettre la découverte de biomarqueurs et de nouvelles stratégies thérapeutiques.

L'étude du répertoire TCR par séquençage de nouvelles générations nécessite en amont une étape d'amplification ciblée sur le TCR. De nombreuses méthodes permettant de capter la diversité du répertoire TCR sont disponibles, néanmoins il est connu que certaines stratégies moléculaires (5'RACE et PCR multiplexe) peuvent avoir un impact distordant sur la diversité des répertoires TCR (Rosati et al., 2017). Cette première étude comparative ne traitait que trois méthodes (2 PCR multiplexe et 1 5'RACE) et n'incluait pas ou seulement deux réplicas. Nous avons entrepris une comparaison large des différentes méthodes disponibles dans les laboratoires académiques et auprès de fournisseurs industriels. L'objectif de ce projet était d'identifier parmi un large choix, une stratégie d'amplification nous offrant une définition approfondie du répertoire TCR. C'est au travers d'une collaboration internationale avec 5 laboratoires experts dans l'analyse du répertoire immunitaire que nous avons pu évaluer les différences entre les 9 méthodes en ce qui concerne la reproductibilité, la sensibilité et la fiabilité, pour la résolution des chaînes TCRα et TCRβ. Pour cela, l'ensemble des laboratoires ont reçu 9 aliquotes provenant d'un échantillon standardisé de lymphocytes T effecteurs. Nous avons également introduit dans cet échantillon standardisé, le matériel génétique provenant d'un clone issu de la lignée Jurkat (https://www.atcc.org) ayant un TCR connu, à différents ratios (1/1000 ; 1/100 ; 1/10) afin d'estimer la capacité des méthodes à détecter des clones de faible à forte concentration. De plus, nous avons anticipé des variations de quantité de matériel génique entre les échantillons. En effet, dans la perspective des études sur les Tregs, nous avons donc étudié l'impact de la quantité de matériel biologique de départ (ARN). Pour cela, à partir d'un autre échantillon de lymphocytes T effecteurs nous avons utilisé deux conditions : (i) 10ng d'ARN de départ et (ii) 100ng d'ARN de départ. Notre étude

a révélé d'importantes différences en termes de fiabilité et de reproductibilité intra et inter-méthodes pour les chaînes TCRα et TCRβ.

En résumé, (i) les résultats obtenus avec les méthodes basées sur la stratégie 5′RACE PCR sont homogènes entre eux ce qui n'est pas le cas entre les méthodes multiplexes PCR. (ii) Seules deux méthodes ont montré de bons résultats dans la capture de la diversité du TCRα. (iii) L'ensemble des méthodes ont détecté le clone Jurkat introduit dans les échantillons, néanmoins nous avons pu observer des variations importantes intra et inter-methodes concernant le rapport entre la quantité d'ARN et la fréquence observée du clone. (iv) La fiabilité fournie par l'utilisation d'UMI est au prix d'une réduction de capture de la diversité à quantité de séquence initiale équivalente. Finalement (v), une faible quantité d'ARN utilisée pour la préparation des librairies a généré des répertoires non représentatifs pour la quasi-totalité des méthodes prouvant une sensibilité variable.

En conclusion, ce projet m'a permis de mettre en évidence l'influence considérable des méthodologies sur la représentativité des données du répertoire et par conséquent, sur leur interprétation, mais également d'identifier une méthode qui répond aux contraintes imposées par la nature des échantillons du laboratoire. Il constitue un point essentiel pour assurer la fiabilité des observations sur mon projet de thèse, mais aussi pour l'ensemble des projets du laboratoire portant sur l'analyse du répertoire TCR par séquençage de nouvelle génération. Enfin, ce projet fournit pour la communauté scientifique une orientation exhaustive et concise pour le choix de la méthode la plus appropriée en fonction de la question scientifique. Il va servir de base à la communauté AIRR pour la mise en œuvre d'une étude comparable sur la résolution du répertoire BCR.

Par ailleurs, bien que large, des questions demeurent sur la calibration et le contrôle de ces approches d'analyse massive du répertoire. C'est dans cette perspective que j'ai contribuer avec le groupe de travail AIRR *biological ressources* de la communauté AIRR à une étude sur les besoins et défis en termes de contrôles, standards et calibreurs pour l'étude des répertoires TCR et BCR. Ce travail, publié sous la forme d'une revue dans eLife, est adossée à ce manuscrit en Annexe 3.

Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases

---

Pierre Barennes, Valentin Quiniou, Mikhail Shugay, Evgeniy S. Egorov, Alexey N. Davydov, Dmitriy M. Chudakov, Imran Uddin, Mazlina Ismail, Theres Oakes, Benny Chain, Anne Eugster, Karl Kashofer, Peter P. Rainer, Samuel Darko, Amy Ransier, Daniel C. Douek, David Klatzmann and Encarnita Mariotti-Ferrandiz

---

NATURE BIOTECHNOLOGY - 2021

nature biotechnology

ANALYSIS

https://doi.org/10.1038/s41587-020-0656-3

Check for updates

Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases

Pierre Barennes[1,2], Valentin Quiniou[1,2], Mikhail Shugay[3,4,5], Evgeniy S. Egorov[4], Alexey N. Davydov[6], Dmitriy M. Chudakov[3,4,5,6], Imran Uddin[7], Mazlina Ismail[7], Theres Oakes[7], Benny Chain[7], Anne Eugster[8], Karl Kashofer[9], Peter P. Rainer[10,11], Samuel Darko[12], Amy Ransier[12], Daniel C. Douek[12], David Klatzmann[1,2] and Encarnita Mariotti-Ferrandiz[1,2] ✉

Monitoring the T cell receptor (TCR) repertoire in health and disease can provide key insights into adaptive immune responses, but the accuracy of current TCR sequencing (TCRseq) methods is unclear. In this study, we systematically compared the results of nine commercial and academic TCRseq methods, including six rapid amplification of complementary DNA ends (RACE)-polymerase chain reaction (PCR) and three multiplex-PCR approaches, when applied to the same T cell sample. We found marked differences in accuracy and intra- and inter-method reproducibility for T cell receptor α (TRA) and T cell receptor β (TRB) TCR chains. Most methods showed a lower ability to capture TRA than TRB diversity. Low RNA input generated non-representative repertoires. Results from the 5′ RACE-PCR methods were consistent among themselves but differed from the RNA-based multiplex-PCR results. Using an in silico meta-repertoire generated from 108 replicates, we found that one genomic DNA-based method and two non-unique molecular identifier (UMI) RNA-based methods were more sensitive than UMI methods in detecting rare clonotypes, despite the better clonotype quantification accuracy of the latter.

Sequencing of the TCR repertoire is increasingly used to measure lymphocyte dynamics in health, in pathological contexts such as autoimmune disease, infections and cancer[1–6] and after interventions such as vaccination[4,7–9] and immunotherapy[10–12], with the goal of identifying TCR biomarkers of disease or of clinical response to treatment and to stratify patients for precision medicine[13]. These diverse applications have different requirements in terms of sensitivity, specificity and depth. Therefore, accurately capturing the TCR repertoire presents great challenges. Challenges include reproducibility, replicability and sensitivity because the diversity of the repertoire is, by essence, unknown when studying a new sample.

TCRs, which drive T cell activation by antigenic peptide recognition, are heterodimers formed by an α and a β chain[14] produced by somatic V(D)J rearrangements during thymopoiesis[15] of 47V and 61J functional TRA genes and 48V, 2D and 12J functional TRB genes[16]. The stochastic V(D)J recombination generates a combinatorial diversity that is further increased by random nucleotide excision and addition at the V(D)J junctions. The independent recombination and subsequent pairing of TRA and TRB chains add an additional level of combinatorial diversity. Recently, computational chain-pairing experiments suggested that the potential diversity of the paired repertoire is $\sim 2 \times 10^{19}$ TCRs[17], whereas the number of different TRB clonotypes in an individual has been estimated to range from 10^6 to 10^8(refs.[18–20]). The TCR repertoire is dynamic, as lymphocytes are continuously generated, die and expand in response to stimulation, and reflects both an individual's immune potential and history.

A large number of TCRseq methods have been developed. They are all complex, multi-step protocols, and each step might have a profound effect on the TCRseq data and, hence, on their interpretation[21]. Methods can be broadly classified as DNA or RNA based, and the latter can be categorized as using multiplex PCR (mPCR) with panels of V and J primers[18,22,23] or using RACE-PCR[6,24–26] optionally incorporating UMIs to limit PCR amplification bias and sequencing errors[6,26–28]. Each method has potential advantages and limitations[24,29–32]. Specifically, DNA-based methods are thought to be more quantitative and can be used on samples where RNA quality might not be guaranteed. In contrast, RNA-based methods are thought to be more sensitive because of the presence of multiple messenger RNA copies per cell and also are more amenable to UMI incorporation[33]. In addition, although TCR germline gene variability between individuals has not been extensively explored, the requirement of efficient primers for mPCR might also be an additional limitation. However, the relative robustness and accuracy of the different approaches have not been systematically compared. In this study, we compared nine different TCRseq library preparation protocols by analyzing the TCR repertoire of aliquots of the same T cell sample. Nine protocols for TCR library preparation were selected according to at least one of the following criteria: 1) published by groups other than the one that developed it[6,18,22,24–26], 2) associated with well-known analysis tools[23,27,34] and 3) commer-

[1]Sorbonne Université, INSERM, Immunology-Immunopathology-Immunotherapy (i3), Paris, France. [2]AP-HP, Hôpital Pitié-Salpêtrière, Biotherapy (CIC-BTi) and Inflammation-Immunopathology-Biotherapy Department (i2B), Paris, France. [3]Center of Life Sciences, Skoltech, Moscow, Russia. [4]Genomics of Adaptive Immunity Department, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia. [5]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Pirogov Russian National Research Medical University, Moscow, Russia. [6]Adaptive Immunity Group, Central European Institute of Technology, Brno, Czechia. [7]Division of Infection and Immunity, University College London, London, UK. [8]DFG-Centre for Regenerative Therapies Dresden, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. [9]Diagnostic and Research Institute of Pathology, Medical University of Graz, Graz, Austria. [10]Division of Cardiology, Medical University of Graz, Graz, Austria. [11]BioTechMed-Graz, Graz, Austria. [12]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. ✉e-mail: encarnita.mariotti@sorbonne-universite.fr

cially available[18,22,26,35]. Our results clearly show that 1) RACE-PCR and mPCR perform differently in TCR clonotype detection; 2) few of the evaluated methods accurately capture the TRA diversity; and 3) UMI-based methods are likely appropriate to study major clonotypes with a high accuracy, whereas robust non-UMI-based methods allow the capture of a larger repertoire.

Results

Experimental design to evaluate the robustness of human TCR repertoire analysis. We set out to compare nine different academic or commercial protocols for library preparation and sequencing (Supplementary Methods and Supplementary Table 1) based on either RACE-PCR (RACE-1 to RACE-6) or mPCR (mPCR-1 to mPCR-3). We sequenced nucleic acids from $CD4^+CD25^-CD127^+$ effector T cells (Supplementary Fig. 1a) sorted from two healthy donors (Experiments A and B). In Experiment A, we evaluated the accuracy and sensitivity of the different methods by spiking Donor A T cell RNA (RACE-1 to RACE-6 and mPCR-3) or DNA (mPCR-1 and mPCR-2) aliquots with different amounts of RNA or DNA from Jurkat cells (Supplementary Fig. 1b). In Experiment B, we analyzed the effect of decreasing amounts of the input material quantity by processing Donor B RNA aliquots of 100 ng and 10 ng (Supplementary Fig. 1c). In both experiments, the $CD4^+CD25^-CD127^+$ T cells were sorted, and the RNA and DNA were extracted and aliquoted in a single laboratory. Triplicates of aliquots were distributed to service providers and academic laboratories. Raw and/or pre-filtered sequences data were all processed using MiXCR[36].

We obtained 5.10^5–2.10^6 reads per aliquot depending on the method (Supplementary Fig. 2a,b). Numbers of unique V, J and VJ sequences as well as UMI distribution for RACE-1 and RACE-2 (Supplementary Fig. 2a–c) were similar between all the methods. Numbers of TCR sequences and clonotypes were correlated in a method-dependent manner but not globally, suggesting that the sequencing depth required for a given number of clonotypes is method dependent (Supplementary Fig. 2d).

Replicability and reproducibility differ among methods. For each method, we first analyzed the proportion of reads that were identified as TCRs (Fig. 1a and Supplementary Fig. 2). For seven of nine methods, we observed 20–60% of non-aligned reads, which were mainly explained by no V and/or J sequence identification. TCR sequences had a high quality score (phred > 30; Fig. 1b) and contained less than 1% PCR errors (Fig. 1c), except for RACE-2, RACE-6, mPCR-2 and mPCR-3. Note that these parameters could not be assessed for one of the commercialized mPCR-1 for which undisclosed proprietary pre-processing of the data is performed.

Using a VDJ rearrangement model (Methods), we computed 17 rearrangement parameters for TRA and TRB sequences from Experiments A and B (Supplementary Fig. 3) and calculated Jensen–Shannon divergence (JSD) distances between samples per parameter. Multi-dimensional scaling (MDS; Fig. 1d) showed that, within each experiment, samples obtained with the same method clustered together, suggesting that each method imposed its methodological imprint on the repertoire profile.

We further compared the different library methods' replicability (that is, the similarity among data obtained with the same method) and reproducibility (that is, the similarity among data obtained with different methods) using JSD as a measure of the distance between data sets[37]. Figure 1e shows that, for TRB, both the replicability and reproducibility of RACE-6 and mPCR-2 were lower than for all the other methods tested. However, when considering TRA, replicability was higher for RACE-3 and RACE-5, and reproducibility was higher for RACE-3, RACE-5 and RACE-2 (with and without UMI). Because RACE-6 showed extremely low replicability for TRB samples and was not reproduced by any other methods, we excluded it from further analysis. Altogether, our results showed that many fundamental parameters of the TCR repertoire, as well as inter-sample replicability and reproducibility, vary between the different methods tested.

The observed T cell receptor β variable gene usage varies between RACE- and multiplex-PCR RNA-based methods. We compared the T cell receptor β variable (TRBV) gene usage obtained from the sequencing data with the percentage of TRBV protein expression quantified by flow cytometry (Fig. 2a and Supplementary Fig. 4a,b). mPCR-1 data were highly correlated with flow cytometry data (Fig. 2b; $R^2 > 0.9$, $2.14 \times 10^{-12} < P < 5.10–5.38 \times 10^{-11}$), which likely reflects the undisclosed proprietary filtering by the provider. All other methods also showed a significant R^2 Pearson correlation score, ranging from 0.4 to 0.8 ($P < 0.05$), with TRBV protein expression (Fig. 2a,b), except for mPCR-3 ($R^2 < 0.2$ and $P > 0.05$). The Pearson correlation of TRBV gene usage within replicates prepared with the same method (Fig. 2c) was high ($R^2 > 0.9$). However, clustering showed that mPCR-3 formed a distinct cluster with a low correlation score ($R^2 < 0.5$) with other methods. The RACE methods data were highly correlated between each other ($R^2 > 0.8$), except RACE-1 and RACE-1_U, which had a lower correlation ($0.6 < R^2 < 0.7$). mPCR-1 and mPCR-2 formed an independent 'DNA cluster' with an $R^2 > 0.6$ when compared to RACE replicates and a low correlation with mPCR-3 ($R^2 < 0.4$). This low correlation with mPCR-3 could, in part, be explained by a skewed TRBV9, TRBV29-1 and TRBV20-1 usage (Supplementary Fig. 4c). Spearman correlation scores were higher between flow cytometry data and mPCR-3 as well as RACE-1 and globally between the methods (Supplementary Fig. 4d,e). In summary, RACE-PCR methods and genomic DNA (gDNA)-based mPCR methods showed similar TRBV usage results, in contrast with the mPCR-3 RNA-based method.

Robustness of TRA and TRB detection is method dependent. We compared the similarity and composition of the 1% most predominant clonotypes (1%_MPC) detected by each method. The Morisita–Horn (MH) similarity index was calculated for each replicate across all the methods for both TRA (Fig. 3a, left) and TRB (Fig. 3a, right) sequences. TRA repertoires from RACE-3 and RACE-5 clustered together, inter- and intra-replicates having a high degree of similarity ($MH \approx 0.8$). RACE-1, RACE-2 and RACE-4 had a lower inter- and intra-method similarity ($0.2 < MH < 0.5$) but a higher similarity with RACE-3 and RACE-5. Similar clustering was obtained with the Jaccard similarity index (JSI), a measure independent of clonotype frequency (Supplementary Fig. 5a). For the TRB repertoires, MH scores were low when comparing RACE and mPCR protocols ($MH \approx 0.36$) but high within the RACE cluster ($0.6 > MH > 0.9$). There was poor similarity between the results of the three mPCR methods, regardless of the template. Differences between RACE and mPCR methods disappeared when calculating the JSI, suggesting a bias in clonotype frequency, as expected when comparing RNA- with DNA-based methods, but less when comparing RNA-based methods. Similar results were obtained by iteratively increasing the percentage of clonotypes (Supplementary Fig. 5b). Rényi diversity profiles (Supplementary Fig. 5c) showed similar results for TRB with all the methods, but the diversity of TRA varied depending on the method. However, the potential diversity estimated using Chao extrapolation (ChaoE) was variable between methods (Supplementary Fig. 5d).

To test a possible bias in capturing the TRA diversity for some methods, we pooled and compared the three spiking replicates per method from Experiment A, as suggested by Greiff et al.[21]. The MH similarity significantly increased for all the RACE-based methods for TRA (Fig. 3b, top) (except RACE-3) and for TRB (Fig. 3b, bottom), with the TRA MH similarity remaining lower than that of TRB. Similar observations were made for mPCR replicates. This suggests that, for a given depth of sequencing, the TRB diversity is better captured than that of TRA.

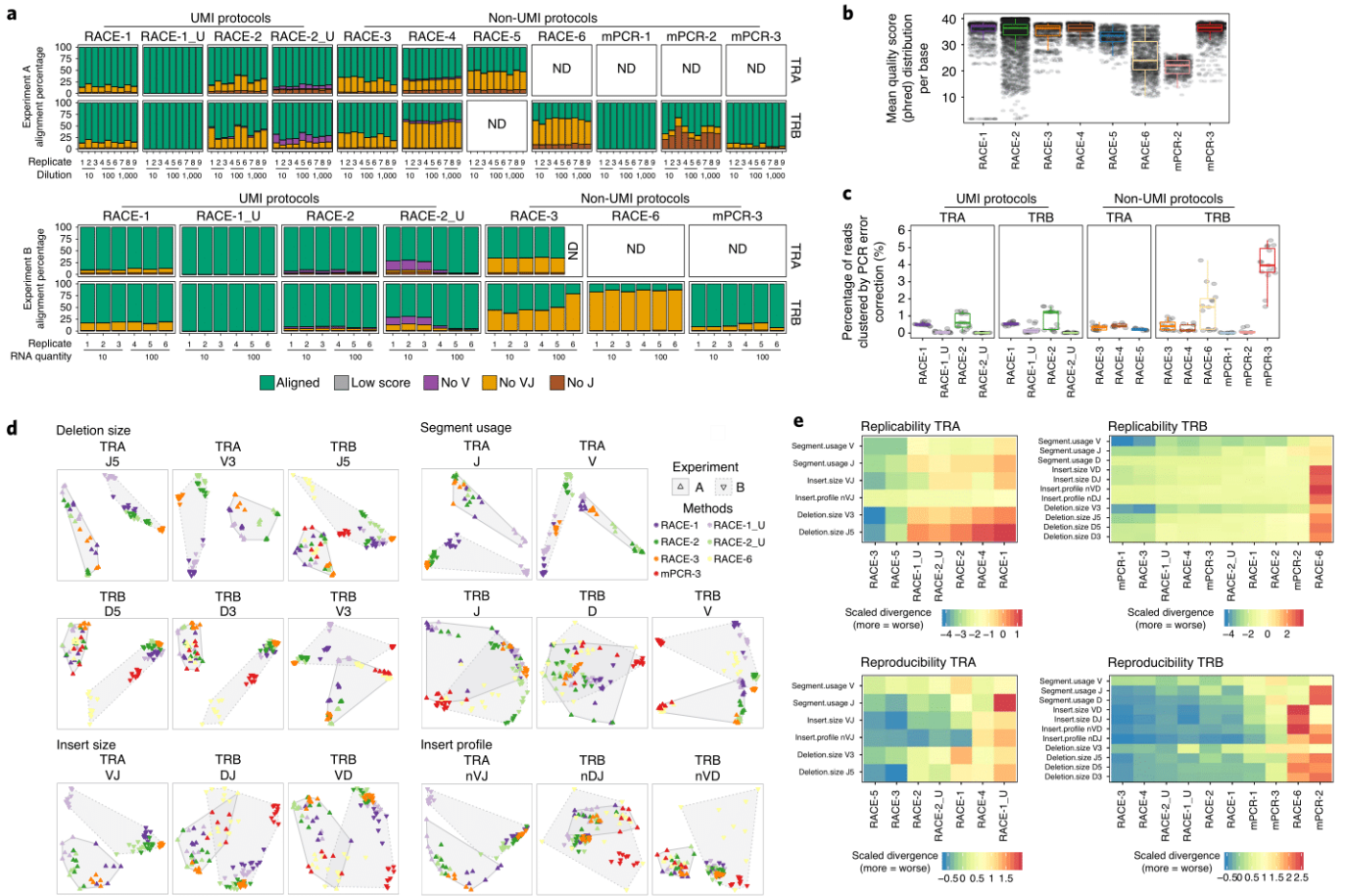


Fig. 1 | Performance statistics and VDJ rearrangement model of each method for Experiments A and B. a, The proportion of sequence reads aligned for TRA or TRB genes for each TCRseq replicate per experiment (Experiment A, top; Experiment B, bottom). The bars represent the percentage of TRA and TRB alignment, and the reason for alignment failure is color coded. **b**, Distribution of the reads quality control per base for each method over all data sets ($n = 15$ (nine technical replicates from Experiment A + six technical replicates from Experiment B)), computed with fastQC software (www.bioinformatics.babraham.ac.uk/projects/fastqc). **c**, Percentage of reads collapsed after PCR error correction for all samples in the study ($n = 15$ as in **b**). For each method, the MiXCR clustering strategy was applied to correct for PCR errors and collapse reads. Each box plot represents the percentage of clustered reads. **d**, MDS of V(D)J recombination parameters. MDS was performed based on the JSD calculated between replicates on weighted VDJ segment usage (Segment usage), non-template nucleotide insert size distributions (Insert size), V/D/J segment trimming distributions (Deletion size) and nucleotide frequencies in N-inserts (Insert profile). JSD values were transformed to rank for better visualization. Solid and dotted polygons outline samples from Experiments A and B, respectively. Colors represents the different methods as in **b** (only methods used in both experiments are presented). **e**, Replicability and reproducibility of the TRA and TRB repertoires for each method. The average JSD calculated in **d** (rows) for TRA (left) and TRB (right) measured between replicates produced by the same method (replicability, top) or replicates of a given method and all other protocols (reproducibility, bottom) was used as distance metric to compare different protocols (columns). Columns are sorted according to the mean scaled distance (averaged over all rows) from the lowest (best replicability/reproducibility) to the highest (worst replicability/reproducibility). Distance values are shown using a color scale. Jurkat TCR sequences were removed from data sets for this analysis. Box plots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5× IQR from the hinge; the lower whisker extends from the hinge to the smallest value at most 1.5× IQR of the hinge). IQR, interquartile range.

Detection sensitivity of rare TCRs depends on the method. To determine the accuracy of the different library amplifications for different clonotype frequencies, we compared the observed frequencies of the TCR from the Jurkat spike-in to their theoretical frequencies of 1/10, 1/100 and 1/1,000 (Supplementary Fig. 1b). TRA-observed frequencies were, on average, three times lower than expected (Fig. 4a, top, Supplementary Table 2 and Supplementary Fig. 6a). In contrast, TRB frequencies were, on average, three times higher than the theoretical percentage, except for mPCR-1 (Fig. 4a, bottom, Supplementary Table 2 and Supplementary Fig. 6a). For most of the methods, except RACE-1_U, RACE-4 and mPCR-3, the ratio between the different dilutions was maintained, as shown by the mean slope values close to 1 (Fig. 4b).

We then compared the inter-sample variation in clonal frequency for those TCR sequences shared between all replicates of an individual method (excluding the Jurkat clone). Figure 4c represents the s.d. of the frequency of each shared clonotype (dots) per method (see details in Supplementary Fig. 6b–d). For TRA, RACE-3 and RACE-5 had the highest number of clonotypes shared between the nine replicates and the lowest s.d. For TRB, all the methods captured a high number of shared clonotypes, and mPCR-1 and RACE-3 had the lowest s.d. Finally, pooling all the clonotypes from all the replicates, we identified nine TRA and 31 TRB clonotypes shared by all the replicates of all methods, corresponding to the most predominant clonotypes (Supplementary Fig. 7). RACE-3, RACE-5 (both RNA based) and mPCR-1

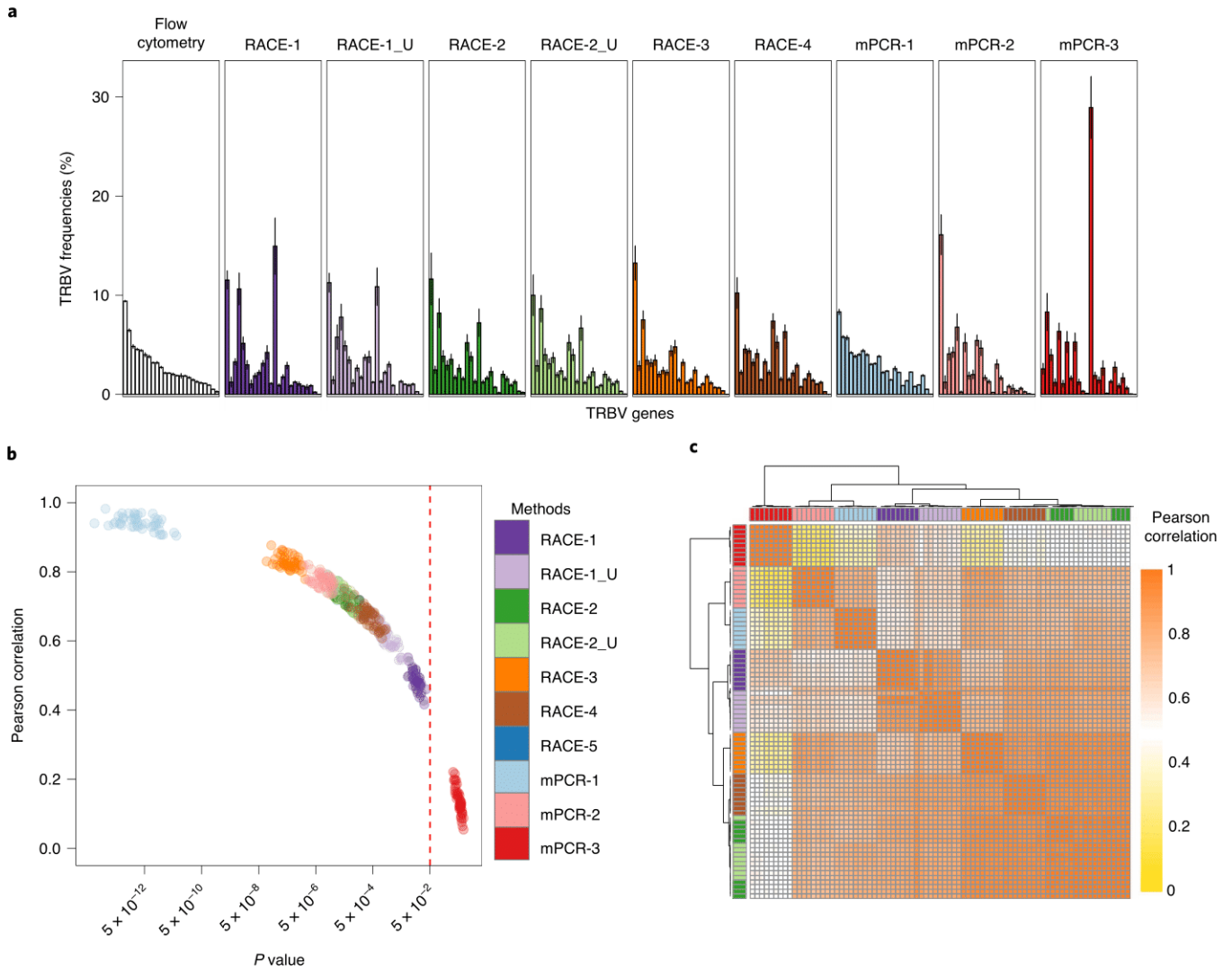


Fig. 2 | TRBV usage comparison between flow cytometry and TCRseq. a, Flow cytometry and TCRseq TRBV frequencies. Bar plots represent the TRBV frequencies calculated from flow cytometry-stained CD4+ T effector cells for the 24 TRBV for which antibodies are available and from the TCRseq data ($n=9$ technical replicates from Experiment A), considering only clonotypes annotated for the same 24 TRBV (original TRBV frequencies were used accordingly). Histograms of the 24 TRBV frequencies are presented as mean values ± s.d. and organized by decreasing order using frequencies obtained by flow cytometry as a reference (TRBV20-1, TRBV19, TRBV12-3/4, TRBV28, TRBV2, TRBV3-1, TRBV30, TRBV6-5/9, TRBV9, TRBV5-1, TRBV4-1/2, TRBV27, TRBV29-1, TRBV6-6, TRBV11-2, TRBV10-3, TRBV25-1, TRBV6-2, TRBV18, TRBV5-5, TRBV14, TRBV5-6, TRBV13 and TRBV4-3). **b**, TRBV usage correlation between flow cytometry and TCRseq. Pearson correlation of the TRBV frequencies between the flow cytometry data sets ($n=5$) and the TCRseq replicates ($n=9$ from Experiment A) was calculated for each method. The plot displays the correlation score (y axis) and the P value (x axis) of the correlation, allowing the classification of the methods. **c**, Heat map of the Pearson correlations between each replicate for the distribution of TRBV gene usage ($n=62$ unique TRBV genes). The Euclidean distance was used for hierarchical clustering as a color-coded matrix ranging from 0 (yellow, maximum dissimilarity) to 1 (orange, maximum similarity). Jurkat TCR sequences were removed from data sets for this analysis.

(DNA based) showed the lowest inter-sample variability in TCR frequency.

The quantity of starting material affects TCR diversity capture. One major limitation when analyzing TCR repertoire is the number of T cells that can be analyzed. Focusing on four RNA-based methods, we analyzed the influence of input RNA quantity on TRA and TRB repertoires (Supplementary Fig. 1c). We compared two sets of samples, one containing 10 ng or 100 ng (corresponding to 10^4 and 10^5 cells, respectively). For all the methods, the richness was higher with large (100-ng) than with small (10-ng) samples (Supplementary Fig. 8a). Rényi diversity profiles (Supplementary Fig. 8b) showed that, when $\alpha < 2$ (that is, when the diversity metric is influenced by rare clones), the diversity of small samples is less than that of larger ones. In contrast, at $\alpha = 2$ (Simpson index) or above, diversity profiles of both samples overlap. Thus, a low RNA input influences the number of rare TCR sequences detected but not the distribution of the more abundant TCRs.

Finally, we evaluated the inter-sample similarity as a function of RNA input quantity by calculating the MH index with the TRVJ combination usage (VJ_usage), all clonotype frequencies (Overall) or with the frequencies of the 1% MPC (Supplementary Fig. 8c, middle). For TRA, the similarity between 10-ng replicates was lower at the level of VJ usage and of all clonotypes compared with that between 100-ng replicates (Supplementary Fig. 8c, top and bottom). For TRB, the results were similar regardless of the quantity (MH > 0.5). When

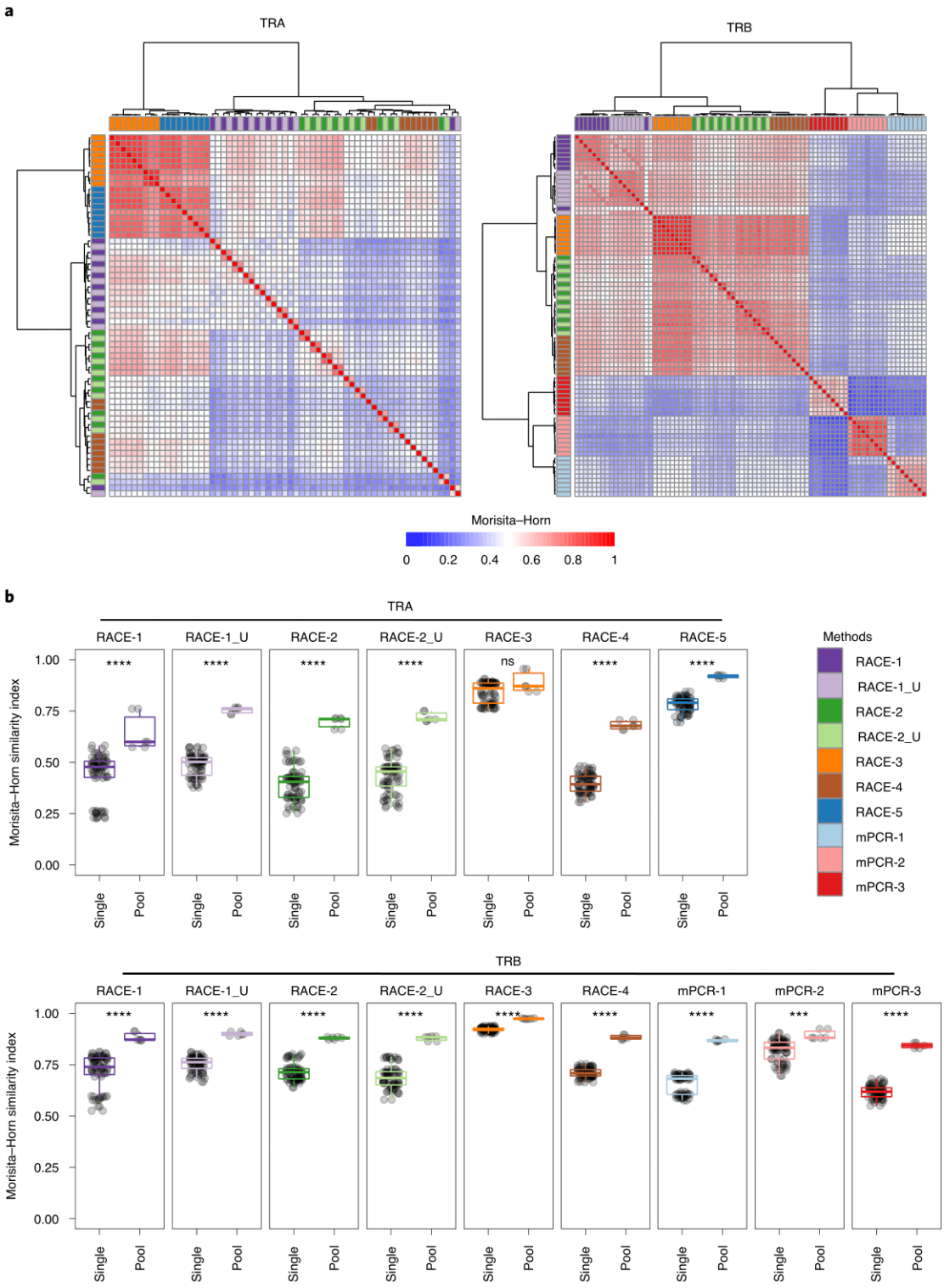


Fig. 3 | The reproducibility of detection of major TCR clonotypes by different methods. a, Heat maps of the MH similarity index. MH scores were calculated between each replicate across all methods for the top 1% of MPCs for TRA (left) and TRB (right). The Euclidean distance was used for hierarchical clustering as a color-coded matrix ranging from 0 (blue, maximum dissimilarity) to 1 (red, maximum similarity). **b**, Comparison between individual replicates (Single, $n=9$ from Experiment A) and pooled replicates (Pool, $n=3$) by the MH similarity index. Data sets from replicates of the same dilution were pooled for each method to get one pooled sample per dilution. Singletons (count = 1) were removed, and MH similarity scores were calculated for the top 1% of MPCs for TRA (left) and TRB (right). Box plots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5× IQR from the hinge; the lower whisker extends from the hinge to the smallest value at most 1.5× IQR of the hinge). The Mann–Whitney statistical test was performed to compare Single versus Pool results by method and by chain. *$P<0.05$; **$P<0.01$; ***$P<0.001$; ****$P<0.0001$. Jurkat TCR sequences were removed from data sets for this analysis. IQR, interquartile range; NS, not significant.

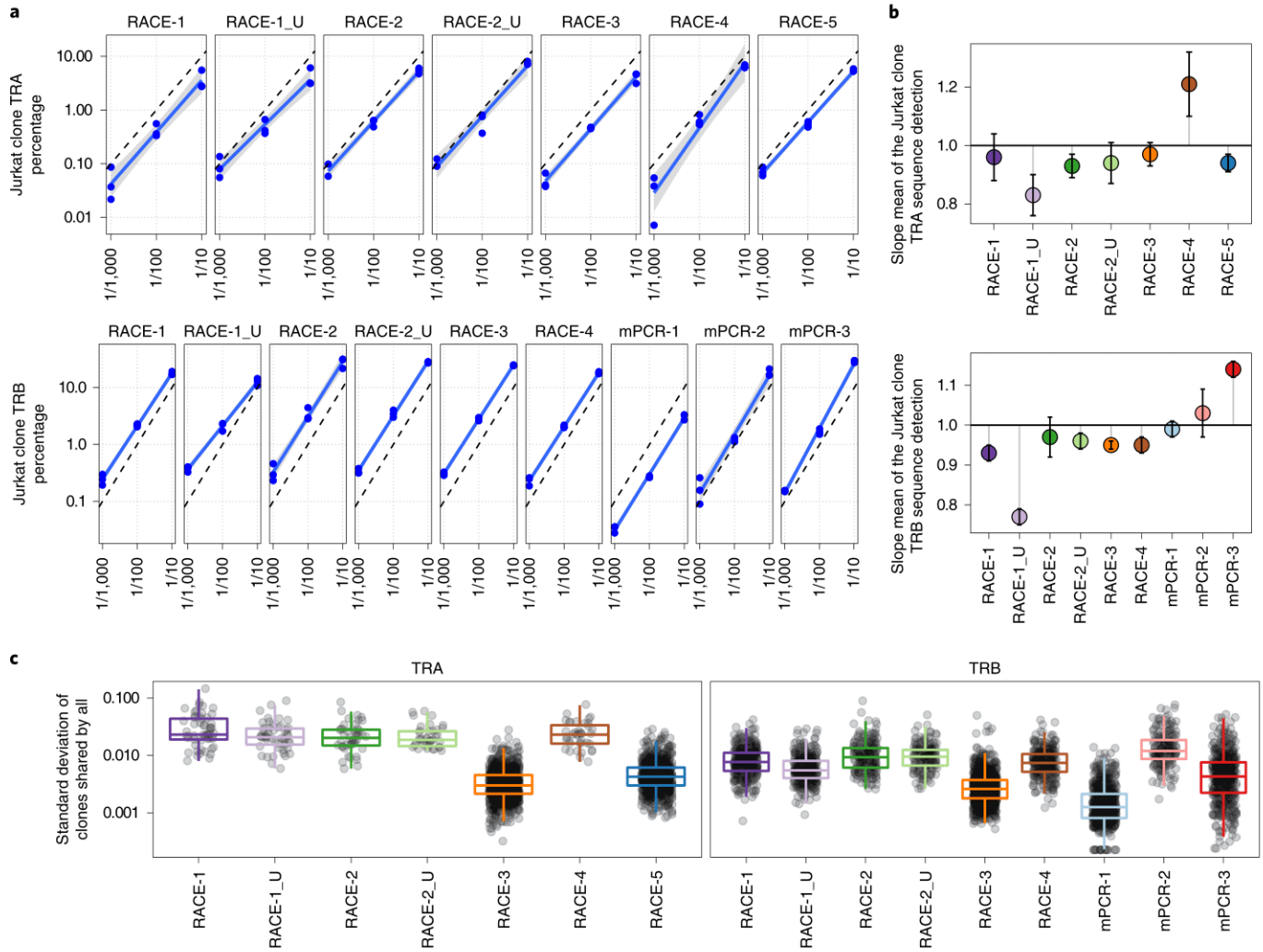


Fig. 4 | Sensitivity of TCR sequence detection by different methods. a, Jurkat clone percentage. Jurkat TRA (top) and TRB (bottom) clonotype percentages were calculated for each experiment per dilution (1/10, 1/100 and 1/1,000 spike-in) and are represented by the blue dots. The blue line represents linear regression; the black dashed line represents the theoretically expected percentage. Errors bands in light gray represent the 95% confidence interval. **b**, Slope of the Jurkat tracking linear regression. Slope was computed between dilutions. Data are represented as mean values of the slope ± s.d. by method ($n = 9$ technical replicates from Experiment A) for TRA (top) and TRB (bottom). **c**, s.d. of the clonotypes shared among all the replicates ($n = 9$ from Experiment A), after excluding the Jurkat clone, calculated per method, for TRA (left) and TRB (right). Box plots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5× IQR from the hinge; the lower whisker extends from the hinge to the smallest value at most 1.5× IQR of the hinge). IQR, interquartile range.

focusing on the 1% MPC, the similarity was comparable irrespective of the quantity for both TRA and TRB. These results indicated that RNA quantity affects rare clonotype detection.

Reliability and sensitivity of each method highlighted using an in silico meta-repertoire. One unavoidable issue when aiming at capturing the diversity of a repertoire is sampling—that is, only a fraction of the cells are analyzed and then a fraction of their nucleic acids[21]. To better assess the ability of each method to robustly capture rare and frequent clonotypes, we took advantage of the fact that, altogether, we generated 45 TRA and 63 TRB replicates of the same cell sample. We aggregated these results to generate an in silico meta-repertoire. To ensure the accuracy of the TCR sequences composing this meta-repertoire, we removed singletons and kept clonotypes found by at least three methods.

We first analyzed how many of the clonotypes present in this meta-repertoire were detected by each method. For TRA (Fig. 5a, left), RACE-3 and RACE-5 data sets included up to 50% of the meta-repertoire clonotypes (MRCs) compared to 10–20% for the other RACE method data sets. Similar results were found for TRB (Fig. 5a, right). We then computed, for each method, the fraction of MRCs found in 0, 1, 2, 3, etc, up to 9 replicates. The dot heat maps (Fig. 5b) showed that, for TRA, RACE-3 and RACE-5 clearly outperformed the other methods, capturing up to 40% of the MRCs in all nine replicates (Fig. 5b, left; replicate number = 9) and missing (that is, never captured in any of the nine replicates) less than 1% of the MRCs (Fig. 5b, left; replicate number = 0). The other RACE protocols detected only 1% of MRCs in all nine replicates and missed 15–20% of the MRCs (Fig. 5b, left). In contrast, there was much less difference between the methods for TRB (Fig. 5b, right).

Finally, we analyzed the frequency of MRC TCRs that were detected or not by each method (Fig. 5c and Supplementary Fig. 9). For TRA (Fig. 5c, left), the frequency of MRCs found in nine replicates (red box plots) ranged from 1% to 0.001% for

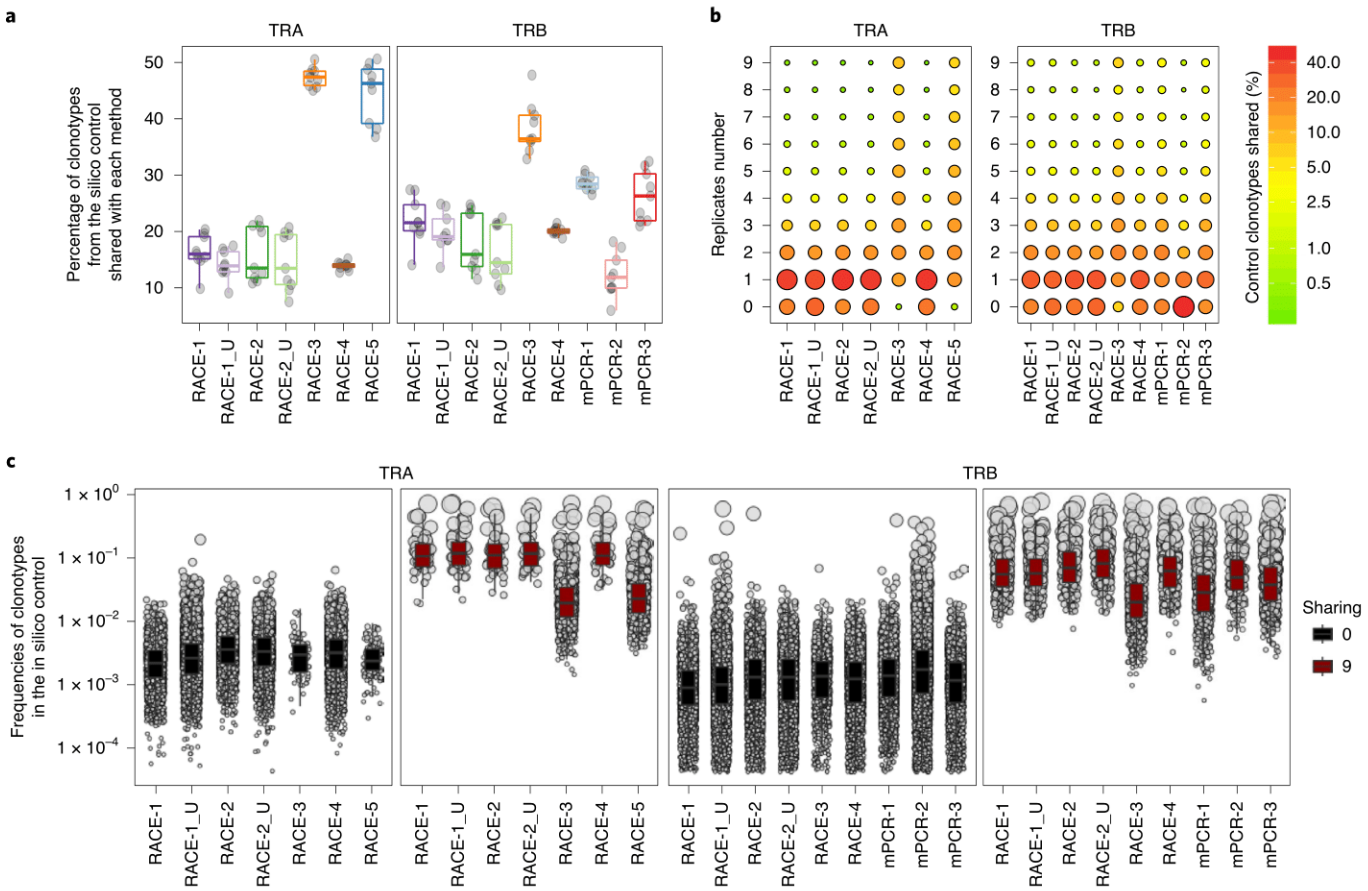


Fig. 5 | Sharing with robust and representative meta-repertoire. a, Replicate sharing fraction in meta-repertoire for TRA (left) and TRB (right). The values represented correspond to the percentage of clonotypes from each replicate ($n=9$ technical replicates from Experiment A) per method found in the meta-repertoire; the median and the 1st and 3rd quartiles are shown. **b**, Replicability of methods with meta-repertoire for TRA (left) and TRB (right). Heatmaps represent the percentage of meta-repertoire (control) clonotypes found in 1 to 9 replicates per method (0: unseen in any of the replicates). **c**, Distribution of meta-repertoire clonotypes in the replicates ($n=9$ from Experiment A) by methods for TRA (left) and TRB (right). Each dot represents a meta-repertoire clonotype, and the box plot represents the average frequencies. Black box plots with corresponding gray dots represent the unseen clonotypes (seen in $n=0$ replicates), and red box plots with corresponding gray dots represent the clonotypes found by all the replicates (seen in $n=9$ replicates). Each method is represented independently. Jurkat TCR sequences were removed from data sets for this analysis. Box plots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5× IQR from the hinge; the lower whisker extends from the hinge to the smallest value at most 1.5× IQR of the hinge). IQR, interquartile range.

RACE-3 and RACE-5 and from 1% to 0.05% for the other methods. In contrast, clonotypes not detected in any replicates (black box plots) were present at 10- to 100-fold lower abundance. A similar overall pattern was seen for TRB, although the frequencies were shifted to a lower range. This analysis suggested that RACE-3 and RACE-5 had increased sensitivity and, hence, were able to detect a larger proportion of clonotypes at lower abundances. These differences were more evident for TRA than for TRB (Fig. 5c, right). The other methods compared behaved very similarly to each other. Notably, those results were independent of sample size (Supplementary Fig. 10).

Discussion

Interpreting the TCR repertoire is an increasingly important tool in understanding the underlying causes of immune-mediated diseases and in assisting the development of new immunotherapeutic strategies. However, despite hundreds of TCRseq studies in the last decade using a variety of different methodologies, there has been no systematic study comparing them.

In this work, we compared methods developed by academics, at a time when there was little or no reliable commercial service provision, with some currently available commercial methods. Both RNA- and gDNA-based methods were included. To avoid mis-implementation of protocols, each method (including appropriate pre-processing of sequence data) was performed by the laboratory or commercial provider (except for kit providers) that developed them.

Unexpectedly, some consistent differences were observed in TRBV usage when compared to flow cytometry measurement of TRBV-encoded proteins, especially for RNA-based profiling. This might reflect bias in amplification of RNA transcripts according to their expression levels, more efficient transcription of some V genes or differences in nonsense-mediated decay[38]. Further studies using single-cell RNA sequencing might shed light on this phenomenon.

Working with human samples often imposes limits on the number of available T cells. Notably, lymphopenia is a common feature in patients undergoing treatment (transplantation and immunosuppressive therapy) or with autoimmune disease[39] and infections.

Table 1 | Comparative performance of the nine TCRseq molecular methods

TR chain	Method	Replicability	Reliability	Sensitivity	Cost per sample ($)	Controls and standards	Format type	fastq data availability
TRA	RACE-1	7	4	4	~230	-	Lab protocol	Yes
	RACE-1_U	4	5	4	~230	UMI	Lab protocol	Yes
	RACE-2	5	4	5	230–280	-	Service or kit	Yes
	RACE-2_U	4	5	5	230–280	UMI	Service or kit	Yes
	RACE-3	3	2	3	~150	-	Kit	Yes
	RACE-4	5	6	4	~150	-	Lab protocol	Yes
	RACE-5	2	3	3	~300	-	Lab protocol	Yes
TRB	mPCR-1	3	3	3	~350–550[a]	Synthetic TCRs	Service or kit	No
	mPCR-2	6	7	7	~25	-	Lab protocol	Yes
	mPCR-3	5	5	3	~350–550[a]	-	Service or kit	Yes
	RACE-1	6	5	4	~230	-	Lab protocol	Yes
	RACE-1_U	4	6	5	~230	UMI	Lab protocol	Yes
	RACE-2	6	6	6	230–280	-	Service or kit	Yes
	RACE-2_U	6	6	7	230–280	UMI	Service or kit	Yes
	RACE-3	2	2	3	~150	-	Kit	Yes
	RACE-4	3	5	4	~150	-	Lab protocol	Yes

For each method, an average rank score for TRA (top) and TRB (bottom) sequencing was calculated for replicability, reliability and sensitivity (first three columns), and practical information was summarized (last four columns). Ranks were calculated as the average of the ranks for results from Figs. 1e, 2c, 3b and 4c for replicability; Figs. 1e, 2b, 4b and 5a,b for reliability; and Figs. 4c and 5b and Supplementary Figs. 2a and 5c for sensitivity. Rank values range from 2 (best) to 7 (worst). Details are provided as Supplementary Data 1. Cost per sample is expressed in US dollars as per current prices for a depth of 1 million TCR sequences per sample on a 25-million-reads sequencing format. The costs cover reagents for library preparation to sequencing. [a]mPCR1 and mPCR3 price ranges correspond to the cost for purchasing either kits (lowest price) or service up to sequencing and basic data analyses from the provider.

Additionally, T cell subsets of interest, as well as available counts of tumor-infiltrating T cells, might be limited. Therefore, it is important to identify which methods provide reliable TCRseq profiles for small numbers of T cells. In this context, we observed that, regardless of the method, starting from a highly polyclonal population, the initial amount of material is critical to obtaining representative results, notably in terms of diversity and rare clone detection.

Although our study focused on polyclonal CD4 T cells from healthy repertoires, we analyzed a wide range of global and sequence-specific repertoire parameters, including V(D)J gene usage, junctional diversity, repertoire diversity and sequence sharing. These parameters are all relevant to any other α/β T cell populations, as indeed are all parameters routinely used to analyze repertoires of samples from pathological and clinical human samples[40].

Because our study incorporated multiple replicates tested with each method, we were able to explore method replicability—that is, the ability of each method to reproduce the same repertoire from different sub-samples from the same individual. Our results showed that, except mPCR-3, all the methods provided consistent results among replicates. We also evaluated the reproducibility—that is, the extent to which different methods record the same results when applied to the same sample. We observed a low degree of TRB clonotype overlap between repertoires amplified from gDNA and RNA (complementary DNA), perhaps reflecting differences in gDNA and RNA copy numbers. The four RACE methods produced relatively similar repertoires as revealed by the MH index. The mPCR on gDNA showed low reproducibility between methods, suggesting that the choice of multiplexing primers might bias the amplification of some clonotypes, as suggested previously[31]. However, most RACE methods (not tested for mPCR) had a lower efficiency in capturing TRA rather than TRB diversity, which might reflect the two- to three-fold lower number of TRA transcripts than TRB transcripts[28].

Finally, sensitivity is important for the study of circulating blood T cells, especially when the goal is to track a few expanded clones associated with infection or autoimmunity or in response to treatment. However, assessing sensitivity based on sample overlap is a complex performance metric, because it is affected by experimental variability but also by sampling. To tackle this problem directly, we generated an in silico meta-repertoire that provided a more robust platform with which to directly compare the sensitivity performance of the different methods. Interestingly, using this standard, we found that two non-UMI methods (RACE-3 and RACE-5) had greater sensitivity than UMI-based methods (RACE-1 and RACE-2) and were able to detect clonotypes at a ten-fold lower frequency. In part, this results from the reads-per-UMI cutoff, which might lead to a decrease in observed TCR diversity if sequencing coverage is not sufficient. For example, introducing a hard cutoff that discards all UMIs with fewer than five reads leads to a decrease in observed TCR diversity. UMI-based methods might be more accurate for assessing clonotype frequency, in line with their use to quantify and correct for PCR errors and bias[41]. Furthermore, a threshold of 2–4 reads per UMI efficiently protects against artifacts and cross-sample contamination[42], which become critical with tighter cluster density on modern Illumina machines. UMI-based methods might require several replicates or higher sequencing coverage to consistently and unambiguously identify rare TCR sequence clonotypes. Notably, both RACE-1 and RACE-2 methods performed better after UMI correction (see Table 1).

Such in silico standards might be of value in further comparative TCRseq method evaluation, although, ideally, synthetic repertoires recapitulating at least the extent of the TRAVJ and TRBVJ combinations and distributions might provide an even more robust alternative. Two such approaches have been proposed for specific clone detection in minimal residual diseases[43,44] as well as for the B cell receptor, but not the T cell receptor, repertoire[45], still at a very low diversity level. The construction of such gold standard repertoires is currently very costly and remains a major challenge that the Adaptive Immune Receptor Repertoire (AIRR) Community[46], engaged in AIRR sequencing standardization[47,48], might tackle in the future. Finally, in this study, some data were pre-processed using

proprietary (mPCR-1 and mPCR-3) or published[27,34] (RACE-1_U and RACE-2_U) tools and then aligned and error corrected using MiXCR (v2.1.10)[36]. To further optimize TCR data accuracy, it would also be interesting to benchmark available software analysis tools, especially regarding UMI analysis and sequence alignment. Our data sets generated using different methods should be a valuable complement to using data sets generated purely in vitro[49,50].

In conclusion, the take-home messages from this work are the following. First, there are satisfactory TCRseq methods based on either DNA or RNA input, and the amount of material affects both diversity and the detection of rare clones. Second, various methods are optimal for detecting maximal diversity, whereas others most accurately quantify the abundance of specific clonotypes. For the latter, UMI-based methods are potentially more accurate, although they could miss relevant, but rare, clones. In contrast, non-UMI RACE methods are more sensitive in capturing rare clones, especially for TRA. Third, the availability of raw data is crucial in allowing reliable and reproducible in-depth analyses of TCR repertoires. The mPCR-1 service provider does not provide access to raw sequence data, whereas mPCR-1 and mPCR-3 service providers do not disclose the proprietary pre-processing filters. In contrast, the RACE-2 provider provides raw data and all pre-processing algorithms. We summarized our results, as well as practical aspects, in Table 1. To this end, we calculated, for each method, a rank value for replicability, reliability and sensitivity based on various measures (Table 1 and Supplementary Data 1). We also summarized cost per sample, presence of controls or standards, format of the method and raw data availability. Table 1 highlights the advantages and disadvantages of the different methods that could serve as guidance for end users. Improved and more sophisticated data analyses are essential to extract the full power of TCR repertoire data. We anticipate that, now that TCR sequencing has come of age, the next key developments in the field will come from new methods of data analysis, as has been the case in the related field of global transcriptomics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0656-3.

Received: 29 May 2019; Accepted: 28 July 2020;
Published online: 07 September 2020

References

1. Cui, J.-H. et al. TCR repertoire as a novel indicator for immune monitoring and prognosis assessment of patients with cervical cancer. *Front. Immunol.* **9**, 2729 (2018).
2. Davis, M. M. The αβ T cell repertoire comes into focus. *Immunity* **27**, 179–180 (2007).
3. Lindau, P. & Robins, H. S. Advances and applications of immune receptor sequencing in systems immunology. *Curr. Opin. Syst. Biol.* **1**, 62–68 (2017).
4. Miles, J. J., Douek, D. C. & Price, D. A. Bias in the αβ T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol. Cell Biol.* **89**, 375 (2011).
5. Schrama, D., Ritter, C. & Becker, J. C. T cell receptor repertoire usage in cancer as a surrogate marker for immune responses. *Semin. Immunopathol.* **39**, 255–268 (2017).
6. Heather, J. M. et al. Dynamic perturbations of the T-cell receptor repertoire in chronic HIV infection and following antiretroviral therapy. *Front. Immunol.* **6**, 644 (2016).
7. Howson, L. J. et al. MAIT cell clonal expansion and TCR repertoire shaping in human volunteers challenged with *Salmonella* Paratyphi A. *Nat. Commun.* **9**, 253 (2018).
8. Pogorelyy, M. V. et al. Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. *Proc. Natl Acad. Sci. USA* **115**, 12704–12709 (2018).
9. Sycheva, A. L. et al. Quantitative profiling reveals minor changes of T cell receptor repertoire in response to subunit inactivated influenza vaccine. *Vaccine* **36**, 1599–1605 (2018).
10. Hogan, S. A. et al. Peripheral blood TCR repertoire profiling may facilitate patient stratification for immunotherapy against melanoma. *Cancer Immunol. Res.* **7**, 77–85 (2019).
11. Jin, Y. et al. TCR repertoire profiling of tumors, adjacent normal tissues, and peripheral blood predicts survival in nasopharyngeal carcinoma. *Cancer Immunol. Immunother.* **67**, 1719–1730 (2018).
12. Wieland, A. et al. T cell receptor sequencing of activated CD8 T cells in the blood identifies tumor-infiltrating clones that expand after PD-1 therapy and radiation in a melanoma patient. *Cancer Immunol. Immunother.* **67**, 1767–1776 (2018).
13. Six, A. et al. The past, present, and future of immune repertoire biology – the rise of next-generation repertoire analysis. *Front. Immunol.* **4**, 413 (2013).
14. Chien, Y. H., Gascoigne, N. R., Kavaler, J., Lee, N. E. & Davis, M. M. Somatic recombination in a murine T-cell receptor gene. *Nature* **309**, 322–326 (1984).
15. Davis, M. M. & Bjorkman, P. J. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**, 395–402 (1988).
16. Lefranc, M.-P. Nomenclature of the human T cell receptor genes. *Curr. Protoc. Immunol.* **40**, A.1O.1–A.1O.23 (2000).
17. Dupic, T., Marcou, Q., Walczak, A. M. & Mora, T. Genesis of the αβ T-cell receptor. *PLoS Comput. Biol.* **15**, e1006874 (2019).
18. Robins, H. S. et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood* **114**, 4099–4107 (2009).
19. Warren, R. L. et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res.* **21**, 790–797 (2011).
20. Qi, Q. et al. Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl Acad. Sci. USA* **111**, 13139–13144 (2014).
21. Greiff, V., Miho, E., Menzel, U. & Reddy, S. T. Bioinformatic and statistical analysis of adaptive immune repertoires. *Trends Immunol.* **36**, 738–749 (2015).
22. Wang, C. et al. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Natl Acad. Sci. USA* **107**, 1518–1523 (2010).
23. Zhang, W. et al. IMonitor: a robust pipeline for TCR and BCR repertoire analysis. *Genetics* **201**, 459–472 (2015).
24. Douek, D. C. et al. A novel approach to the analysis of specificity, clonality, and frequency of HIV-specific T cell responses reveals a potential mechanism for control of viral escape. *J. Immunol.* **168**, 3099–3104 (2002).
25. Eugster, A. et al. Measuring T cell receptor and T cell gene expression diversity in antigen-responsive human CD4$^+$ T cells. *J. Immunol. Methods* **400–401**, 13–22 (2013).
26. Mamedov, I. Z. *et al.* Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Front. Immunol.* **4**, 456 (2013).
27. Shugay, M. et al. Towards error-free profiling of immune repertoires. *Nat. Methods* **11**, 653–655 (2014).
28. Oakes, T. et al. Quantitative characterization of the T cell receptor repertoire of naïve and memory subsets using an integrated experimental and computational pipeline which is robust, economical, and versatile. *Front. Immunol.* **8**, 1267 (2017).
29. Liu, X. et al. Systematic comparative evaluation of methods for investigating the TCRβ repertoire. *PLoS ONE* **11**, e0152464 (2016).
30. Rosati, E. et al. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* **17**, 61 (2017).
31. Dunn-Walters, D., Townsend, C., Sinclair, E. & Stewart, A. Immunoglobulin gene analysis as a tool for investigating human immune responses. *Immunol. Rev.* **284**, 132–147 (2018).
32. Doenecke, A., Winnacker, E.-L. & Hallek, M. Rapid amplification of cDNA ends (RACE) improves the PCR-based isolation of immunoglobulin variable region genes from murine and human lymphoma cells and cell lines. *Leukemia* **11**, 1787–1792 (1997).
33. Nielsen, S. C. A. & Boyd, S. D. Human adaptive immune receptor repertoire analysis—past, present, and future. *Immunol. Rev.* **284**, 9–23 (2018).
34. Thomas, N., Heather, J., Ndifon, W., Shawe-Taylor, J. & Chain, B. Decombinator: a tool for fast, efficient gene assignment in T-cell receptor sequences using a finite state machine. *Bioinformatics* **29**, 542–550 (2013).
35. Taylor, S., Yasuyama, N. & Farmer, A. A SMARTer approach to profiling the human T-cell receptor repertoire. *J. Immunol.* **196**, 209.5 (2016).
36. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
37. Yokota, R., Kaminaga, Y. & Kobayashi, T. J. Quantification of inter-sample differences in T-cell receptor repertoires using sequence-based information. *Front. Immunol.* **8**, 1500 (2017).
38. Gudikote, J. P. & Wilkinson, M. F. T-cell receptor sequences that elicit strong down-regulation of premature termination codon-bearing transcripts. *EMBO J.* **21**, 125–134 (2002).

39. Schulze-Koops, H. Lymphopenia and autoimmune diseases. *Arthritis Res. Ther.* **6**, 178–180 (2004).
40. Miho, E. et al. Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front. Immunol.* **9**, 224 (2018).
41. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).
42. Britanova, O. V. et al. Dynamics of individual T cell repertoires: from cord blood to centenarians. *J. Immunol.* **196**, 5005–5013 (2016).
43. Brüggemann, M. et al. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* **33**, 2241–2253 (2019).
44. Knecht, H. et al. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. *Leukemia* **33**, 2254–2265 (2019).
45. 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).
46. Breden, F. et al. Reproducibility and reuse of adaptive immune receptor repertoire data. *Front. Immunol.* **8**, 1418 (2017).
47. Rubelt, F. et al. Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data. *Nat. Immunol.* **18**, 1274–1278 (2017).
48. Vander Heiden, J. A. et al. AIRR community standardized representations for annotated immune repertoires. *Front. Immunol.* **9**, 2206 (2018).
49. Zhang, Y. et al. Tools for fundamental analysis functions of TCR repertoires: a systematic comparison. *Brief. Bioinform.* https://doi-org.eres.qnl.qa/10.1093/bib/bbz092 (2019).
50. Weber, C. R. et al. immuneSIM: tunable multi-feature simulation of B- and T-cell receptor repertoires for immunoinformatics benchmarking. *Bioinformatics* 36, 3594–3596 *(2020).*

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Blood effector T cell isolation. Peripheral blood mononuclear cells from two healthy blood donors (Etablissement français du sang; French Blood Center) were obtained with written informed consent for biomedical research. The experiments were carried out in conformity with the Helsinki Declaration on Biomedical Research. Donor A (Experiment A) and Donor B (Experiment B) were both men, 36 and 54 years old, respectively. $CD3^+CD4^+CD127^+CD25^-$ cells ($CD4^+$ T effector cells) were sorted at the Sorbonne Université laboratory as follows: $CD4^+$ T cells were isolated by Lymphoprep (Stemcell Technologies) density gradient and positive selection using the Dynabeads CD4 Positive Isolation Kit (Invitrogen). Enriched $CD4^+$ T cells were then labeled with anti-$CD3^+$, $CD4^+$, $CD127^+$ and $CD25^+$ antibodies, and effector T cells were sorted on a FACS ARIA II with a purity greater than 95% (Supplementary Fig. 1a).

Jurkat cell culture. The Jurkat cell line with a known TCR (TRAV8-4-CAVSDLEPNSSASKIIF-TRAJ3; TRBV12-3-CASSFSTCSANYGYTF-TRBJ1-2) (Clone E6-1), from the American Type Culture Collection (ATCC), was grown in 5% CO_2, in RPMI 1640 medium, supplemented with 10% (vol/vol) fetal bovine serum, 2 mM L-glutamine, 50 U ml^{-1} penicillin and 50 µg ml^{-1} streptomycin at the Sorbonne Université laboratory.

RNA and DNA extraction. In Experiment A, DNA and RNA were both extracted using TRIzol Reagent (Invitrogen) from 5 million Jurkat cells and 20 million $CD4^+$ T effector cells. In Experiment B, only RNA was extracted using the RNAqueous Kit (Invitrogen) from 7.2 million $CD4^+$ T effector cells following the manufacturer's recommendations. DNA concentration and RNA concentration were measured on a NanoDrop1000 (Thermo Fisher Scientific), and RNA integrity was determined on a Bioanalyzer (Agilent) with measurements higher than 8. RNA and DNA extraction and validation were performed at the Sorbonne Université laboratory.

Aliquot preparation for method comparison. In Experiment A, RNA or DNA from the $CD4^+$ effector T cells sorted from Donor A was split into three aliquots that were spiked with different amounts of RNA or DNA from the Jurkat cell line, at ratios of 1/10, 1/100 and 1/1,000. Each spiked aliquot was further split into three, and all replicates of 100 ng each were processed by all methods tested (seven for RNA and two for DNA; Supplementary Fig. 1b). In Experiment B, we analyzed the effect of the input material quantity. RNA from sorted $CD4^+$ effector T cells of Donor B was extracted, split into 15 aliquots of 100 ng each and 15 aliquots of 10 ng each and processed in triplicate using five of the RNA-based methods (Supplementary Fig. 1c). Aliquots were prepared at the Sorbonne Université laboratory and sent to the partners.

Flow cytometry. Vβ identification was performed on enriched $CD4^+$ effector T cells from Experiment A (see 'Blood effector T cell isolation' for enrichment procedure) stained with the IOTest Beta Mark Kit (Beckman Coulter) according to the manufacturer's protocol, as well as with CD4-APC, CD127-BV421 and CD25-PECy7. Data acquisition was performed on a CytoFLEX (Beckman Coulter) using CytExpert software. FlowJo was used for data analysis. Vβ frequencies were calculated on $CD4^+CD25^-CD127^+$ gated cells (Supplementary Fig. 4a,b). Staining was performed at the Sorbonne Université laboratory.

TCR library preparation and sequencing. The nine protocols for TCR library preparation compared in this study were selected according to at least one of the following criteria: 1) published by groups other than the one that developed it (mPCR-1, mPCR-3, RACE-1, RACE-2, RACE-4 and RACE-5); 2) associated with well-known analysis tools (RACE-1, RACE-2, mPCR-2); and 3) commercially available (RACE-2, RACE-3, mPCR-1 and mPCR-3). Sequencing protocols were harmonized, taking into account published recommendations or recommendations provided by the manufacturer of commercial kits or by the owner or users of the protocol. All protocols are detailed in Supplementary Methods.

TCR deep sequencing data processing. FASTQ raw data files were obtained from each method, except for Multiplex-1 and -2, for which we obtained, respectively, FASTA files and FASTQ files after proprietary pre-processing. For RACE-1 and RACE-2, UMI pre-processing was performed following protocols published elsewhere[27,28,34]. FASTQ and FASTA files were then processed for TRB and TRA sequence annotation using MiXCR software[36] (v2.1.10) with RNA-sequencing default parameters (-p rna-seq -s hsa) as available online. MiXCR extracts TRA and TRB repertoires, providing correction of PCR and sequencing errors.

Data analysis. Statistical comparisons and multivariate analyses were performed using R software version 3.5.0 (www.r-project.org). We used the ggplot2 package to generate figures[51], except heat maps. More complex analyses are detailed in the next section.

Comparing VDJ rearrangement statistics. An empirical VDJ rearrangement model for each method was built as follows. We analyzed clonotype tables to obtain comprehensive statistics of VDJ rearrangements, including the frequencies of V/D/J segment usage, number of added N bases (namely 'insert profile'—that is, the probability distribution of having A/T/G/C inserted in the N-region of CDR3 given that we observe a certain base inserted before it) and V/J segment trimming bases, with the IGoR package[52]. This model is built in a 'greedy' way in the sense that it uses best alignments provided by MiXCR rather than running expectation maximization procedures as described in Murugan et al.[53]. We used the JSD between distributions of VDJ usage to define the following two statistics that we use for comparative analysis of different TCRseq methods: 1) replicability, measured as the distance between different samples produced by the same protocol and 2) reproducibility, measured as the distance between samples produced by two different protocols. MDS used for sample mapping was performed on rank-transformed distances to avoid the distorting effect of outliers. All the analyses involve VDJ usage inferred from weighted data (TCR clonotype is weighted by its frequency in the sample) to account for TCRseq method amplification biases.

Similarity analysis. Pearson and Spearman correlations, the MH index[54] and the JSI[55] were used to assess the similarity between samples. The MH index takes into account the relative abundance of species in the sample, whereas the JSI is a measure of the intersection between two populations relative to the size of their union and is independent of relative abundances. Both indices vary between 0 (no overlap) and 1 (perfect overlap). JSI and MH were calculated using the divo package[56] on R. To discriminate indices represented by a heat map with the pheatmap package[57], we used a different set of colors. The Pearson and Spearman correlations are presented as yellow/white/orange (Fig. 2c and Supplementary Fig. 4e), MH is presented as blue/white/red (Fig. 3a) and JSI is presented as purple/yellow/green (Supplementary Fig. 5a).

Diversity profiling. The diversity was analyzed using two indices. Rényi entropy[58] is a generalization of Shannon entropy, which increases when both species richness and evenness are high. Rényi entropy is a function of a parameter α spanning from 1) the species richness ($\alpha = 0$), which corresponds to the number of clonotypes regardless of their abundance, to 2) the clonal dominance ($\alpha \to +\infty$), corresponding to the frequency of the most predominant clonotype. For $\alpha = 1$, the Shannon diversity index is computed. The exponential of the Rényi entropy corresponds to the actual number of clonotypes in the data sets[59] and is used to build a diversity profile[60]. It was computed using the entropy package[61] on R. The ChaoE[62] index was calculated with the iNEXT package[63] as a measure of extrapolation of the possible number of clonotypes based on the observed clonotypes. Rarefaction curves were interpolated from zero to the current sample size and then extrapolated to the size of the largest of samples, allowing comparison of diversity estimates. Interpolation and extrapolation of Chao index were computed using multinomial models[64].

Meta-repertoire construction. We generated an in silico meta-repertoire from the sequences obtained from the 108 replicates (45 for TRA and 63 for TRB). This meta-repertoire, for each chain, was designed to minimize biases by 1) pooling all clonotypes from the nine data sets and removed singletons to avoid introducing noise due to PCR errors; (ii) selecting non-reprocessed data sets, meaning before UMI; 3) keeping only clonotypes found by at least three different methods to avoid bias toward one particular method—the threshold was defined to reach a data set size as close as possible to the original data sets to avoid additional sampling; and 4) normalizing the size of each data set to the lowest data set to ensure the same weighting for each method. Completion of the representative meta-repertoire was achieved by pooling all the data sets. This generated a pooled data set of 14 458 TRA and 18 735 TRB clonotypes.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All the fastq data obtained in this study, including the Jurkat Clone E6-1 (ATCC TIB 152) cell line TCR α and β sequences, were deposited in the National Center for Biotechnology Information Sequence Read Archive repository following MiAIRR standard recommendations[47] under the BioProject ID PRJNA548335. The aligned sequence data will be stored in an iReceptor Repository at Sorbonne Université as a repository in the AIRR Data Commons and can be explored and downloaded through the iReceptor Gateway[65] (https://gateway.ireceptor.org). Source data for TCRVβ flow cytometry data are provided as Supplementary Fig. 4a,b. All other data are available from the corresponding author upon reasonable request.

Code availability

All software packages and programs are publicly available and open source. Scripts used to analyze the data with MiXCR are available from https://mixcr.milaboratory.com. Decombinator is available from https://github.com/innate2adaptive/Decombinator. MiGEC is available from https://github.com/mikessh/migec. Detailed VDJ rearrangement statistics scripts are available from https://github.com/antigenomics/repseq-protocol-comparison. There is no restriction on the use of the code or data.

References

51. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
52. Marcou, Q., Mora, T. & Walczak, A. M. High-throughput immune repertoire analysis with IGoR. *Nat. Commun.* **9**, 561 (2018).
53. Murugan, A., Mora, T., Walczak, A. M. & Callan, C. G. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc. Natl Acad. Sci. USA* **109**, 16161–16166 (2012).
54. Horn, H. S. Measurement of "overlap" in comparative ecological studies. *Am. Nat.* **100**, 419–424 (1966).
55. Jaccard, P. The distribution of the flora in the Alpine zone. *New Phytol.* **11**, 37–50 (1912).
56. Sadee, C., Pietrzak, M., Seweryn, M. & Rempala, G. divo: Tools for Analysis of Diversity and Similarity in Biological Systems (Diversity and Overlap Analysis Package). https://rdrr.io/cran/divo/ (2017).
57. Kolde, R. pheatmap: pretty heatmaps. https://cran.r-project.org/web/packages/pheatmap/index.html (2019).
58. Renyi, A. On measures of information and entropy. In *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability* 547–561 (University of California Press, 1961).
59. Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 427–432 (1973).
60. Chaara, W. et al. RepSeq data representativeness and robustness assessment by shannon entropy. *Front. Immunol.* **9**, 1038 (2018).
61. Hausser, J. & Strimmer, K. *Estimation of Entropy, Mutual Information and Related Quantities*. (2014).
62. Colwell, R. K. et al. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. *J. Plant Ecol.* **5**, 3–21 (2012).
63. Hsieh, T. C., Ma, K. H. & Chao, A. *Package iNEXT: Interpolation and Extrapolation for Species Diversity* (2019).
64. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
65. Corrie, B. D. et al. iReceptor: a platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol. Rev.* **284**, 24–41 (2018).

Acknowledgements

We are grateful to M. Barbié for providing the human samples. This work benefited from equipment and services from the iGenSeq core facility at ICM. This work was supported the ERC-Advanced TRiPoD (322856), LabEx Transimmunom (ANR-11-IDEX-0004-02) and RHU iMAP (ANR-16-RHUS-0001) grants to D.K. E.M.F. is funded by the European Research Area Network-Cardiovascular Diseases (JCT2018 and ANR-18-ECVD-0001) and iReceptorPlus (H2020 Research and Innovation Programme 825821) grants. M.S. and D.M.C. were supported by a grant from the Ministry of Science and Higher Education of the Russian Federation (075–15–2019-1789). This work was funded, in part, by the intramural program of the National Institute of Allergy and Infectious Diseases (to D.C.D.). B.C. was supported by the National Institute for Health Research UCL Hospitals Biomedical Research. A.E. was supported by DFG CRTD (FZ 111). A.N.D. was supported by the Ministry of Education, Youth and Sports of the Czech Republic under project CEITEC 2020 (LQ1601). K.K. and P.P.R. were supported by the European Research Area Network-Cardiovascular Diseases (JCT2018 and AIR-MI Consortium) program.

Author contributions

P.B., V.Q., E.S.E., A.N.D., I.U., M.I., T.O., A.E., S.D., A.R., K.K. and P.R. performed the experiments and raw data pre-processing. P.B., V.Q., M.S. and M.I. analyzed the data. E.M.F., D.M.C., B.C., D.C.D. and D.K. designed the experiments. P.B., D.K. and E.M.F. wrote the manuscript with input from all authors. D.K. and E.M.F. conceived the study, which was supervised by E.M.F. D.K., B.C., A.E., D.C.D., D.M.C., K.K. and M.S. obtained funding for the study.

Competing interests

D.M.C. and M.S. are cofounders of MiLaboratory LLC. A.E., A.N.D., A.R., B.C., D.C.D., D.K., E.M.F., E.S.E., I.U., K.K., M.I., P.B., P.R., S.D., T.O. and V.Q. declare no conflicts of interest.

Additional information

Supplementary information Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0656-3.

Correspondence and requests for materials should be addressed to E.M.-F.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

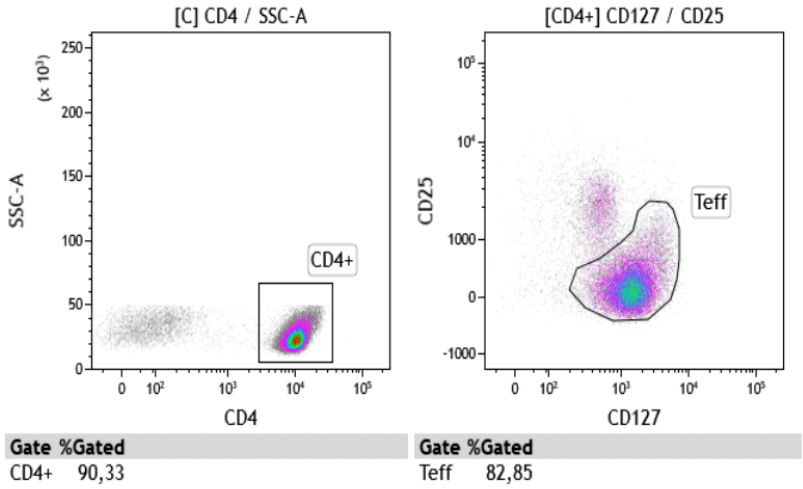
https://doi.org/10.1038/s41587-020-0656-3

Supplementary information

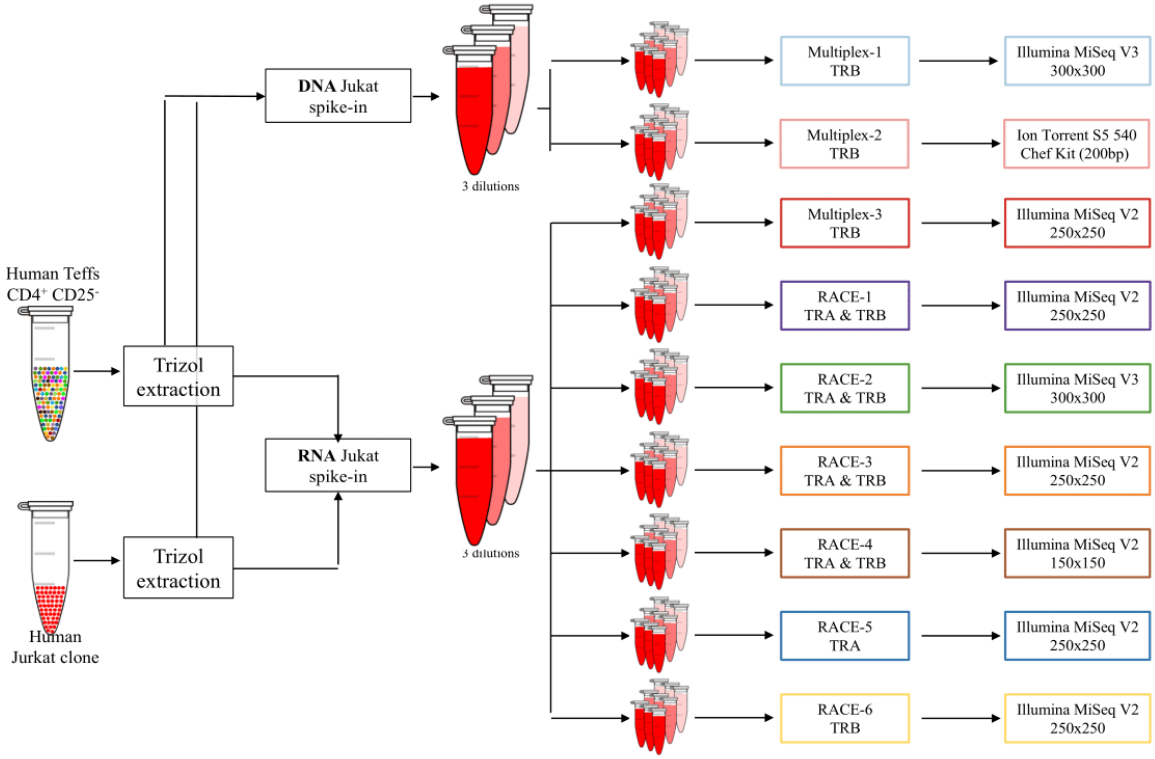
Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases

In the format provided by the authors and unedited

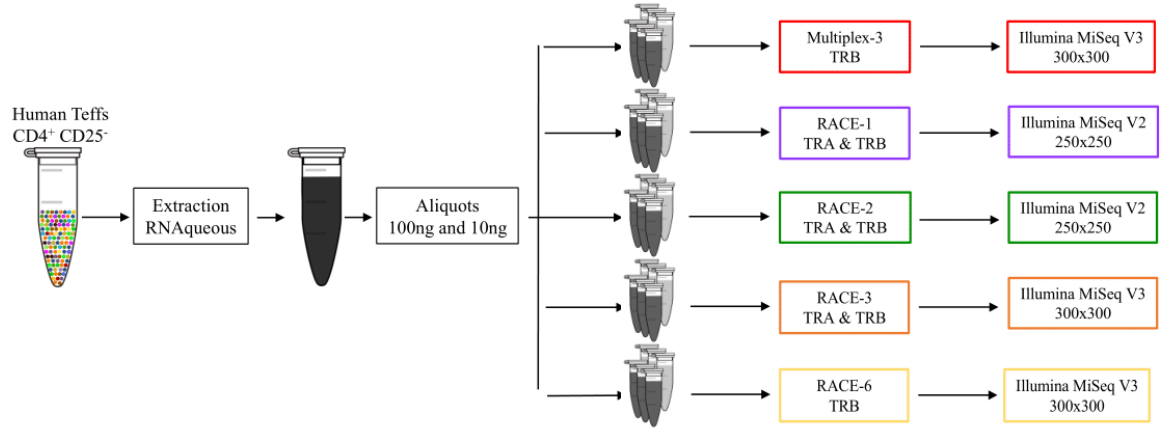
a



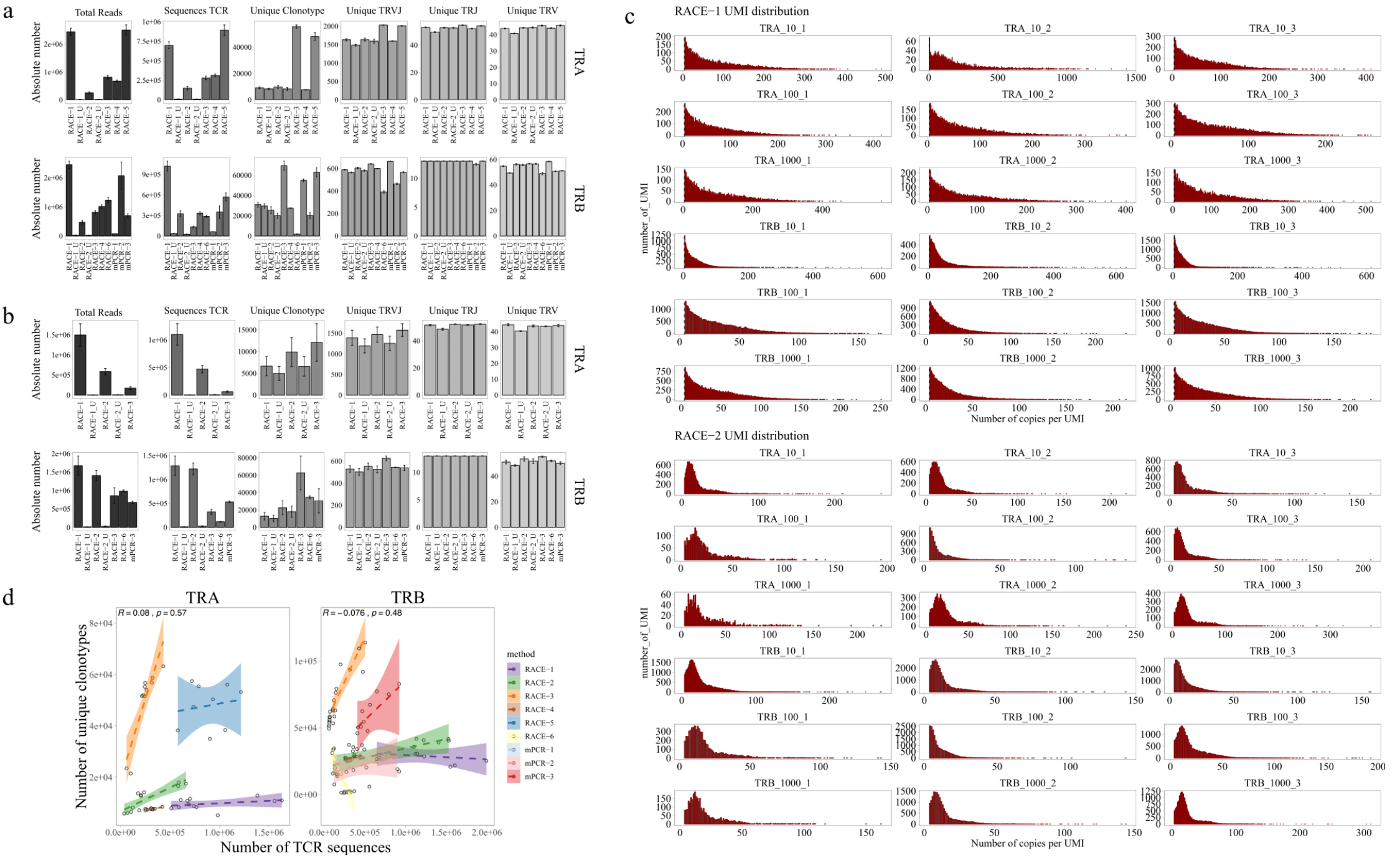
b



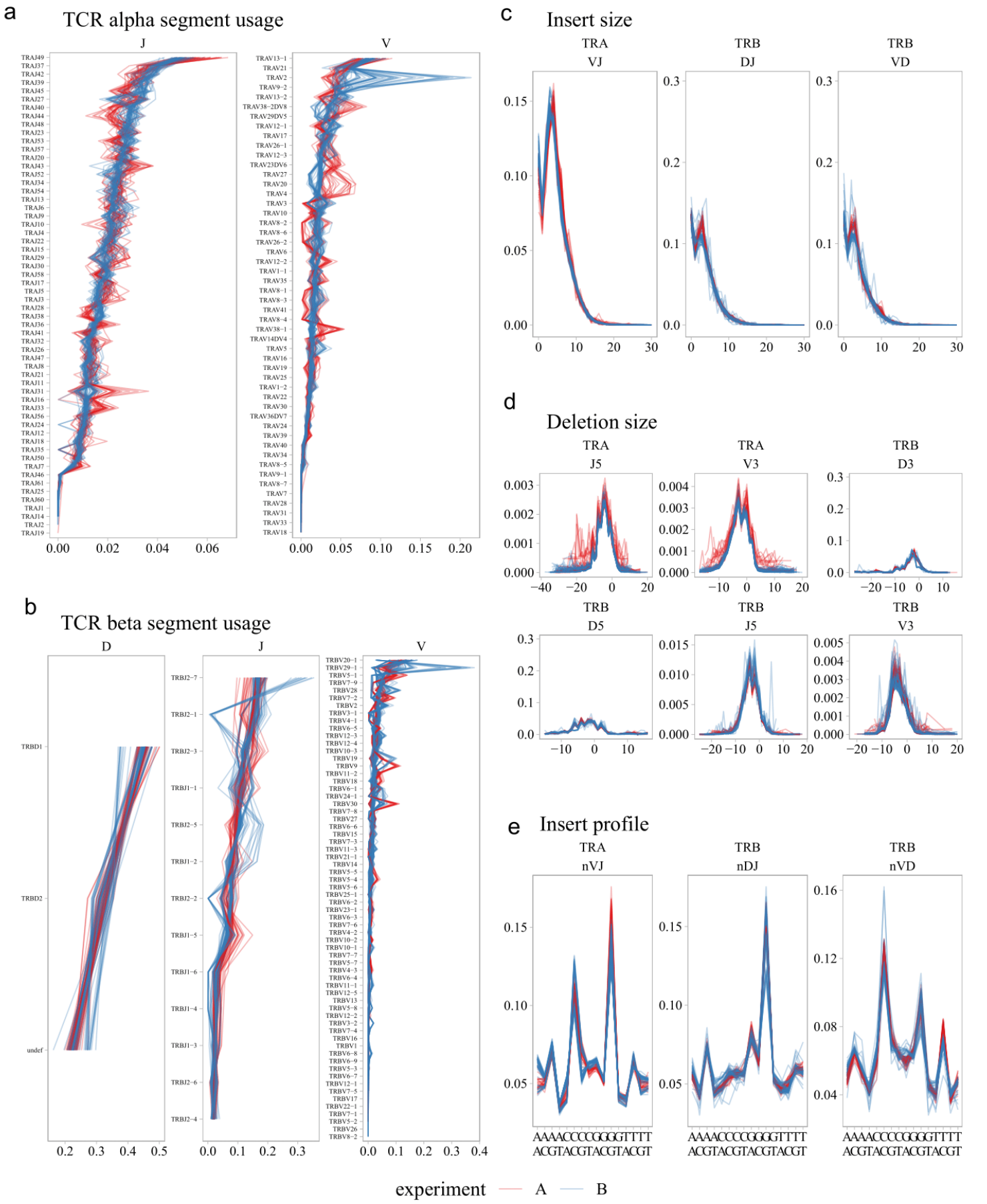
c



Supplementary Fig. 1: Sorting cell strategy and experimental design. **a**, Gating strategy for effector T cell sorting. CD4+CD25- T cells were gated on previously enriched lymphocyte (left panel). Effector T cells were gated and sorted based on $CD127^{+}CD25^{-}$ expression (right panel). **b**, Experiment A. DNA and RNA of FACS-sorted effector T cells ($CD4^{+}CD25^{-}CD127^{+}$) from donor 1 and cultured Jurkat leukemic T-cells were extracted. Jurkat DNA or RNA were then added to effector T cells DNA or RNA, respectively, at the following ratios: 1/1000, 1/100, 1/10 to a final quantity of 100ng. Triplicate of each ratio condition were processed by 9 TCR library preparation methods (7 with RNA and 2 with DNA). **c**, Experiment B. RNA of FACS-sorted effector T cells ($CD4^{+}CD25^{-}CD127^{+}$) from donor 2 was extracted. Triplicate of aliquots prepared from two quantity of RNA (10ng and 100ng) were processed by 5 TCR library preparation methods..

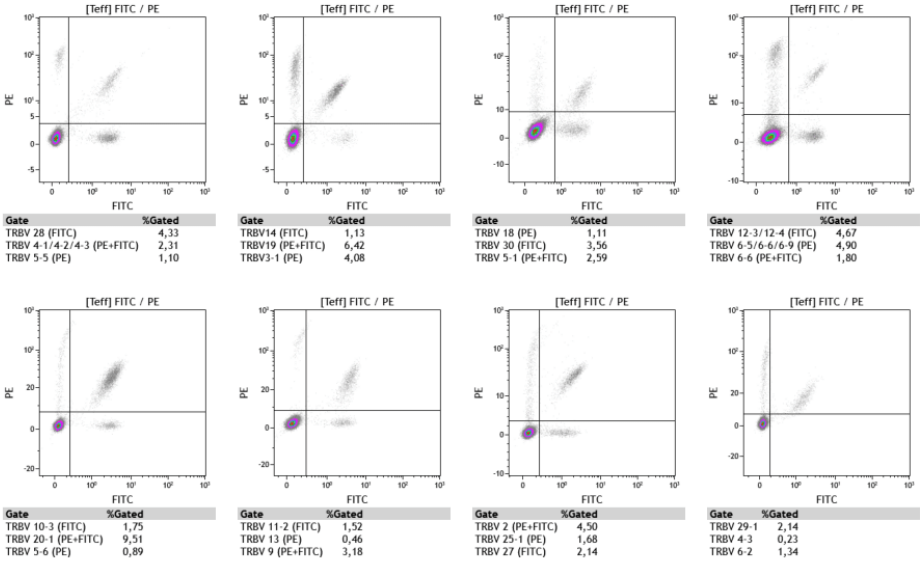


Supplementary Fig. 2: Sequencing statistics summary per method and experiment. a-b, For experiment A (a, n = 9) and experiment B (b, n = 6), histogram plots represent from left to right the mean values +/- SEM of raw reads (Total Reads), TCR Sequences, unique clonotype (V-CDR3aa-J), V-J combination (Unique TRVJ), J gene (Unique TRJ) and V gene (Unique TRV) obtained by each method for both TRA (top) and TRB (bottom) chains. **c**, UMI distribution per sample for RACE-1 and RACE-2 methods from experiment A. X-Axis represents the number of copies of each UMI and y-axis represents the number of UMI for each copy-number. **d**, Correlation between the number of unique clonotype and the number of TCR sequences per method and chain from experiment A. Error bands represent the 95% confidence interval.



Supplementary Fig. 3: VDJ rearrangement statistic distributions. VDJ rearrangement parameter distribution have been assessed as described in material and methods using "weighted" data, i.e. the distributions are scaled by the number of reads associated with each unique clonotype. Results are represented for experiment A (red) and experiment B (blue) data as follows: **a-b**, Variable (V), Diversity (D) and Joining (J) segment usage distributions for TCR alpha (a) and beta (b) chains. **c**, Insert size distributions for VJ junctions (TRA) and VD/DJ junctions (TRB). **d**, Number of nucleotides trimmed from 5' and 3' ends of V, D and J segments for TRA and TRB chains. **e**, Non-template nucleotide frequencies for VJ (nVJ), VD/DJ (nVD and nDJ) junctions. Insertion probabilities of an observed base (top letter) given previous base (bottom letter) are given, 5'->3' direction is considered for nVJ and nVD, 3'->5' is considered for nDJ. Jurkat TCR sequences were removed from datasets for this analysis.

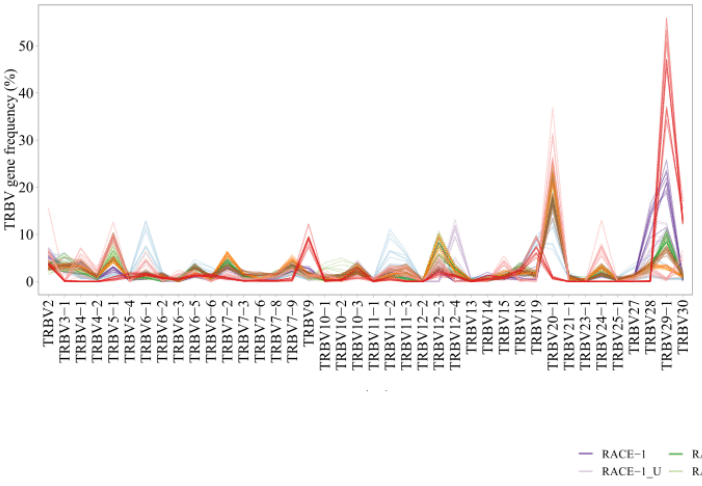
a



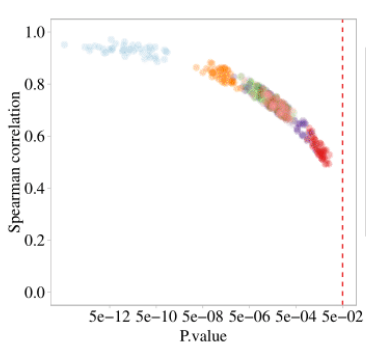
b

TRBV	Cytometry 1	Cytometry 2	Cytometry 3	Cytometry 4	Cytometry 5	Standard deviation
TRBV2	4,5	4,31	4,55	4,31	4,44	0,109
TRBV3-1	4,02	3,89	3,89	3,82	4,34	0,207
TRBV4-1/2	2,08	2,33	2,11	2,15	2,09	0,103
TRBV4-3	0,23	0,21	0,27	0,18	0,44	0,102
TRBV5-1	2,59	2,78	2,63	2,8	2,83	0,108
TRBV5-5	1,1	1,32	1,11	1,18	1,05	0,105
TRBV5-6	0,89	1,01	0,95	0,93	0,88	0,052
TRBV6-2	1,34	1,54	1,56	1,52	1,32	0,116
TRBV6-5/9	3,09	3,31	3,13	3,19	3,2	0,084
TRBV6-6	1,8	1,94	1,93	1,93	1,85	0,062
TRBV9	3,18	3,04	3,18	3,39	3,13	0,129
TRBV10-3	1,75	1,98	1,82	1,89	1,86	0,085
TRBV11-2	1,52	2,18	1,86	1,78	1,97	0,243
TRBV12-3/4	4,67	4,99	4,62	5,05	4,82	0,190
TRBV13	0,46	0,52	0,52	0,52	0,51	0,026
TRBV14	1,07	1,06	1,08	1,14	1,15	0,042
TRBV18	1,11	1,34	1,2	1,3	1,39	0,113
TRBV19	6,35	6,62	6,54	6,24	6,53	0,156
TRBV20-1	9,51	9,29	9,32	9,46	9,4	0,092
TRBV25-1	1,68	1,65	1,87	1,7	1,54	0,119
TRBV27	2,14	2,17	2,05	2,22	2,1	0,065
TRBV28	4,33	4,54	4,57	4,59	4,66	0,124
TRBV29-1	2,14	2	2	2,15	1,82	0,134
TRBV30	3,56	3,76	3,81	3,96	3,91	0,156

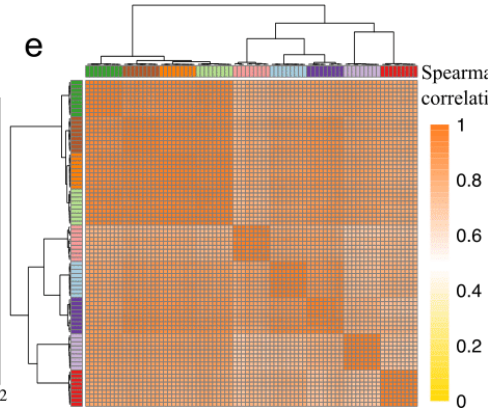
c



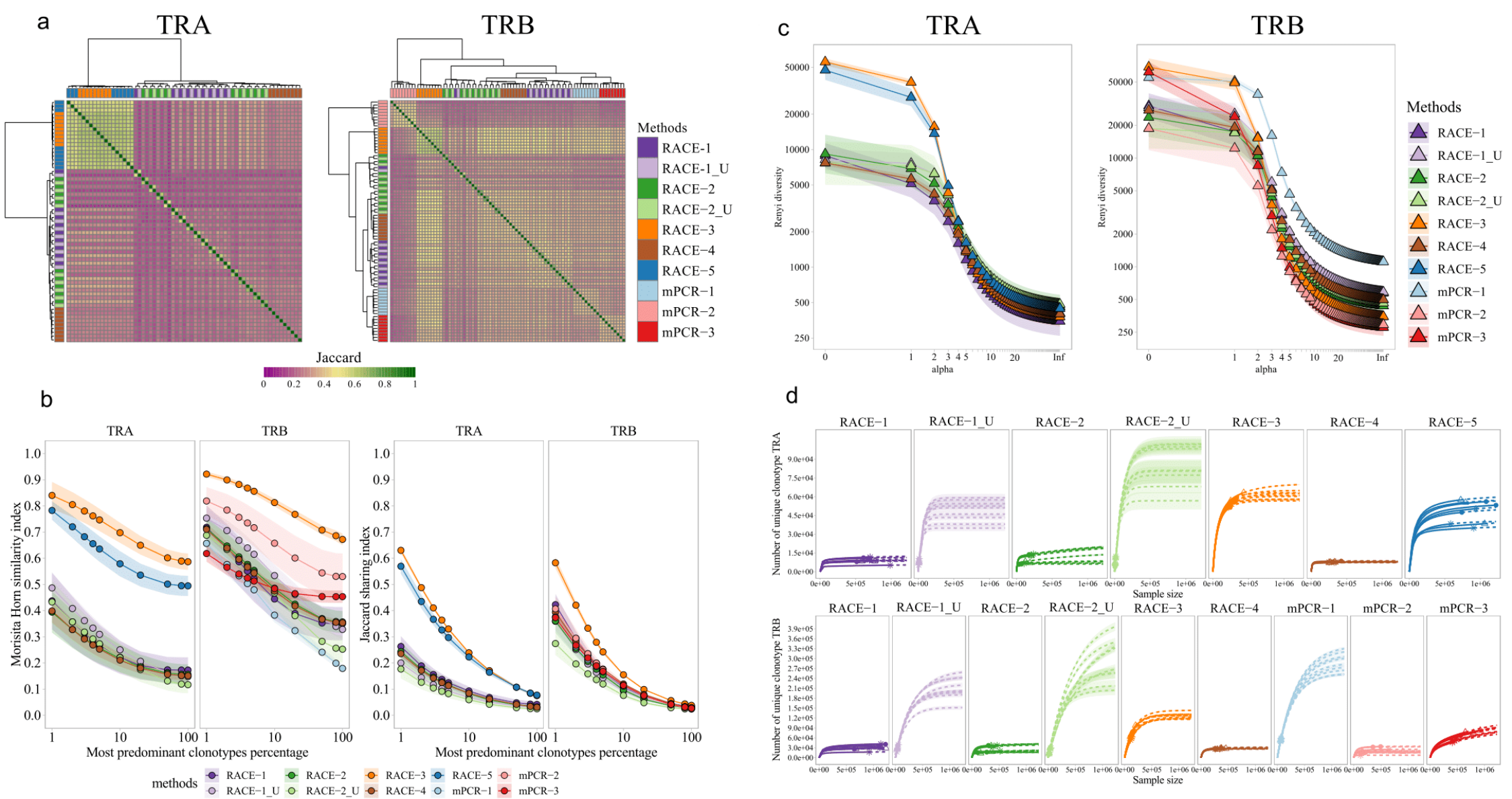
d



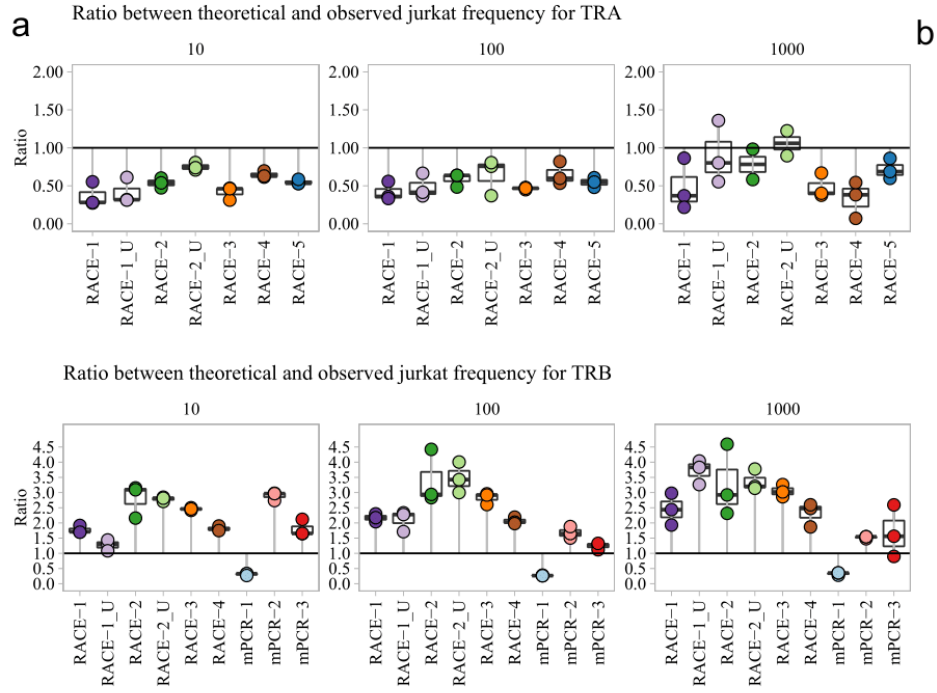
e



Supplementary Fig. 4: Correlation assessment between flow cytometry and TCRseq TRBV expression level. **a**, TRBV gene repertoire of effector T cells by flow cytometry analysis. Each dotplot corresponds to 3 TRBVs: one TRBV is conjugated to FITC, another one with PE, and the third to both FITC and PE. Percentage of each TRBV subset are shown below the dotplot. Total CD4⁺ lymphocyte TRBV coverage is 64,03%. **b**, Summary table of the TRBV gene usage and standard deviation analyzed by flow-cytometry for the 5 cellular replicas, each replica is composed of 50 000 cells from the previously harvested effector T cells. c, TRBV distribution using experiment A (after Jurkat sequences removal) TCRseq datasets. TRBV frequencies were re-calculated on the top 5% MPC for each method. Only TRBV with significant differences between at least two methods are represented (Two way ANOVA, P-values < 0.05). **d**, Spearman's correlation of the TRBV frequencies between the 5 flow cytometry datasets and the 9 TCRseq replicates was calculated for each method. The plot represents the correlation score (y-axis) and the P-value of the correlation (x- axis), allowing the classification of the methods. **e**, Heatmap of the Spearman's correlations between each replicate for the distribution of TRBV gene usage (n=62 unique TRBV genes). The Euclidean distance was used for hierarchical clustering as a color-coded matrix ranging from 0 (yellow, maximum dissimilarity) to 1 (orange, maximum similarity). Jurkat TCR sequences were removed from datasets for this analysis. Jurkat TCR sequences were removed from datasets for this analysis.

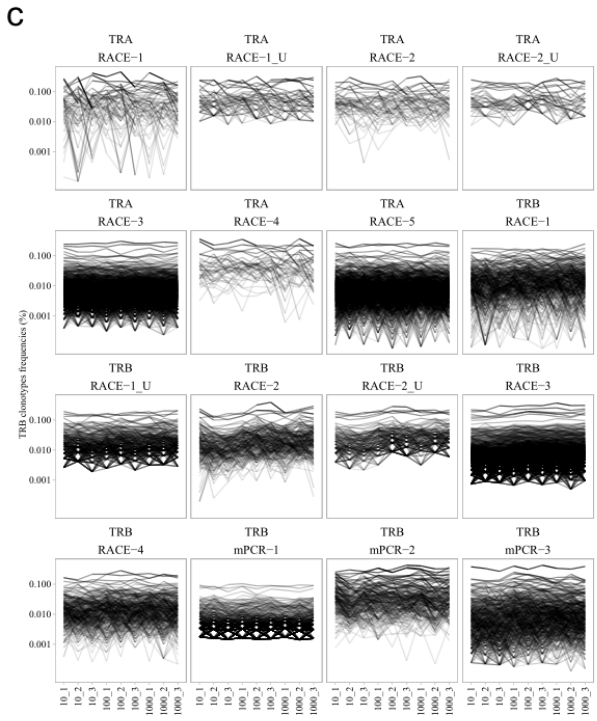


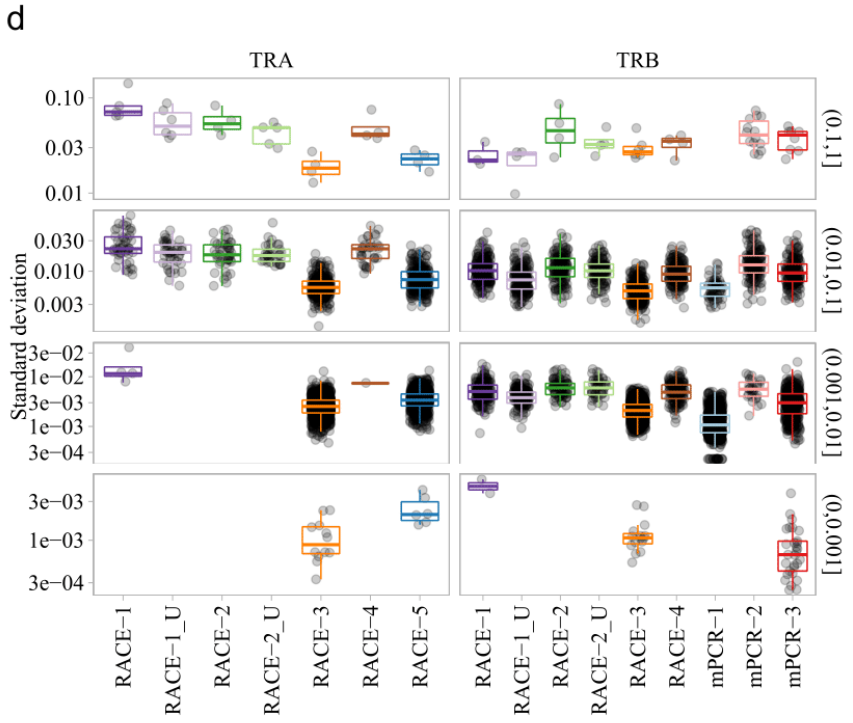
Supplementary Fig. 5: Most predominant clonotype similarity indices and clonotype diversity. **a**, Heatmaps of Jaccard similarity index (JSI). JSI scores were calculated between each replicate across all methods on the top 1% most predominant clonotypes (MPC) for TRA (left) and TRB (right). Euclidean distance was used for hierarchical clustering as a color-coded matrix ranging from 0 (violet, maximal dissimilarity) to 1 (green, maximal similarity). **b**, Morisita-Horn and Jaccard similarity profiles. Similarity profiles computed on increasing percentages of most predominant clonotypes (MPC): from 1% MPC to 100% MPC (the latter corresponding to elsewhere mentioned as “Overall”). For each method, mean similarity metrics were calculated between replicates (n = 9 technical replicates). Shaded areas represent the standard deviation of the mean. **c**, Rényi diversity profile. For each dataset (n = 9 technical replicates), diversity metrics using clonotype frequencies were calculated for increasing values of Rényi order α until stabilization of the resulting diversity (see material and methods) and displayed as a Rényi diversity profile. α varies from 0 (Richness) to ∞ (Berger-Parker); for α = 1, the Shannon entropy was computed as described in Rényi1. Shaded areas represent the standard deviation of the mean. **d**, Rarefaction curves for TRA (top) and TRB (bottom), representing the number unique of clonotype as a function of the number of TCR sequences (Sample size), are displayed for each method. Solid and dashed lines represent interpolated and extrapolated values, respectively. Dots indicate the observed sample size and richness (number of unique TRA or TRB clonotype). Shaded areas represent 95% confidence intervals. Jurkat TCR sequences were removed from datasets for this analysis.



b

methods	TRA	TRB
RACE-1	55	395
RACE-1_U	50	378
RACE-2	46	228
RACE-2_U	39	187
RACE-3	1252	1249
RACE-4	39	330
RACE-5	958	ND
mPCR-1	ND	772
mPCR-2	ND	205
mPCR-3	ND	563





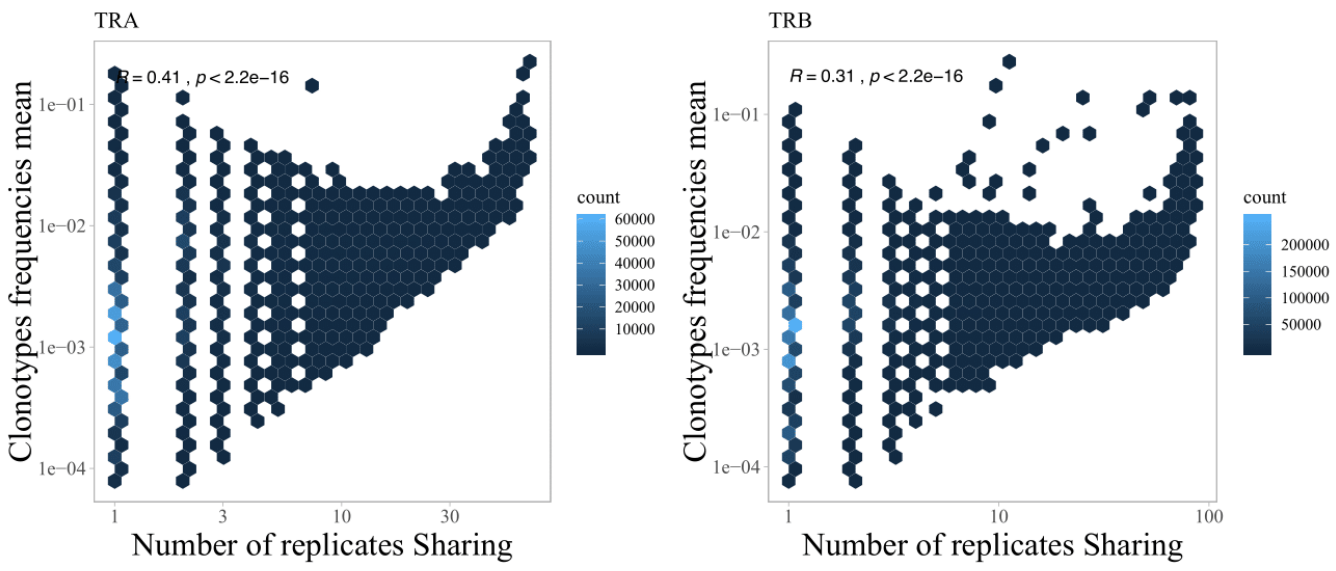
Supplementary Fig. 6: Clonotypes shared by all replicates per method details. **a**, Quantification of Jurkat TRA and TRB clonotype detection per method. The plots represent for each method and each dilution (n = 3 technical replicates per dilution), the ratio between the observed and the expected frequencies of TRA (top) and TRB (bottom) clonotypes from the Jurkat spike-in. **b**, Number of unique clonotypes shared by all replicates per method, for TRA and TRB. **c**, Tracking frequencies distribution of these shared clonotypes. **d**, Standard deviations of frequencies of clonotypes shared between all the replicates (n=9) per method, split into frequency ranks, for TRA (left) and TRB (right). Jurkat TCR sequences were removed from datasets for this analysis. Boxplots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5*IQR from the hinge (where IQR is the interquartile range). The lower whisker extends from the hinge to the smallest value at most 1.5*IQR of the hinge).

a

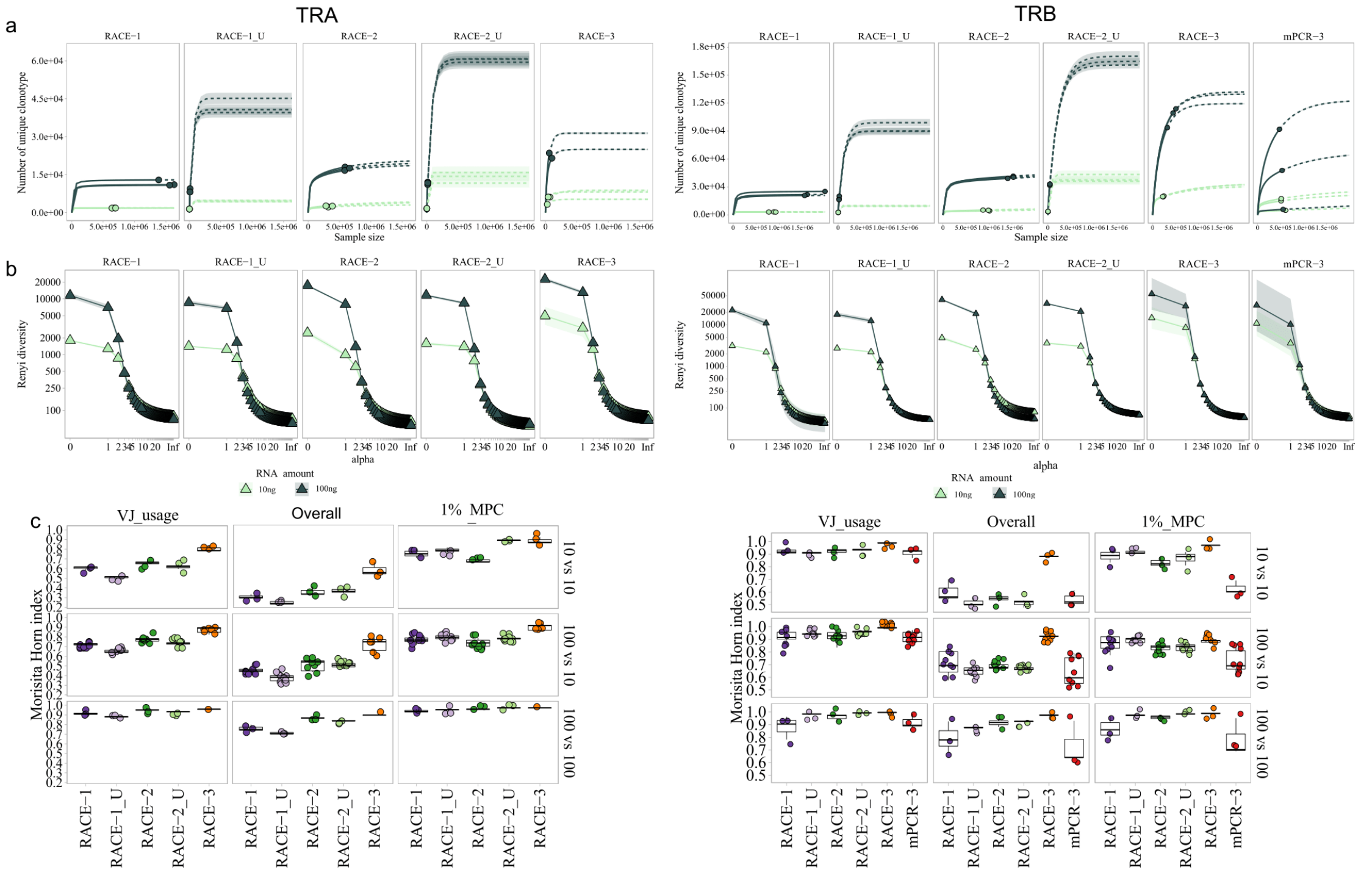
Number of Replicates Sharing	TRA		TRB	
	Number of clonotypes	Percentage of clonotypes	Number of clonotypes	Percentage of clonotypes
1	473758	68,2571	1476606	74,0494
2	118976	17,1416	358103	17,9583
3	38420	5,5354	81067	4,0654
4	19641	2,8298	31626	1,5860
5	10951	1,5778	13765	0,6903
6	7348	1,0587	7790	0,3907
7	5061	0,7292	4907	0,2461
8	3726	0,5368	3395	0,1703
9	2783	0,4010	2435	0,1221
10	2138	0,3080	1887	0,0946
11	1712	0,2467	1500	0,0752
12	1444	0,2080	1128	0,0566
13	1056	0,1521	998	0,0500
14	899	0,1295	792	0,0397
15	776	0,1118	668	0,0335
16	610	0,0879	588	0,0295
17	592	0,0853	524	0,0263
18	496	0,0715	482	0,0242
19	416	0,0599	430	0,0216
20	381	0,0549	382	0,0192
21	306	0,0441	297	0,0149
22	251	0,0362	303	0,0152
23	244	0,0352	255	0,0128
24	203	0,0292	245	0,0123
25	197	0,0284	228	0,0114
26	169	0,0243	212	0,0106
27	159	0,0229	183	0,0092
28	141	0,0203	184	0,0092
29	109	0,0157	177	0,0089
30	107	0,0154	152	0,0076
31	93	0,0134	143	0,0072
32	89	0,0128	138	0,0069
33	83	0,0120	120	0,0060
34	60	0,0086	117	0,0059
35	76	0,0109	120	0,0060
36	60	0,0086	116	0,0058
37	55	0,0079	100	0,0050
38	48	0,0069	106	0,0053
39	43	0,0062	73	0,0037
40	29	0,0042	98	0,0049
41	37	0,0053	89	0,0045

Number of Replicates Sharing	TRA		TRB	
	Number of clonotypes	Percentage of clonotypes	Number of clonotypes	Percentage of clonotypes
42	35	0,0050	69	0,0035
43	30	0,0043	69	0,0035
44	29	0,0042	55	0,0028
45	21	0,0030	59	0,0030
46	23	0,0033	62	0,0031
47	24	0,0035	70	0,0035
48	14	0,0020	59	0,0030
49	15	0,0022	53	0,0027
50	18	0,0026	48	0,0024
51	23	0,0033	46	0,0023
52	9	0,0013	44	0,0022
53	13	0,0019	50	0,0025
54	7	0,0010	33	0,0017
55	6	0,0009	49	0,0025
56	13	0,0019	42	0,0021
57	12	0,0017	43	0,0022
58	8	0,0012	37	0,0019
59	8	0,0012	38	0,0019
60	7	0,0010	40	0,0020
61	5	0,0007	32	0,0016
62	7	0,0010	37	0,0019
63	9	0,0013	38	0,0019
64	X	X	27	0,0014
65	X	X	29	0,0015
66	X	X	37	0,0019
67	X	X	36	0,0018
68	X	X	32	0,0016
69	X	X	28	0,0014
70	X	X	41	0,0021
71	X	X	23	0,0012
72	X	X	41	0,0021
73	X	X	19	0,0010
74	X	X	24	0,0012
75	X	X	15	0,0008
76	X	X	13	0,0007
77	X	X	22	0,0011
78	X	X	18	0,0009
79	X	X	19	0,0010
80	X	X	24	0,0012
81	X	X	31	0,0016
Total	694079	100	1994081	100

b

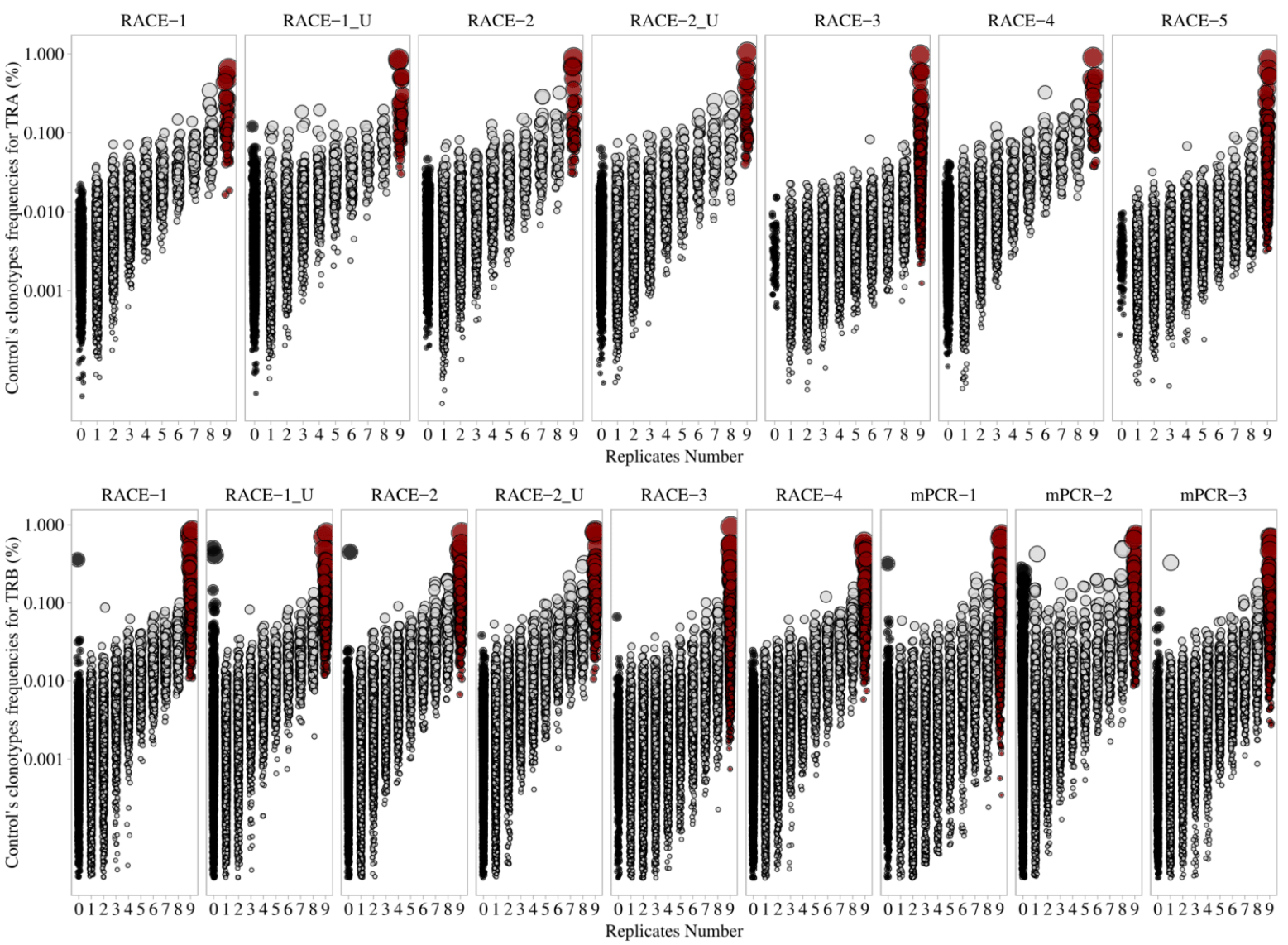


Supplementary Fig. 7: Clonotypes sharing by all replicates and abundance correlation with clonotype sharing. **a**, Table of the clonotypes sharing distribution. From all the clonotypes identified across all the replicates for all the methods without any filter on clone size (694 079 TRA clonotypes and 1 994 081 TRB clonotypes), the number and percentage of TRA and TRB clonotypes shared between increasing number of replicates are shown. **b**, Correlation statistics between the average frequency of clonotypes among all replicates and their sharing. To avoid minimizing clonotype frequencies found in single replicates, for each clonotype, the average frequency was calculated as a function of the number of replicates in which it was identified (For example, if clonotype A is found at a frequency of 10% in 1 replicate, the average frequency is 10; while if B is found at 10% in replicate 1 and 7% in replicate 2, the average is 8.5%).

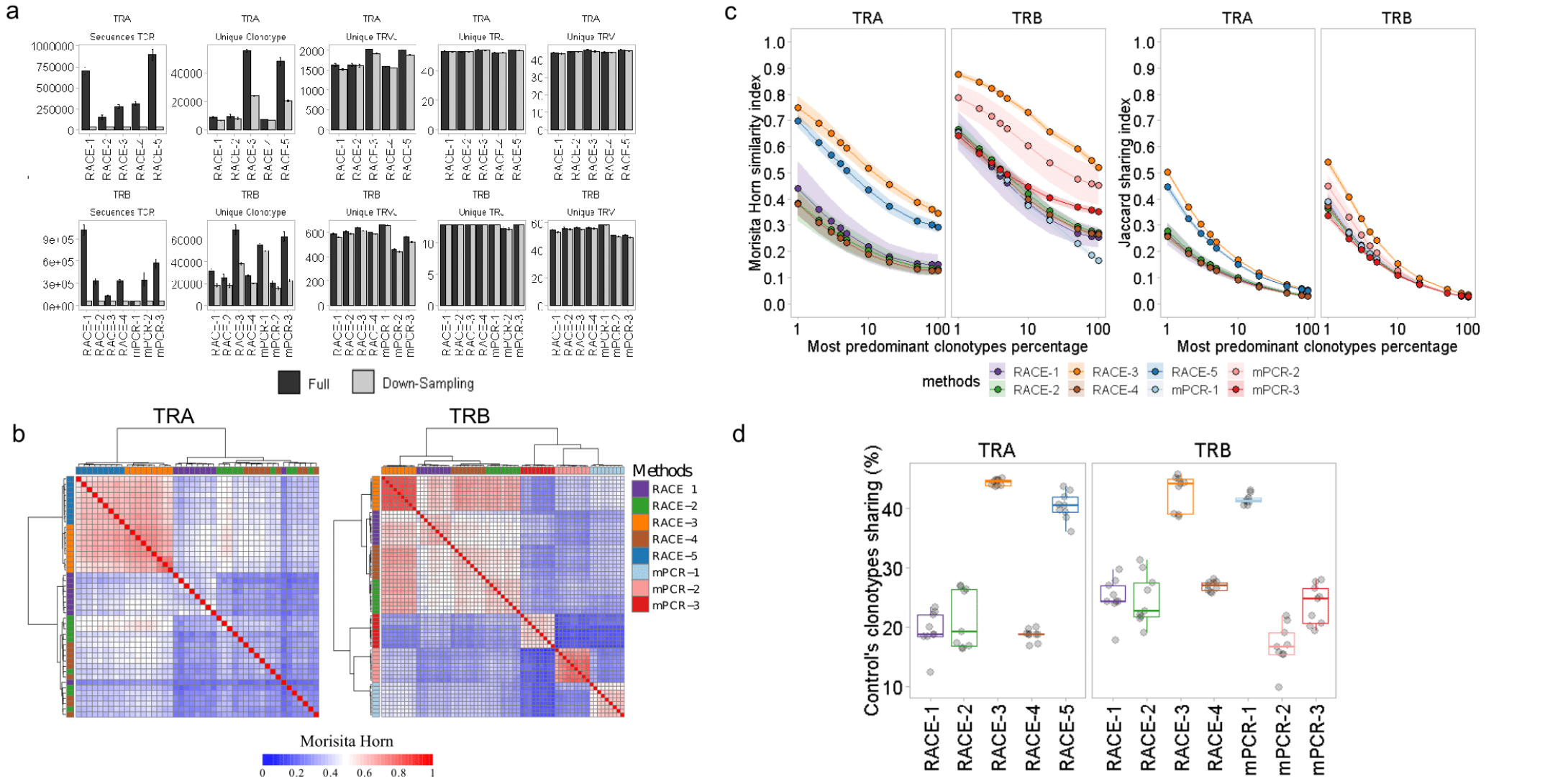


Supplementary Fig. 8: The impact of RNA quantity on TRA and TRB repertoires. a, Rarefaction curves, representing the number of unique clonotype as a function of the number of TCR sequences (Sample size), are displayed for each method, with 100ng (dark green, n=3 replicates from experiment B) or 10ng (light green, n=3 replicates from experiment B) as input RNA. Solid and dashed lines represent interpolated and extrapolated values, respectively. Dots indicate the observed sample size and richness (number of unique clonotype). Shaded areas represent the 95% confidence interval.**b**, Rényi diversity profile. For each dataset, diversity metrics using clonotype frequencies

were calculated for increasing values of Rényi order α until stabilization of the resulting diversity (see material and methods) and displayed as a Rényi diversity profile. A varies from 0 (Richness) to ∞ (Berger-Parker); for α = 1, the Shannon entropy was computed as described in Rényi[1]. Each graph represents the profiles obtained from data generated from 100ng (dark green, n=3 replicates from experiment B) or 10ng (light green, n=3 replicates from experiment B) of input RNA. Shaded areas represent the standard deviation of the mean. **c**, Morisita-Horn scores were calculated between 10ng (top, n=3 replicates from experiment B), 100 ng (middle, n=3 replicates from experiment B) and 10 vs. 100ng (bottom, n=6 replicates from experiment B) replicates on V-J usage, overall clonotypes and the 1% MPC. Results are presented for TRA in the left panel and for TRB in right panel. Boxplots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5*IQR from the hinge (where IQR is the inter-quartile range). The lower whisker extends from the hinge to the smallest value at most 1.5*IQR of the hinge).



Supplementary Fig. 9: Detailed distribution of meta-repertoire clonotypes in the replicates by method for TRA and TRB. Distribution of meta-repertoire clonotypes by replicates sharing number for TRA (top) and TRB (bottom). Black dots (0) represent the unseen clonotypes and red dots (9) represent the clonotypes found by the 9 replicates. Each method is represented independently. Jurkat TCR sequences were removed from datasets for this analysis.



Supplementary Fig.10: Down-sampling analysis. We performed such a down-sampling based on the smallest sample (except UMI-processed methods) in terms of TCR sequences for TRA and TRB separately (37599 TRA and 57468 TRB). **a**, Histogram plots represent from left to right the mean values +/- SD of TCR sequences, unique clonotype (V-CDR3aa-J), V-J combination (Unique TRVJ), J gene (Unique TRJ) and V gene (Unique TRV) obtained by each method for both TRA (top) and TRB (bottom) chains, for the full datasets (n = 9) in dark and the down-sampled datasets (n = 9) in grey. **b**, Heatmaps of the Morisita-Horn similarity index (MH). MH scores were calculated between each replicate across all methods for the top 1% of most predominant clonotypes (MPC) for TRA (left) and TRB (right). The Euclidean distance was used for hierarchical clustering as a color-coded matrix ranging from 0 (blue, maximum dissimilarity) to 1 (red, maximum similarity). **c**, Morisita-Horn and Jaccard similarity profiles. Similarity profiles computed on increasing percentages of most predominant clonotypes (MPC): from 1% MPC to 100% MPC (the latter corresponding to elsewhere mentioned as “Overall”). For each method, similarity metrics were calculated between replicates (n = 9). Shaded areas represent the standard deviation of the mean. **d**, We also generated a meta-repertoire using the sampled repertoires. Replicate (n = 9) sharing fraction in meta-repertoire (focus on meta-repertoire clonotypes) for TRA (left) and TRB (right). The values represented correspond to the percentage of clonotypes from each replicate per method found in the meta-repertoire, median and the 1st and 3rd quartiles are shown. As you can see, despite this drastic downsampling for some of the methods, the results are very much comparable with those obtained on the full datasets. Boxplots visualize five summary statistics: the median, two hinges (the lower and upper hinges correspond to the first and third quartiles) and two whiskers (the upper whisker extends from the hinge to the largest value no further than 1.5*IQR from the hinge (where IQR is the inter-quartile range). The lower whisker extends from the hinge to the smallest value at most 1.5*IQR of the hinge).

Protocol	Multiplex-1	Multiplex-2	Multiplex-3	RACE-1	RACE-2	RACE-3	RACE-4	RACE-5	RACE-6
Methods identification	Adaptive Technology Robins, H.S. et al. Blood (2009)	iMonitor Zhang, W. et al. Genetics (2015)	iRepertoire Wang, C. et al. PNAS (2010)	Chain, B team Oakes, T. et al. Front. Immunol. (2017) Uddin, I. et al. Methods Mol. Biol. (2018)	MiLaboratory LLC Mamedov, I.Z. et al. Front. Immunol. (2013)	Takara Taylor, S. et al. J. Immunol. (2016)	Douek, D team Douek, D.C. et al. J. Immunol. (2002)	Eugster, A. team Eugster, A. et al. J. Immunol. Methods (2013)	Invitrogen
TCR chain	Beta	Beta	Beta	Alpha & Beta (independent reactions)	Alpha & Beta (independent reactions)	Alpha & Beta (Same reaction)	Alpha & Beta (independent reactions)	Alpha	Beta
library_layout	paired	single	paired	paired	paired	single	paired	paired	paired
Starting material	gDNA	gDNA	RNA	RNA	RNA	RNA	RNA	RNA	RNA
PCR method	m-PCR	m-PCR	arm-PCR	TRAB RACE	TRAB RACE	TRAB RACE	TRAB RACE	TRA RACE	TRB RACE
Forward Primer targets	TRBV genes n=24	TRBV genes n=24	TRBV genes n=24	Anchor primer (target added by ligation)	NN Oligo (target added by template switching) - MiLaboratory LLC	SMART-Seq v4 Oligonucleotide (target added by template switching) - Takara Bio	5PIIA anchor primer (target added by template switching)	Template Switch Oligo (target added by template switching)	Abridged Anchor Primer (target added by nucleotidylexotransferance (TdT enzyme))
TSO*-seq lenght (nt)	Not applicable	Not applicable	Not applicable	58	Proprietary (Not disclosed)	Proprietary (Not disclosed)	23	37	30
Reverse primer target	TRBJ n=12	TRBJ n=13	TRBC	TRAC & TRBC	TRAC & TRBC	TRAC & TRBC	TRAC & TRBC	TRAC	TRBC
UMI (yes/no)	No	No	No	Yes	Yes	No	No	No	No
UMI-size (nt)	NA	NA	NA	12	12	NA	NA	NA	NA
RT (Yes/no)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reverse Transcriptase	Not applicable	Not applicable	Proprietary (Not disclosed)	Superscript IV (Invitrogen)	SMARTScribe (Takara Bio)	SMARTScribe (Takara Bio)	SuperScript II (Invitrogen)	SuperScript II (Invitrogen)	SuperScript II (Invitrogen)
Number of PCR	2	1	2	3	2	2	3	3	1
PCR1 cycle number	30	30	25	4	18	20	20	15	35
PCR2 cycle number	7	NA	30	6	10	20	5	11	NA
PCR3 cycle number	NA	NA	NA	10-15 (real time PCR)	NA	NA	3	25	NA
Polymerase type	HotStarTaq Plus DNA Polymerase 5 (Qiagen)	Ion AmpliSeq HiFi Mix, NEBNext Fast DNA Library Prep Set	Proprietary (Not disclosed)	Phusion High Fidelity (New England Biolabs)	Q5 HotStart High-Fidelity DNA Polymerase (New England Biolabs)	SeqAmp DNA polymerase (Takara Bio)	KAPA HiFi HotStart (Roche)	PrimeSTAR HS DNA Polymerase (Takara Bio)	*Taq* DNA Polymerase (Invitrogen)
Purification method	PCR Cleanup Beads (Adaptive Biotechnologies)	AMPure XP beads	AMPure XP beads + gel migration	AMPure XP beads	AMPure XP beads	AMPure XP beads	AMPure XP beads	QIAquick PCR Purification Kit	AMPure XP beads
Amplicons length	400bp	100-200bp	520bp	670-710 bp	670-710 bp	700bp	550	670bP	620bp
Sequencing chemistry used	MiSeq v3 2x300	Ion Torrent S5 540 Chef Kit (200bp)	MiSeq v2 2x250** and v3 2x300***	MiSeq v2 2x250	MiSeq v3 2x300** and v2 2x250***	MiSeq v2 2x250** and v3 2x300***	MiSeq v2 2x150	MiSeq v2 2x250	MiSeq v2 2x250** and v3 2x300***
PhiX percentage	5%	NA	10%	15-20%	20%	10%	20%	30%	10%
Pre-processing method (Software)	Proprietary (Not disclosed)	none	Proprietary (Not disclosed)	Decombinator for UMI	MIGEC for UMI	NA	NA	NA	NA
UMI threshold (Meanning UMI cut-off (reads/UMI))	NA	NA	NA	NONE	4	NA	NA	NA	NA
Alignment/annotation tool	MiXCR (rna-seq alignment parameter)								

***TSO : Template Swithing Oligo**
**** Sequencing chemestry used in Experiment A**
***** Sequencing chemestry used in Experiment B**

Supplementary Table 1: Methods summary table. Table summarizing each protocol by method.

Dilution	Replicates	TRA							TRB								
		RACE-1	RACE-1_U	RACE-2	RACE-2_U	RACE-3	RACE-4	RACE-5	mPCR-1	mPCR-2	mPCR-3	RACE-1	RACE-1_U	RACE-2	RACE-2_U	RACE-3	RACE-4
1/1000	1	0,0369	0,0554	0,0583	0,0896	0,0376	0,0382	0,0598	0,0278	0,1479	0,2594	0,1931	0,4038	0,2315	0,3203	0,3262	0,2477
	2	0,0217	0,0802	0,0982	0,1224	0,0402	0,0543	0,0687	0,0352	0,1548	0,0897	0,2977	0,3264	0,2923	0,3140	0,2861	0,1867
	3	0,0864	0,1359	0,0000	0,0000	0,0668	0,0071	0,0861	0,0358	0,1545	0,1562	0,2437	0,3829	0,4592	0,3774	0,3021	0,2597
1/100	1	0,3600	0,3673	0,6380	0,8051	0,4493	0,8192	0,4801	0,2798	1,5019	1,2813	2,0399	2,3331	2,8327	2,9997	2,6046	2,1898
	2	0,5575	0,6650	0,4841	0,3696	0,4783	0,5336	0,6064	0,2652	1,6500	1,1181	2,2980	1,7128	2,9392	3,4259	2,9580	2,0251
	3	0,3328	0,4149	0,6372	0,7570	0,4683	0,5995	0,5498	0,2616	1,8794	1,3268	2,1735	2,2762	4,4213	4,0048	2,9174	1,9785
1/10	1	2,7366	3,1731	6,0191	8,0655	3,1161	6,1469	5,3301	3,2976	27,3554	21,1914	17,1010	12,9059	21,5958	27,0760	24,1411	17,9631
	2	5,5243	6,1062	4,7304	7,0884	4,6755	6,9603	5,2839	3,3055	29,7710	16,4809	19,1637	14,3869	31,5247	28,4652	25,0016	18,9881
	3	2,8056	3,1257	5,3871	7,3826	4,6158	6,2865	5,8529	2,6962	29,6532	16,5424	16,9448	10,8251	30,9180	28,1424	24,5969	17,4689

Supplementary Table 2: Jurkat clone percentages per method. Each value represents the percentage of the Jurkat TRA (left) and TRB (right) clonotype observed in each replicate from the 1/1000 (top), 1/100 (middle) and 1/10 (bottom) dilution condition for each method (column).

Supplementary Material and methods

- Multiplex-1

DNA was sent to Adaptive technology® (USA) for TCR amplification following their Survey protocol[1,2]. Briefly, from genomic DNA (gDNA), each TCR rearrangement is amplified using a set of 24 forward primers, each targeting one TRBV human gene, and 12 reverse TRBJ primers targeting each TRBJ gene to ensure a complete coverage of the CDR3 sequence. A 30 cycle PCR1 leads to amplicons further purified. 2µL of such purified amplicons are amplified in a second round of PCR (7 cycles) where sample barcodes are added. PCR2 amplicons are further purified. The sequencing is then carried out on a MiSeq V3 Illumina sequencer using a 2x300bp read length protocol by the provider.

- Multiplex-2

TCRB sequencing was performed using primers from the IMonitor publication[3] and a protocol adapted for Ion Torrent sequencing. Primer stocks were made by adding 2µl of each 100µM TCR-J primer to 176µl of PCR grade water and 2µl of each 100µM TCR-V Primer to 136µl of PCR grade water creating a stock solution 1µM for each primer. The PCR reactions contained 2µl Ion Ampliseq 5x HiFi PCR Mastermix (Thermo Fisher Scientific), 2µl of each primer stock and 40ng DNA with water to a final volume of 10µl. PCR cycling conditions were 99°C for 2 min, then 25 cycles of 99°C for 15 sec and 65°C for 60sec, followed by a hold at 10°C. PCR products were purified by AMPure XP beads using a 1,8x volume of beads and elution in 25µl of water. Subsequently the PCR products were end-repaired, phosphorylated and ligated to IonXpress adapters using the NEBNext Fast DNA Library Prep Set for Ion torrent (NEB CatNr E6270L) according to manufacturer's recommendations. A post-amplification was applied with 8 cycles followed by another round of AMPure cleanup. Molarity of the library was determined by Ion Library TaqMan Quantitation Kit (Thermo Fisher CatNr 4468802). The library was sequenced

on Ion Torrent S5XL using the 530 Chip Kit and the Ion 510 & Ion 520 & Ion 530 Kit – Chef (Thermo Fisher CatNr A34019) to read sequences up to 400bp length and to obtain 0.5-1.5x106 reads per sample.

- Multiplex-3

RNA has been processed with the iProfile-human TRB kit (iRepertoire®). Briefly, each TRB is reverse-transcribed using a set of 24 forward primers, each targeting one TRBV human gene, and a reverse primer located in the TRBC gene to ensure a complete coverage of the TRB sequence. PCR1 primers include barcodes to allow sample identification after the sequencing. PCR1 products are then purified on magnetic beads and a second round of PCR is performed. Amplified libraries are excised from agarose gel and purified[4]. The sequencing is then carried out on a MiSeq V2 Illumina sequencer using a 2x250bp read length protocol for the experiment A and a MiSeq V3 Illumina sequencer using a 2x300bp read length protocol for the experiment B at "Institut du Cerveau et de Moelle epinière" (ICM) to obtain 0.5-1.5x106 reads per sample. Details can be found on the iRepertoire resources document available online (https://docs.wixstatic.com/ugd/c9f231_a02ccf9aa124475e86a7c384c810c077.pdf)

- RACE-1

The alpha and beta chains of the TCR repertoire were sequenced by amplifying cDNA generated from the 10ng and 100ng RNA samples provided. Details of the method are published[5,6]. Briefly, the RNA is reverse transcribed from a TCR constant region primer. A oligonucleotide containing an anchor primer, together with a 12 base pair UMI is ligated onto the 3' end of the cDNA using the single strand T4 RNA ligase 1 (NEB, #MO204). Subsequent PCR reactions are then used to amplify the TCR cDNA (from the 3' end of the constant region until the 3' end of the RNA), and to introduce adaptors for Illumina sequencing and indices for sample multiplexing. The

sequencing is then carried out on a MiSeq V3 Illumina sequencer using a 2x300bp read length protocol for both experiment to obtain 0.5-1.5x10^6 reads per sample.

- RACE-2

TCR alpha and beta libraries were prepared on 10 or 100ng of RNA from each sample with Human TCR kit (MiLaboratory® LLC) following provider protocol. Briefly, it is a 5'RACE protocol with template switch oligo that carries 12 random "N" nucleotides (UMI). Reverse transcription starts from the primers located in the TRBC and TRAC genes, in one tube. 1st PCR amplifies cDNA with a forward primer annealing on the template switch oligo, and reverse nested primers annealing on TRBC and TRAC genes. 2nd PCR is performed separately for TCR alpha and beta chains and introduces Illumina adapters with indexing. PCR product is purified using AMPUre XP beads (Beckman-Coulter®), analyzed with real-time PCR. The sequencing is then carried out on a MiSeq V3 Illumina sequencer using a 2x300bp read length protocol for the experiment A and a MiSeq V2 Illumina sequencer using a 2x250bp read length protocol for the experiment B to obtain 0.5-1.5x10^6 reads per sample.

- RACE-3

T cell receptor (TCR) alpha/beta libraries were prepared on 10 or 100ng of RNA from each sample with SMARTer Human TCR a/b Profiling Kit (Takarabio®) following provider protocol[7]. 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 is then carried out on a

MiSeq V2 Illumina sequencer using a 2x250bp read length protocol for the experiment A and a MiSeq V3 Illumina sequencer using a 2x300bp read length protocol for the experiment B at "Institut du Cerveau et de Moelle epinière" (ICM) to obtain 0.5-1.5x10^6 reads per sample.

- RACE-4

T cell receptor genes were sequenced using a method similarly as described in Gros et al. 2014[8]. In brief, extracted RNA was denatured and cDNA was synthesized using an oligo (dT) primer and a template-switching oligo during an incubation at 42°C for 90 minutes then 70°C for 10 minutes. Following purification with AMPure XP beads (Beckman-Coulter®), cDNA was amplified using the 5PIIA primer (5'- AAGCAGTGGTATCAACGCAGAGT-3') and constant region primers TCRb (5'-TGCTTCTGATGGCTCAAACACAGCGACCT-3') or TCRa (5'-TCTCAGCTGGTACACGGCAGGGTCAGGGT-3') with the following cycling conditions: (95°C 5 min, 5 cycles of 98°C 15 sec, 72°C 1 min, 5 cycles of 98°C 15 sec, 70°C 10 sec, 72°C 1 min, 10-15 cycles of 98°C 15 sec, 68°C 10 sec, 72°C 1 min). The amplicons were purified using AMPure XP beads, amplified further to incorporate Illumina sequences. The sequencing is then carried out on a MiSeq V2 Illumina sequencer using a 2x150bp read length protocol for both experiment to obtain 0.5-1.5x10^6 reads per sample.

- RACE-5

First-strand cDNA was synthesized by 5' RACE. RNA, 1 mM dNTP and 0.125µM final of the 3' primer binding to the TCRa C region (5'-CACTGTTGCTCTTGAAGTCC-3') were denatured for 5', 65°C. A mix containing 2.5 µM final of the template switching primer (5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTrGrGrG-3'), 2mM DTT, 20 U rRNAsin (Promega®, Madison, WI), 50 mM TrisHCl (pH 8), 187 mM KCl, 3 mM MnCl2, and 0.05% Tween 20 was added to a final volume of 20µl. After preincubation (2', 42°C), 200 U SuperScript II Reverse Transcriptase (Invitrogen®, Life Technologies) were added, incubation was continued for 90',

42°C, and 15', 70°C. cDNA was purified using the MinElute PCR Purification Kit (Qiagen®). Whole cDNA was amplified over 3 rounds PCR with PrimeSTAR HS DNA Polymerase (TAKARA®, Japan), allowing the addition of barcodes and adaptors for Illumina sequencing. Primers: 1st: 5'-TCGGTGAATAGGCAGACAGA-3' and 5'-GTGACTGGAGTTCAGACGTG-3', 2nd: 5'ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGCAGGGTCAGGGTTCTGGAT-3' and 5'-CAAGCAGAAGACGGCATACGAGATindexGTGACTGGAGTTCAGAC-3' and 3rd: 5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACAC-3' and 5'-CAAGCAGAAGACGGCATACGAGATindexGTGACTGGAGTTCAGAC-3'. Amplification was for 15, 12, and 25 cycles with the following conditions: 1st: 50µl total (98°C 10min, 56°C 7min, 72°C 1h50min), 2nd: 1µl in 20 µl total (98°C 10min, 49*°C 7min, 72°C 1h50min) (asterisk [*] represents 2°C increment per cycle starting at 49°C), and 3rd: 1µl 1:100 in 50µl total (98°C 10min, 58°C 7min, 72°C 1h50 min). The final product was purified using AMPUre XP beads (Beckman-Coulter®). The sequencing is then carried out on a MiSeq V2 Illumina sequencer using a 2x250bp read length protocol for both experiment to obtain 0.5-1.5x10^6 reads per sample.

- RACE-6

This method of amplification was performed using the 5'RACE kit, Version 2.0 (Invitrogen®) according to the supplier's recommendations. Briefly, the first strand cDNA was synthesized using a constant region-specific TRBC-RT primer (5'-CACGTGGTCGGGGWAGAAGC-3'). RT and PCR amplification were performed using manufacturer recommendations on 10 or 100ng RNA. Libraries were purified by using AMPUre XP beads (Beckman-Coulter®). The sequencing is then carried out on a MiSeq V2 Illumina sequencer using a 2x250bp read length protocol for the experiment A and a MiSeq V3 Illumina sequencer using a 2x300bp read length protocol for the

experiment B at “Institut du Cerveau et de Moelle epinière” (ICM) to obtain 0.5-1.5x10^6 reads per sample.

Supplementary references

1. Carlson, C. S. *et al.* Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* **4**, (2013).
2. Robins, H. S. *et al.* Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* **114**, 4099–107 (2009).
3. Zhang, W. *et al.* IMonitor: A Robust Pipeline for TCR and BCR Repertoire Analysis. *Genetics* **201**, 459–472 (2015).
4. Wang, C. *et al.* High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Natl. Acad. Sci.* **107**, 1518–1523 (2010).
5. Oakes, T. *et al.* Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile. *Front. Immunol.* **8**, (2017).
6. Uddin, I. *et al.* An Economical, Quantitative, and Robust Protocol for High-Throughput T Cell Receptor Sequencing from Tumor or Blood. in *Cancer Immunosurveillance: Methods and Protocols* (eds. López-Soto, A. & Folgueras, A. R.) 15–42 (Springer New York, 2019). doi:10.1007/978-1-4939-8885-3_2.
7. Taylor, S., Yasuyama, N. & Farmer, A. A SMARTer approach to profiling the human T-cell receptor repertoire. *J. Immunol.* **196**, 209.5-209.5 (2016).
8. Gros, A. *et al.* PD-1 identifies the patient-specific $CD8^+$ tumor-reactive repertoire infiltrating human tumors. *J. Clin. Invest.* **124**, 2246–2259 (2014).

III.2. Article 2 - « Autoimmune disease-specific TCR repertoires in peripheral blood T cells »

Les maladies auto-immunes (MAI) regroupent un large éventail de maladies caractérisées par un déséquilibre de la balance d'immuno-régulation entre les cellules $CD4^+$ Teffs et $CD4^+$ Tregs. Ces populations lymphocytaires expriment à leur surface un TCR qui détermine la spécificité et l'activation de chaque cellule. Dans le contexte des MAI tels que le diabète de type 1 (DT1) et la polyarthrite rhumatoïde (PR), on observe une atteinte tissulaire spécifique. Les caractéristiques des variations de la composition du répertoire TCR et leur signification clinique restent encore mal connues, en particulier pour les cellules $CD4^+$ Tregs, principalement impliquées dans les troubles auto-immuns.

Dans le cadre de cette étude, nous souhaitons évaluer si, au niveau du sang, il est possible d'identifier des cellules T associées au pronostic de ces MAI. Dans ce but, nous avons collecté le sang périphérique de 54 volontaires sains (VS), 68 patients atteints de PR et 52 patients atteints de DT1. Le répertoire des chaines α et β du TCR de chaque sous-type cellulaire ($CD4^+$ Tregs et $CD4^+$ Teffs) de chaque sujet de l'étude ont été amplifié en utilisant la méthodologie décrite en matériel et méthode et validée dans l'étude comparative (section III.1).

Nous avons utilisé pour la première fois dans l'étude du répertoire TCR une approche statistique adaptée aux données dispersées pour identifier les clones TCR associés aux maladies, à savoir la *sparse Partial Least Squares Discriminant Analysis* (sPLSDA). Les signatures TCR associées à chaque sous-type cellulaire de chaque MAI ont été identifiées et utilisées pour la classification des maladies en comparaison aux VS, ce qui a donné une fiabilité de prédiction comprise entre 81.1%% et 96,6%. Les clones associés aux maladies ont montré des caractéristiques différentes entre les deux MAI. Les signatures significativement associées aux pathologies ont été validées sur des ensembles de données externes.

Nous avons réalisé l'analyse la plus complète et contrôlée des répertoires TCR$\alpha\beta$ des cellules $CD4^+$ Tregs démontrant que dans des pathologies chroniques, le sang périphérique, contient une information à l'échelle du TCR, pertinente et permettant la classification des patients entre eux et par rapports à des VS. Ces biomarqueurs

caractéristiques des pathologies pourront être utiles pour appuyer un diagnostic parfois compliqué et incertain, mais aussi pour mieux comprendre l'hétérogénéité entre les patients, notamment en ce qui concerne la réponse aux traitements.

Autoimmune disease-specific TCR repertoires in peripheral blood T cells

---

Pierre Barennes, Paul Stys, Roberta Lorenzon, Michèle Barbié-Sastre, Nicolas Coatnoan, Hélène Vantomme, Kenz Le Gouge, Michelle Rosenzwagj, Encarnita Mariotti-Ferrandiz and David Klatzmann

---

Autoimmune disease-specific TCR repertoires in peripheral blood T cells

Authors:

Pierre Barennes[1,2,*] , Paul Stys[1,*], Roberta Lorenzon[2], Michèle Barbié-Sastre[2], Nicolas Coatnoan[2], Hélène Vantomme[1,2], Kenz Le Gouge[1], Michelle Rosenzwagj[1,2], Encarnita Mariotti-Ferrandiz[1,2,$], David Klatzmann[1,2,$,#]

Affiliations:

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

[2] AP-HP, Hôpital Pitié-Salpêtrière, Biotherapy (CIC-BTi) and Inflammation-Immunopathology-Biotherapy Department (i2B), Paris, France.

* and $ equally contributed to the study

[#]corresponding author: David Klatzmann

Hôpital Pitié-Salpêtrière, 83 bd de l'Hôpital, 75651 Paris, France.

Phone:+33142177461

Email: david.klatzmann@sorbonne-universite.fr

Abstract

T cells play a major role in autoimmune disorders, in particular the effector (Teff) and regulatory (Treg) $CD4^+$ cell subpopulations. The study of the diversity and specificity of their T-cell receptor (TCR) is necessary to better understand the mechanisms underlying autoimmune diseases (AIDs). For this purpose, we analyzed the TCR$\alpha\beta$ repertoire of type 1 diabetes and rheumatoid arthritis patients in comparison with healthy volunteers. We identified signatures of TCRs of each AID, in each of the Teffs and Tregs subpopulations compared to healthy volunteers, allowing us to classify patients according to their pathophysiological state. We have identified clusters of convergent and non-convergent signatures depending on the AIDs and the cell subtype. This work supports the interest of studying the TCR repertoire for the identification of biomarkers in the context of immune diseases and paves the way for future clinical use to improve diagnosis and identify new therapeutic targets.

INTRODUCTION

T cells represent the core population of the antigen-specific cellular immune response. This specificity is provided by their T cell receptors (TCR) expressed on the surface of each T cell allowing their activation and proliferation by the specific recognition of antigens. The TCR is composed of two chains, the α and the β chain, each resulting from a somatic recombination between a variable gene (V), a diversity gene (D, only for the β chain) and a junctional gene (J) (Chien et al., 1984; Davis and Bjorkman, 1988). The recombination process generates a junctional highly diverse region, namely the complementarity determining region 3 (CDR3), involved in the specificity of TCR binding to peptide MHC complexes (pMHC) through direct peptide recognition (Borg et al., 2005; Garcia et al., 1999). Altogether, this evolutionary selected machinery produces a functional and highly diverse TCR repertoire (Chothia et al., 1988; Dupic et al., 2019) reaching up to 10^{19} TCRs in humans (Bradley and Thomas, 2019). This massive diversity of receptors can be studied by adaptive immune receptor repertoire sequencing (AIRR-seq) based on high-throughput sequencing (HTS) and applying several modelling strategies, including diversity metrics, clonal tracking, repertoire overlap and sequence similarity network (Miho et al., 2018; Six et al., 2013; Valkiers et al., 2022). More recently, searching for sets of TCRs, also named signatures, defining a pathological condition has been proposed in infectious (Emerson et al., 2017) or autoimmune (Liu et al., 2019) contexts. However, those studies analyzed the total TCR, disregarding the T-cell subset origin, such as CD4 vs CD8, but also effector vs regulatory T-cells and therefore ignoring the functional relevance of such TCRs.

This is particularly important when studying autoimmune diseases (AIDs) since it is well established that a qualitative and/or a quantitative regulatory T-cell (Treg) insufficiency leads to AIDs (Bonelli et al., 2008; Boyer et al., 2004; Kukreja et al., 2002; Lindley et al., 2005; Noack and Miossec, 2014). Indeed, previous studies have shown that reversing the Treg defect using targeted immunotherapies such as low-dose IL-2 improve the clinical condition of patients with AID, such as type 1 diabetes (T1D) or systemic lupus erythematosus (SLE) as well as in other indications, including graft versus host disease or Alzheimer (Alves et al., 2017; He et al., 2020, 2016; Humrich and Riemekasten, 2019; Rosenzwajg et al., 2019, 2015; Saadoun et al., 2011; von Spee-Mayer et al., 2016). These results clearly suggest the importance of Tregs in AID control. Yet, the TCR repertoire of such cells remains poorly understood. In a recent study, Liu et al showed differences in terms of TCRβ diversity of the total T-cell TCR repertoire between rheumatoid arthritis (RA) and SLE patients compared with healthy volunteers (HV). However, as they analysed the total TCR repertoire, without separating the different T-cells subsets, it is impossible to correlate the observations with the cell function and therefore the role of such repertoire alterations.

In this study, we aimed at identifying Treg TCR signature that can classify patients according to their disease. For that, we analyzed a TCR$\alpha\beta$ repertoire from $CD4^+$ T effector (Teffs) and Tregs separated from the peripheral blood of T1D and RA patients recruited as part of the Transimmunom trial (NCT02466217). The TCR$\alpha\beta$ repertoire has been obtained with a previously validated and established method (Barennes et al., 2021). Using an original method based on machine learning, including sparse Partial Least Square Discriminant Analysis (sPLS-DA) for feature selection as well as leave-one-out (LOO) for patient sampling and classification, we identified Teff and Treg TCR$\alpha\beta$ signatures that classify T1D, RA and HV with a high accuracy. Furthermore, we found that the Treg signature was of particular biological relevance in RA, but not in T1D, by (i) comparing the performance of the Treg signature classification to that of a random selection of TCRs and (ii) identifying our TCR signature in a validation cohort.

RESULTS

TCR repertoire from sorted cell subsets is a robust disease classifier

From the observational Transimmunom trial, we recruited 54 HV, 52 T1D and 68 RA patients to study their TCR repertoire by AIRR-seq (***Figure 1a ;*** **Supplementary Fig. 1**). $CD4^+$ Teffs and $CD4^+$ Tregs were sorted from the whole blood. AIRR-Seq using a validated and robust experimental protocol has been performed to obtain both TRA and TRB repertoires (***Figure 1b***). After alignment and pre-processing steps that include quality control and singletons removal (see material and method), we totally obtained 180 million sequences, among which 4.6 million TRA and 7.2 million TRB unique clonotypes (**Supplementary Table 1**).

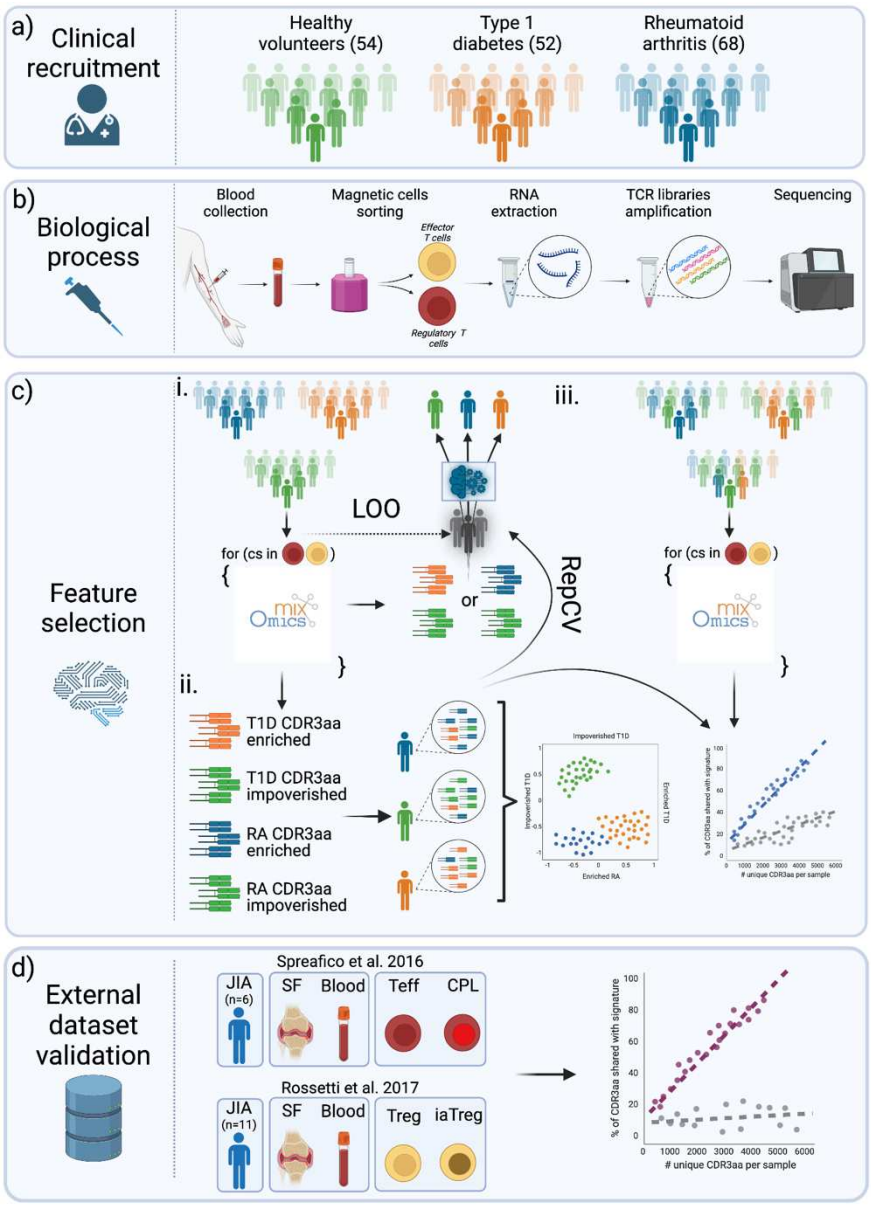


Figure 1 : Experimental and statistical study design. a*. Clinical recruitment. From the observational Transimmunom trial, we recruited 54 HV (green), 52 T1D (orange) and 68 RA (blue).* ***b****. Biological process. From the patients' blood we sorted CD4 Teffs (red cells) and CD4 Tregs (yellow cells). We extracted the RNA from these cells and performed the TCR libraries creation and sequencing.* ***c****. Feature selection. (i) Signatures classification by Leave One Out (LOO) resampling. (ii) Final signatures clustering power and classification power by Repeated Cross-Validation (RepCV) resampling. (iii) Sharing strength compared to permutation signatures.* ***d****. External dataset validation. Signatures sharing comparison with the TCRβ of published datasets. Each dataset is presented according to the number of patients with the pathology, the origin of the samples and the cell type. An analysis against a collection of databases is also performed (not shown on this figure).*

To identify TCR signatures of each AIDs, we applied supervised analysis on CDR3 sequences to extract discriminant features followed by validation using machine learning (***Figure 1c***). More precisely, we used the sparse Partial Least Squares Discriminant Analysis (sPLSDA) method combined with a Leave One Out (LOO) resampling method (**Figure 1c, i**). Briefly, for each patient group and cell subset TRA and TRB repertoires, we set aside one subject (patient or HV), as a test sample, and applied the sPLSDA method between the

disease group (T1D or RA) and the HV group, on the training dataset to identify differentially expressed CDR3aa. We then assessed the classification power of these CDR3aa on each sample set aside. To validate the approach, we performed a random selection of CDR3aa and the classification. The features selected by the sPLSDA approach allowed us to obtain a good power of classification, higher than 80% for all groups, chains and subpopulations (min: 81.1% T1D TRB Treg and max: 96.6% RA TRB Treg, **Supplementary Fig. 2a**). This classification is significantly higher than that obtained following a random selection of public CDR3aa, showing the benefit provided by the feature selection approach (**Figure 2a**). Nevertheless, the classification with CDR3aa randomly drawn in the dataset is higher than a “random” classification (expected: 50%), ranging from 50% to 75% good classification. This indicates that the TCR repertoire contains, among the 1000 randomly selected public CDR3aa, enough information to allow for a "correct" classification in some cases. We then compared the number of different CDR3aa selected across all LOO iterations and found a stronger selection homogeneity with the sPLSDA approach (about 2000 unique CDR3aa out of >100,000 potentials) compared to the random CDR3aa (>100,000 unique CDR3aa; **Figure 2b**). Similar observations were drawn when looking at the details of the classification per patient, which also showed the homogeneity of patient classification with the sPLS-DA approach as compared with the random selection (**Figure 2c**).

In order to define the final signatures, we used the medians of the variable importance (VIs) computed across all LOO iterations and selected, for each TRA and TRB chain, the 10% of CDR3aa with the highest VI median (**Supplementary Fig. 2b**). This strategy allowed selecting the most differentially shared CDR3aa across all LOO iterations and therefore to keep only the most robust CDR3aa (**Supplementary Fig. 2c**). This resulted in a list of enriched and impoverished CDR3aa for each cell subset and disease. To evaluate the classification robustness of our final signature, we applied a larger sampling approach, using repeated cross validation (RepCV) (10 folds). To gain insights into the contribution of the different parameters studied, we performed Rep CV considering each chain, each signature type (enriched vs. impoverished) and each cell population (Treg and Teff) separately as well as the effect of joint signatures (e.g. TRA & TRB; **Figure 2d**). To do so, we evaluated the average F1 score (a

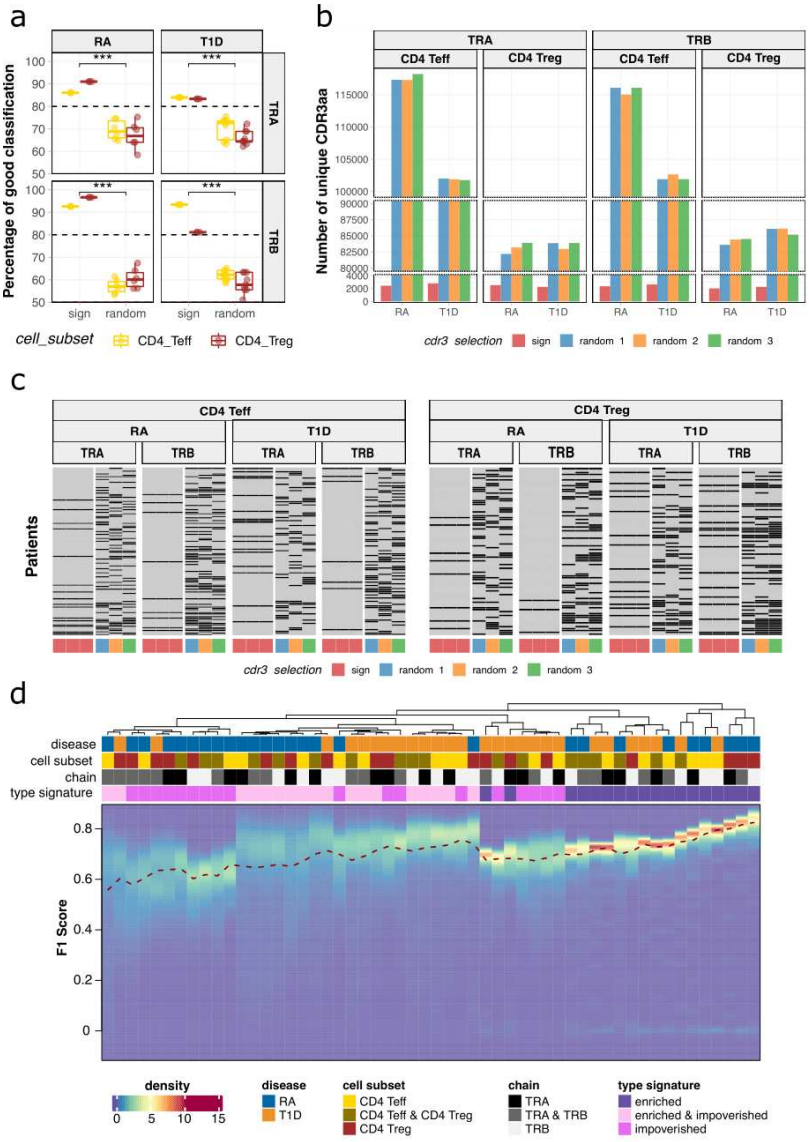


Figure 2 : Classification power of Teff and Treg CDR3aa. a*. Percentage of good classification. Each boxplot is the mean of classification results over all the LOO by disease (RA on left and T1D on right), cell subset (CD4 Treg in brown and CD4 Teff in yellow) and chain (TRA on top and TRB on bottom) over the HVs for sPLSDA CDR3aa (left in the frame) and selection random CDR3aa selection (right in the frame).* ***b****. Number of unique CDR3aa by "signature". Barplot of the distinct number of CDR3aa selected over all the LOO iterations by disease, cell subset and chain for 3 independent sPLSDA CDR3aa selection (red) and 3 random CDR3aa selections (blue, orange, green).* ***c****. Heatmap of the classification results. Each row is a patient colored by the classification results (red: bad classification, gray: good classification) over all the LOO iterations by disease, cell subset and chain for 3 independent sPLSDA CDR3aa selection (red) and 3 random CDR3aa selections (blue, orange, green).* ***d****. Densitymap of the F1 score for the RepCV (10 folds) for each signature combination (disease, cell subset, chain and type of signature). The mean is represented by a red dashed line. Hierarchical clustering is performed using Euclidean distance and complete method. Wilcoxon test was performed: P > 0.05 is NS; *P ≤ 0.05; **P ≤ 0.01; ****P ≤ 0.0001.*

metric for evaluating the performance of classification models) across all RepCVs. The results are summarized in the heatmap (Figure 2d). We observed that the classification power is very different depending on the considered signature. Our results showed that enriched RA Treg signatures, regardless of the considered chain, display a good median F1 score > 0.8. For all the other conditions, the prediction have a F1 score <0,8. Impoverished TCR signatures being

the worst (F1<0.7). We also found that combining chains, subpopulations and/or signature types did not increase the prediction power. We therefore focused on enriched signatures.

T1D and RA TCR signatures are cell-subset and disease specific

The strategy used allowed us to obtain signatures of comparable size between groups (**Figure 3a**). We then aimed at characterizing these signatures. First, we found that TRB signatures for both subsets and patient groups were significantly shorter as compared with the corresponding full datasets (**Figure 3b**), typical of public clonotypes. We then evaluated the sharing between RA and T1D signatures, as well as between cell subsets and found that most of the CDR3aa forming each signature are unique to the condition they were identified. Few

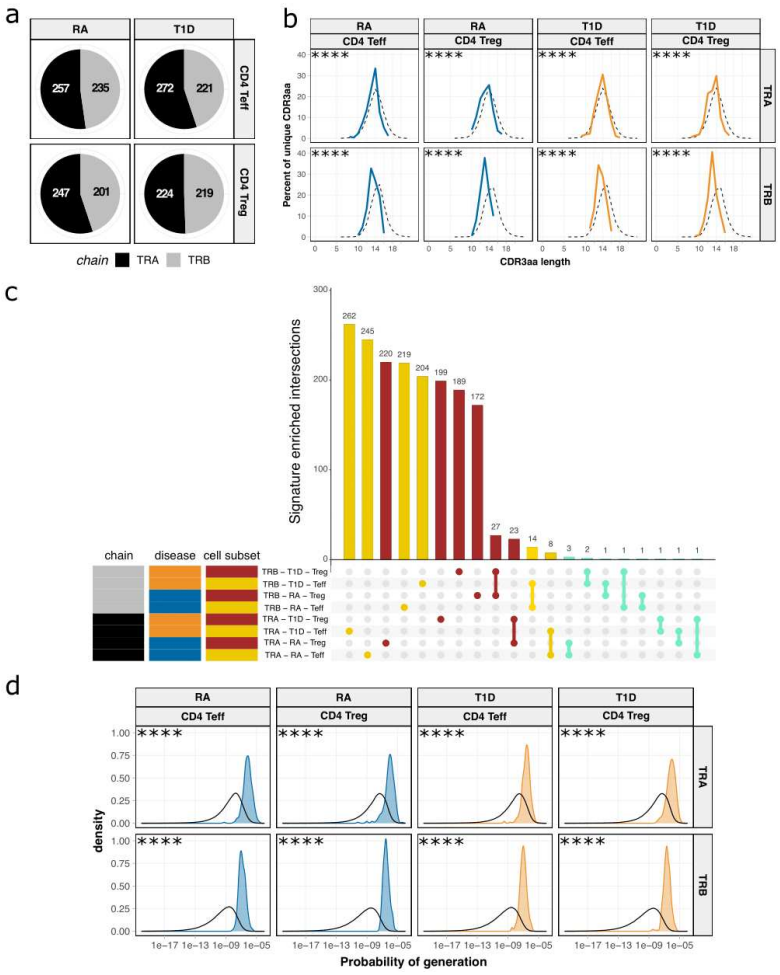


***Figure 3 : Signature description. a**. Signature size distribution. Pie-chart represent by chain (TRA: black and TRB: grey) the signatures size, for RA patients (left part) and T1D patients (right part); CD4 Teff (top part) and CD4 Treg (bottom). **b**. Signature CDR3 length distribution. Distribution of the unique CDR3aa by length for enriched signatures from RA patients (blue lines) and T1D patients (orange lines) compared with the CDR3aa length of the full dataset per condition (dashed lines); TRA chain (top) and TRB chain (bottom). **c**. Signatures sharing. Number of identical CDR3aa sequences between the different signatures for the alpha chain (left) and the beta chain (right). **d**. Signatures Pgen. Probability of generation (Pgen) of CDR3aa sequences of each signature, enriched (purple) vs. impoverished (pink) compared to the theoretical distribution (black). T-test performed on the mean of the distributions. The student t-test was used to compare the mean of the distributions. P > 0.05 is NS; *P ≤ 0.05; **P ≤ 0.01; ****P ≤ 0.0001.*

(from 14 to 27, representing at most 10% of the original signatures) were shared between diseases, yet belonging to the same cell type (Treg or Teff). Very few CDR3s (at most 3) were shared between Treg and Teff from the same disease or between diseases (**Figure 3c**). Finally, we observed that the signatures were characterized by a higher probability of generation (Pgen) as compared with the full corresponding datasets, further supporting the publicness of the CDR3s found, yet being enriched in a cell-subset and disease specific manner (**Figure 3d**).

The signatures discriminate patients but are not necessarily associated with the disease

In order to evaluate the clustering power of these signatures, we analyzed the presence of each CDR3 from each signature in the different datasets and applied hierarchical clustering, on signatures from Teff and Treg separately (**Figure 4a**). For Teff signatures, we identified 4 sub-clusters (named Te1 to Te4) with an overall hierarchical purity of 0.79. The Te1 cluster is composed of 70% of HV patients, the Te2 of 70% of the RA, the Te3 cluster of 91% of RA and the last Te4 cluster is composed of 90% of T1D (**Figure 4a, left**). For Tregs, we identified also 4 sub-clusters (Tr1 to Tr4) grouping each disease with an overall hierarchical purity of 0.78. The Tr1 and Tr2 clusters are composed of 61% of HV, 20% of RA and 19% of T1D, the Tr3 cluster of 94% of RA patients and the Tr4 cluster is mostly composed by T1D patient (84%) (**Figure 4a, right**). These results revealed that the RA and T1D enriched TCR signatures initially generated in comparison with HV were partially overlapping between datasets, yet with capable of clustering patients with two different AIDs.

This is particularly true for the CDR3aa obtained from the CD4 Treg enriched signature in RA patients, for which we observed a slope of the linear models of partitioning clearly superior to the other signatures (**Supplementary Fig. 3a**). By carefully analyzing the samples from misclustered patients (not belonging to the disease cluster), we found that they were mostly samples for which the number of unique CDR3aa is low. This was not true for T1D misclustered patients in the Te1 cluster (**Supplementary Fig. 3b**).

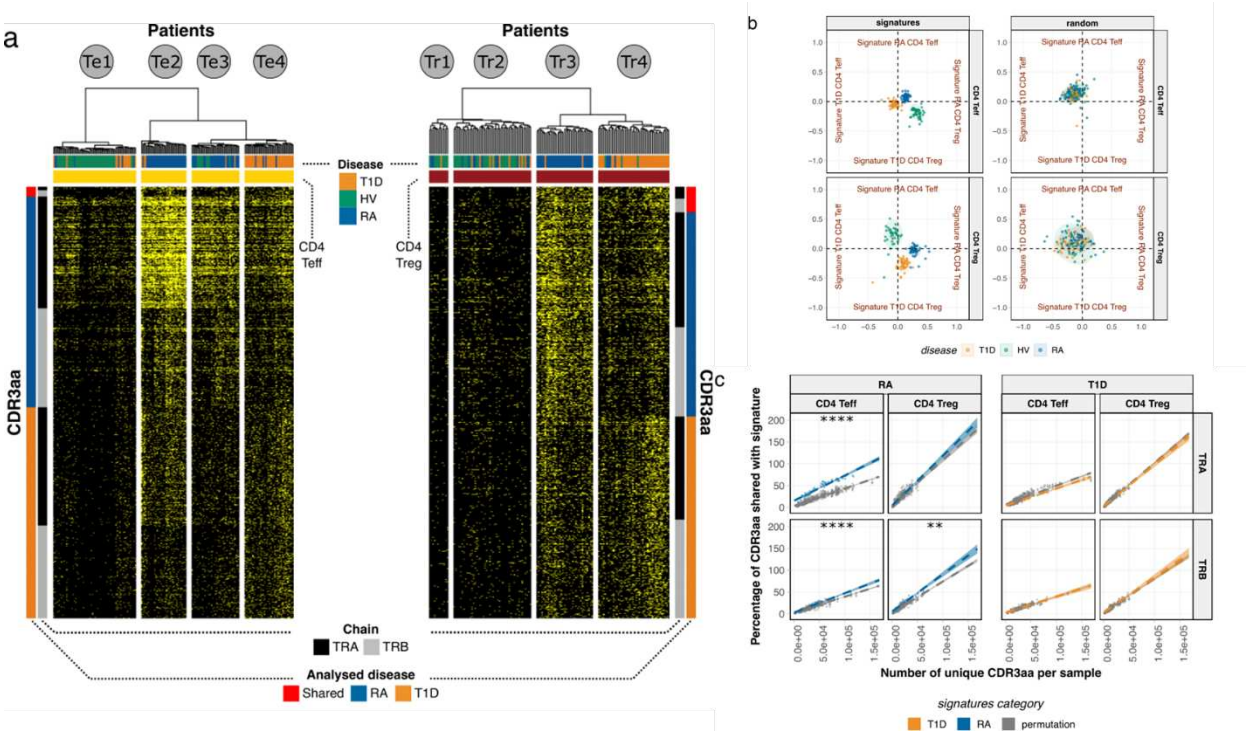


Figure 4 : AIDs signatures descriptions. a*. Signatures' clustering power. Heatmaps represent each CDR3aa (x axis) presen (in yellow) and absent (black) from the Teff (left) and Treg (right) signatures in HV (green), T1D (orange) and RA (blue) patients (y axis). TRA (black) and TRB (gray) CDR3s are included. CDR3aa shared between groups are represented in red.* ***b****. Signatures' proportion separation. The Radviz representation shows the position of each patient according to the cumulative frequency of clones expressing CDR3aa of each signature (Signatures on the left and a random signature on the right), for Teffs (top) and Tregs (bottom). T1D patients are in orange, RAs in blue and HVs in green.* ***c****. Signatures' sharing by patient. Scatterplot comparing the distribution of the number of disease-associated CDR3aa to the total number of unique CDR3aa sampled for disease (RA: blue, T1D: orange) and permutation signatures (gray). Each point represents a patient depending of the signature origin. Regression linear models were analyzed using Anova with Holm multiple-testing correction. P > 0.05 is NS; *P ≤ 0.05; **P ≤ 0.01; ****P ≤ 0.0001.*

To better understand the classification power, we evaluated the clustering power using the cumulative frequency of all the clonotypes from each signature. The Radviz plot showed a good separation between patients compared to the random signature (randomly selected CDR3aa) for CD4 Teff and especially for CD4 Treg (**Figure 4b**). This clustering for both chain and cell-subset informed us that the sum of the clonotypes frequencies of each signature is a perfect index to cluster patients by physio-pathological state.

To exclude the possibility that the signature power was due to a group-effect, we applied a permutation strategy. (**Figure 1c, iii**). Briefly, we performed 7 iterations of random permutations between samples from patients (RA or T1D) and from HVs. Permutated groups being then composed of patients and HV. For each permutation, we selected a "permutation signature" and evaluated the percentage of signature sharing by patients and compared to the sharing with the non-permutated signature. By comparing the linear coefficient of the models,

we observed significantly higher coefficient for the enriched (non-permutated) signatures from RA CD4 Teff TRA and TRB as well as from RA CD4 Treg TRB as compared to the permutated signatures (**Figure 4c, left**). This was not true for T1D derived signatures, suggesting that a signature identification requires solid validation using various controls. Altogether, these results revealed that using sPLSDA, we can identify TCR signatures that define clinical status. This is true in our dataset for TCR signatures derived from RA Teff and Treg but not from T1D samples.

RA Treg enriched signatures is composed of clones involved in the inflammation site

Finally, to assess whether the CDR3aa contained in the TCR RA signatures originate from T cells involved in the immune response, we evaluated their enrichment using external data sets (**Figure 1d**). For the following steps, we will focus only on the significant signatures from RA (**Figure 4c**) and only on the TRB chain due to the lack of TRA sequences in publically available datasets. In particular, we used the datasets from Spreafico and coll. and Rossetti and coll. (Spreafico et al., 2016; Rossetti et al., 2017) who analyzed the TRB repertoires of respectively CD4 Teff and CD4 Treg cells from blood and synovial fluid of patients with juvenile idiopathic arthritis (JIA).

We evaluated the sharing of our CD4 Teff TRB signature of our RA patients with the TRB repertoire of the CD4 Teffs from the Spreafico dataset (**Figure 5a**). We observed a significantly higher sharing of our enriched signature RA TRB Teff signature with the CD4 Teff from JIA patients' blood compared to random CDR3aa but also with the impoverished CD4 Teff signature (**Figure 5a, left**). A similar trend, yet not statistically significant was found when we assessed the sharing with TRB repertoires from Circulating Pathogenic-like Lymphocyte (CPL) from blood (**Figure 5a, center**). When we compared with the TRB repertoire from the synovium, we found a significant enrichment of our signature as compared with the random signature (**Figure 5a, right**). Similarly, we evaluated the sharing of our RA CD4 Treg TRB signature with the TRB repertoire of CD4 Treg from Rossetti's dataset (**Figure 5b**). We

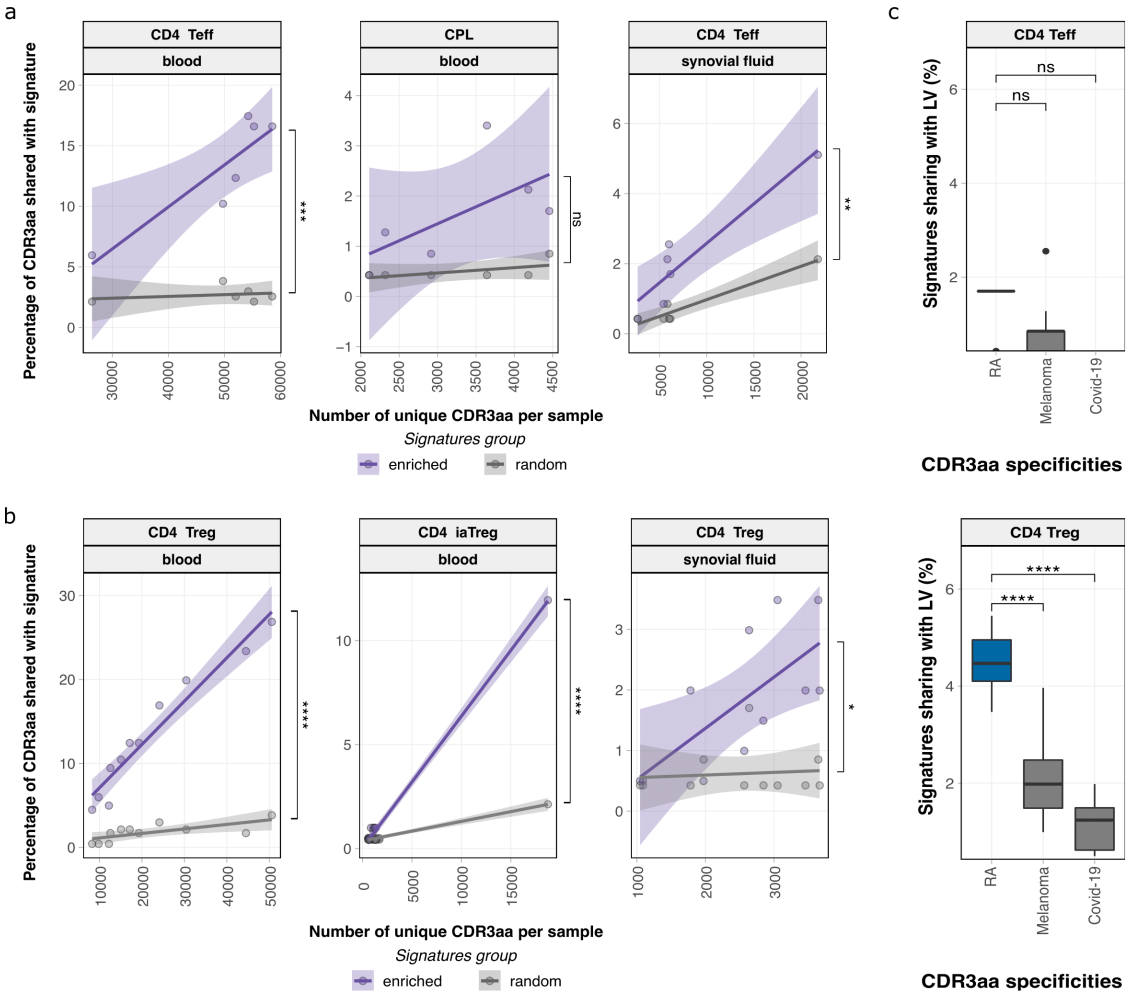


Figure 5 : RA enriched signatures are involved in the pathology. a*. CD4 Teff RA Signatures shared by JIA Teff patients from Spreafico et al. 2016. Scatter plot comparing the distribution of the number of CDR3aa associated with the CD4 Teff signatures of RA patients to the total number of unique CDR3aa per sample from the Spreafico dataset (Enriched: purple and random CDR3aa in grey). Each point represents a patient and the origin of the samples is presented by the facet. The comparison with CD4 Teff from blood is on the left, with CPL from blood in the center and with CD4 Teff from synovial fluid on the right.* ***b****. CD4 Treg RA Signatures shared by JIA Treg patients from Rossetti et al. 2017. Scatter plot comparing the distribution of the number of CDR3aa associated with the CD4 Treg signatures of RA patients to the total number of unique CDR3aa per sample from the Rossetti dataset (Enriched: purple and random CDR3aa in grey). Each point represents a patient and the origin of the samples is presented by the facet. The comparison with CD4 Treg from blood is on the left, with iaTreg from blood in the center and with CD4 Treg from synovial fluid on the right.* ***c****. Boxplots show the proportion of RA-, melanoma-, and Covid-19- associated CDR3aa TRB found in RA CD4 Teff and CD4 Treg signatures. CDR3s with one of these 3 specificities (RA, Melanoma, and Covid-19) were connected by LD = 1, allowing the assignment of this specificity to CDR3aa from the RA CD4 Teff and CD4 Treg CDR3 signatures. Regression linear models were analyzed using Anova with Holm multiple-testing correction. Wilcoxon test was performed with the RA group as reference: P > 0.05 is NS; *P ≤ 0.05; **P ≤ 0.01; ****P ≤ 0.0001.*

observed a significantly higher sharing of our enriched signature with CD4 Treg from the blood of JIA patients compared to random CDR3aa (**Figure 5b, left**). This is also the case when looking at inflammation-associated Treg (iaTreg) from blood (**Figure 5b, center**). Finally, the comparison with Treg from the site of inflammation showed a significant difference with random CDR3aa (**Figure 5b, right**). Altogether, these results suggest that the CD4 Teff and Treg TCR

blood signature found in our RA patients contains public CDR3s that might be enriched in the inflammation site.

It seems that these enriched RA signatures are associated with the RA pathology. To support these observations, we compared the sharing between RA signatures and 10 CDR3aa random selection iterations (n=30) among TRB CDR3aa associated with different specificities (RA, Melanoma and Covid-19). Given the low number of CDR3aa in each iteration, we assigned the specificity between sequences with known specificity and the CDR3aa from our signatures using the Levenshtein distance (Dash et al., 2017; Klinger et al., 2015; Madi et al., 2017; **Figure 5c**). This analysis revealed a significant enrichment of the CDR3aa of the RA Treg signature with published CDR3aa's associated with RA.

DISCUSSION

AIRR-sequencing is now believed to provide not only insights in pathophysiology of immune related diseases, but also to become a source of biomarkers or novel therapeutic targets (Aran et al., 2022; Liu et al., 2019; Nakayama and Michels, 2021; Wang et al., 2022). Yet, to achieved such challenge, it is important to have access to biologically relevant datasets and to propose validated approach. This is of particular importance for autoimmune diseases. Indeed, such diseases are highly heterogeneous, with patients showing various outcomes and response to treatment that could be link to different T-cell subsets. Machine learning has shown to provide such power, yet so far none of the proposed methods were fully validated.

In this study, we developed a method to identify TCR signatures of AIDs based on sPLSDA, a simple to use, efficient and reproducible approach offering a predictive power of the pathophysiological state of patients with AIDs. Importantly, we designed our study so that we could applied our strategy on TCR repertoire obtained by HTS from CD4 Teff and CD4 Treg cells, two critical T-cell subsets in AIDs. The patient classification of our TCR signatures was significantly higher as to compare with a random selection of CDR3aa. This first control was important as it furthermore showed that our feature selection process identified very homogenously CDR3aa. When we further focused on CDR3aa that significantly contributed to the prediction power based on the variable importance, we observed a significant decrease in classification for some signatures, except for enriched Treg TCRs from RA patients, highlighting the interest of sorting Treg cells in the context of AIDs. To confirm the reliability of our signatures, we built permutation groups of patients and HV. While our Treg and Teff signatures from RA patients was always significantly more enriched in RA patients compared with a patient/HV permutated signature. This was not the case for the T1D derived signatures. These results again highlight the need for sophisticated controls to validate such approach.

The absence of robustness of the T1D derived signature suggest that the CDR3s identified were not related to the pathology unlike the one from RA patients. This may be due to the difference in pathogenesis development between these two AIDs. Indeed, adult T1D is characterized by an early and silent initiation. When diagnosed in adult, the pancreas is largely destroyed and patient disease is controlled with insulin administration. In contrast, RA is an active disease, characterized with flare and remission states, suggestive of a partially active immune response. These differences may explain the different power of classification of TCR repertoire in these two indications.

Finally, using external datasets, we validated our CD4 Teff and CD4 Treg signatures from the RA patients. Indeed, we found that our signatures were enriched in the blood of JIA patients but most importantly in the synovial tissue. This was true for both Teff and Treg derived signatures. This is, to our knowledge, the first demonstration in AID that a blood TCR signature is also, at least partially, identified in the inflamed tissue. Further analyses will be necessary to confirm this observation, such as the single cell immunoprofiling of the TCR$\alpha\beta$ from the synovium. Nevertheless, using similarity sequence with annotated TCRs for their antigen specificity stored in public databases, we found that our RA signatures were enriched for TCR identified as specific for RA. These results again highlight the reliability of our method, to identify biologically relevant TCR signatures.

Altogether, we propose here an innovative and reliable method to identify TCR signature of diagnostic value that could be further used to follow-up patient response to treatment or disease progression. This method highlighted a TCR$\alpha\beta$ signature that is shared with an external validation cohort and enriched in CDR3s with known specificities for RA derived antigens.

MATERIAL AND METHODS

Patients' recruitment and cell isolation

Patients were recruited within the Transimmunom observational trial (NCT02466217, Lorenzon et al., 2018). More precisely, we collected 30ml of blood from 54 Healthy Volonteers, 52 Type 1 Diabetes (T1D) and 68 rheumatoid arthritis (RA). Peripheral blood mononuclear cells (PBMC) are isolated by density gradient (Ficoll) and enrichment for $CD4^+$ T cells using EasySep^{TM} magnetic beads, Tregs ($CD4^+$, $CD25^+$) and Teff ($CD4^+$, $CD25^-$) were separate based on CD25 magnetic beads. Purity was performed by cytometry using CD4 and FoxP3 antibody and sample were validated with purity > 80%.

Isolation of and extraction of RNA

RNA was isolated from each cell-subset unsing the RNAqueous-Kit (Invitrogen®) extraction kit, according to the manufacturer's protocol. The RNA concentration and integrity were determined on Tapestation (Agilent®).

TCR repertoire library preparation and sequencing

TCR repertoire library preparation and sequencing were performed. Briefly, T cell receptor (TCR) α and β libraries were prepared on 100 ng for Teff and 10ng for Treg of RNA from each sample with the SMARTer Human TCR a/b Profiling Kit (Takarabio®) following the provider's protocol, has previously described (Barennes et al., 2020). TCR libraries were quantified and their purity was checked using a Tapestation (Agilent®). Sequencing was performed with Hiseq 2500 (Illumina®) SR-300 protocols plus 10% PhiX by LIGAN-PM Genomics platform (Lille, France).

TCR deep sequencing data processing

Quality control using fastQC (Andrews, 2015) were performed and TCR sequences annotation for TRA and TRB using MiXCR software (Bolotin et al., 2015) (v3.0.13) with RNA-sequencing default parameters (-p rna-seq -s hsa) as available online and based on IMGT references (https://www.imgt.org). MiXCR extracts TRA and TRB repertoires, providing correction of PCR and sequencing errors.

Data Analysis

Statistical comparisons and multivariate analyses were performed using R software version 4.1.1 (www.r-project.org).

Machine learning

The feature selection was performed using sPLSDA algorithm from the MixOmics R package (Rohart et al., 2017). Briefly, to create the final signatures of CDR3aa, we performed LOO re-sampling strategy. For each LOO, we selected the 1,000 most important variables per chain of the public CDR3aa. Finally, after all the iterations, we computed the mean variable importance for each CDR3aa and selected the 10% most robust of them. Classifications were performed using the R package mlr (Bischl et al., 2016) with Leave-one-out (LOO) and Repeated Cross-Validation (RepCV) resampling and the "glmnet" learner was used with the only adjusted setting, the threshold setting.

Database

Human CDR3aa β associated with RA (n = 51) were extracted from Liu's dataset of 206 RA patients (Liu et al., 2019). These RA-specific clones were significantly associated with multiple

clinical factors (Disease activity score 28; Erythrocyte sedimentation rate; Antinuclear antibody; Immunoglobulin G). To select specific CDR3aa β controls, Melanoma (n=113) and Covid-19 (n=63) associated sequences were collected from the McPAS-TCR and VDJdatabase (Bagaev et al., 2020; Shugay et al., 2018; Tickotsky et al., 2017). These sequences were derived from $CD4^+$ T cells or total T cells and although these databases are curated, we selected the CDR3aa identified by the methods with the highest confidence of specificities.

Network specificities assignment

The Levenshtein distances (LD<1) was computed between the CDR3aa from database and CDR3aa from RA signatures and each connected CDR3aa or exact matched sequences was considered as same specificity. A down-sampling to 30 unique CDR3aa of each specificity (RA, melanoma, Covid-19) was randomly sampled and LD distance was performed and we repeat 10 times this strategy. LD distance were performed using "stringdist" R packages (van der Loo, 2014).

Data representation

Experimental design

The experimental and statistical study design was performed with Biorender® (https://biorender.com/).

Density-Heatmap and heatmap

Density-heatmap was performed with the ComplexHeatmap R package (Gu et al., 2016), using Euclidean distance and complete method. Heatmap was performed with pheatmap R package ((Kolde, 2019)), using binary distance and Ward.D2 method. Clusters purity represent the percent of the total number of patients that were classified correctly, in the unit range [0.1].

$$Purity=1/N\sum i=1kmaxj|ci\cap tj|$$

where N = number of patients, k = number of clusters, ci is a cluster in C, and tj is the classification which has the max count for cluster ci.

Radviz plot

The Radviz plot, a multivariate data visualization algorithm that plots each feature dimension uniformly around the circumference of a circle, and then plots points within the circle so that the point normalizes its values on the central axes at each arc. For each sample, per cell subset and per chain, the plot was performed with only 4 values corresponding to the sum of the frequencies of each clonotype using CDR3aa of the (i) RA CD4 Teff or CD4 Treg enriched signature, (ii) RA CD4 Teff or CD4 Treg impoverished signature, (iii) T1D CD4 Teff or CD4 Treg enriched signature, (iv) T1D CD4 Teff or CD4 Treg impoverished signature.

Other representations

All the other representations were generated with the ggplot2 R package (Wickham, 2016).

Statistical Analyses

The statistical analyses were performed using the nonparametric Wilcoxon test with Holm multiple-testing correction. Regression linear models were analyzed using Anova with Holm multiple-testing correction using "rstatix" R package. A P-value of <0.05 was considered statistically significant: *P<0.05; **P<0.01; ***P<0.001; ****P<0.0001; not significant is P >0.05.

Data availability

All fastq data obtained in this study were deposited in the National Center for Biotechnology Information Sequence Read Archive following standard MiAIRR recommendations (https://doi.org/10.1038/ni.3873).

Acknowledgements

All the authors thank the patients for having accepted to enter in this clinical protocol. We thank all the staff members involved in TRANSIMMUNOM consortium: Florence Levenez and Mylène Robinet of INRA; Adela Mekouo-Tagne and Olivia Tran for the monitoring and the regulatory aspects of the study; Anne-Marie Chenier, Eric Dasque, Edith Guilloux, Françoise Jouhannet, Celine Kocher, Laurence Koehl, Martine Lebot from Pitié-Salpêtrière CIC and the nurses of the Rheumatology Department of the Saint Antoine Hospital for management of the participants. The authors thank the UMR 8199 LIGAN- PM Genomics platform (Lille, France) for efficient sequencing.

Funding

This work benefited from equipment and services from the iGenSeq core facility at ICM. This work was supported the ERC-Advanced TRiPoD (322856), LabEx Transimmunom (ANR-11-IDEX-0004-02) and RHU iMAP (ANR-16-RHUS-0001) grants to D.K. E.M.F. is funded by the European Research Area Network-cardiovascular diseases (JCT2018 and ANR-18-ECVD-0001) and iReceptorPlus (H2020 Research and Innovation Programme 825821) grants.

Author contributions

R.L. recruited patients. P.B, P.S, M.B.S., K.L.G., V.H., and N.C. performed the experiments and raw data pre-processing. P.B., and P.S. analyzed the data. M.R., E.M.F. and D.K. designed the experiments. P.B., P.S., E.M.F. and D.K. wrote the manuscript with input from all authors. E.M.F. and D.K. conceived the study and obtained funding for the study.

Competing interests

D.K., E.M.F, H.V., K.L.G., M.B.S., M.R., N.C., P.B., P.S., R.L. and V.H. declare no conflicts of interest.

References and notes

Alves, S., Churlaud, G., Audrain, M., Michaelsen-Preusse, K., Fol, R., Souchet, B., Braudeau, J., Korte, M., Klatzmann, D., Cartier, N., 2017. Interleukin-2 improves amyloid pathology, synaptic failure and memory in Alzheimer's disease mice. Brain J. Neurol. 140, 826–842. https://doi.org/10.1093/brain/aww330

Andrews, S., 2015. FastQC : A quality control tool for high throughput sequence data.

Aran, A., Garrigós, L., Curigliano, G., Cortés, J., Martí, M., 2022. Evaluation of the TCR Repertoire as a Predictive and Prognostic Biomarker in Cancer: Diversity or Clonality? Cancers 14, 1771. https://doi.org/10.3390/cancers14071771

Bagaev, D.V., Vroomans, R.M.A., Samir, J., Stervbo, U., Rius, C., Dolton, G., Greenshields-Watson, A., Attaf, M., Egorov, E.S., Zvyagin, I.V., Babel, N., Cole, D.K., Godkin, A.J., Sewell, A.K., Kesmir, C., Chudakov, D.M., Luciani, F., Shugay, M., 2020. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res. 48, D1057–D1062. https://doi.org/10.1093/nar/gkz874

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

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z., 2016. mlr: Machine Learning in R. J. Mach. Learn. Res. 1–5.

Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., Chudakov, D.M., 2015. MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods 12, 380–381. https://doi.org/10.1038/nmeth.3364

Bonelli, M., Savitskaya, A., von Dalwigk, K., Steiner, C.W., Aletaha, D., Smolen, J.S., Scheinecker, C., 2008. Quantitative and qualitative deficiencies of regulatory T cells in patients with systemic lupus erythematosus (SLE). Int. Immunol. 20, 861–868. https://doi.org/10.1093/intimm/dxn044

Borg, N.A., Ely, L.K., Beddoe, T., Macdonald, W.A., Reid, H.H., Clements, C.S., Purcell, A.W., Kjer-Nielsen, L., Miles, J.J., Burrows, S.R., McCluskey, J., Rossjohn, J., 2005. The CDR3 regions of an immunodominant T cell receptor dictate the "energetic landscape" of peptide-MHC recognition. Nat. Immunol. 6, 171–180. https://doi.org/10.1038/ni1155

Boyer, O., Saadoun, D., Abriol, J., Dodille, M., Piette, J.-C., Cacoub, P., Klatzmann, D., 2004. CD4+CD25+ regulatory T-cell deficiency in patients with hepatitis C-mixed cryoglobulinemia vasculitis. Blood 103, 3428–3430. https://doi.org/10.1182/blood-2003-07-2598

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

Chien, Y.H., Gascoigne, N.R., Kavaler, J., Lee, N.E., Davis, M.M., 1984. Somatic recombination in a murine T-cell receptor gene. Nature 309, 322–326. https://doi.org/10.1038/309322a0

Chothia, C., Boswell, D.R., Lesk, A.M., 1988. The outline structure of the T-cell alpha beta receptor. EMBO J. 7, 3745–3755.

Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H.O., Kedzierska, K., La Gruta, N.L., Bradley, P., Thomas, P.G., 2017. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature 547, 89–93. https://doi.org/10.1038/nature22383

Davis, M.M., Bjorkman, P.J., 1988. T-cell antigen receptor genes and T-cell recognition. Nature 334, 395–402. https://doi.org/10.1038/334395a0

Dupic, T., Marcou, Q., Walczak, A.M., Mora, T., 2019. Genesis of the αβ T-cell receptor. PLoS Comput. Biol. 15, e1006874. https://doi.org/10.1371/journal.pcbi.1006874

Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., Rieder, M., Robins, H.S., 2017. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat. Genet. 49, 659–665. https://doi.org/10.1038/ng.3822

Garcia, K.C., Teyton, L., Wilson, I.A., 1999. Structural basis of T cell recognition. Annu. Rev. Immunol. 17, 369–397. https://doi.org/10.1146/annurev.immunol.17.1.369

Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313

He, J., Zhang, R., Shao, M., Zhao, X., Miao, M., Chen, J., Liu, J., Zhang, Xiaoying, Zhang, Xia, Jin, Y., Wang, Y., Zhang, S., Zhu, L., Jacob, A., Jia, R., You, X., Li, X., Li, C., Zhou, Y., Yang, Y., Ye, H., Liu, Y., Su, Y., Shen, N., Alexander, J., Guo, J., Ambrus, J., Lin, X., Yu, D., Sun, X., Li, Z., 2020. Efficacy and safety of low-dose IL-2 in the treatment of systemic lupus erythematosus: a randomised, double-blind, placebo-controlled trial. Ann. Rheum. Dis. 79, 141. https://doi.org/10.1136/annrheumdis-2019-215396

He, J., Zhang, Xia, Wei, Y., Sun, X., Chen, Y., Deng, J., Jin, Y., Gan, Y., Hu, X., Jia, R., Xu, C., Hou, Z., Leong, Y.A., Zhu, L., Feng, J., An, Y., Jia, Y., Li, C., Liu, X., Ye, H., Ren, L., Li, R., Yao, H., Li, Y., Chen, S., Zhang, Xuewu, Su, Y., Guo, J., Shen, N., Morand, E.F., Yu, D., Li, Z., 2016. Low-dose interleukin-2 treatment selectively modulates CD4(+) T cell subsets in patients with systemic lupus erythematosus. Nat. Med. 22, 991–993. https://doi.org/10.1038/nm.4148

Humrich, J.Y., Riemekasten, G., 2019. Low-dose interleukin-2 therapy for the treatment of systemic lupus erythematosus. Curr. Opin. Rheumatol. 31, 208–212. https://doi.org/10.1097/BOR.0000000000000575

Klinger, M., Pepin, F., Wilkins, J., Asbury, T., Wittkop, T., Zheng, J., Moorhead, M., Faham, M., 2015. Multiplex Identification of Antigen-Specific T Cell Receptors Using a Combination of Immune Assays and Immune Receptor Sequencing. PLOS ONE 10, e0141561. https://doi.org/10.1371/journal.pone.0141561

Kolde, R., 2019. pheatmap: Pretty Heatmaps. R Package.

Kukreja, A., Cost, G., Marker, J., Zhang, C., Sun, Z., Lin-Su, K., Ten, S., Sanz, M., Exley, M., Wilson, B., Porcelli, S., Maclaren, N., 2002. Multiple immuno-regulatory defects in type-1 diabetes. J. Clin. Invest. 109, 131–140. https://doi.org/10.1172/JCI13605

Lindley, S., Dayan, C.M., Bishop, A., Roep, B.O., Peakman, M., Tree, T.I.M., 2005. Defective suppressor function in CD4(+)CD25(+) T-cells from patients with type 1 diabetes. Diabetes 54, 92–99. https://doi.org/10.2337/diabetes.54.1.92

Liu, X., Zhang, W., Zhao, M., Fu, L., Liu, L., Wu, J., Luo, S., Wang, L., Wang, Z., Lin, L., Liu, Yan, Wang, S., Yang, Y., Luo, L., Jiang, J., Wang, X., Tan, Y., Li, T., Zhu, B., Zhao, Y., Gao, X., Wan, Z., Huang,

C., Fang, M., Li, Q., Peng, H., Liao, X., Chen, J., Li, F., Ling, G., Zhao, H., Luo, H., Xiang, Z., Liao, J., Liu, Yu, Yin, H., Long, H., Wu, H., Yang, huanming, Wang, J., Lu, Q., 2019. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. Ann. Rheum. Dis. 78, 1070–1078. https://doi.org/10.1136/annrheumdis-2019-215442

Lorenzon, R., Mariotti-Ferrandiz, E., Aheng, C., Ribet, C., Toumi, F., Pitoiset, F., Chaara, W., Derian, N., Johanet, C., Drakos, I., Harris, S., Amselem, S., Berenbaum, F., Benveniste, O., Bodaghi, B., Cacoub, P., Grateau, G., Amouyal, C., Hartemann, A., Saadoun, D., Sellam, J., Seksik, P., Sokol, H., Salem, J.-E., Vicaut, E., Six, A., Rosenzwajg, M., Bernard, C., Klatzmann, D., 2018. Clinical and multi-omics cross-phenotyping of patients with autoimmune and autoinflammatory diseases: the observational TRANSIMMUNOM protocol. BMJ Open 8, e021037. https://doi.org/10.1136/bmjopen-2017-021037

Madi, A., Poran, A., Shifrut, E., Reich-Zeliger, S., Greenstein, E., Zaretsky, I., Arnon, T., Laethem, F.V., Singer, A., Lu, J., Sun, P.D., Cohen, I.R., Friedman, N., 2017. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. eLife 6, e22057. https://doi.org/10.7554/eLife.22057

Miho, E., Yermanos, A., Weber, C.R., Berger, C.T., Reddy, S.T., Greiff, V., 2018. Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. Front. Immunol. 9, 224. https://doi.org/10.3389/fimmu.2018.00224

Nakayama, M., Michels, A.W., 2021. Using the T Cell Receptor as a Biomarker in Type 1 Diabetes. Front. Immunol. 12, 777788. https://doi.org/10.3389/fimmu.2021.777788

Noack, M., Miossec, P., 2014. Th17 and regulatory T cell balance in autoimmune and inflammatory diseases. Autoimmun. Rev. 13, 668–677. https://doi.org/10.1016/j.autrev.2013.12.004

Rohart, F., Gautier, B., Singh, A., Lê Cao, K.-A., 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Comput. Biol. 13, e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Rosenzwajg, M., Churlaud, G., Mallone, R., Six, A., Dérian, N., Chaara, W., Lorenzon, R., Long, S.A., Buckner, J.H., Afonso, G., Pham, H.-P., Hartemann, A., Yu, A., Pugliese, A., Malek, T.R., Klatzmann, D., 2015. Low-dose interleukin-2 fosters a dose-dependent regulatory T cell tuned milieu in T1D patients. J. Autoimmun. 58, 48–58. https://doi.org/10.1016/j.jaut.2015.01.001

Rosenzwajg, M., Lorenzon, R., Cacoub, P., Pham, H.P., Pitoiset, F., El Soufi, K., RIbet, C., Bernard, C., Aractingi, S., Banneville, B., Beaugerie, L., Berenbaum, F., Champey, J., Chazouilleres, O., Corpechot, C., Fautrel, B., Mekinian, A., Regnier, E., Saadoun, D., Salem, J.-E., Sellam, J., Seksik, P., Daguenel-Nguyen, A., Doppler, V., Mariau, J., Vicaut, E., Klatzmann, D., 2019. Immunological and clinical effects of low-dose interleukin-2 across 11 autoimmune diseases in a single, open clinical trial. Ann. Rheum. Dis. 78, 209–217. https://doi.org/10.1136/annrheumdis-2018-214229

Rossetti, M., Spreafico, R., Consolaro, A., Leong, J.Y., Chua, C., Massa, M., Saidin, S., Magni-Manzoni, S., Arkachaisri, T., Wallace, C.A., Gattorno, M., Martini, A., Lovell, D.J., Albani, S., 2017. TCR repertoire sequencing identifies synovial Treg cell clonotypes in the bloodstream during active inflammation in human arthritis. Ann. Rheum. Dis. 76, 435–441. https://doi.org/10.1136/annrheumdis-2015-208992

Saadoun, D., Rosenzwajg, M., Joly, F., Six, A., Carrat, F., Thibault, V., Sene, D., Cacoub, P., Klatzmann, D., 2011. Regulatory T-Cell Responses to Low-Dose Interleukin-2 in HCV-Induced Vasculitis. N. Engl. J. Med. 365, 2067–2077. https://doi.org/10.1056/NEJMoa1105143

Shugay, M., Bagaev, D.V., Zvyagin, I.V., Vroomans, R.M., Crawford, J.C., Dolton, G., Komech, E.A., Sycheva, A.L., Koneva, A.E., Egorov, E.S., Eliseev, A.V., Van Dyk, E., Dash, P., Attaf, M., Rius, C., Ladell, K., McLaren, J.E., Matthews, K.K., Clemens, E.B., Douek, D.C., Luciani, F., van Baarle, D., Kedzierska, K., Kesmir, C., Thomas, P.G., Price, D.A., Sewell, A.K., Chudakov, D.M., 2018. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res. 46, D419–D427. https://doi.org/10.1093/nar/gkx760

Six, A., Mariotti-Ferrandiz, M.E., Chaara, W., Magadan, S., Pham, H.-P., Lefranc, M.-P., Mora, T., Thomas-Vaslin, V., Walczak, A.M., Boudinot, P., 2013. The Past, Present, and Future of Immune Repertoire Biology – The Rise of Next-Generation Repertoire Analysis. Front. Immunol. 4. https://doi.org/10.3389/fimmu.2013.00413

Spreafico, R., Rossetti, M., van Loosdregt, J., Wallace, C.A., Massa, M., Magni-Manzoni, S., Gattorno, M., Martini, A., Lovell, D.J., Albani, S., 2016. A circulating reservoir of pathogenic-like CD4+ T cells shares a genetic and phenotypic signature with the inflamed synovial micro-environment. Ann. Rheum. Dis. 75, 459–465. https://doi.org/10.1136/annrheumdis-2014-206226

Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E., Friedman, N., 2017. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinforma. Oxf. Engl. 33, 2924–2929. https://doi.org/10.1093/bioinformatics/btx286

Valkiers, S., de Vrij, N., Gielis, S., Verbandt, S., Ogunjimi, B., Laukens, K., Meysman, P., 2022. Recent advances in T-cell receptor repertoire analysis: Bridging the gap with multimodal single-cell RNA sequencing. ImmunoInformatics 5, 100009. https://doi.org/10.1016/j.immuno.2022.100009

van der Loo, M., 2014. The stringdist package for approximate string matching. R J. 111–122.

von Spee-Mayer, C., Siegert, E., Abdirama, D., Rose, A., Klaus, A., Alexander, T., Enghard, P., Sawitzki, B., Hiepe, F., Radbruch, A., Burmester, G.-R., Riemekasten, G., Humrich, J.Y., 2016. Low-dose interleukin-2 selectively corrects regulatory T cell defects in patients with systemic lupus erythematosus. Ann. Rheum. Dis. 75, 1407–1415. https://doi.org/10.1136/annrheumdis-2015-207776

Wang, X., Muzaffar, J., Kirtane, K., Song, F., Johnson, M., Schell, M.J., Li, J., Yoder, S.J., Conejo-Garcia, J.R., Guevara-Patino, J.A., Bonomi, M., Bhateja, P., Rocco, J.W., Steuer, C.E., Saba, N.F., Chung, C.H., 2022. T cell repertoire in peripheral blood as a potential biomarker for predicting response to concurrent cetuximab and nivolumab in head and neck squamous cell carcinoma. J. Immunother. Cancer 10, e004512. https://doi.org/10.1136/jitc-2022-004512

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, 2nd ed. 2016. ed, Use R! Springer International Publishing : Imprint: Springer, Cham. https://doi.org/10.1007/978-3-319-24277-4

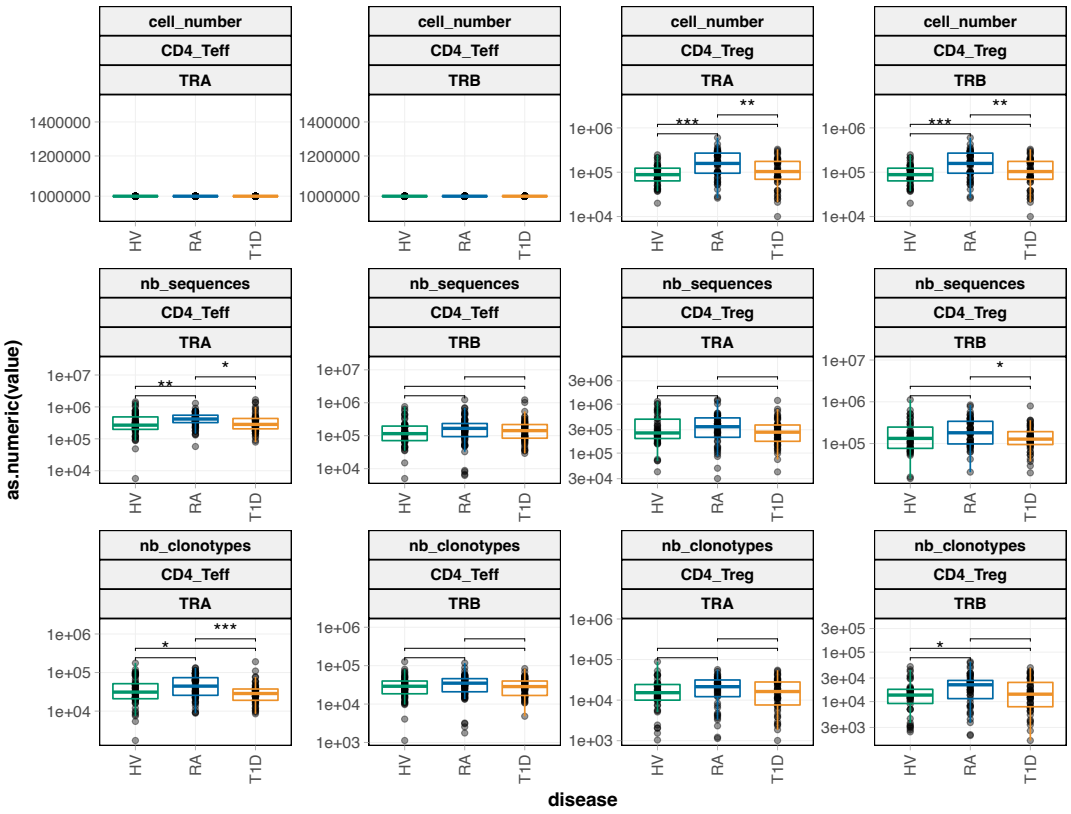
SUPPLEMENTARY

Supplementary table 1

Project	Disease	other category	Cell subset	Number of sample	Number of total TCR sequences		Number of unique clonotypes	
					TRA	TRB	TRA	TRB
Transimmunom	HV		CD4 Teff	54	19 957 156	9 177 647	1 446 015	1 698 416
			CD4 Treg	40	15 209 413	8 553 179	520 411	633 216
	RA		CD4 Teff	68	29 656 503	14 056 153	2 070 290	2 247 557
			CD4 Treg	49	18 173 574	11 597 655	738 028	1 032 477
	T1D		CD4 Teff	52	20 397 495	10 129 715	1 186 514	1 541 443
			CD4 Treg	50	15 184 744	7 936 162	682 894	812 502
	Total			**313**	**118 578 885**	**61 450 511**	**4 631 095**	**7 213 448**

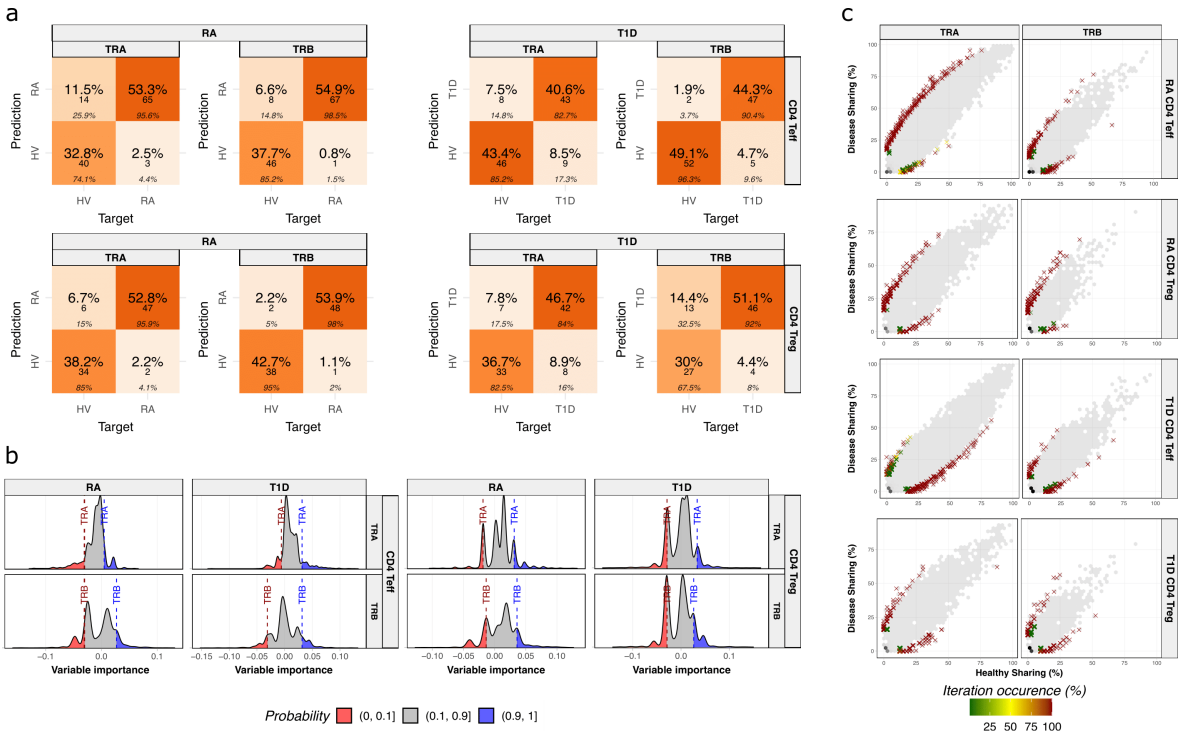
Table 1 : Data summary

Supplementary fig.1



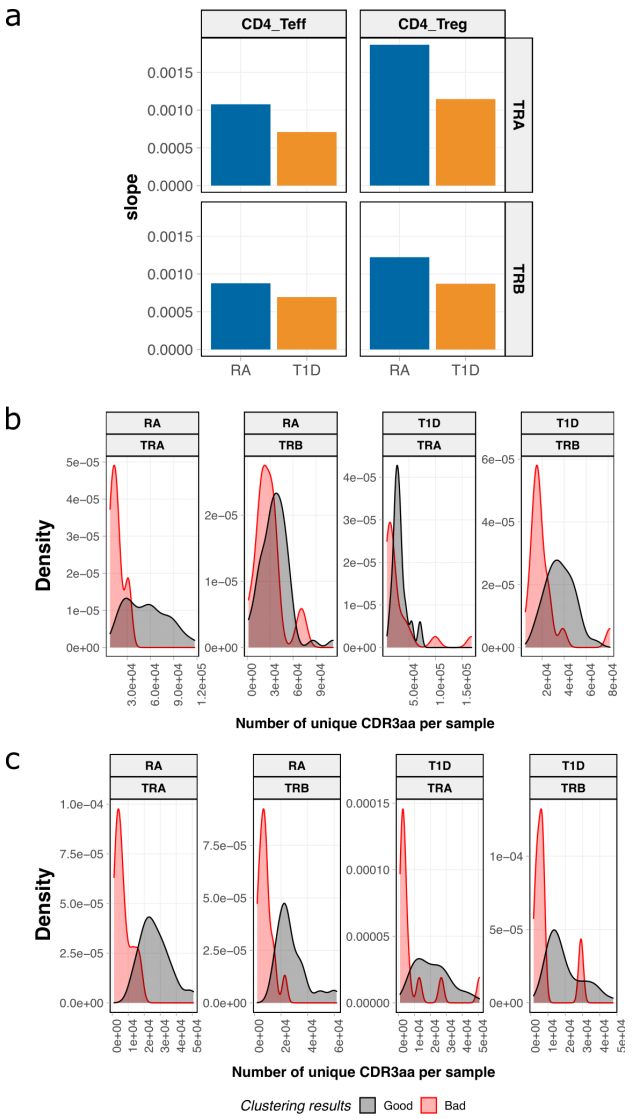
Supplementary Fig. 1. Dataset description. *For each patient by groups (HV: green, RA: blue and T1D: orange) the cell number (top), TCR sequences number (middle) and clonotypes number (bottom) are presented for Teff (left) and Treg (right)*

Supplementary fig.2



***Supplementary Fig. 2: Details of the features selection. a**. Confusion matrix over the entire LOO classification for RA vs HV (left), T1D vs HV (right), Teff (top) Treg (bottom). Values represent the percentage of patients in the entire classification (middle percentage), the number of patients (number under the percentage), and the percentage by disease group (small percentage at the bottom). **b**. Variables importance distribution. The distribution represents the average variables importance over all the LOO iterations. Red and blue distributions represent the most important selected variables. Results are presented for Teff on the left, Treg on the right for each disease (RA: first panel, T1D second panel). **c**. Signatures sharing distribution. By group (disease (y-axis) vs. HV (x-axis)), for the TRA (left) and TRB (right) chain, CDR3aa sharing by patients is represented by the light gray shape for CDR3aa that are not selected by sPLSDA selection and by the color from green to red (depending on the number of occurrences overall theLOO iterations) for CDR3aa signatures, for RA on top and T1D on bottom.*

Supplementary Fig.3



Supplementary Fig. 3: Signature sharing strength. ***a***. *Slope of signature sharing scatterplot. Results are presented for Teff signatures on the left, Treg on the right, for the TRA chain on top and TRB on the bottom, from each disease (RA; blue, T1D: orange).* ***b*** *and* ***c****. Density curve according to the number of unique CDR3aa in samples considered to be classified in the wrong cluster of the heatmap (Figure 4a) in red and well classified in gray. For Teff in* ***c****, Treg in* ***c****, RA on the left and T1D on the right.*

III.3. Résultats complémentaires

En compléments des deux articles précédents, j'ai également eu l'occasion de travailler sur deux autres sujets, actuellement en cours d'analyse.

III.3.1. Le répertoire TCR, un biomarqueur de la réponse au traitement

Le lupus érythémateux systémique (LES) appartient à la catégorie des maladies auto-immunes systémiques, caractérisées par une atteinte de plusieurs organes. Cette maladie reste difficile à traiter, une rémission durable n'est obtenue que chez une faible fraction des patients (Franklyn et al., 2014). Sur le plan immunologique, elle se caractérise par la production d'auto-anticorps dirigés contre de nombreux auto-antigènes, notamment contre l'ADN double-brins. Cette surproduction d'auto-anticorps est la conséquence d'un important dérèglement immunitaire, notamment cytokinique avec un rôle central des cellules T $CD4^+$. Il a été démontré que les lymphocytes de patients LES produisent de faibles taux d'IL-2 (Linker-Israeli et al., 1983), cytokine auto- et paracrine, indispensable à la survie et la fonction suppressive des Tregs, notamment via le renforcement de leur fonctions immuno-régulatrices ainsi qu'un affaiblissement de la réponse Teffs par privatisation d'IL-2 nécessaire à leurs fonctions pro-inflammatoires. Ce déficit en IL-2 induit une dérégulation de la balance Tregs/Teffs (Humrich et al., 2010; Mellor-Pita, 2006; Suen et al., 2009). Dans le cadre d'essais cliniques, le traitement de patients atteints de MAI par l'IL-2 à faible dose ($IL\text{-}2_{fd}$) a montré une augmentation des Tregs dans le sang, associé à un bénéfice clinique (He et al., 2020; Rosenzwajg et al., 2019, 2015b; Saadoun et al., 2011). En effet, rétablir la balance anti-inflammatoire dans une pathologie auto-immune est une stratégie capable de réduire les effets indésirables infectieux souvent observés avec les immunosuppresseurs non spécifiques (Kang and Park, 2003). Le rééquilibre de cette balance immunologique a été associé à une diminution du score clinique et d'une rémission chez certains patients atteints de LES (He et al., 2016).

Il a été démontré que le répertoire TCR des cellules T est altéré chez les patients atteints de LES par rapport aux VS (Costa et al., 2013; Liu et al., 2019), mais sans distinction entre les répertoires des cellules $CD4^+$ Tregs et les cellules $CD4^+$ Teffs. Nous souhaitons évaluer distinctement la diversité des sous-populations ayant des rôles

antagonistes et des régulations différentes, avec dans un premier temps l'évaluation de l'impact précoce (v1 et v5) du traitement par IL-2_{fb} sur les répertoires. Ceci nous permettra de cibler la population d'intérêt pour identifier des biomarqueurs TCR prédictifs de succès ou échecs thérapeutiques.

La diversité des répertoires TCR des patients LES avant le début du traitement est analysée grâce à l'entropie de Shannon (Shannon, 1948). Nous observons des perturbations de la diversité des patients le jour du recrutement par rapport aux VS (données issues de Transimmunom, section II.1.1) ; les patients LES ayant une diversité des répertoires $CD4^+$ Teffs significativement supérieure et à l'inverse une diversité $CD4^+$ Tregs significativement inférieure (**Figure 17a**). Une analyse longitudinale révèle que l'IL-2_{fd} engendre une modulation de la diversité du répertoire

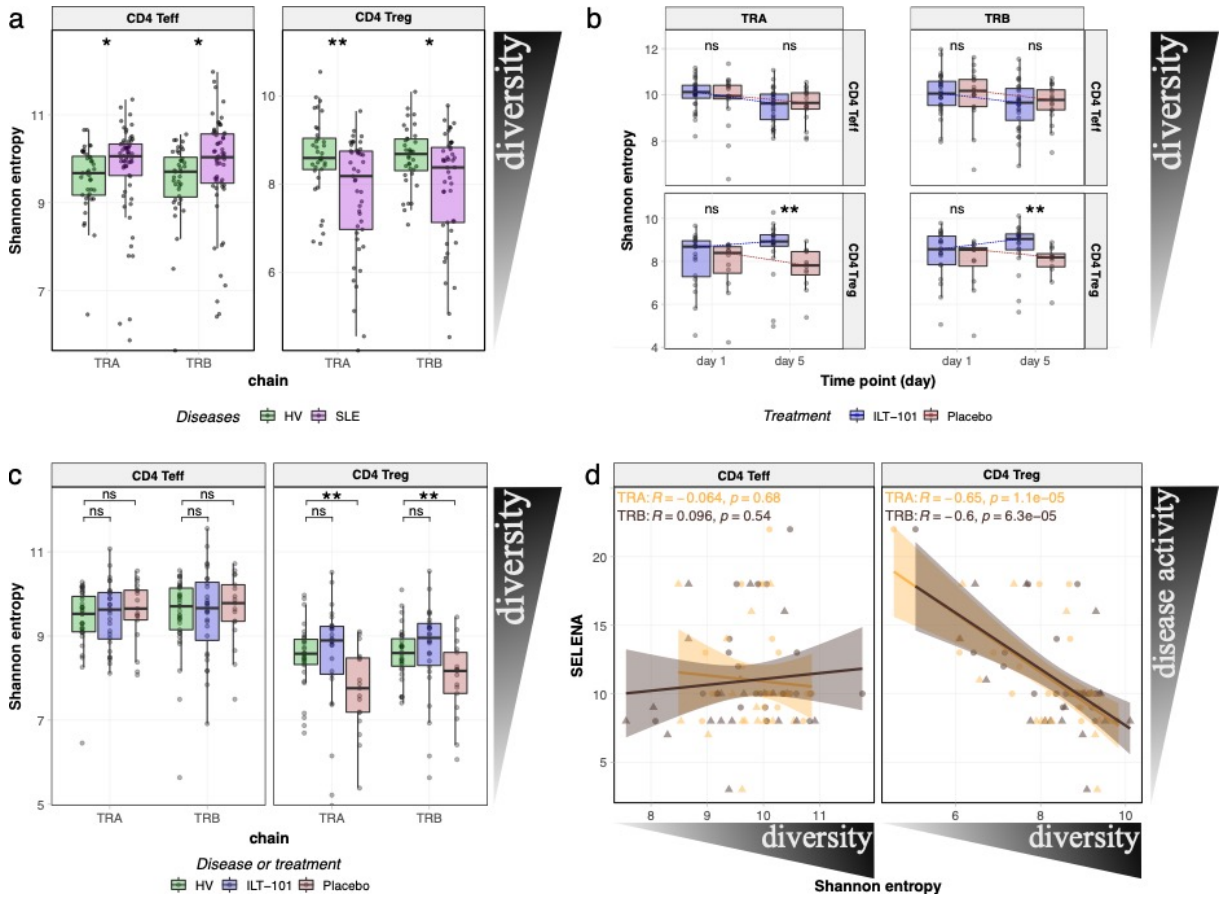


Figure 17 : Modulation de la diversité des répertoires TCR Teff et Treg sous IL-2_{fd}.
a., b., c. L'entropie de Shannon est une mesure de diversité globale du répertoire. a. Shannon entropie des volontaires sains (VS : vert) et les patients lupiques (LES) à baseline. b. Modulation de l'entropie de shannon en fonction du traitement (IL-2 en bleu, Placebo en rouge) par point de prélèvement. c. Entropie de Shannon des VS (vert) comparée à celles des LES à jours 5 (IL-2 en bleu, Placebo en rouge). d. Corrélation entre l'entropie de Shannon et l'activité de la maladie (SELENA) chez les patients IL-2 répondeurs pour la chaine TRA (orange) et TRB (marron).

TCR spécifiquement aux CD4⁺ Tregs (**Figure 17b**) et que cette modulation permet de rétablir une diversité similaire à celle des volontaires sains seulement 5 jours après le début du traitement (**Figure 17a&c**). Ce qui n'est pas le cas pour les patients traités par un placebo (**Figure 17c**). La modulation de cette diversité des CD4⁺ Tregs est elle-même corrélée avec le score clinique de la maladie, ce qui n'est pas le cas pour les CD4⁺ Teffs (**Figure 17d**). En revanche nous n'observons pas de différences significatives lorsque l'on compare la modulation de la diversité des patients traités par l'IL-2_{fd} répondeurs et ceux non répondeurs (données non présentées). Cette modulation de la diversité n'est donc pas une valeur prédictive de l'efficacité du traitement.

Bien que l'IL-2_{fd} impacte principalement la diversité des CD4⁺ Tregs, de façon similaire chez les répondeurs et les non répondeurs, une analyse du partage des expansions (clonotypes ayant une augmentation de 2 fois sa fréquence entre v1 et v5) sous l'impact du traitement nous montre une convergence (Jaccard, 1912) significativement supérieure (spécifiquement pour la chaine TRA) chez les patients répondeurs (**Figure 18a**). Ensemble, ces résultats nous indiquent que les CD4⁺ Tregs

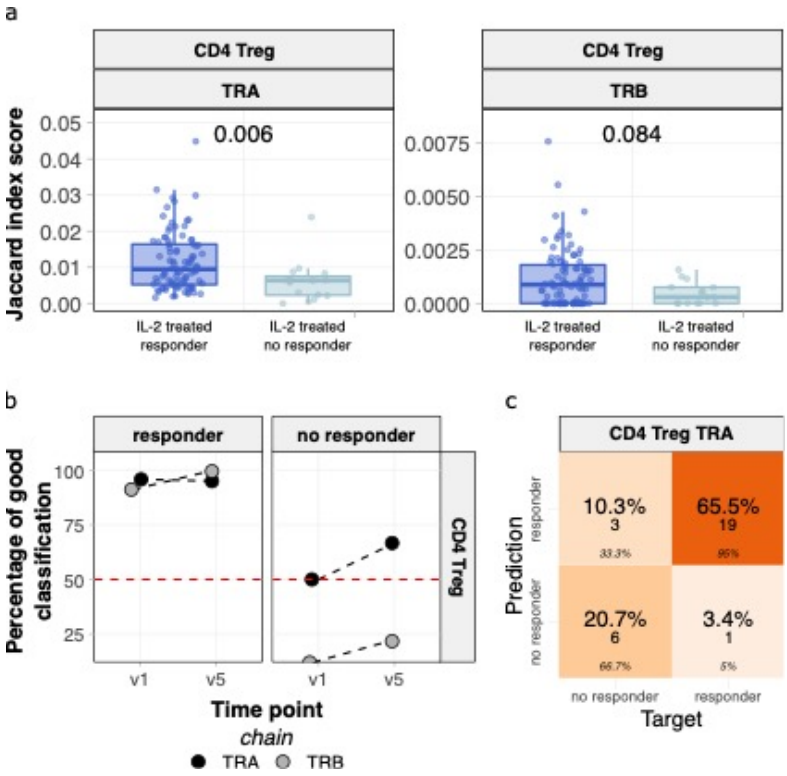


Figure 18 : Évaluation de la valeur prédictive des répertoires TCR des Tregs face à la réponse au traitement chez les patients traités par IL-2_{fd}.
a. Indice de jaccard entre répertoires 2 à 2 des clonotypes dont la fréquence a doublé entre la v1 et la v5. b. Pourcentage de bonne classification des patients répondeurs à l'IL-2_{fd} (gauche) et les patients non répondeurs (droite) grâce à l'analyse des CD4⁺ Tregs à v1 et v5. c. Matrice de confusion entre les IL-2 répondeurs et IL-2 non répondeurs à jour 5.

sont la population cellulaire pouvant contenir une information prédictive sur la réponse au traitement.

Afin de tester cette hypothèse nous avons appliqué la stratégie d'analyse présentée précédemment dans les sections III.2 à savoir l'utilisation de l'approche sPLSDA pour la sélection des CDR3 de chaque itération de LOO et une classification entre le groupe de patients IL-2_{fb} répondeurs et non répondeurs pour le premier prélèvement (v1) et pour le deuxième prélèvement (v5).

Nous observons lors du premier prélèvement que le pourcentage de bonne classification est de 95,8% pour les répondeurs (23/24) et de 50% (4/8) pour les non répondeurs donnant une classification globale de 84,4% avec la chaine TRA. Avec la chaine TRB, ce pourcentage est de 91,7% pour les répondeurs (22/24) mais de seulement 12,5% (1/8) pour les non répondeurs donnant une classification globale de 71,9% (**Figure 18b**). Malgré des scores de classification en apparence élevés, ils ne sont pas représentatifs des résultats réels dû à la différence du nombre de patient entre les deux groupes. Le répertoire TCR des $CD4^+$ Tregs avant le début du traitement n'est pas une valeur prédictive suffisamment fiable. Nous observons en revanche une augmentation des pourcentages de classification correcte avec les prélèvements au jours 5 après le début du traitement, et ceci notamment pour les non répondeurs, passant ainsi de 50% à 66,7% (6/9) pour la chaine TRA et de 12,5% à 22,2% (2/9) pour la chaine TRB (**Figure 18b**). Nous observons un pourcentage de classification correcte globale de 86,7% avec la chaine TRA des cellules CD4+ Tregs lors de la visite à jours 5, classification mieux équilibrée que les précédentes, avec 95% (19/20) des patients répondeurs bien classés et 66,7% (6/9) des patients non répondeurs bien classés (**Figure 18c**).

Au bilan, nous avons vu que l'IL-2_{fd} module la diversité des $CD4^+$ Tregs, pour rétablir une diversité comparable à celle des VS. Cette modulation est corrélée à l'activité de la maladie. Plus la diversité des $CD4^+$ Tregs augmente, moins la maladie est sévère, ce qui n'est pas observé dans la population $CD4^+$ Teffs. Cette modulation de la diversité chez les patients traités par l'IL-2_{fd} est comparable entre les répondeurs et les non répondeurs, néanmoins, les clonotypes qui prolifèrent entre les deux points de prélèvements convergent chez les répondeurs. A partir de ces observations, nous avons établi que la population $CD4^+$ Tregs avait la plus forte valeur prédictive de la réponse au traitement, ce qui n'a pas été confirmé avec les classifications à *baseline*. En

revanche une augmentation de ce pouvoir prédictif a été observé au deuxième prélèvement permettant de classer correctement 25/29 patients. Ensemble ces résultats justifient l'intérêt de distinguer les populations cellulaires dans l'analyse du répertoire TCR.

L'IL-2_{fd} permettant la survie et la prolifération des cellules $CD4^+$ Tregs (He et al., 2016), nous avons montré que cette augmentation avait une répercussion sur la diversité du répertoire TCR de cette population dans le sang. Même si nous avons montré chez la souris NOD l'apparition d'expansions clonales suite à l'injection d'IL-2_{fd}, la diversité a été étudiée plus tardivement (3 semaines). Il conviendrait d'analyser les prélèvements postérieurs pour infirmer ou confirmer cette observation chez l'homme (Mhanna et al., 2021). Néanmoins, des résultats similaires à ceux présentés ont été obtenus dans une étude sur l'impact de l'IL-2_{fd} dans la maladie du greffon contre hôte. Ils ont également identifié une augmentation du partage des répertoires TCR entre les populations $CD4^+$ Tregs et $CD4^+$ Teffs en lien avec le traitement, notamment chez leurs patients répondeurs, suggérant que la conversion des $CD4^+$ Teffs en $CD4^+$ Tregs contribue à l'efficacité clinique (Whangbo et al., 2018). Il nous faudrait évaluer le partage de notre signature $CD4^+$ Tregs (notamment TRA) identifié à v5 avec les répertoires TCR des $CD4^+$ Teffs à v1. Si un partage significatif est observé, suggérant une conversion des $CD4^+$ Teffs, alors une analyse cytokinique complémentaire de l'Il-10 et du TGF-β pourrait appuyer cette hypothèse. Enfin, il est intéressant d'observer que dans nos résultats, la chaine α du TCR est la plus informative, car, dans le postulat d'une conversion de la population effectrice vers la population régulatrice sous impact de l'IL-2_{fd}, l'exclusion allélique incomplète de la chaine α du TCR considérée comme une source potentielle d'auto-immunité est un argument en faveur de cette hypothèse (Elliott and Altmann, 1995; Padovan et al., 1993; Schuldt et al., 2017).

Finalement, notre étude comporte pour le moment plusieurs limites. Basé sur l'observation faite chez la souris NOD, l'IL-2_{fd} favorise des expansions clonales significativement enrichies en TCR spécifiques du DT1 dans les cellules $CD4^+$ Tregs activées (Mhanna et al., 2021). Or, nous n'avons pas la distinction des phénotypes d'activation des CD4+ Tregs et pourrions ne pas détecter cette information. Également, l'efficacité du traitement n'est plus à démontrer mais créait un déséquilibre entre le nombre de patients répondeurs et non répondeurs. Ces résultats encourageants sont à

confirmer sur une cohorte plus grande. Pour terminer, nous observons une augmentation du pouvoir de prédiction entre v1 et v5 or certaines études suggèrent que l'augmentation maximale des $CD4^+$ Tregs est observée entre le jour 5 et le jour 8 (Rosenzwajg et al., 2015a), il aurait pu être intéressant d'avoir un prélèvement plus tardif pour évaluer le pouvoir de classification.

III.3.2. Le répertoire TCR comme biomarqueur pronostique face à la Covid-19

L'issue d'une infection virale est largement déterminée par la qualité de la réponse immunitaire adaptative (Bertoletti and Ferrari, 2016; Peteranderl et al., 2016). En effet, plusieurs études suggèrent qu'une réponse coordonnée des cellules T est essentielle pour une issue favorable de l'état pathologique des patients infectés par le Sars-Cov-2 (Mathew et al., 2020; Rydyznski Moderbacher et al., 2020). Cette maladie infectieuse émergente est caractérisée par une forte mortalité des individus immunodéprimés, touchant également gravement des individus considérés sains. La composition du répertoire TCR du sang périphérique de patients en phase COVID-19 aigüe de l'infection est différente du point de vue de la diversité, ainsi que de l'expression de certains gènes et clonotypes en comparaison aux patients guéris (Schultheiß et al., 2020). Ainsi, afin de tester l'association de ces perturbations avec l'évolution de la maladie, nous avons récupéré un ensemble de données publiques de séquences TCRβ provenant du sang périphérique de patients infectés par la COVID-19 (Snyder et al., 2020). Parmi ces patients, nous avions l'information de l'évolution médicale et avons identifié 35 patients décédés suite à l'infection. Nous avons fait une sélection de patients ayant récupérés face à l'infection (n = 38) comparables en termes d'âge (**Figure 19a**) et de délais entre le diagnostic et le prélèvement sanguin (**Figure 19b**). L'objectif étant d'évaluer l'évolution de la maladie chez des patients ce qui permettrait d'orienter les soins à apporter.

Pour cela, nous appliquons une stratégie d'analyse similaire à celle présentée précédemment dans la section III.2 et III.3.1 à savoir l'utilisation de l'approche sPLSDA pour la sélection des CDR3 pour chaque itération de LOO et une classification entre le groupe de patients décédés et ceux qui se sont rétablis de l'infection.

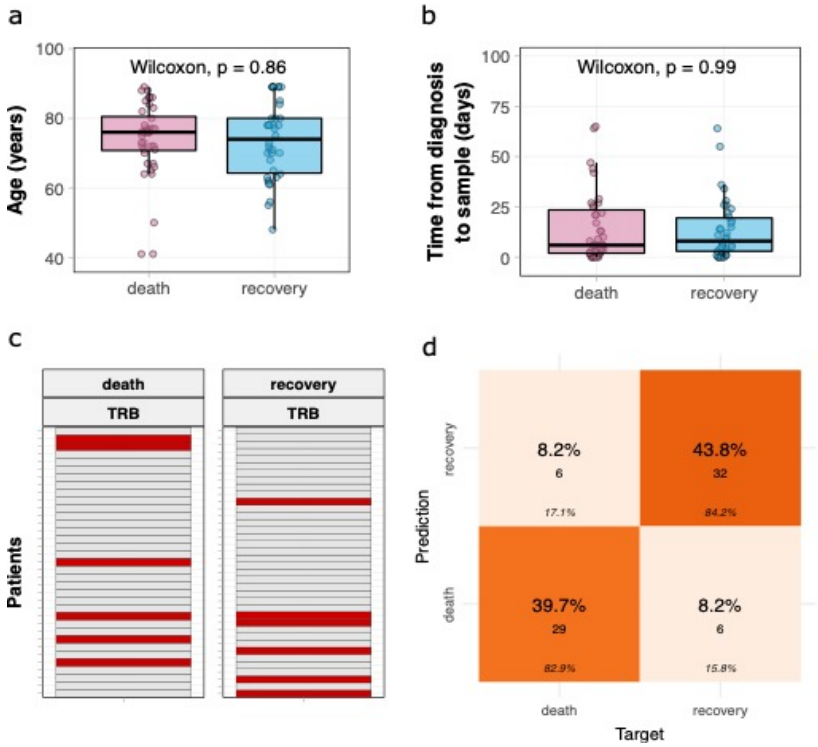


Figure 19 : La signature TCR prédit l'issue du COVID-19.
a. Boxplot descriptif de l'âge des patients de la cohorte de patients décédés du Covid-19 en rose et la cohorte de patients ayant guéris en bleu. b. Boxplot descriptif du nombre de jours entre le diagnostic et le prélèvement des patients de la cohorte de patients décédés du Covid-19 en rose et la cohorte de patients ayant guéris en bleu. c. Heatmap des résultats de la classification par patient. Chaque ligne est un patient coloré en rouge s'il a été classifié dans la mauvaise catégorie et en gris s'il a été classifié dans la bonne catégorie. Les résultats des patients décédés de l'infection sont sur la gauche et ceux ayant récupérés sur la droite. d. Matrice de confusion montrant les résultats de la prédiction de l'issue COVID-19.

Nous avons ainsi identifié des signatures TCRβ ayant une valeur pronostic de 83,5% (61/73) de bonne classification distribuée de façon équitable entre les groupes (**Figure 19c&d**). Ainsi, le répertoire TCRβ analysé à partir de cellules T totales du sang prélevé peu de temps après le diagnostic de l'infection (moyenne à 7 jours) nous permet de prédire avec précision l'issue de l'infection.

Concernant les patients décédés, nous avons eu accès à l'information du nombre de jours entre le prélèvement et la mort. Cette information était disponible pour 28/35 patients avec une moyenne de 10 jours. Parmi ceux classés comme pouvant avoir une forme sévère seulement 11/29 patients décédés (38%) ont été transférés en unité de soins intensifs, ainsi cette étude est une preuve de concept que l'analyse du répertoire TCR peut être un outil permettant d'orienter la prise en charge d'un patient.

Comme dit précédemment, le nombre de patients mal classés est de 6 dans chacun des groupes. Il est donc essentiel de comprendre pourquoi ces patients ont été

classés dans la catégorie inverse. En effet, concernant un patient ayant récupéré, ceci n'aurait pas de grande conséquence, en revanche pour les 6 patients décédés prédits comme allant récupérer face au COVID-19, cette fois-ci la décision serait lourde de conséquence. Une analyse fine des variables cliniques serait un élément essentiel pour tenter d'expliquer ces résultats mais ces prélèvements venant de divers laboratoires à travers le monde, les variables cliniques ne sont pas toujours disponibles.

En conclusion, la réponse immunitaire produit une signature de TCR associée à la COVID-19 et probablement à la survie. Le devenir des patients est une première information, il nous faudra par la suite évaluer le rôle biologique de ces cellules en combinant les données des patients infectés avec des données publiques caractérisées par leur spécificité antigénique. Les signatures du répertoire associées à la guérison peuvent être précieuses pour guider le développement et l'évaluation des candidats vaccins. Une des limites de cette étude préliminaire, ainsi que celle de la majorité des données publiques disponibles, est que les données ont été générées à partir de sang total ou PBMC. Par conséquent, nous ne pouvons pas savoir si la signature prédisant une bonne issue du COVID-19 est associée aux cellules Teffs ($CD4^+$ ou $CD8^+$) qui peuvent éliminer efficacement le virus ou celle des cellules Tregs qui préviennent l'immunopathologie, ou les deux.

Chapitre IV. DISCUSSION ET PERSPECTIVES

Compte tenu des fonctions cruciales des cellules Tregs dans le maintien de la tolérance périphérique, cette population constitue la population d'intérêt à étudier. Malgré de nombreuses études sur les défauts du nombre et/ou de la fonction de ces cellules dans le contexte de la MAI, la spécificité antigénique des cellules Tregs déficientes reste insuffisamment caractérisée. Or, le potentiel clinique de ces cellules par l'identification de biomarqueurs spécifiques d'antigènes à des fins thérapeutiques est prometteur. Actuellement, la majorité des MAI est traitée avec des médicaments immunosuppresseurs permettant de contrôler l'auto-immunité en limitant l'activité globale du SI, ce qui entraîne une susceptibilité accrue au risque d'infection chez ces patients, nécessitant une surveillance régulière. Induire une tolérance périphérique ciblée afin d'inverser un processus pathologique tout en permettant au SI de fonctionner normalement est un objectif que l'immunothérapie personnalisée spécifique d'antigène peut atteindre.

L'ensemble de ce travail est organisé autour de la découverte d'un biomarqueur allant de la validation de la méthode de génération des données, à la mise en œuvre d'une stratégie d'identification de variables discriminantes, jusqu'à la validation en tant que biomarqueurs dans d'autres études et finalement la preuve de concept dans différents contextes biologiques.

IV.1. L'importance de la méthode de production

Dans la première partie nous avons choisi d'établir une comparaison internationale de différentes méthodes de biologie moléculaire pour l'analyse du répertoire TCR par séquençage de nouvelle génération. Au travers de cette étude, nous avons identifié une méthode fondée sur le principe de PCR 5'RACE (nommée « RACE-3 » dans l'article) comme étant celle répondant le mieux aux contraintes imposées par les échantillons du laboratoire. Néanmoins, il existe plusieurs limites à ce projet.

Dans l'étude, nous avons comparé 9 méthodes, chacune basée sur un protocole d'amplification (5'RACE ou mPCR) élaboré à partir d'un matériel génétique défini (ADN ou ARN). La répartition des méthodes dans ces groupes de "protocole - matériel biologique" est la suivante : 6 méthodes "5'RACE - ARN", 2 méthodes "mPCR - ADN" et une seule méthode "mPCR - ARN". Cette distribution déséquilibrée entre les groupes ne semble pas être représentative de l'utilisation par la population scientifique

(Annexe3, Figure 2), représentant ainsi un biais à prendre en compte dans l'interprétation de la reproductibilité des méthodes mPCR entre elles.

Lors de la conception de l'expérience, nous aurions dû modifier certains paramètres afin d'évaluer la sensibilité maximale de chaque méthode. L'identification du clone Jurkat, n'a pas été poussée à des concentrations suffisamment fiables. Bien que les dilutions évaluées aient fourni des informations sur la fiabilité et la distorsion de représentativité d'un clone au travers diverses concentrations, toutes les méthodes ont été capables de détecter ce clone à la plus faible concentration. Dans le contexte des MAI, la plus faible dilution du clone Jurkat (1/1000) n'est pas représentative des conditions de détection des cellules auto-réactives spécifiques d'un antigène circulant dans le sang entre les organes lymphoïdes secondaires (Roep and Peakman, 2012).

Dans cette étude, les méthodes ont été évaluées sur une seule méthodologie de séquençage correspondant à la chimie la plus courante recommandée par les fournisseurs au moment de ce projet. Ainsi, les biais potentiels, les avantages et les inconvénients dans la diversité de la compatibilité des protocoles avec les différentes chimies et plateformes de séquençage n'ont pas pu être évalués. Par la suite, en lien avec l'évolution intensive des plateformes et méthodologies de séquençage, nous avons évalué la compatibilité de la méthode choisie avec d'autres plateformes de séquençage afin d'augmenter le nombre d'échantillons par séquençage et ainsi diminuer les prix. Nous avons ainsi mis en évidence l'incompatibilité de notre méthodologie avec la technologie NextSeq d'Illumina. La diversité nucléotidique des séquences soumises à séquençage est essentielle pour une performance optimale de génération de données de haute qualité. Cette diversité faisant référence à la proportion relative de nucléotides A, C, G et T de chaque cycle. Elle est particulièrement importante au cours

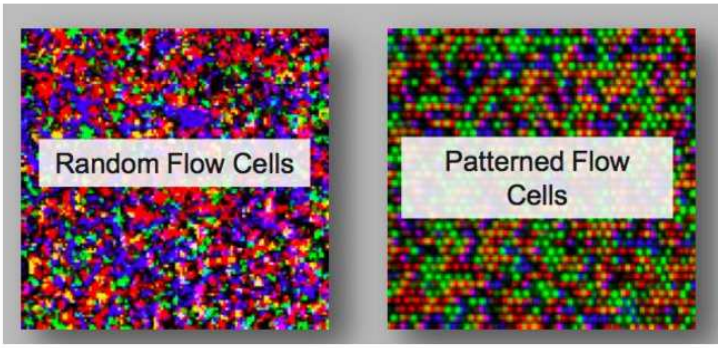


Figure 20 : Illustration d'une cellule aléatoire et d'une cellule à motif.

des premiers cycles de séquençage pour l'identification des *clusters* de banques d'ADN. Le système HiSeq comme le MiSeq (utilisé pour la validation de la méthode) utilise la chimie à 4 couleurs (chaque nucléotide émet une longueur d'onde unique) associée à une cellule de séquençage aléatoire (**Figure 20** ; *Random Flow Cells*). Les systèmes NextSeq et NovaSeq 6000 (dernière génération de NGS) utilisent la chimie à 2 couleurs (T = vert, C = rouge, A = vert/rouge, G = pas de signal), permettant un séquençage plus rapide et un traitement des données plus efficace. Cependant, cette chimie plus limitée est davantage sensible à la diversité. Par ailleurs, ces deux systèmes n'ont pas le même support de cellules. Quand le NextSeq est associé à une cellule aléatoire, le NovaSeq 6000 utilise une cellule à motifs (*Patterned Flow Cells ;* **Figure 20**), facilitant la détection des *clusters*. Ainsi, la faible diversité des bibliothèques TCR (dû à la région TCR constante présente au début de nos amplicons) entraine une impossibilité d'identification des *clusters* avec le système bicolore combiné à une cellule aléatoire proposé par le NextSeq. C'est un inconvénient non abordé dans cet article, et que nous avons évalué uniquement avec la méthode RACE-3. Après discussion avec les co-auteurs de l'article, il semble que ce problème peut être « résolu » à l'aide d'UMI. Ces identifiants moléculaires aléatoires fournissent la diversité nécessaire à l'identification des clusters.

La méthode RACE-3 disponible lors du lancement de ce projet ne proposait pas d'UMI, or quelques années plus tard, à la demande des utilisateurs, une version 2 (RACE-3 v2) de leur kit a été proposée. Nous avons évalué l'impact de l'ajout d'UMI sur la représentation du répertoire TCR (**Figure 21**). Brièvement, nous avons analysé les répertoires TCR obtenus à partir d'un même échantillon d'ARN produit avec les différentes versions du kit proposé par Takara® ; Version 1 (sans UMI) : RACE-3 v1 ; Version 2 (avec UMI, sans prise en compte des UMI pour la correction) : RACE-3 v2 ; Version 2 (avec UMI, avec prise en compte des UMI pour la correction) : RACE-3 v2-U. Nous n'observons pas de différence dans le pourcentage d'erreur PCR entre la version 1 et 2 et comme attendu, le traitement des UMI permet de diminuer considérablement le pourcentage d'erreur (**Figure 21a**). En revanche l'ajout d'UMI dans la structure des banques TCR diminue la richesse de capture du nombre de clonotypes des chaines α et β (RACE-3 v1 vs RACE-3 v2). Cette diminution est d'autant plus marquée suite à la correction apportée par les UMI (**Figure 21b**, alpha = 0). Comme pour les méthodes analysées dans l'article présenté en section III.1 (RACE-1 et RACE-2) nous n'observons pas de différence entre les protocoles pour ce qui est de la

diversité des expansions clonales (**Figure 21b**, alpha = +∞). Finalement, nous observons un partage des clonotypes expandus plus important entre les réplicas de la version 1 en comparaison aux réplicas des versions 2, avec et sans traitement UMI (**Figure 21c**).

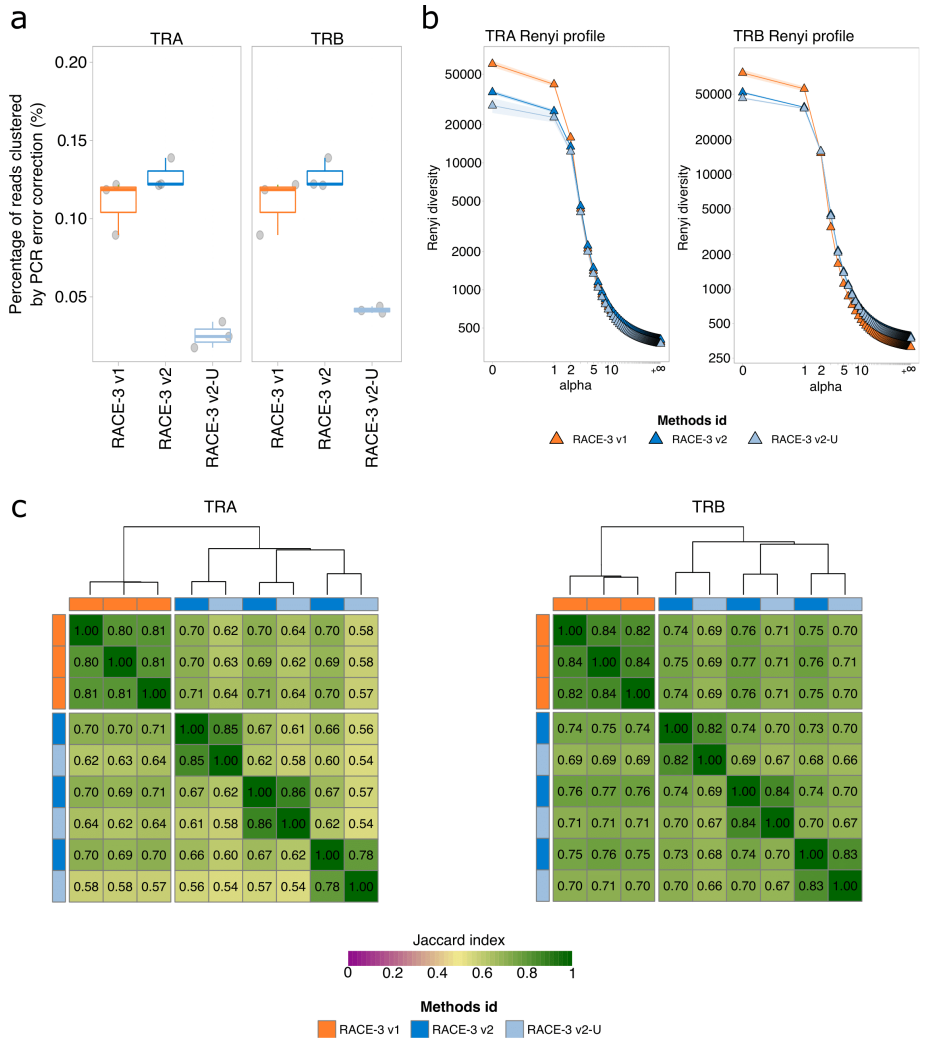


Figure 21 : Impact de l'ajout d'UMI sur la méthode RACE-3.
a. Pourcentage de lectures regroupées après correction des erreurs de PCR. Chaque boîte représente le pourcentage de lectures regroupées. b. Profil de diversité de Rényi. Pour chaque ensemble de données (n = 3 réplicas), les mesures de diversité utilisant les fréquences des clonotypes ont été calculées pour des valeurs croissantes de l'ordre Rényi a jusqu'à stabilisation de la diversité (section II.4) et affichées sous forme de profil de diversité de Rényi. Les zones avec transparences représentent l'écart-type de la moyenne. c. Heatmap de l'indice de similarité Jaccard. Les scores Jaccard ont été calculés entre chaque réplicas, toutes méthodes confondues, pour le top 1% des clonotypes les plus expandus pour TRA (à gauche) et TRB (à droite). La distance euclidienne a été utilisée pour le regroupement hiérarchique sous forme de matrice à code couleur allant de 0 (violet, dissimilarité maximale) à 1 (vert, similarité maximale). Pour l'ensemble des parties la méthode RACE-3 v1 est en orange, la méthode RACE-3 v2 en bleu foncé et la méthode RACE-3 v2-U en bleu clair.

Au bilan, l'incorporation des UMI dans la structure des amplicons TCR produit par la méthode RACE-3 diminue la sensibilité de capture de la diversité accompagnée d'une plus faible reproductibilité entre les réplicas. Sur la base de ces résultats, nous avons choisi de conserver la version 1 du protocole proposé par Takara®.

Illustré par les résultats présentés ci-dessus, on constate que toutes les étapes de production des données (amplification et séquençage) évoluent rapidement. De ce fait, en lien avec la nécessité de contrôle biologique pour la standardisation du répertoire TCR (section I.4.4), il aurait été intéressant dans ce projet de comparaison de méthodes d'amplification de créer un échantillon initial en très grande quantité (à travers un regroupement de cellules provenant de plusieurs individus). L'échantillon « contrôle » nous aurait permis d'évaluer l'impact de ces changements afin d'anticiper un changement de stratégie, de savoir si les données sont comparables et d'évaluer de nouveaux protocoles. De plus, cet échantillon séquencé en 108 réplicas représente un répertoire clonotypique dont l'abondance relative est définie. Celle-ci aurait pu être intégré dans toutes les expériences de séquençage comme échantillon de référence. Grâce à la capture de la diversité de cet échantillon, nous aurions pu évaluer et distinguer les biais techniques non biologiques dus par exemple à (i) la fluctuation de l'efficacité des différentes enzymes ; (ii) la variabilité de la collection des échantillons intégrés dans une l'expérience (variation du nombre de cellules notamment) mais également (iii) la différence de profondeur de séquençage. En prenant en compte ces paramètres, les effets dus à l'expérience elle-même pourraient être corrigés et assurer ainsi une interprétation fiable des résultats.

IV.2. Les différents types de réponse immunitaire

La deuxième partie des résultats de ma thèse porte sur l'identification de cellules exprimant des TCR associés à une réponse immunitaire discriminante. Pour cela, nous avons mis en place une stratégie d'identification des CDR3 (stratégie utilisée au laboratoire permettant de s'abstraire du biais HLA) combinée à une série de rééchantillonnage d'un patient permettant d'évaluer la capacité de classification de ces CDR3 sélectionnés. Cette stratégie ayant fait ses preuves, nous l'avons appliquée pour toutes nos hypothèses.

Tout d'abord, nous avons montré que le répertoire TCR des $CD4^+$ Tregs et des $CD4^+$ Teffs de patients atteints de MAI permettait de discriminer un état pathologique

d'un état sain mais également les MAI entre elles. Grâce aux contrôles de permutation de patients à mon avis trop peu effectués dans les études de classification, nous avons mis en évidence des différences de réponse immunitaire entre les MAI. La convergence des signatures du diabète de type 1 (DT1) est similaire à celles issues des permutations, reflétant une absence de convergence ou une trop grande hétérogénéité. La dynamique de la réponse immunitaire dans le DT1 nous offre des éléments pour expliquer cette observation. Tout d'abord, la cohorte de patients DT1 recrutée dans le cadre de cet essai se composait principalement de patients de long terme, stables dont la maladie est considérée comme peu active.

Plusieurs études démontrent que la dynamique de la réponse immunitaire chez les patients varie considérablement en fonction de la durée de la maladie. Des expansions significatives spécifiques d'auto-antigènes pancréatiques ont été détectées chez des patients ayant reçu un diagnostic récent, laissant place à une réponse plus diversifiée au fil du temps jusqu'à une absence d'auto-réactivité (basée uniquement sur les épitopes testés dans cette étude) à 1 an du diagnostic (Coppieters et al., 2012). De plus, l'étude de la recirculation des cellules diabétogéniques, observée dans le sang par rapport aux cellules pancréatiques suggère que les LT du sang périphérique sont en trop faible quantité pour caractériser le répertoire auto-immun du DT1 et quand bien même une cohorte suffisamment importante pourrait contrecarrer cette limite, il semblerait que les LT $CD4^+$ soient essentiellement restreints aux tissus avec seulement 4% de partage entre les clones LT $CD4^+$ du sang et des ganglions pancréatiques contre 20% pour les LT $CD8^+$ (Seay et al., 2016). Cependant, ces observations pourraient être modérées, car cette étude se concentre uniquement sur la chaine β du TCR et contrairement à leurs observations, Nakayama M. et al ont mis en évidence un partage d'expansion clonale de TCRα provenant LT $CD4^+$ du sang de patient DT1. Résultat également observé dans les LT $CD8^+$ montrant un partage entre trois patients d'un clonotype TCRα associé à différentes chaines β (Fuchs et al., 2017). Malgré notre étude sur les deux chaines du TCR nous n'avons pas confirmé ces observations.

En revanche, nous avons observé chez les patients atteints de polyarthrite rhumatoïde (PR) une réponse immunitaire convergente significativement supérieure aux signatures résultantes des permutations. Nous avons validé ces signatures avec des données extérieures au laboratoire et fait un lien avec des sous-populations identifiées par les auteurs comme étant associées à la maladie (bien que les études aient été faites

chez des patients d'arthrite juvénile, ils ont montré une corrélation entre la proportion de ces cellules dans le sang de patients atteints de PR et la sévérité de la maladie). D'autres études ont également détecté des perturbations dans le répertoire de patients atteints de PR (Jiang et al., 2020; Liu et al., 2019). Contrairement au DT1, la PR est une maladie dans laquelle la destruction auto-immune peut être active sur différents tissus articulaires, ce qui n'est pas le cas chez les patients DT1, dont la plupart ont une destruction totale des cellules β pancréatiques au moment du diagnostic.

Pour les autres résultats nous avons mis en évidence la capacité de prédiction fiable dans la réponse au traitement et dans la réponse face à une infection. Néanmoins, même si l'on peut s'attendre à une forte réponse immunitaire dans le contexte d'une infection par le virus Sars-Cov-2 et d'un traitement immunomodulateur, ces étapes de validation par comparaison des signatures résultant de permutations de patients sont essentielles pour différencier un marqueur (non associé à une pathologie) d'un biomarqueur.

Compte tenu des éléments énumérés ci-dessus il est concevable que dans le cadre du diabète, si les LT $CD4^+$ ne circulent pas dans le sang, car restreints à l'organe cible (Eugster et al., 2015), la stratégie d'identification d'une réponse immunitaire convergente au niveau du sang soit inefficace. De ce fait, il est nécessaire de se concentrer sur le site de l'inflammation. Deux stratégies sont envisageables, l'une serait d'analyser le répertoire TCR des ganglions et/ou des insulites pancréatiques, mais celle-ci reste difficilement réalisable ; l'autre, plus hypothétique, pourrait être que, sachant que des études ont révélé que les clones diabétogènes sont observables chez des VS (Seay et al., 2016; Yu et al., 2015), ces cellules seraient absentes de la circulation chez les patients atteints de DT1, car présentes dans le site pancréatique. Pour cela, ces cellules seraient alors appauvries dans le sang des patients diabétiques et considérées comme « enrichies » chez les VS.

De plus, nous avons constaté que l'information la plus fiable provenait des $CD4^+$ Tregs uniquement chez les patients PR. Nous pouvons imaginer que dans le DT1, nous n'avons pu saisir cette information prédictive, car au niveau sanguin, elle semble être détenue principalement par les $CD8^+$, qui semblent être une source de biomarqueur plus fiable, ce qui illustre l'importance de caractériser et discriminer les sous-population impliquées dans la pathogenèse.

IV.3. Discriminer les sous-types cellulaires pour mieux observer

La recherche de biomarqueurs nécessite d'en connaitre la source phénotypique pour en comprendre le processus biologique qui en découle. Bien que l'implication des sous-types de cellules T dans le contexte des MAI soit connu (Sakaguchi et al., 2008) la grande majorité des études se concentre sur les cellules T totales ou une simple séparation entre les $CD4^+$ et/ou les $CD8^+$ (Liu et al., 2019; Musters et al., 2018; Sui et al., 2015). Les caractéristiques anormales qui surviennent dans la MAI peuvent être négligées lorsqu'on se concentre sur un ensemble de cellules non pertinentes qui dilue le signal provenant de sous-ensembles potentiellement petits mais pertinents. Cette question est cruciale, mais il reste difficile dans l'organisation d'un essai clinique de mettre en place une séparation fine des nombreuses sous-populations phénotypiquement distinctes.

Sur la base de mon projet de thèse principalement dédié à l'étude du répertoire TCR dans le contexte des MAI, j'ai choisi de me concentrer exclusivement sur la population des LT régulateurs, bien que les propos que je vais aborder par la suite s'appliquent également à la population effectrice.

Diverses sous-populations de Tregs peuvent impliquer différents mécanismes de suppression et contrôler spécifiquement certains types de cellules effectrices (Shevach and Thornton, 2014; Wing et al., 2019). La forte hétérogénéité de cette population est souvent la source de résultats incohérents et contradictoires (Zhang et al., 2018). Ces aspects seront nécessaires pour la compréhension des mécanismes impliqués dans l'induction des MAI.

Dans notre étude sur l'analyse des répertoires TCR du sang de patients atteints de PR et de DT1 face aux VS, nous avons montré que les cellules $CD4^+$ Tregs fournissent le meilleur pouvoir de classification, notamment chez les patients PR (section III.2). Ces cellules principalement impliquées dans le contrôle des désordres auto-immuns ne représentent qu'une faible proportion des LT $CD4^+$ (5%), sans ce tri cellulaire nous n'aurions pas obtenu ces résultats. Il en est de même au sujet de l'étude sur l'impact de l'$IL\text{-}2_{fd}$ sur le répertoire TCR de patient LES (section III.3.1).

Malgré cette séparation il s'avère que de multiples phénotypes plus fins de Tregs ont été identifiés, associés à des pathologies, comme en atteste les deux études

(Rossetti et al., 2017; Spreafico et al., 2016) qui nous ont permis de valider nos signatures dans l'étude des biomarqueurs de diagnostique (section III.2). Ils ont identifié des sous-populations présentant des phénotypes d'activation, de migration et de division cellulaires. Ces $CD4^+$ Tregs (iaTregs) et ces $CD4^+$ Teffs (CPL) ne représentent qu'une faible proportion des cellules T du sang (respectivement 5-10% des $CD4^+$ Tregs et 1-2% des $CD4^+$ Teffs) mais ils ont démontré que ces sous-populations sont associés à l'activité de la maladie. D'autres études ont également montré, dans des contextes différents, que l'expression d'un seul marqueur permet d'identifier la population de Tregs d'intérêt dans un contexte physiopathologique (Kidani et al., 2022). Nous pouvons donc nous questionner sur l'hétérogénéité des phénotypes exprimant les clonotypes de nos signatures. Une des stratégies pour accéder à cette information sera discutée plus tard lorsque nous aborderons la question du *Single-cell* (SC).

Nous avons justement mentionné la population des pTregs, ces Tregs induits en périphérie, émanant de $CD4^+$ Teffs. Les répertoires TCR des $CD4^+$ Tregs et des $CD4^+$ Teffs sont en grande partie distincts, malgré un partage observé principalement entre les cellules $CD4^+$ Tregs et les $CD4^+$ Teffs auto-réactifs (Fazilleau et al., 2007; Hsieh et al., 2006; Lagattuta et al., 2022). La séparation entre les sous-populations $CD4^+$ Teffs et $CD4^+$ Tregs devrait nous permettre d'évaluer ce degré de partage à l'état « basal » dans plusieurs MAI (PR, DT1 et LES v1) ainsi que l'induction de la conversion sous l'effet de l'$IL\text{-}2_{fd}$ chez les patients LES (Zheng et al., 2007). Par la suite il serait intéressant de caractériser phénotypiquement ces pTregs et mettre ainsi en avant un des mécanismes potentiels de l'$IL\text{-}2_{fd}$. Cette conversion des répertoires TCR de populations ayant des rôles antagonistes dans la réponse immunitaire pourrait être un élément quantitatif prédictif de la réponse au traitement.

Bien que leurs implications dans le cadre des MAI n'est pas encore bien établi, plusieurs études suggèrent que les $CD8^+$ Tregs (Shimokawa et al., 2020; Tennakoon et al., 2006) contribuent à la tolérance périphérique. La réserve émise sur l'étude de cette population est probablement due à la faible proportion qu'elle semble représenter dans le sang (0,4% des cellules T) ainsi qu'à leur caractérisation encore incertaine (Churlaud et al., 2015; Mishra et al., 2021). Malgré tout, il a été montré que bien qu'ayant une sensibilité à l'IL-2 2 à 4 fois inferieure en comparaison aux cellules $CD4^+$ Tregs, les $CD8^+$ Tregs prolifèrent et s'activent plus que leurs homologues $CD4^+$ sous l'impact de

la cytokine (Churlaud et al., 2015). Il serait donc intéressant de caractériser cette sous-population dans le cadre d'essai clinique tel que LUPIL-2, et notamment évaluer l'effet combiné de ces populations régulatrices comme biomarqueurs TCR de la réponse au traitement.

A la lumière de ces exemples, nous voyons que l'hétérogénéité des populations de cellules T, phénotypiquement distinctes, constituant les répertoires TCR d'un individu, est un élément essentiel pour la compréhension des mécanismes de tolérance. Sans une séparation cellulaire en amont de l'amplification il nous est impossible avec la méthode d'analyse que nous faisons actuellement de caractériser fonctionnellement la population étudiée. C'est pourquoi les approches par analyse de cellule-unique ou *single-cell* et l'amélioration de ces techniques est un élément majeur pour la compréhension des mécanisme étudiés.

Le *single-cell* est une méthode d'isolement de cellule unique dans des microréacteurs (nanopuits ou gouttelettes) permettant l'analyse individuelle du matériel génétique (De Simone et al., 2018). Cette technologie offre désormais l'information du TCRαβ qui, en combinaison à la transcriptomique fournit une vision complète du rôle fonctionnel et physiologique des cellules T dans les réponses immunitaires (Valkiers et al., 2022). Néanmoins, à l'heure actuelle, les méthodes de production unicellulaire ne permettent d'analyser qu'une faible proportion de cellules (~10 000), adressant une limite sur la relevance des études sur les LT totaux issus du sang périphérique. En attendant le développement de technologies offrant une plus grande profondeur cellulaire, il semble essentiel de comprendre que les études de cellules « en vrac » ou *bulk* et de cellule unique ne sont pas opposées, mais complémentaires. Nous avons montré que notre signature TCR provenant de patients atteints de PR était composée de cellules iaTregs caractérisées par Rossetti (Rossetti et al., 2017). L'utilisation des marqueurs cellulaires définissant cette sous-population de $CD4^+$ Tregs pourrait nous permettre d'isoler ces cellules pour une analyse ciblée en cellule-unique et ainsi obtenir toutes les informations nécessaires pour comprendre la spécificité des clones exprimant la signature.

IV.4. De la séquence à la spécificité TCR des cellules T

La compréhension de spécificité du TCR pour son/ses antigène(s) reste l'une des tâches les plus compliquées à appréhender encore aujourd'hui. Bien que de nombreux outils de prédiction de spécificité voient le jour (section I.4.2.7) un élément essentiel est à prendre en compte, la validation. En effet, nombreux de ces algorithmes mettent à profit des séquences pour lesquelles la/les spécificité(s) sont connues pour leurs résultats/validations/apprentissages. Ces séquences listées dans les bases de données (section I.4.2.7), sont en grande majorité des séquences seulement « associées » à une maladie, c'est-à-dire trouvées dans certaines conditions expérimentales, sans validation de la reconnaissance par des tests in vitro complémentaires. C'est pourquoi, dans une perspective d'identification de spécificités, il est primordial d'effectuer une curation pour extraire uniquement celles provenant de méthodes d'identifications de la spécificité robuste (Tétramères/Dextramères).

De plus, à l'heure actuelle une proportion importante de ces séquences provient de séquençages totaux permettant d'étudier le répertoire TCR en profondeur, comme nous l'avons fait. Il n'en demeure pas moins que l'information sur l'appariement des chaînes TCRα et TCRβ est absente. Bien que certaines études mettent en avant la dominance de la chaine TRB dans la spécificité antigénique (Springer et al., 2021), d'autres soulignent l'intérêt d'étudier la chaine α (Dietrich et al., 2003; Heikkilä et al., 2021). Ainsi, dans l'objectif d'identifier non plus une signature composée de clonotypes TRA et TRB mais bien une signature de clones TCRαβ, il est nécessaire de connaitre l'hétérodimère TCRαβ, déterminant de la spécificité antigénique d'une cellule T (Carter et al., 2019). Bien que certaines études tendent à montrer qu'il est possible de prédire cette association dans un contexte HLA restreint (Mösch and Frishman, 2021), la recombinaison entre les deux chaines est, pour l'heure, considérée comme indépendante permettant d'exploiter au maximum la diversité d'association. Par ailleurs, l'association α/β peut être réalisée par inférence computationnelle sur la base des fréquences de co-occurrences des clones dans des échantillons indépendants. Différentes stratégies ont été proposées : soit en utilisant une stratégie d'échantillonnage répété d'un même échantillon (Howie et al., 2015; Lee et al., 2017) soit une stratégie longitudinale/temporelle de co-trajectoire/dynamique de clone (Minervina et al., 2020). Cette dernière approche pourrait être utilisée dans l'essai clinique longitudinal LUPIL-2 à travers la co-dynamique de nos signatures de CDR3 α

et β identifiés séparément, l'alternative étant également le séquençage *single-cell*. Ainsi la détection des signatures TCR du sang dans une analyse en cellule unique, comme proposée ci-dessus, devrait permettre de caractériser le phénotype, la fonction et la pertinence biologique de tels biomarqueurs.

Nous serons alors en mesure d'identifier les spécificités potentielles de nos TCR grâce à des outils d'inférences de spécificité (section I.4.2.7) mais aussi à des outils d'apprentissage et de modélisation structurel permettant de prédire l'interaction entre un TCR et un épitope à la surface d'un CMH (Teraguchi et al., 2020). Il nous faudra ensuite effectuer une validation en deux étapes. La première étant la validation in vitro de la spécificité des clones identifiés en utilisant des tests d'activation de TCR transduits dans des lignées rapportrices, comme décrit dans Mann et al., 2020 et également utilisé en Annexe 2, Figure 5c, face à des peptides identifiés. La seconde étant la validation de la pertinence biologique et de l'efficacité in vivo de nos clones protecteurs. Pour cela, bien qu'à ma connaissance, ceci n'ait pas encore été évalué dans les MAI, l'utilisation de modèles de souris humanisées nous permettra d'évaluer le rôle et l'efficacité protectrice des clones sur le développement des MAI (Koboziev et al., 2015). Ces preuves de concept ont déjà été abordées dans le cancer et le VIH, démontrant la faisabilité de l'approche (Carrillo et al., 2018). Ces avancées dans l'application des thérapies spécifiques aux antigènes, qui présentent un grand intérêt, pourraient être étendues aux maladies auto-immunes.

IV.5. Les thérapies spécifiques d'antigènes

Les thérapies spécifiques d'antigènes représentent un enjeu prometteur pour supprimer les cellules T effectrices et induire les cellules T régulatrices spécifiques d'auto-antigènes avec l'hypothèse qu'en ciblant un antigène particulier, ces cellules pourraient avoir une action plus efficace sans pour autant altérer l'immunité de l'hôte contre les agressions infectieuses.

L'un des moyens le plus simple est d'utiliser l'auto-antigène impliqué comme cible antigénique permettant d'expandre les $CD4^+$ Tregs spécifiques. La tolérance de ces antigènes ainsi que l'efficacité de cette méthode n'est plus à prouver chez la souris, néanmoins la translation à l'homme reste encore aujourd'hui compliquée et l'administration d'un antigène potentiellement pathogène comporte un certain risque

(Serra and Santamaria, 2019). Il a été montré que l'insuline entière et le peptide 9-23 de la chaine β de l'insuline entrainent des modifications épigénétiques dans les cellules T $CD4^+$. Alors que le peptide induit une activation des $CD4^+$ Tregs, la protéine entière active les deux populations de $CD4^+$ (Iwaszkiewicz-Grzes et al., 2021).

C'est pourquoi en parallèle à cela, plusieurs équipes s'affairent depuis 20 ans à évaluer la tolérance et l'efficacité de l'immunothérapie par injection de $CD4^+$ Tregs autologues polyclonaux expandus ex-vivo. Cette immunothérapie personnalisée a fait ses preuves dans le cadre de la maladie du greffon contre hôte pour lesquelles l'injection offrait une protection vis-à-vis de la létalité de la maladie chez la souris (Cohen et al., 2002; Taylor et al., 2002). En 2014, des résultats encourageants issu d'un essai de phase I chez l'homme ont montré la bonne tolérance associée à une rémission clinique de 8/12 patients diabétiques suite à la perfusion de Tregs polyclonaux expandus ex-vivo (Marek-Trzonkowska et al., 2014). Ces Tregs polyclonaux se sont également avérés être d'un grande stabilité et toujours détectable dans le sang de patients un an après le transfert (Bluestone et al., 2015). Thérapie évaluée plus tard chez un patient atteint d'un LES, démontrant également la tolérance et la sureté du protocole dans ce contexte pathologique (Dall'Era et al., 2019). Or en 2004, deux équipes ont montrés quasi-simultanément chez des souris NOD, qu'un nombre 4 à 20 fois inférieur de cellules Tregs spécifiques d'un antigène pancréatique en comparaison à des cellules Tregs polyclonales permettait une protection de l'apparition de la maladie comparable et même meilleure (Tang et al., 2004b; Tarbell et al., 2004), observation confirmée plus tard dans un autre contexte auto-immun (Stephens et al., 2009). Ces stratégies mettent en lumière le besoin essentiel vis-à-vis de la connaissance de la spécificité antigénique porté par le TCR des cellules T. Ces deux approches sont néanmoins difficiles à mettre en œuvre chez l'homme, car les protocoles d'isolement et d'expansion de ces cellules rares sont délicats à mettre en œuvre et nécessitent au préalable une connaissance des antigènes associés à la maladie et une compréhension de la manière dont les patients, à différents stades de la progression de la maladie, peuvent répondre.

Ainsi, le champ de bataille s'est élargi, notamment grâce à la bio-ingénierie, explorant une variété d'approches pour utiliser la spécificité des LT $CD4^+$ et notamment les $CD4^+$ Tregs dans le cadre des MAI comme des thérapies personnalisées. Tout d'abord, les premières études se sont orientées vers le transfert de gènes, en

manipulant les cellules $CD4^+$ Teffs spécifiques pour faire exprimer le facteur de transcription FOXP3, leur permettant d'acquérir un phénotype régulateur. Jaeckel et al. ont montré, chez la souris NOD, que les cellules $CD4^+$ Teffs spécifiques transduites avec le gène FOXP3, étaient activées dans les ganglions lymphatiques pancréatiques, ralentissant le développement du DT1 (Jaeckel et al., 2005). L'identification de signatures TCR, comme nous l'avons fait, provenant de $CD4^+$ Teffs permettrait d'isoler ces cellules qui sont en quantité plus importante dans le sang que les $CD4^+$ Tregs. Dans la même idée, le transfert de TCRs spécifiques directement dans une cellule $CD4^+$ Tregs (TCR-Tregs) a également montré des résultats intéressants (Kim et al., 2018). L'un des avantages de cette approche TCR-Treg en comparaison à la sélection de Tregs spécifiques est qu'elle permet de générer un plus grand nombre de cellules. De plus, Brusko et al. ont montré la possibilité de transduction d'un TCR restreint au CMH-I dans une cellule $CD4^+$ Tregs n'affectant pas la capacité de suppression spécifique de l'antigène du TCR transduit (Brusko et al., 2010). Cette technologie prometteuse doit encore être développée avant de pouvoir être appliquée à la thérapie humaine cependant contrairement aux MAI spécifiques d'organes tel que la PR et le DT1, les MAI systémiques tel que le LES, les auto-antigènes ne sont pas définis, ne permettant pas, jusqu'à présent, l'utilisation des stratégies précédemment décrites.

Pour contourner ce problème plusieurs approches peuvent être envisagées. L'une peut être, dans le cadre des MAI spécifiques sans auto-antigènes immuno-dominants de diriger les cellules vers l'organe cible plutôt que vers un antigène spécifique pas ou mal caractérisé. Les cellules pourront alors exercer indirectement leur action tolérogénique dans le site de l'inflammation (Umeshappa et al., 2020). L'autre pourrait être de s'intéresser à d'autres cibles antigéniques de l'organisme pouvant promouvoir les $CD4^+$ Tregs, tel que les bactéries de la flore intestinale. Il existe des preuves solides que les cellules T, notamment les $CD4^+$ Tregs et le microbiote intestinal sont étroitement interconnectés et liés à la physiopathologie des MAI (McLean et al., 2015). Des souris NOD élevées dans des environnements exempts de germes montrent une incidence plus élevée de diabète par rapport à celles élevées dans un environnement normal, suggérant que les microbes peuvent influencer le développement de maladies auto-immune (Wen et al., 2008). Dans le même ordre d'idées, l'absence d'exposition au microbiote vaginal maternel, dû à un accouchement par césarienne, impact la colonisation intestinale du nouveau-né conférant un risque accru de plus de 20 % de développer un DT1 (Cardwell et al., 2008). Enfin,

l'observation de différences d'expressions d'espèces bactériennes entre les microbiotes de sujets sains comparés à ceux de patients atteints de MAI fournit un élément à étudier comme source de cibles antigéniques permettant la modulation du répertoire TCR. A titre d'exemple, l'étude de Shimokawa a mis en avant une association entre l'abondance de la bactérie *Ruminococcus* et l'induction de $CD8^+$ Tregs, permettant de prévenir l'apparition du diabète chez la souris. Ces observations ont été extrapolées à l'homme à l'aide d'une corrélation positive entre la présence de la bactérie et le nombre de $CD8^+$ Tregs (Shimokawa et al., 2020). Ainsi à l'aide de test d'activation de cellules T exprimant les signatures identifiées nous pourrions déterminer des cibles antigéniques permettant de promouvoir des clones tolérogéniques.

Dans le même ordre d'idée, l'enthousiasme autour des nouvelles possibilités offertes par l'étude du microbiote intestinal sur la modulation du SI pourrait prendre une nouvelle dimension avec l'étude complémentaire du virome intestinal. Bien qu'aujourd'hui peu de chose sont connues sur le rôle du virobiote dans les MAI, il semblerait qu'il faille imaginer un axe en tri-dimensions autour du microbiote intestinal, du virobiote intestinal et du système immunitaire (Tomofuji et al., 2022)

IV.6. Conclusion générale

En résumé, ce travail m'a permis d'entrevoir le large éventail d'informations sur la compréhension du SI offert par l'étude du répertoire TCR ainsi que le chemin restant à parcourir pour une utilisation clinique. Nous avons anticipé et appréhendé les biais technologiques et méthodologiques dans l'étude de cet élément biologique. Nous avons contribué à une meilleure compréhension du rôle des récepteurs des cellules T dans le contexte des MAI et des infections. Ensemble la communauté scientifique spécialisée dans l'étude AIRR-seq avance dans la bonne direction vers l'utilisation du répertoire TCR à des fins cliniques. En témoigne l'approbation en 2021 par la *Food and Drug Administration* (FDA) du premier outil de diagnostic basé sur le TCR-seq permettant de détecter une infection récente ou antérieure par le SARS-CoV-2. (T-Detect COVID, https://www.t-detect.com/).

Finalement, bien que peu de $CD4^+$ Tregs spécifiques de l'antigène soient nécessaires pour l'immunothérapie des MAI, l'identification de TCR spécifiques reste un défi. La grande diversité des TCR, associée à la faible proportion de cellules

spécifiques circulant dans le sang à quoi s'ajoute des cibles antigéniques mal définies rend l'identification des TCR appropriés à des fins thérapeutiques ardues. Néanmoins, nous sommes à même d'imaginer dans le cadre de l'essai clinique LUPIL-2 que l'identification et la validation des clones TCR, avec connaissance de leurs cibles antigéniques pourront donner lieu à une combo-thérapie IL-2_{fd} associée à l'induction de la spécificité des Tregs protecteur par (i) injection d'auto-antigènes ; (ii) isolation et expansion ex-vivo des Tregs spécifiques ; (iii) bio-ingénierie de TCR-Treg et finalement (iv) insémination bactérienne. Ceci permettra de répondre au mieux à la grande variabilité inter- et intra-pathologies auto-immunes. La probabilité de répondre à un traitement sera évalué pour chaque patient permettant de proposer de réelles thérapies personnalisées.

Références bibliographiques

Adeegbe, D., Bayer, A.L., Levy, R.B., Malek, T.R., 2006. Cutting edge: allogeneic CD4+CD25+Foxp3+ T regulatory cells suppress autoimmunity while establishing transplantation tolerance. J. Immunol. Baltim. Md 1950 176, 7149–7153. https://doi.org/10.4049/jimmunol.176.12.7149

Adeegbe, D., Levy, R.B., Malek, T.R., 2010. Allogeneic T regulatory cell-mediated transplantation tolerance in adoptive therapy depends on dominant peripheral suppression and central tolerance. Blood 115, 1932–1940. https://doi.org/10.1182/blood-2009-08-238584

Afzal, S., Gil-Farina, I., Gabriel, R., Ahmad, S., von Kalle, C., Schmidt, M., Fronza, R., 2019. Systematic comparative study of computational methods for T-cell receptor sequencing data analysis. Brief. Bioinform. 20, 222–234. https://doi.org/10.1093/bib/bbx111

Ahmed, R., Akondy, R.S., 2011. Insights into human CD8(+) T-cell memory using the yellow fever and smallpox vaccines. Immunol. Cell Biol. 89, 340–345. https://doi.org/10.1038/icb.2010.155

Alexander, T., Sattler, A., Templin, L., Kohler, S., Groß, C., Meisel, A., Sawitzki, B., Burmester, G.-R., Arnold, R., Radbruch, A., Thiel, A., Hiepe, F., 2013. Foxp3+ Helios+ regulatory T cells are expanded in active systemic lupus erythematosus. Ann. Rheum. Dis. 72, 1549–1558. https://doi.org/10.1136/annrheumdis-2012-202216

Alpert, A., Pickman, Y., Leipold, M., Rosenberg-Hasson, Y., Ji, X., Gaujoux, R., Rabani, H., Starosvetsky, E., Kveler, K., Schaffert, S., Furman, D., Caspi, O., Rosenschein, U., Khatri, P., Dekker, C.L., Maecker, H.T., Davis, M.M., Shen-Orr, S.S., 2019. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. Nat. Med. 25, 487–495. https://doi.org/10.1038/s41591-019-0381-y

Anderson, M.S., Venanzi, E.S., Klein, L., Chen, Z., Berzins, S.P., Turley, S.J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C., Mathis, D., 2002. Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. Science 298, 1395–1401. https://doi.org/10.1126/science.1075958

Andrews, S., 2015. FastQC : A quality control tool for high throughput sequence data.

Archbold, J.K., Macdonald, W.A., Gras, S., Ely, L.K., Miles, J.J., Bell, M.J., Brennan, R.M., Beddoe, T., Wilce, M.C.J., Clements, C.S., Purcell, A.W., McCluskey, J., Burrows, S.R., Rossjohn, J., 2009. Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. J. Exp. Med. 206, 209–219. https://doi.org/10.1084/jem.20082136

Armstrong, K.M., Piepenbrink, K.H., Baker, B.M., 2008. Conformational changes and flexibility in T-cell receptor recognition of peptide-MHC complexes. Biochem. J. 415, 183–196. https://doi.org/10.1042/BJ20080850

Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., Kourilsky, P., 1999. A direct estimate of the human alphabeta T cell receptor diversity. Science 286, 958–961. https://doi.org/10.1126/science.286.5441.958

Asano, M., Toda, M., Sakaguchi, N., Sakaguchi, S., 1996. Autoimmune disease as a consequence of developmental abnormality of a T cell subpopulation. J. Exp. Med. 184, 387–396. https://doi.org/10.1084/jem.184.2.387

Bach, J.-F., 2005. Infections and autoimmune diseases. J. Autoimmun. 25 Suppl, 74–80. https://doi.org/10.1016/j.jaut.2005.09.024

Bach, J.-F., 2002. The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. N. Engl. J. Med. 347, 911–920. https://doi.org/10.1056/NEJMra020100

Bagaev, D.V., Vroomans, R.M.A., Samir, J., Stervbo, U., Rius, C., Dolton, G., Greenshields-Watson, A., Attaf, M., Egorov, E.S., Zvyagin, I.V., Babel, N., Cole, D.K., Godkin, A.J., Sewell, A.K., Kesmir, C., Chudakov, D.M., Luciani, F., Shugay, M., 2020. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res. 48, D1057–D1062. https://doi.org/10.1093/nar/gkz874

Bálint, Š., Müller, S., Fischer, R., Kessler, B.M., Harkiolaki, M., Valitutti, S., Dustin, M.L., 2020. Supramolecular attack particles are autonomous killing entities released from cytotoxic T cells. Science 368, 897–901. https://doi.org/10.1126/science.aay9207

Banchereau, J., Steinman, R.M., 1998. Dendritic cells and the control of immunity. Nature 392, 245–252. https://doi.org/10.1038/32588

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

Basile, G. de S., Geissmann, F., Flori, E., Uring-Lambert, B., Soudais, C., Cavazzana-Calvo, M., Durandy, A., Jabado, N., Fischer, A., Deist, F.L., 2004. Severe combined immunodeficiency caused by deficiency in either the δ or the ε subunit of CD3. J. Clin. Invest. 114, 1512–1517. https://doi.org/10.1172/JCI200422588

Baumgartner, C.K., Yagita, H., Malherbe, L.P., 2012. A TCR affinity threshold regulates memory CD4 T cell differentiation following vaccination. J. Immunol. Baltim. Md 1950 189, 2309–2317. https://doi.org/10.4049/jimmunol.1200453

Bautista, J.L., Lio, C.-W.J., Lathrop, S.K., Forbush, K., Liang, Y., Luo, J., Rudensky, A.Y., Hsieh, C.-S., 2009. Intraclonal competition limits the fate determination of regulatory T cells in the thymus. Nat. Immunol. 10, 610–617. https://doi.org/10.1038/ni.1739

Bertoletti, A., Ferrari, C., 2016. Adaptive immunity in HBV infection. J. Hepatol. 64, S71–S83. https://doi.org/10.1016/j.jhep.2016.01.026

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. Sci. Transl. Med. 12, eaaz3738. https://doi.org/10.1126/scitranslmed.aaz3738

Bjornevik, K., Cortese, M., Healy, B.C., Kuhle, J., Mina, M.J., Leng, Y., Elledge, S.J., Niebuhr, D.W., Scher, A.I., Munger, K.L., Ascherio, A., 2022. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. Science 375, 296–301. https://doi.org/10.1126/science.abj8222

Bluestone, J.A., Buckner, J.H., Fitch, M., Gitelman, S.E., Gupta, S., Hellerstein, M.K., Herold, K.C., Lares, A., Lee, M.R., Li, K., Liu, W., Long, S.A., Masiello, L.M., Nguyen, V., Putnam, A.L., Rieck, M., Sayre, P.H., Tang, Q., 2015. Type 1 diabetes immunotherapy using polyclonal regulatory T cells. Sci. Transl. Med. 7, 315ra189. https://doi.org/10.1126/scitranslmed.aad4134

Boissonnas, A., Scholer-Dahirel, A., Simon-Blancal, V., Pace, L., Valet, F., Kissenpfennig, A., Sparwasser, T., Malissen, B., Fetler, L., Amigorena, S., 2010. Foxp3+ T Cells Induce Perforin-Dependent Dendritic Cell Death in Tumor-Draining Lymph Nodes. Immunity 32, 266–278. https://doi.org/10.1016/j.immuni.2009.11.015

Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., Chudakov, D.M., 2015. MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods 12, 380–381. https://doi.org/10.1038/nmeth.3364

Bonelli, M., Savitskaya, A., von Dalwigk, K., Steiner, C.W., Aletaha, D., Smolen, J.S., Scheinecker, C., 2008. Quantitative and qualitative deficiencies of regulatory T cells in patients with systemic lupus erythematosus (SLE). Int. Immunol. 20, 861–868. https://doi.org/10.1093/intimm/dxn044

Bopp, T., Becker, C., Klein, M., Klein-Hessling, S., Palmetshofer, A., Serfling, E., Heib, V., Becker, M., Kubach, J., Schmitt, S., Stoll, S., Schild, H., Staege, M.S., Stassen, M., Jonuleit, H., Schmitt, E., 2007. Cyclic adenosine monophosphate is a key component of regulatory T cell-mediated suppression. J. Exp. Med. 204, 1303–1310. https://doi.org/10.1084/jem.20062129

Borbulevych, O.Y., Piepenbrink, K.H., Gloor, B.E., Scott, D.R., Sommese, R.F., Cole, D.K., Sewell, A.K., Baker, B.M., 2009. T Cell Receptor Cross-reactivity Directed by Antigen-Dependent Tuning of Peptide-MHC Molecular Flexibility. Immunity 31, 885–896. https://doi.org/10.1016/j.immuni.2009.11.003

Borg, N.A., Ely, L.K., Beddoe, T., Macdonald, W.A., Reid, H.H., Clements, C.S., Purcell, A.W., Kjer-Nielsen, L., Miles, J.J., Burrows, S.R., McCluskey, J., Rossjohn, J., 2005. The CDR3 regions of an immunodominant T cell receptor dictate the "energetic landscape" of peptide-MHC recognition. Nat. Immunol. 6, 171–180. https://doi.org/10.1038/ni1155

Bortone, D.S., Woodcock, M.G., Parker, J.S., Vincent, B.G., 2021. Improved T-cell Receptor Diversity Estimates Associate with Survival and Response to Anti–PD-1 Therapy. Cancer Immunol. Res. 9, 103–112. https://doi.org/10.1158/2326-6066.CIR-20-0398

Bottazzo, G.F., Dean, B.M., McNally, J.M., MacKay, E.H., Swift, P.G., Gamble, D.R., 1985. In situ characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulitis. N. Engl. J. Med. 313, 353–360. https://doi.org/10.1056/NEJM198508083130604

Boudinot, P., Marriotti-Ferrandiz, M.E., Pasquier, L.D., Benmansour, A., Cazenave, P.-A., Six, A., 2008. New perspectives for large-scale repertoire analysis of immune receptors. Mol. Immunol. 45, 2437–2445. https://doi.org/10.1016/j.molimm.2007.12.018

Bougneres, P.F., Carel, J.C., Castano, L., Boitard, C., Gardin, J.P., Landais, P., Hors, J., Mihatsch, M.J., Paillard, M., Chaussain, J.L., 1988. Factors associated with early remission of type I diabetes in children treated with cyclosporine. N. Engl. J. Med. 318, 663–670. https://doi.org/10.1056/NEJM198803173181103

Boulé, J.B., Rougeon, F., Papanicolaou, C., 2001. Terminal deoxynucleotidyl transferase indiscriminately incorporates ribonucleotides and deoxyribonucleotides. J. Biol. Chem. 276, 31388–31393. https://doi.org/10.1074/jbc.M105272200

Bousso, P., Wahn, V., Douagi, I., Horneff, G., Pannetier, C., Le Deist, F., Zepp, F., Niehues, T., Kourilsky, P., Fischer, A., de Saint Basile, G., 2000. Diversity, functionality, and stability of the T cell repertoire derived *in vivo* from a single human T cell precursor. Proc. Natl. Acad. Sci. 97, 274–278. https://doi.org/10.1073/pnas.97.1.274

Boyer, O., Saadoun, D., Abriol, J., Dodille, M., Piette, J.-C., Cacoub, P., Klatzmann, D., 2004. CD4+CD25+ regulatory T-cell deficiency in patients with hepatitis C-mixed cryoglobulinemia vasculitis. Blood 103, 3428–3430. https://doi.org/10.1182/blood-2003-07-2598

Boyman, O., Surh, C.D., Sprent, J., 2006. Potential use of IL-2/anti-IL-2 antibody immune complexes for the treatment of cancer and autoimmune disease. Expert Opin. Biol. Ther. 6, 1323–1331. https://doi.org/10.1517/14712598.6.12.1323

Brady, B.L., Steinel, N.C., Bassing, C.H., 2010. Antigen Receptor Allelic Exclusion: An Update and Reappraisal. J. Immunol. 185, 3801–3808. https://doi.org/10.4049/jimmunol.1001158

Britanova, O.V., Putintseva, E.V., Shugay, M., Merzlyak, E.M., Turchaninova, M.A., Staroverov, D.B., Bolotin, D.A., Lukyanov, S., Bogdanova, E.A., Mamedov, I.Z., Lebedev, Y.B., Chudakov, D.M., 2014. Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling. J. Immunol. 192, 2689–2698. https://doi.org/10.4049/jimmunol.1302064

Brogdon, J.L., Leitenberg, D., Bottomly, K., 2002. The potency of TCR signaling differentially regulates NFATc/p activity and early IL-4 transcription in naive CD4+ T cells. J. Immunol. Baltim. Md 1950 168, 3825–3832. https://doi.org/10.4049/jimmunol.168.8.3825

Brusko, T.M., Koya, R.C., Zhu, S., Lee, M.R., Putnam, A.L., McClymont, S.A., Nishimura, M.I., Han, S., Chang, L.-J., Atkinson, M.A., Ribas, A., Bluestone, J.A., 2010. Human Antigen-Specific Regulatory T Cells Generated by T Cell Receptor Gene Transfer. PLoS ONE 5, e11726. https://doi.org/10.1371/journal.pone.0011726

Cabaniols, J.-P., Fazilleau, N., Casrouge, A., Kourilsky, P., Kanellopoulos, J.M., 2001. Most α/β T Cell Receptor Diversity Is Due to Terminal Deoxynucleotidyl Transferase. J. Exp. Med. 194, 1385–1390. https://doi.org/10.1084/jem.194.9.1385

Cabello-Kindelan, C., Mackey, S., Sands, A., Rodriguez, J., Vazquez, C., Pugliese, A., Bayer, A.L., 2020. Immunomodulation Followed by Antigen-Specific Treg Infusion Controls Islet Autoimmunity. Diabetes 69, 215–227. https://doi.org/10.2337/db19-0061

Cappuccilli, M., Donati, G., Comai, G., Baraldi, O., Conte, D., Capelli, I., Aiello, V., Pession, A., La Manna, G., 2017. Identification of expanded T-cell clones by spectratyping in nonfunctioning kidney transplants. J. Inflamm. Res. Volume 10, 41–47. https://doi.org/10.2147/JIR.S124944

Cardwell, C.R., Stene, L.C., Joner, G., Cinek, O., Svensson, J., Goldacre, M.J., Parslow, R.C., Pozzilli, P., Brigis, G., Stoyanov, D., Urbonaite, B., Sipetić, S., Schober, E., Ionescu-Tirgoviste, C., Devoti, G., de Beaufort, C.E., Buschard, K., Patterson, C.C., 2008. Caesarean section is associated with an increased risk of childhood-onset type 1 diabetes mellitus: a meta-analysis of observational studies. Diabetologia 51, 726–735. https://doi.org/10.1007/s00125-008-0941-z

Carrillo, M.A., Zhen, A., Kitchen, S.G., 2018. The Use of the Humanized Mouse Model in Gene Therapy and Immunotherapy for HIV and Cancer. Front. Immunol. 9, 746. https://doi.org/10.3389/fimmu.2018.00746

Carter, J.A., Preall, J.B., Grigaityte, K., Goldfless, S.J., Jeffery, E., Briggs, A.W., Vigneault, F., Atwal, G.S., 2019. Single T Cell Sequencing Demonstrates the Functional Role of αβ TCR Pairing in Cell Lineage and Antigen Specificity. Front. Immunol. 10, 1516. https://doi.org/10.3389/fimmu.2019.01516

Chaudhry, A., Samstein, R.M., Treuting, P., Liang, Y., Pils, M.C., Heinrich, J.-M., Jack, R.S., Wunderlich, F.T., Brüning, J.C., Müller, W., Rudensky, A.Y., 2011. Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. Immunity 34, 566–578. https://doi.org/10.1016/j.immuni.2011.03.018

Chen, E.L.Y., Brauer, P.M., Martinez, E.C., Huang, X., Yu, N., Anderson, M.K., Li, Y., Zúñiga-Pflücker, J.C., 2021. Cutting Edge: TCR-β Selection Is Required at the CD4+CD8+ Stage of Human T Cell Development. J. Immunol. Baltim. Md 1950 206, 2271–2276. https://doi.org/10.4049/jimmunol.2100141

Chen, W., Jin, W., Hardegen, N., Lei, K.-J., Li, L., Marinos, N., McGrady, G., Wahl, S.M., 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. https://doi.org/10.1084/jem.20030152

Chien, Y.H., Gascoigne, N.R., Kavaler, J., Lee, N.E., Davis, M.M., 1984. Somatic recombination in a murine T-cell receptor gene. Nature 309, 322–326. https://doi.org/10.1038/309322a0

Chiou, S.-H., Tseng, D., Reuben, A., Mallajosyula, V., Molina, I.S., Conley, S., Wilhelmy, J., McSween, A.M., Yang, X., Nishimiya, D., Sinha, R., Nabet, B.Y., Wang, C., Shrager, J.B., Berry, M.F., Backhus, L., Lui, N.S., Wakelee, H.A., Neal, J.W., Padda, S.K., Berry, G.J., Delaidelli, A., Sorensen, P.H., Sotillo, E., Tran, P., Benson, J.A., Richards, R., Labanieh, L., Klysz, D.D., Louis, D.M., Feldman, S.A., Diehn, M., Weissman, I.L., Zhang, J., Wistuba, I.I., Futreal, P.A., Heymach, J.V., Garcia, K.C., Mackall, C.L., Davis, M.M., 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. https://doi.org/10.1016/j.immuni.2021.02.014

Choi, J.-Y., Ho, J.H., Pasoto, S.G., Bunin, V., Kim, S.T., Carrasco, S., Borba, E.F., Gonçalves, C.R., Costa, P.R., Kallas, E.G., Bonfa, E., Craft, J., 2015. Circulating follicular helper-like T cells in systemic lupus erythematosus: association with disease activity. Arthritis Rheumatol. Hoboken NJ 67, 988–999. https://doi.org/10.1002/art.39020

Chothia, C., Boswell, D.R., Lesk, A.M., 1988. The outline structure of the T-cell alpha beta receptor. EMBO J. 7, 3745–3755.

Christophersen, A., Ráki, M., Bergseng, E., Lundin, K.E., Jahnsen, J., Sollid, L.M., Qiao, S.-W., 2014. Tetramer-visualized gluten-specific CD4+ T cells in blood as a potential diagnostic marker for coeliac disease without oral gluten challenge. United Eur. Gastroenterol. J. 2, 268–278. https://doi.org/10.1177/2050640614540154

Chronister, W.D., Crinklaw, A., Mahajan, S., Vita, R., Koşaloğlu-Yalçın, Z., Yan, Z., Greenbaum, J.A., Jessen, L.E., Nielsen, M., Christley, S., Cowell, L.G., Sette, A., Peters, B., 2021. TCRMatch: Predicting T-Cell Receptor Specificity Based on Sequence Similarity to Previously Characterized Receptors. Front. Immunol. 12, 640725. https://doi.org/10.3389/fimmu.2021.640725

Churlaud, G., Pitoiset, F., Jebbawi, F., Lorenzon, R., Bellier, B., Rosenzwajg, M., Klatzmann, D., 2015. Human and Mouse CD8+CD25+FOXP3+ Regulatory T Cells at Steady State and during Interleukin-2 Therapy. Front. Immunol. 6. https://doi.org/10.3389/fimmu.2015.00171

Ciupe, S.M., Devlin, B.H., Markert, M.L., Kepler, T.B., 2013. Quantification of total T-cell receptor diversity by flow cytometry and spectratyping. BMC Immunol. 14, 35. https://doi.org/10.1186/1471-2172-14-35

Codina-Busqueta, E., Scholz, E., Muñoz-Torres, P.M., Roura-Mir, C., Costa, M., Xufré, C., Planas, R., Vives-Pi, M., Jaraquemada, D., Martí, M., 2011. TCR Bias of In Vivo Expanded T Cells in Pancreatic Islets and Spleen at the Onset in Human Type 1 Diabetes. J. Immunol. 186, 3787–3797. https://doi.org/10.4049/jimmunol.1002423

Cohen, J.L., Trenado, A., Vasey, D., Klatzmann, D., Salomon, B.L., 2002. CD4(+)CD25(+) immunoregulatory T Cells: new therapeutics for graft-versus-host disease. J. Exp. Med. 196, 401–406. https://doi.org/10.1084/jem.20020090

Collison, L.W., Workman, C.J., Kuo, T.T., Boyd, K., Wang, Y., Vignali, K.M., Cross, R., Sehy, D., Blumberg, R.S., Vignali, D.A.A., 2007. The inhibitory cytokine IL-35 contributes to regulatory T-cell function. Nature 450, 566–569. https://doi.org/10.1038/nature06306

Coppieters, K.T., Dotta, F., Amirian, N., Campbell, P.D., Kay, T.W.H., Atkinson, M.A., Roep, B.O., von Herrath, M.G., 2012. Demonstration of islet-autoreactive CD8 T cells in insulitic lesions from recent onset and long-term type 1 diabetes patients. J. Exp. Med. 209, 51–60. https://doi.org/10.1084/jem.20111187

Corneo, B., Moshous, D., Güngör, T., Wulffraat, N., Philippet, P., Le Deist, F.L., Fischer, A., de Villartay, J.P., 2001. Identical mutations in RAG1 or RAG2 genes leading to defective V(D)J recombinase activity can cause either T-B-severe combined immune deficiency or Omenn syndrome. Blood 97, 2772–2776. https://doi.org/10.1182/blood.v97.9.2772

Corthay, A., Nandakumar, K.S., Holmdahl, R., 2001. Evaluation of the percentage of peripheral T cells with two different T cell receptor alpha-chains and of their potential role in autoimmunity. J. Autoimmun. 16, 423–429. https://doi.org/10.1006/jaut.2001.0504

Costa, N., Pires, A.E., Gabriel, A.M., Goulart, L.F., Pereira, C., Leal, B., Queiros, A.C., Chaara, W., Moraes-Fontes, M.F., Vasconcelos, C., Ferreira, C., Martins, J., Bastos, M., Santos, M.J., Pereira, M.A., Martins, B., Lima, M., João, C., Six, A., Demengeot, J., Fesel, C., 2013. Broadened T-cell repertoire diversity in ivIg-treated SLE patients is also related to the individual status of regulatory T-cells. J. Clin. Immunol. 33, 349–360. https://doi.org/10.1007/s10875-012-9816-7

Cutolo, M., Capellino, S., Sulli, A., Serioli, B., Secchi, M.E., Villaggio, B., Straub, R.H., 2006. Estrogens and autoimmune diseases. Ann. N. Y. Acad. Sci. 1089, 538–547. https://doi.org/10.1196/annals.1386.043

D'Agati, V.D., Appel, G.B., Estes, D., Knowles, D.M., Pirani, C.L., 1986. Monoclonal antibody identification of infiltrating mononuclear leukocytes in lupus nephritis. Kidney Int. 30, 573–581. https://doi.org/10.1038/ki.1986.223

Dall'Era, M., Pauli, M.L., Remedios, K., Taravati, K., Sandova, P.M., Putnam, A.L., Lares, A., Haemel, A., Tang, Q., Hellerstein, M., Fitch, M., McNamara, J., Welch, B., Bluestone, J.A., Wofsy, D., Rosenblum, M.D., Autoimmunity Centers of Excellence, 2019. Adoptive Treg Cell Therapy in a Patient With Systemic Lupus Erythematosus. Arthritis Rheumatol. Hoboken NJ 71, 431–440. https://doi.org/10.1002/art.40737

Danska, J.S., Livingstone, A.M., Paragas, V., Ishihara, T., Fathman, C.G., 1990. The presumptive CDR3 regions of both T cell receptor alpha and beta chains determine T cell specificity for myoglobin peptides. J. Exp. Med. 172, 27–33. https://doi.org/10.1084/jem.172.1.27

Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H.O., Kedzierska, K., La Gruta, N.L., Bradley, P., Thomas, P.G., 2017. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature 547, 89–93. https://doi.org/10.1038/nature22383

Davis, M.M., Bjorkman, P.J., 1988. T-cell antigen receptor genes and T-cell recognition. Nature 334, 395–402. https://doi.org/10.1038/334395a0

De Neuter, N., Bartholomeus, E., Elias, G., Keersmaekers, N., Suls, A., Jansens, H., Smits, E., Hens, N., Beutels, P., Van Damme, P., Mortier, G., Van Tendeloo, V., Laukens, K., Meysman, P., Ogunjimi, B., 2019. Memory CD4+ T cell receptor repertoire data mining as a tool for identifying cytomegalovirus serostatus. Genes Immun. 20, 255–260. https://doi.org/10.1038/s41435-018-0035-y

De Simone, M., Rossetti, G., Pagani, M., 2018. Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges. Front. Immunol. 9, 1638. https://doi.org/10.3389/fimmu.2018.01638

Delves, P.J., Roitt, I.M., 2000. The Immune System. N. Engl. J. Med. 343, 37–49. https://doi.org/10.1056/NEJM200007063430107

Di Sante, G., Tolusso, B., Fedele, A.L., Gremese, E., Alivernini, S., Nicolò, C., Ria, F., Ferraccioli, G., 2015. Collagen Specific T-Cell Repertoire and HLA-DR Alleles: Biomarkers of Active Refractory Rheumatoid Arthritis. EBioMedicine 2, 2037–2045. https://doi.org/10.1016/j.ebiom.2015.11.019

Dietrich, P.-Y., Le Gal, F.-A., Dutoit, V., Pittet, M.J., Trautman, L., Zippelius, A., Cognet, I., Widmer, V., Walker, P.R., Michielin, O., Guillaume, P., Connerotte, T., Jotereau, F., Coulie, P.G., Romero, P., Cerottini, J.-C., Bonneville, M., Valmori, D., 2003. Prevalent Role of TCR α-Chain in the Selection of the Preimmune Repertoire Specific for a Human Tumor-Associated Self-Antigen. J. Immunol. 170, 5103–5109. https://doi.org/10.4049/jimmunol.170.10.5103

Ding, Y.H., Smith, K.J., Garboczi, D.N., Utz, U., Biddison, W.E., Wiley, D.C., 1998. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. Immunity 8, 403–411. https://doi.org/10.1016/s1074-7613(00)80546-4

Elhanati, Y., Sethna, Z., Callan, C.G., Mora, T., Walczak, A.M., 2018. Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. Immunol. Rev. 284, 167–179. https://doi.org/10.1111/imr.12665

Elliott, J.I., Altmann, D.M., 1995. Dual T cell receptor alpha chain T cells in autoimmunity. J. Exp. Med. 182, 953–959. https://doi.org/10.1084/jem.182.4.953

Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., Rieder, M., Robins, H.S., 2017. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat. Genet. 49, 659–665. https://doi.org/10.1038/ng.3822

Eugster, A., Lindner, A., Catani, M., Heninger, A.-K., Dahl, A., Klemroth, S., Kühn, D., Dietz, S., Bickle, M., Ziegler, A.-G., Bonifacio, E., 2015. High Diversity in the TCR Repertoire of GAD65 Autoantigen-Specific Human $CD4^+$ T Cells. J. Immunol. 194, 2531–2538. https://doi.org/10.4049/jimmunol.1403031

Fazilleau, N., Bachelez, H., Gougeon, M.-L., Viguier, M., 2007. Cutting Edge: Size and Diversity of $CD4^+$ $CD25^{high}$ $Foxp3^+$ Regulatory T Cell Repertoire in Humans: Evidence for Similarities and Partial Overlapping with $CD4^+$ $CD25^-$ T Cells. J. Immunol. 179, 3412–3416. https://doi.org/10.4049/jimmunol.179.6.3412

Fisson, S., Darrasse-Jèze, G., Litvinova, E., Septier, F., Klatzmann, D., Liblau, R., Salomon, B.L., 2003. Continuous activation of autoreactive CD4+ CD25+ regulatory T cells in the steady state. J. Exp. Med. 198, 737–746. https://doi.org/10.1084/jem.20030686

Fontenot, J.D., Gavin, M.A., Rudensky, A.Y., 2003. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nat. Immunol. 4, 330–336. https://doi.org/10.1038/ni904

Franklyn, K., Hoi, A., Nikpour, M., Morand, E.F., 2014. The need to define treatment goals for systemic lupus erythematosus. Nat. Rev. Rheumatol. 10, 567–571. https://doi.org/10.1038/nrrheum.2014.118

Froelich, C.J., Orth, K., Turbov, J., Seth, P., Gottlieb, R., Babior, B., Shah, G.M., Bleackley, R.C., Dixit, V.M., Hanna, W., 1996. New paradigm for lymphocyte granule-mediated cytotoxicity. Target cells bind and internalize granzyme B, but an endosomolytic agent is necessary for cytosolic delivery and subsequent apoptosis. J. Biol. Chem. 271, 29073–29079. https://doi.org/10.1074/jbc.271.46.29073

Fuchs, Y.F., Eugster, A., Dietz, S., Sebelefsky, C., Kühn, D., Wilhelm, C., Lindner, A., Gavrisan, A., Knoop, J., Dahl, A., Ziegler, A.-G., Bonifacio, E., 2017. CD8+ T cells specific for the islet autoantigen IGRP are restricted in their T cell receptor chain usage. Sci. Rep. 7, 44661. https://doi.org/10.1038/srep44661

Fujinami, R.S., Oldstone, M.B., Wroblewska, Z., Frankel, M.E., Koprowski, H., 1983. Molecular mimicry in virus infection: crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. Proc. Natl. Acad. Sci. U. S. A. 80, 2346–2350. https://doi.org/10.1073/pnas.80.8.2346

Furie, R.A., Petri, M.A., Wallace, D.J., Ginzler, E.M., Merrill, J.T., Stohl, W., Chatham, W.W., Strand, V., Weinstein, A., Chevrier, M.R., Zhong, Z.J., Freimuth, W.W., 2009. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Rheum. 61, 1143–1151. https://doi.org/10.1002/art.24698

Garcia, K.C., Teyton, L., Wilson, I.A., 1999. Structural basis of T cell recognition. Annu. Rev. Immunol. 17, 369–397. https://doi.org/10.1146/annurev.immunol.17.1.369

Germain, R.N., 1994. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. Cell 76, 287–299. https://doi.org/10.1016/0092-8674(94)90336-0

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. Front. Immunol. 10, 2820. https://doi.org/10.3389/fimmu.2019.02820

Gittelman, R.M., Lavezzo, E., Snyder, T.M., Zahid, H.J., Elyanow, R., Dalai, S., Kirsch, I., Baldo, L., Manuto, L., Franchin, E., Vecchio, C.D., Pacenti, M., Boldrin, C., Cattai, M., Saluzzo, F., Padoan, A., Plebani, M., Simeoni, F., Bordini, J., Lorè, N.I., Lazarevic, D., Cirillo, D.M., Ghia, P., Toppo, S., Carlson, J.M., Robins, H.S., Tonon, G., Crisanti, A., 2020. Diagnosis and Tracking of SARS-CoV-2 Infection By T-Cell Receptor Sequencing (preprint). Infectious Diseases (except HIV/AIDS). https://doi.org/10.1101/2020.11.09.20228023

Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., Haas, N., Arlehamn, C.S.L., Sette, A., Boyd, S.D., Scriba, T.J., Martinez, O.M., Davis, M.M., 2017. Identifying specificity groups in the T cell receptor repertoire. Nature 547, 94–98. https://doi.org/10.1038/nature22976

Gomez-Tourino, I., Kamra, Y., Baptista, R., Lorenc, A., Peakman, M., 2017. T cell receptor β-chains display abnormal shortening and repertoire sharing in type 1 diabetes. Nat. Commun. 8, 1792. https://doi.org/10.1038/s41467-017-01925-2

Gough, S.C.L., Simmonds, M.J., 2007. The HLA Region and Autoimmune Disease: Associations and Mechanisms of Action. Curr. Genomics 8, 453–465. https://doi.org/10.2174/138920207783591690

Grossman, W.J., Verbsky, J.W., Tollefsen, B.L., Kemper, C., Atkinson, J.P., Ley, T.J., 2004. Differential expression of granzymes A and B in human cytotoxic lymphocyte subsets and T regulatory cells. Blood 104, 2840–2848. https://doi.org/10.1182/blood-2004-03-0859

Gui, J., Mustachio, L.M., Su, D.-M., Craig, R.W., 2012. Thymus Size and Age-related Thymic Involution: Early Programming, Sexual Dimorphism, Progenitors and Stroma. Aging Dis. 3, 280–290.

Guo, L., Wang, G., Wang, Yeming, Zhang, Q., Ren, L., Gu, X., Huang, T., Zhong, J., Wang, Ying, Wang, X., Huang, L., Xu, L., Wang, C., Chen, L., Xiao, X., Peng, Y., Knight, J.C., Dong, T., Cao, B., Wang, J., 2022. SARS-CoV-2-specific antibody and T-cell responses 1 year after infection in people recovered from COVID-19: a longitudinal cohort study. Lancet Microbe 3, e348–e356. https://doi.org/10.1016/S2666-5247(22)00036-2

Harty, J.T., Tvinnereim, A.R., White, D.W., 2000. CD8+ T cell effector mechanisms in resistance to infection. Annu. Rev. Immunol. 18, 275–308. https://doi.org/10.1146/annurev.immunol.18.1.275

Haskins, K., McDuffie, M., 1990. Acceleration of diabetes in young NOD mice with a CD4+ islet-specific T cell clone. Science 249, 1433–1436. https://doi.org/10.1126/science.2205920

Hassler, T., Urmann, E., Teschner, S., Federle, C., Dileepan, T., Schober, K., Jenkins, M.K., Busch, D.H., Hinterberger, M., Klein, L., 2019. Inventories of naive and tolerant mouse CD4 T cell repertoires reveal a hierarchy of deleted and diverted T cell receptors. Proc. Natl. Acad. Sci. 116, 18537–18543. https://doi.org/10.1073/pnas.1907615116

He, J., Zhang, R., Shao, M., Zhao, X., Miao, M., Chen, J., Liu, J., Zhang, Xiaoying, Zhang, Xia, Jin, Y., Wang, Y., Zhang, S., Zhu, L., Jacob, A., Jia, R., You, X., Li, X., Li, C., Zhou, Y., Yang, Y., Ye, H., Liu, Y., Su, Y., Shen, N., Alexander, J., Guo, J., Ambrus, J., Lin, X., Yu, D., Sun, X., Li, Z., 2020. Efficacy and safety of low-dose IL-2 in the treatment of systemic lupus erythematosus: a randomised, double-blind, placebo-controlled trial. Ann. Rheum. Dis. 79, 141. https://doi.org/10.1136/annrheumdis-2019-215396

He, J., Zhang, Xia, Wei, Y., Sun, X., Chen, Y., Deng, J., Jin, Y., Gan, Y., Hu, X., Jia, R., Xu, C., Hou, Z., Leong, Y.A., Zhu, L., Feng, J., An, Y., Jia, Y., Li, C., Liu, X., Ye, H., Ren, L., Li, R., Yao, H., Li, Y., Chen, S., Zhang, Xuewu, Su, Y., Guo, J., Shen, N., Morand, E.F., Yu, D., Li, Z., 2016. Low-dose interleukin-2 treatment selectively modulates CD4(+) T cell subsets in patients with systemic lupus erythematosus. Nat. Med. 22, 991–993. https://doi.org/10.1038/nm.4148

Heather, J.M., Ismail, M., Oakes, T., Chain, B., 2017. High-throughput sequencing of the T-cell receptor repertoire: pitfalls and opportunities. Brief. Bioinform. bbw138. https://doi.org/10.1093/bib/bbw138

Heikkilä, N., Sormunen, S., Mattila, J., Härkönen, T., Knip, M., Ihantola, E.-L., Kinnunen, T., Mattila, I.P., Saramäki, J., Arstila, T.P., 2021. Generation of self-reactive, shared T-cell receptor α chains in the human thymus. J. Autoimmun. 119, 102616. https://doi.org/10.1016/j.jaut.2021.102616

Hellerstein, M., 2020. What are the roles of antibodies versus a durable, high quality T-cell response in protective immunity against SARS-CoV-2? Vaccine X 6, 100076. https://doi.org/10.1016/j.jvacx.2020.100076

Holler, P.D., Kranz, D.M., 2003. Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. Immunity 18, 255–264. https://doi.org/10.1016/s1074-7613(03)00019-0

Hori, S., Nomura, T., Sakaguchi, S., 2003. Control of regulatory T cell development by the transcription factor Foxp3. Science 299, 1057–1061. https://doi.org/10.1126/science.1079490

Howie, B., Sherwood, A.M., Berkebile, A.D., Berka, J., Emerson, R.O., Williamson, D.W., Kirsch, I., Vignali, M., Rieder, M.J., Carlson, C.S., Robins, H.S., 2015. High-throughput pairing of T cell receptor α and β sequences. Sci. Transl. Med. 7. https://doi.org/10.1126/scitranslmed.aac5624

Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J.D., Rudensky, A.Y., 2006. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. Nat. Immunol. 7, 401–410. https://doi.org/10.1038/ni1318

Huang, C.-T., Workman, C.J., Flies, D., Pan, X., Marson, A.L., Zhou, G., Hipkiss, E.L., Ravi, S., Kowalski, J., Levitsky, H.I., Powell, J.D., Pardoll, D.M., Drake, C.G., Vignali, D.A.A., 2004. Role of LAG-3 in regulatory T cells. Immunity 21, 503–513. https://doi.org/10.1016/j.immuni.2004.08.010

Humrich, J.Y., Morbach, H., Undeutsch, R., Enghard, P., Rosenberger, S., Weigert, O., Kloke, L., Heimann, J., Gaber, T., Brandenburg, S., Scheffold, A., Huehn, J., Radbruch, A., Burmester, G.-R., Riemekasten, G., 2010. Homeostatic imbalance of regulatory and effector T cells due to IL-2 deprivation amplifies murine lupus. Proc. Natl. Acad. Sci. 107, 204–209. https://doi.org/10.1073/pnas.0903158107

Humrich, J.Y., Riemekasten, G., 2019. Low-dose interleukin-2 therapy for the treatment of systemic lupus erythematosus. Curr. Opin. Rheumatol. 31, 208–212. https://doi.org/10.1097/BOR.0000000000000575

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945. https://doi.org/10.1038/nature03001

Isacchini, G., Sethna, Z., Elhanati, Y., Nourmohammad, A., Walczak, A.M., Mora, T., 2020. Generative models of T-cell receptor sequences. Phys. Rev. E 101, 062414. https://doi.org/10.1103/PhysRevE.101.062414

Iwaszkiewicz-Grzes, D., Piotrowska, M., Gliwinski, M., Urban-Wójciuk, Z., Trzonkowski, P., 2021. Antigenic Challenge Influences Epigenetic Changes in Antigen-Specific T Regulatory Cells. Front. Immunol. 12, 642678. https://doi.org/10.3389/fimmu.2021.642678

Jaccard, P., 1912. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. New Phytol. 11, 37–50. https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jaeckel, E., von Boehmer, H., Manns, M.P., 2005. Antigen-specific FoxP3-transduced T-cells can control established type 1 diabetes. Diabetes 54, 306–310. https://doi.org/10.2337/diabetes.54.2.306

Janeway, C.A., 1992. The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol. Today 13, 11–16. https://doi.org/10.1016/0167-5699(92)90198-G

Jenkins, M.K., Chu, H.H., McLachlan, J.B., Moon, J.J., 2010. On the composition of the preimmune repertoire of T cells specific for Peptide-major histocompatibility complex ligands. Annu. Rev. Immunol. 28, 275–294. https://doi.org/10.1146/annurev-immunol-030409-101253

Jerne, N.K., 1955. THE NATURAL-SELECTION THEORY OF ANTIBODY FORMATION. Proc. Natl. Acad. Sci. 41, 849–857. https://doi.org/10.1073/pnas.41.11.849

Jiang, X., Wang, S., Zhou, C., Wu, J., Jiao, Y., Lin, L., Lu, X., Yang, B., Zhang, Wei, Xiao, X., Li, Y., Wu, X., Wang, X., Chen, Hua, Zhao, L., Fei, Y., Yang, Huaxia, Zhang, Wen, Zhang, F., Chen, Hui, Zhang, J., Li, B., Yang, Huanming, Wang, J., Liu, X., Zhang, X., 2020. Comprehensive TCR repertoire analysis of CD4+ T-cell subsets in rheumatoid arthritis. J. Autoimmun. 109, 102432. https://doi.org/10.1016/j.jaut.2020.102432

Joffre, O., Gorsse, N., Romagnoli, P., Hudrisier, D., van Meerwijk, J.P.M., 2004. Induction of antigen-specific tolerance to bone marrow allografts with CD4+CD25+ T lymphocytes. Blood 103, 4216–4221. https://doi.org/10.1182/blood-2004-01-0005

Jokinen, E., Huuhtanen, J., Mustjoki, S., Heinonen, M., Lähdesmäki, H., 2021. Predicting recognition between T cell receptors and epitopes with TCRGP. PLOS Comput. Biol. 17, e1008814. https://doi.org/10.1371/journal.pcbi.1008814

Jörg, S., Grohme, D.A., Erzler, M., Binsfeld, M., Haghikia, A., Müller, D.N., Linker, R.A., Kleinewietfeld, M., 2016. Environmental factors in autoimmune diseases and their role in multiple sclerosis. Cell. Mol. Life Sci. CMLS 73, 4611–4622. https://doi.org/10.1007/s00018-016-2311-1

Jurtz, V.I., Jessen, L.E., Bentzen, A.K., Jespersen, M.C., Mahajan, S., Vita, R., Jensen, K.K., Marcatili, P., Hadrup, S.R., Peters, B., Nielsen, M., 2018. NetTCR: sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks (preprint). Bioinformatics. https://doi.org/10.1101/433706

Kadowaki, K.M., Matsuno, H., Tsuji, H., Tunru, I., 1994. CD4+ T cells from collagen-induced arthritic mice are essential to transfer arthritis into severe combined immunodeficient mice. Clin. Exp. Immunol. 97, 212–218. https://doi.org/10.1111/j.1365-2249.1994.tb06070.x

Kalimuddin, S., Tham, C.Y.L., Qui, M., de Alwis, R., Sim, J.X.Y., Lim, J.M.E., Tan, H.-C., Syenina, A., Zhang, S.L., Le Bert, N., Tan, A.T., Leong, Y.S., Yee, J.X., Ong, E.Z., Ooi, E.E., Bertoletti, A., Low, J.G., 2021. Early T cell and binding antibody responses are associated with COVID-19 RNA vaccine efficacy onset. Med 2, 682-688.e4. https://doi.org/10.1016/j.medj.2021.04.003

Kang, I., Park, S.H., 2003. Infectious complications in SLE after immunosuppressive therapies. Curr. Opin. Rheumatol. 15, 528–534. https://doi.org/10.1097/00002281-200309000-00002

Kidani, Y., Nogami, W., Yasumizu, Y., Kawashima, A., Tanaka, A., Sonoda, Y., Tona, Y., Nashiki, K., Matsumoto, R., Hagiwara, M., Osaki, M., Dohi, K., Kanazawa, T., Ueyama, A., Yoshikawa, M., Yoshida, T., Matsumoto, M., Hojo, K., Shinonome, S., Yoshida, H., Hirata, M., Haruna, M., Nakamura, Y., Motooka, D., Okuzaki, D., Sugiyama, Y., Kinoshita, M., Okuno, T., Kato, T., Hatano, K., Uemura, M., Imamura, R., Yokoi, K., Tanemura, A., Shintani, Y., Kimura, T., Nonomura, N., Wada, H., Mori, M., Doki, Y., Ohkura, N., Sakaguchi, S., 2022. CCR8-targeted specific depletion of clonally expanded Treg cells in tumor tissues evokes potent tumor immunity with long-lasting memory. Proc. Natl. Acad. Sci. 119, e2114282119. https://doi.org/10.1073/pnas.2114282119

Kim, J.K., Klinger, M., Benjamin, J., Xiao, Y., Erle, D.J., Littman, D.R., Killeen, N., 2009. Impact of the TCR signal on regulatory T cell homeostasis, function, and trafficking. PloS One 4, e6580. https://doi.org/10.1371/journal.pone.0006580

Kim, J.M., Rasmussen, J.P., Rudensky, A.Y., 2007. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. Nat. Immunol. 8, 191–197. https://doi.org/10.1038/ni1428

Kim, Y.C., Zhang, A.-H., Yoon, J., Culp, W.E., Lees, J.R., Wucherpfennig, K.W., Scott, D.W., 2018. Engineered MBP-specific human Tregs ameliorate MOG-induced EAE through IL-2-triggered inhibition of effector T cells. J. Autoimmun. 92, 77–86. https://doi.org/10.1016/j.jaut.2018.05.003

Klarenbeek, P.L., de Hair, M.J.H., Doorenspleet, M.E., van Schaik, B.D.C., Esveldt, R.E.E., van de Sande, M.G.H., Cantaert, T., Gerlag, D.M., Baeten, D., van Kampen, A.H.C., Baas, F., Tak, P.P., de Vries, N., 2012. Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. Ann. Rheum. Dis. 71, 1088–1093. https://doi.org/10.1136/annrheumdis-2011-200612

Klein, L., Hinterberger, M., Wirnsberger, G., Kyewski, B., 2009. Antigen presentation in the thymus for positive selection and central tolerance induction. Nat. Rev. Immunol. 9, 833–844. https://doi.org/10.1038/nri2669

Koboziev, I., Jones-Hall, Y., Valentine, J.F., Reinoso Webb, C., Furr, K.L., Grisham, M.B., 2015. Use of Humanized Mice to Study the Pathogenesis of Autoimmune and Inflammatory Diseases: Inflamm. Bowel Dis. 21, 1652–1673. https://doi.org/10.1097/MIB.0000000000000446

Komech, E.A., Pogorelyy, M.V., Egorov, E.S., Britanova, O.V., Rebrikov, D.V., Bochkova, A.G., Shmidt, E.I., Shostak, N.A., Shugay, M., Lukyanov, S., Mamedov, I.Z., Lebedev, Y.B., Chudakov, D.M., Zvyagin, I.V., 2018. CD8+ T cells with characteristic T cell receptor beta motif are detected in blood and expanded in synovial fluid of ankylosing spondylitis patients. Rheumatol. Oxf. Engl. 57, 1097–1104. https://doi.org/10.1093/rheumatology/kex517

Kretschmer, K., Apostolou, I., Hawiger, D., Khazaie, K., Nussenzweig, M.C., von Boehmer, H., 2005. Inducing and expanding regulatory T cell populations by foreign antigen. Nat. Immunol. 6, 1219–1227. https://doi.org/10.1038/ni1265

Kukreja, A., Cost, G., Marker, J., Zhang, C., Sun, Z., Lin-Su, K., Ten, S., Sanz, M., Exley, M., Wilson, B., Porcelli, S., Maclaren, N., 2002. Multiple immuno-regulatory defects in type-1 diabetes. J. Clin. Invest. 109, 131–140. https://doi.org/10.1172/JCI13605

Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y., Tonegawa, S., 1989. Junctional sequences of T cell receptor γδ genes: Implications for γδ T cell lineages and for a novel intermediate of V-(D)-J joining. Cell 59, 859–870. https://doi.org/10.1016/0092-8674(89)90609-0

Lagattuta, K.A., Kang, J.B., Nathan, A., Pauken, K.E., Jonsson, A.H., Rao, D.A., Sharpe, A.H., Ishigaki, K., Raychaudhuri, S., 2022. Repertoire analyses reveal T cell antigen receptor sequence features that influence T cell fate. Nat. Immunol. 23, 446–457. https://doi.org/10.1038/s41590-022-01129-x

Lampeter, E.F., Homberg, M., Quabeck, K., Schaefer, U.W., Wernet, P., Bertrams, J., Grosse-Wilde, H., Gries, F.A., Kolb, H., 1993. Transfer of insulin-dependent diabetes between HLA-identical siblings by bone marrow transplantation. Lancet Lond. Engl. 341, 1243–1244. https://doi.org/10.1016/0140-6736(93)91148-f

Langrish, C.L., Chen, Y., Blumenschein, W.M., Mattson, J., Basham, B., Sedgwick, J.D., McClanahan, T., Kastelein, R.A., Cua, D.J., 2005. IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. J. Exp. Med. 201, 233–240. https://doi.org/10.1084/jem.20041257

Lê Cao, K.-A., Boitard, S., Besse, P., 2011. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics 12, 253. https://doi.org/10.1186/1471-2105-12-253

Lee, E.S., Thomas, P.G., Mold, J.E., Yates, A.J., 2017. Identifying T Cell Receptors from High-Throughput Sequencing: Dealing with Promiscuity in TCRα and TCRβ Pairing. PLOS Comput. Biol. 13, e1005313. https://doi.org/10.1371/journal.pcbi.1005313

Legoux, F.P., Lim, J.-B., Cauley, A.W., Dikiy, S., Ertelt, J., Mariani, T.J., Sparwasser, T., Way, S.S., Moon, J.J., 2015. CD4 + T Cell Tolerance to Tissue-Restricted Self Antigens Is Mediated by Antigen-Specific Regulatory T Cells Rather Than Deletion. Immunity 43, 896–908. https://doi.org/10.1016/j.immuni.2015.10.011

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., Hoffmann, J.A., 1996. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell 86, 973–983. https://doi.org/10.1016/s0092-8674(00)80172-5

Leung, B.P., McInnes, I.B., Esfandiari, E., Wei, X.Q., Liew, F.Y., 2000. Combined effects of IL-12 and IL-18 on the induction of collagen-induced arthritis. J. Immunol. Baltim. Md 1950 164, 6495–6502. https://doi.org/10.4049/jimmunol.164.12.6495

Ley, K., Laudanna, C., Cybulsky, M.I., Nourshargh, S., 2007. Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nat. Rev. Immunol. 7, 678–689. https://doi.org/10.1038/nri2156

Li, M.O., Wan, Y.Y., Sanjabi, S., Robertson, A.-K.L., Flavell, R.A., 2006. Transforming growth factor-beta regulation of immune responses. Annu. Rev. Immunol. 24, 99–146. https://doi.org/10.1146/annurev.immunol.24.021605.090737

Lindley, S., Dayan, C.M., Bishop, A., Roep, B.O., Peakman, M., Tree, T.I.M., 2005. Defective suppressor function in CD4(+)CD25(+) T-cells from patients with type 1 diabetes. Diabetes 54, 92–99. https://doi.org/10.2337/diabetes.54.1.92

Linker-Israeli, M., Bakke, A.C., Kitridou, R.C., Gendler, S., Gillis, S., Horwitz, D.A., 1983. Defective production of interleukin 1 and interleukin 2 in patients with systemic lupus erythematosus (SLE). J. Immunol. Baltim. Md 1950 130, 2651–2655.

Liu, X., Zhang, W., Zhao, M., Fu, L., Liu, L., Wu, J., Luo, S., Wang, L., Wang, Z., Lin, L., Liu, Yan, Wang, S., Yang, Y., Luo, L., Jiang, J., Wang, X., Tan, Y., Li, T., Zhu, B., Zhao, Y., Gao, X., Wan, Z., Huang, C., Fang, M., Li, Q., Peng, H., Liao, X., Chen, J., Li, F., Ling, G., Zhao, H., Luo, H., Xiang, Z., Liao, J., Liu, Yu, Yin, H., Long, H., Wu, H., Yang, huanming, Wang, J., Lu, Q., 2019. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. Ann. Rheum. Dis. 78, 1070–1078. https://doi.org/10.1136/annrheumdis-2019-215442

Lopez, J.A., Susanto, O., Jenkins, M.R., Lukoyanova, N., Sutton, V.R., Law, R.H.P., Johnston, A., Bird, C.H., Bird, P.I., Whisstock, J.C., Trapani, J.A., Saibil, H.R., Voskoboinik, I., 2013. Perforin forms transient pores on the target cell plasma membrane to facilitate rapid access of granzymes during killer cell attack. Blood 121, 2659–2668. https://doi.org/10.1182/blood-2012-07-446146

Lorenzon, R., Mariotti-Ferrandiz, E., Aheng, C., Ribet, C., Toumi, F., Pitoiset, F., Chaara, W., Derian, N., Johanet, C., Drakos, I., Harris, S., Amselem, S., Berenbaum, F., Benveniste, O., Bodaghi, B., Cacoub, P., Grateau, G., Amouyal, C., Hartemann, A., Saadoun, D., Sellam, J., Seksik, P., Sokol, H., Salem, J.-E., Vicaut, E., Six, A., Rosenzwajg, M., Bernard, C., Klatzmann, D., 2018. Clinical and multi-omics cross-phenotyping of patients with autoimmune and autoinflammatory diseases: the observational TRANSIMMUNOM protocol. BMJ Open 8, e021037. https://doi.org/10.1136/bmjopen-2017-021037

Madi, A., Poran, A., Shifrut, E., Reich-Zeliger, S., Greenstein, E., Zaretsky, I., Arnon, T., Laethem, F.V., Singer, A., Lu, J., Sun, P.D., Cohen, I.R., Friedman, N., 2017. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. eLife 6, e22057. https://doi.org/10.7554/eLife.22057

Malek, T.R., Castro, I., 2010. Interleukin-2 receptor signaling: at the interface between tolerance and immunity. Immunity 33, 153–165. https://doi.org/10.1016/j.immuni.2010.08.004

Mamedov, I.Z., Britanova, O.V., Zvyagin, I.V., Turchaninova, M.A., Bolotin, D.A., Putintseva, E.V., Lebedev, Y.B., Chudakov, D.M., 2013. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. Front. Immunol. 4, 456. https://doi.org/10.3389/fimmu.2013.00456

Mann, S.E., Zhou, Z., Landry, L.G., Anderson, A.M., Alkanani, A.K., Fischer, J., Peakman, M., Mallone, R., Campbell, K., Michels, A.W., Nakayama, M., 2020. Multiplex T Cell Stimulation Assay Utilizing a T Cell Activation Reporter-Based Detection System. Front. Immunol. 11, 633. https://doi.org/10.3389/fimmu.2020.00633

Manojlović, L.M., 2015. Photometry-based estimation of the total number of stars in the Universe. Appl. Opt. 54, 6589. https://doi.org/10.1364/AO.54.006589

Manzel, A., Muller, D.N., Hafler, D.A., Erdman, S.E., Linker, R.A., Kleinewietfeld, M., 2014. Role of "Western diet" in inflammatory autoimmune diseases. Curr. Allergy Asthma Rep. 14, 404. https://doi.org/10.1007/s11882-013-0404-6

Marcou, Q., Mora, T., Walczak, A.M., 2018. High-throughput immune repertoire analysis with IGoR. Nat. Commun. 9, 561. https://doi.org/10.1038/s41467-018-02832-w

Marek-Trzonkowska, N., Myśliwiec, M., Dobyszuk, A., Grabowska, M., Derkowska, I., Juścińska, J., Owczuk, R., Szadkowska, A., Witkowski, P., Młynarski, W., Jarosz-Chobot, P., Bossowski, A., Siebert, J., Trzonkowski, P., 2014. Therapy of type 1 diabetes with CD4(+)CD25(high)CD127-regulatory T cells prolongs survival of pancreatic islets - results of one year follow-up. Clin. Immunol. Orlando Fla 153, 23–30. https://doi.org/10.1016/j.clim.2014.03.016

Martinez-Valdez, H., Cohen, A., 1988. Coordinate regulation of mRNAs encoding adenosine deaminase, purine nucleoside phosphorylase, and terminal deoxynucleotidyltransferase by phorbol esters in human thymocytes. Proc. Natl. Acad. Sci. 85, 6900–6903.

Mason, D., 1998. A very high level of crossreactivity is an essential feature of the T-cell receptor. Immunol. Today 19, 395–404. https://doi.org/10.1016/S0167-5699(98)01299-7

Mathew, D., Giles, J.R., Baxter, A.E., Oldridge, D.A., Greenplate, A.R., Wu, J.E., Alanio, C., Kuri-Cervantes, L., Pampena, M.B., D'Andrea, K., Manne, S., Chen, Z., Huang, Y.J., Reilly, J.P., Weisman, A.R., Ittner, C.A.G., Kuthuru, O., Dougherty, J., Nzingha, K., Han, N., Kim, J., Pattekar, A., Goodwin, E.C., Anderson, E.M., Weirick, M.E., Gouma, S., Arevalo, C.P., Bolton, M.J., Chen, F., Lacey, S.F., Ramage, H., Cherry, S., Hensley, S.E., Apostolidis, S.A., Huang, A.C., Vella, L.A., UPenn COVID Processing Unit, Betts, M.R., Meyer, N.J., Wherry, E.J., 2020. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science 369, eabc8511. https://doi.org/10.1126/science.abc8511

Matsumoto, Y., Matsuo, H., Sakuma, H., Park, I.-K., Tsukada, Y., Kohyama, K., Kondo, T., Kotorii, S., Shibuya, N., 2006. CDR3 Spectratyping Analysis of the TCR Repertoire in Myasthenia Gravis. J. Immunol. 176, 5100–5107. https://doi.org/10.4049/jimmunol.176.8.5100

McGonagle, D., McDermott, M.F., 2006. A proposed classification of the immunological diseases. PLoS Med. 3, e297. https://doi.org/10.1371/journal.pmed.0030297

McHeyzer-Williams, M.G., Davis, M.M., 1995. Antigen-Specific Development of Primary and Memory T Cells in Vivo. Science 268, 106–111. https://doi.org/10.1126/science.7535476

McLean, M.H., Dieguez, D., Miller, L.M., Young, H.A., 2015. Does the microbiota play a role in the pathogenesis of autoimmune diseases? Gut 64, 332–341. https://doi.org/10.1136/gutjnl-2014-308514

Mellor, A.L., Munn, D.H., 2004. IDO expression by dendritic cells: tolerance and tryptophan catabolism. Nat. Rev. Immunol. 4, 762–774. https://doi.org/10.1038/nri1457

Mellor-Pita, S., 2006. Decrease of regulatory T cells in patients with systemic lupus erythematosus. Ann. Rheum. Dis. 65, 553–554. https://doi.org/10.1136/ard.2005.044974

Merkenschlager, M., Graf, D., Lovatt, M., Bommhardt, U., Zamoyska, R., Fisher, A.G., 1997. How many thymocytes audition for selection? J. Exp. Med. 186, 1149–1158. https://doi.org/10.1084/jem.186.7.1149

Mhanna, V., Fourcade, G., Barennes, P., Quiniou, V., Pham, H.P., 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. https://doi.org/10.2337/db20-0896

Miller, J.F., Mitchell, G.F., 1968. Cell to cell interaction in the immune response. I. Hemolysin-forming cells in neonatally thymectomized mice reconstituted with thymus or thoracic duct lymphocytes. J. Exp. Med. 128, 801–820. https://doi.org/10.1084/jem.128.4.801

Minervina, A.A., Pogorelyy, M.V., Komech, E.A., Karnaukhov, V.K., Bacher, P., Rosati, E., Franke, A., Chudakov, D.M., Mamedov, I.Z., Lebedev, Y.B., Mora, T., Walczak, A.M., 2020. Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. eLife 9, e53704. https://doi.org/10.7554/eLife.53704

Mishra, S., Srinivasan, S., Ma, C., Zhang, N., 2021. CD8+ Regulatory T Cell – A Mystery to Be Revealed. Front. Immunol. 12, 708874. https://doi.org/10.3389/fimmu.2021.708874

Mittal, S.K., Roche, P.A., 2015. Suppression of antigen presentation by IL-10. Curr. Opin. Immunol. 34, 22–27. https://doi.org/10.1016/j.coi.2014.12.009

Mombaerts, P., Iacomini, J., Johnson, R.S., Herrup, K., Tonegawa, S., Papaioannou, V.E., 1992. RAG-1-deficient mice have no mature B and T lymphocytes. Cell 68, 869–877. https://doi.org/10.1016/0092-8674(92)90030-G

Mora, T., Walczak, A.M., 2018. Quantifying lymphocyte receptor diversity, in: Systems Immunology. CRC Press, pp. 183–198.

Moraes-Vasconcelos, D., Costa-Carvalho, B.T., Torgerson, T.R., Ochs, H.D., 2008. Primary Immune Deficiency Disorders Presenting as Autoimmune Diseases: IPEX and APECED. J. Clin. Immunol. 28, 11–19. https://doi.org/10.1007/s10875-008-9176-5

Moran, A.E., Holzapfel, K.L., Xing, Y., Cunningham, N.R., Maltzman, J.S., Punt, J., Hogquist, K.A., 2011. T cell receptor signal strength in Treg and iNKT cell development demonstrated by a novel fluorescent reporter mouse. J. Exp. Med. 208, 1279–1289. https://doi.org/10.1084/jem.20110308

Moreews, M., Le Gouge, K., 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., De Guillebon, J.-M., 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. Sci. Immunol. 6, eabh1516. https://doi.org/10.1126/sciimmunol.abh1516

Morgan, D.A., Ruscetti, F.W., Gallo, R., 1976. Selective in vitro growth of T lymphocytes from normal human bone marrows. Science 193, 1007–1008. https://doi.org/10.1126/science.181845

Mösch, A., Frishman, D., 2021. TCRpair: prediction of functional pairing between HLA-A*02:01-restricted T-cell receptor α and β chains. Bioinformatics 37, 3938–3940. https://doi.org/10.1093/bioinformatics/btab573

Mosmann, T.R., Coffman, R.L., 1989. TH1 and TH2 cells: different patterns of lymphokine secretion lead to different functional properties. Annu. Rev. Immunol. 7, 145–173. https://doi.org/10.1146/annurev.iy.07.040189.001045

Murugan, A., Mora, T., Walczak, A.M., Callan, C.G., 2012. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proc. Natl. Acad. Sci. 109, 16161–16166. https://doi.org/10.1073/pnas.1212755109

Musters, A., Klarenbeek, P.L., Doorenspleet, M.E., Balzaretti, G., Esveldt, R.E.E., van Schaik, B.D.C., Jongejan, A., Tas, S.W., van Kampen, A.H.C., Baas, F., de Vries, N., 2018. In Rheumatoid Arthritis, Synovitis at Different Inflammatory Sites Is Dominated by Shared but Patient-Specific T Cell Clones. J. Immunol. 201, 417–422. https://doi.org/10.4049/jimmunol.1800421

Nagahama, K., Nishimura, E., Sakaguchi, S., 2007. Induction of tolerance by adoptive transfer of Treg cells. Methods Mol. Biol. Clifton NJ 380, 431–442. https://doi.org/10.1007/978-1-59745-395-0_27

Nakamura, K., Kitani, A., Strober, W., 2001. Cell Contact–Dependent Immunosuppression by Cd4+Cd25+Regulatory T Cells Is Mediated by Cell Surface–Bound Transforming Growth Factor β. J. Exp. Med. 194, 629–644. https://doi.org/10.1084/jem.194.5.629

Nakayama, M., Michels, A.W., 2021. Using the T Cell Receptor as a Biomarker in Type 1 Diabetes. Front. Immunol. 12, 777788. https://doi.org/10.3389/fimmu.2021.777788

Noack, M., Miossec, P., 2014. Th17 and regulatory T cell balance in autoimmune and inflammatory diseases. Autoimmun. Rev. 13, 668–677. https://doi.org/10.1016/j.autrev.2013.12.004

Nourshargh, S., Alon, R., 2014. Leukocyte migration into inflamed tissues. Immunity 41, 694–707. https://doi.org/10.1016/j.immuni.2014.10.008

Nüsslein-Volhard, C., Wieschaus, E., 1980. Mutations affecting segment number and polarity in Drosophila. Nature 287, 795–801. https://doi.org/10.1038/287795a0

Oderup, C., Cederbom, L., Makowska, A., Cilio, C.M., Ivars, F., 2006. Cytotoxic T lymphocyte antigen-4-dependent down-modulation of costimulatory molecules on dendritic cells in CD4+ CD25+ regulatory T-cell-mediated suppression. Immunology 118, 240–249. https://doi.org/10.1111/j.1365-2567.2006.02362.x

Oettinger, M.A., Schatz, D.G., Gorka, C., Baltimore, D., 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science 248, 1517–1523. https://doi.org/10.1126/science.2360047

Olson, B.J., Moghimi, P., Schramm, C.A., Obraztsova, A., Ralph, D., Vander Heiden, J.A., Shugay, M., Shepherd, A.J., Lees, W., Matsen, F.A., 2019. sumrep: A Summary Statistic Framework for Immune Receptor Repertoire Comparison and Model Validation. Front. Immunol. 10, 2533. https://doi.org/10.3389/fimmu.2019.02533

Omer, A., Peres, A., Rodriguez, O.L., Watson, C.T., Lees, W., Polak, P., Collins, A.M., Yaari, G., 2022. T cell receptor beta germline variability is revealed by inference from repertoire data. Genome Med. 14, 2. https://doi.org/10.1186/s13073-021-01008-4

Padovan, E., Casorati, G., Dellabona, P., Meyer, S., Brockhaus, M., Lanzavecchia, A., 1993. Expression of two T cell receptor alpha chains: dual receptor T cells. Science 262, 422–424. https://doi.org/10.1126/science.8211163

Pandiyan, P., Zheng, L., Ishihara, S., Reed, J., Lenardo, M.J., 2007. CD4+CD25+Foxp3+ regulatory T cells induce cytokine deprivation-mediated apoptosis of effector CD4+ T cells. Nat. Immunol. 8, 1353–1362. https://doi.org/10.1038/ni1536

Parra, Z.E., Baker, M.L., Hathaway, J., Lopez, A.M., Trujillo, J., Sharp, A., Miller, R.D., 2008. Comparative genomic analysis and evolution of the T cell receptor loci in the opossum Monodelphis domestica. BMC Genomics 9, 111. https://doi.org/10.1186/1471-2164-9-111

Pearson, C.M., Wood, F.D., 1964. PASSIVE TRANSFER OF ADJUVANT ARTHRITIS BY LYMPH NODE OR SPLEEN CELLS. J. Exp. Med. 120, 547–560. https://doi.org/10.1084/jem.120.4.547

Peteranderl, C., Herold, S., Schmoldt, C., 2016. Human Influenza Virus Infections. Semin. Respir. Crit. Care Med. 37, 487–500. https://doi.org/10.1055/s-0036-1584801

Pogorelyy, M.V., Minervina, A.A., Chudakov, D.M., Mamedov, I.Z., Lebedev, Y.B., Mora, T., Walczak, A.M., 2018. Method for identification of condition-associated public antigen receptor sequences. eLife 7, e33050. https://doi.org/10.7554/eLife.33050

Pogorelyy, M.V., Shugay, M., 2019. A Framework for Annotation of Antigen Specificities in High-Throughput T-Cell Repertoire Sequencing Studies. Front. Immunol. 10, 2159. https://doi.org/10.3389/fimmu.2019.02159

Postow, M.A., Manuel, M., Wong, P., Yuan, J., Dong, Z., Liu, C., Perez, S., Tanneau, I., Noel, M., Courtier, A., Pasqual, N., Wolchok, J.D., 2015. Peripheral T cell receptor diversity is associated with clinical outcomes following ipilimumab treatment in metastatic melanoma. J. Immunother. Cancer 3, 23. https://doi.org/10.1186/s40425-015-0070-4

Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.-Y., Olshen, R.A., Weyand, C.M., Boyd, S.D., Goronzy, J.J., 2014. Diversity and clonal selection in the human T-cell repertoire. Proc. Natl. Acad. Sci. 111, 13139–13144. https://doi.org/10.1073/pnas.1409155111

Quddus, J., Johnson, K.J., Gavalchin, J., Amento, E.P., Chrisp, C.E., Yung, R.L., Richardson, B.C., 1993. Treating activated CD4+ T cells with either of two distinct DNA methyltransferase inhibitors, 5-azacytidine or procainamide, is sufficient to cause a lupus-like disease in syngeneic mice. J. Clin. Invest. 92, 38–53. https://doi.org/10.1172/JCI116576

Qureshi, O.S., Zheng, Y., Nakamura, K., Attridge, K., Manzotti, C., Schmidt, E.M., Baker, J., Jeffery, L.E., Kaur, S., Briggs, Z., Hou, T.Z., Futter, C.E., Anderson, G., Walker, L.S.K., Sansom, D.M., 2011. Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. Science 332, 600–603. https://doi.org/10.1126/science.1202947

Rankin, A.L., Reed, A.J., Oh, S., Cozzo Picca, C., Guay, H.M., Larkin, J., Panarey, L., Aitken, M.K., Koeberlein, B., Lipsky, P.E., Tomaszewski, J.E., Naji, A., Caton, A.J., 2008. CD4+ T cells recognizing a single self-peptide expressed by APCs induce spontaneous autoimmune arthritis. J. Immunol. Baltim. Md 1950 180, 833–841. https://doi.org/10.4049/jimmunol.180.2.833

Reiser, J.-B., Darnault, C., Grégoire, C., Mosser, T., Mazza, G., Kearney, A., van der Merwe, P.A., Fontecilla-Camps, J.C., Housset, D., Malissen, B., 2003. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. Nat. Immunol. 4, 241–247. https://doi.org/10.1038/ni891

Richardson, B., Sawalha, A.H., Ray, D., Yung, R., 2012. Murine models of lupus induced by hypomethylated T cells (DNA hypomethylation and lupus…). Methods Mol. Biol. Clifton NJ 900, 169–180. https://doi.org/10.1007/978-1-60761-720-4_8

Ritvo, P.-G., Saadawi, A., Barennes, P., Quiniou, V., Chaara, W., El Soufi, K., Bonnet, B., Six, A., Shugay, M., Mariotti-Ferrandiz, E., Klatzmann, D., 2018. High-resolution repertoire analysis reveals a major bystander activation of Tfh and Tfr cells. Proc. Natl. Acad. Sci. U. S. A. 115, 9604–9609. https://doi.org/10.1073/pnas.1808594115

Robinson, D.S., Hamid, Q., Ying, S., Tsicopoulos, A., Barkans, J., Bentley, A.M., Corrigan, C., Durham, S.R., Kay, A.B., 1992. Predominant TH2-like bronchoalveolar T-lymphocyte population in atopic asthma. N. Engl. J. Med. 326, 298–304. https://doi.org/10.1056/NEJM199201303260504

Rodriguez-Calvo, T., Ekwall, O., Amirian, N., Zapardiel-Gonzalo, J., von Herrath, M.G., 2014. Increased immune cell infiltration of the exocrine pancreas: a possible contribution to the pathogenesis of type 1 diabetes. Diabetes 63, 3880–3890. https://doi.org/10.2337/db14-0549

Roep, B.O., Peakman, M., 2012. Antigen targets of type 1 diabetes autoimmunity. Cold Spring Harb. Perspect. Med. 2, a007781. https://doi.org/10.1101/cshperspect.a007781

Rognes, T., Scheffer, L., Greiff, V., Sandve, G.K., 2022. CompAIRR: ultra-fast comparison of adaptive immune receptor repertoires by exact and approximate sequence matching. Bioinformatics btac505. https://doi.org/10.1093/bioinformatics/btac505

Rosati, E., Dowds, C.M., Liaskou, E., Henriksen, E.K.K., Karlsen, T.H., Franke, A., 2017. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol. 17, 61. https://doi.org/10.1186/s12896-017-0379-9

Rosati, E., Pogorelyy, M.V., Dowds, C.M., Moller, F.T., Sorensen, S.B., Lebedev, Y.B., Frey, N., Schreiber, S., Spehlmann, M.E., Andersen, V., Mamedov, I.Z., Franke, A., 2020. Identification of Disease-associated Traits and Clonotypes in the T Cell Receptor Repertoire of Monozygotic Twins Affected by Inflammatory Bowel Diseases. J. Crohns Colitis 14, 778–790. https://doi.org/10.1093/ecco-jcc/jjz179

Rosenberg, S.A., Lotze, M.T., Muul, L.M., Leitman, S., Chang, A.E., Ettinghausen, S.E., Matory, Y.L., Skibber, J.M., Shiloni, E., Vetto, J.T., 1985. Observations on the systemic administration of autologous lymphokine-activated killer cells and recombinant interleukin-2 to patients with metastatic cancer. N. Engl. J. Med. 313, 1485–1492. https://doi.org/10.1056/NEJM198512053132327

Rosenzwajg, M., Churlaud, G., Mallone, R., Six, A., Dérian, N., Chaara, W., Lorenzon, R., Long, S.A., Buckner, J.H., Afonso, G., Pham, H.-P., Hartemann, A., Yu, A., Pugliese, A., Malek, T.R., Klatzmann, D., 2015a. Low-dose interleukin-2 fosters a dose-dependent regulatory T cell tuned milieu in T1D patients. J. Autoimmun. 58, 48–58. https://doi.org/10.1016/j.jaut.2015.01.001

Rosenzwajg, M., Churlaud, G., Mallone, R., Six, A., Dérian, N., Chaara, W., Lorenzon, R., Long, S.A., Buckner, J.H., Afonso, G., Pham, H.-P., Hartemann, A., Yu, A., Pugliese, A., Malek, T.R., Klatzmann, D., 2015b. Low-dose interleukin-2 fosters a dose-dependent regulatory T cell tuned milieu in T1D patients. J. Autoimmun. 58, 48–58. https://doi.org/10.1016/j.jaut.2015.01.001

Rosenzwajg, M., Lorenzon, R., Cacoub, P., Pham, H.P., Pitoiset, F., El Soufi, K., RIbet, C., Bernard, C., Aractingi, S., Banneville, B., Beaugerie, L., Berenbaum, F., Champey, J., Chazouilleres, O., Corpechot, C., Fautrel, B., Mekinian, A., Regnier, E., Saadoun, D., Salem, J.-E., Sellam, J., Seksik, P., Daguenel-Nguyen, A., Doppler, V., Mariau, J., Vicaut, E., Klatzmann, D., 2019. Immunological and clinical effects of low-dose interleukin-2 across 11 autoimmune diseases in a single, open clinical trial. Ann. Rheum. Dis. 78, 209–217. https://doi.org/10.1136/annrheumdis-2018-214229

Rossetti, M., Spreafico, R., Consolaro, A., Leong, J.Y., Chua, C., Massa, M., Saidin, S., Magni-Manzoni, S., Arkachaisri, T., Wallace, C.A., Gattorno, M., Martini, A., Lovell, D.J., Albani, S., 2017. TCR repertoire sequencing identifies synovial Treg cell clonotypes in the bloodstream during active inflammation in human arthritis. Ann. Rheum. Dis. 76, 435–441. https://doi.org/10.1136/annrheumdis-2015-208992

Rubtsov, A.V., Rubtsova, K., Kappler, J.W., Marrack, P., 2010. Genetic and hormonal factors in female-biased autoimmunity. Autoimmun. Rev. 9, 494–498. https://doi.org/10.1016/j.autrev.2010.02.008

Rydyznski Moderbacher, C., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, Christina, Choi, J., Kato, Y., Crotty, E.G., Kim, Cheryl, Rawlings, S.A., Mateus, J., Tse, L.P.V., Frazier, A., Baric, R., Peters, B., Greenbaum, J., Ollmann Saphire, E., Smith, D.M., Sette, A., Crotty, S., 2020. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. Cell 183, 996-1012.e19. https://doi.org/10.1016/j.cell.2020.09.038

Saadoun, D., Rosenzwajg, M., Joly, F., Six, A., Carrat, F., Thibault, V., Sene, D., Cacoub, P., Klatzmann, D., 2011. Regulatory T-Cell Responses to Low-Dose Interleukin-2 in HCV-Induced Vasculitis. N. Engl. J. Med. 365, 2067–2077. https://doi.org/10.1056/NEJMoa1105143

Sadlack, B., Merz, H., Schorle, H., Schimpl, A., Feller, A.C., Horak, I., 1993. Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. Cell 75, 253–261. https://doi.org/10.1016/0092-8674(93)80067-o

Sakaguchi, S., Mikami, N., Wing, J.B., Tanaka, A., Ichiyama, K., Ohkura, N., 2020. Regulatory T Cells and Human Disease. Annu. Rev. Immunol. 38, 541–566. https://doi.org/10.1146/annurev-immunol-042718-041717

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. Baltim. Md 1950 155, 1151–1164.

Sakaguchi, S., Yamaguchi, T., Nomura, T., Ono, M., 2008. Regulatory T Cells and Immune Tolerance. Cell 133, 775–787. https://doi.org/10.1016/j.cell.2008.05.009

Sakano, H., Hüppi, K., Heinrich, G., Tonegawa, S., 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. Nature 280, 288–294. https://doi.org/10.1038/280288a0

Scaviner, D., Lefranc, M.P., 2000. The human T cell receptor alpha variable (TRAV) genes. Exp. Clin. Immunogenet. 17, 83–96. https://doi.org/10.1159/000019128

Schatz, D.G., Oettinger, M.A., Baltimore, D., 1989. The V(D)J recombination activating gene, RAG-1. Cell 59, 1035–1048. https://doi.org/10.1016/0092-8674(89)90760-5

Schlissel, M.S., 1998. Structure of nonhairpin coding-end DNA breaks in cells undergoing V(D)J recombination. Mol. Cell. Biol. 18, 2029–2037. https://doi.org/10.1128/MCB.18.4.2029

Schuldt, N.J., Auger, J.L., Spanier, J.A., Martinov, T., Breed, E.R., Fife, B.T., Hogquist, K.A., Binstadt, B.A., 2017. Cutting Edge: Dual TCRα Expression Poses an Autoimmune Hazard by Limiting Regulatory T Cell Generation. J. Immunol. Baltim. Md 1950 199, 33–38. https://doi.org/10.4049/jimmunol.1700406

Schuldt, N.J., Binstadt, B.A., 2019. Dual TCR T Cells: Identity Crisis or Multitaskers? J. Immunol. 202, 637–644. https://doi.org/10.4049/jimmunol.1800904

Schultheiß, C., Paschold, L., Simnica, D., Mohme, M., Willscher, E., von Wenserski, L., Scholz, R., Wieters, I., Dahlke, C., Tolosa, E., Sedding, D.G., Ciesek, S., Addo, M., Binder, M., 2020. Next-Generation Sequencing of T and B Cell Receptor Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease. Immunity 53, 442-455.e4. https://doi.org/10.1016/j.immuni.2020.06.024

Seay, H.R., Yusko, E., Rothweiler, S.J., Zhang, L., Posgai, A.L., Campbell-Thompson, M., Vignali, M., Emerson, R.O., Kaddis, J.S., Ko, D., Nakayama, M., Smith, M.J., Cambier, J.C., Pugliese, A., Atkinson, M.A., Robins, H.S., Brusko, T.M., 2016. Tissue distribution and clonal diversity of the T and B cell repertoire in type 1 diabetes. JCI Insight 1. https://doi.org/10.1172/jci.insight.88242

Seddon, B., Mason, D., 1999. Peripheral autoantigen induces regulatory T cells that prevent autoimmunity. J. Exp. Med. 189, 877–882. https://doi.org/10.1084/jem.189.5.877

Serra, P., Santamaria, P., 2019. Antigen-specific therapeutic approaches for autoimmunity. Nat. Biotechnol. 37, 238–251. https://doi.org/10.1038/s41587-019-0015-4

Sethna, Z., Isacchini, G., Dupic, T., Mora, T., Walczak, A.M., Elhanati, Y., 2020. Population variability in the generation and selection of T-cell repertoires. PLoS Comput. Biol. 16, e1008394. https://doi.org/10.1371/journal.pcbi.1008394

Sewell, A.K., 2012. Why must T cells be cross-reactive? Nat. Rev. Immunol. 12, 669–677. https://doi.org/10.1038/nri3279

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423.

Shcherbinin, D.S., Belousov, V.A., Shugay, M., 2020. Comprehensive analysis of structural and sequencing data reveals almost unconstrained chain pairing in TCRαβ complex. PLOS Comput. Biol. 16, e1007714. https://doi.org/10.1371/journal.pcbi.1007714

Shevach, E.M., Thornton, A.M., 2014. tTregs, pTregs, and iTregs: similarities and differences. Immunol. Rev. 259, 88–102. https://doi.org/10.1111/imr.12160

Shimokawa, C., Kato, T., Takeuchi, T., Ohshima, N., Furuki, T., Ohtsu, Y., Suzue, K., Imai, T., Obi, S., Olia, A., Izumi, T., Sakurai, M., Arakawa, H., Ohno, H., Hisaeda, H., 2020. CD8+ regulatory T cells are critical in prevention of autoimmune-mediated diabetes. Nat. Commun. 11, 1922. https://doi.org/10.1038/s41467-020-15857-x

Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A.M., 1992. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell 68, 855–867. https://doi.org/10.1016/0092-8674(92)90029-c

Shugay, M., Bagaev, D.V., Zvyagin, I.V., Vroomans, R.M., Crawford, J.C., Dolton, G., Komech, E.A., Sycheva, A.L., Koneva, A.E., Egorov, E.S., Eliseev, A.V., Van Dyk, E., Dash, P., Attaf, M., Rius, C., Ladell, K., McLaren, J.E., Matthews, K.K., Clemens, E.B., Douek, D.C., Luciani, F., van Baarle, D., Kedzierska, K., Kesmir, C., Thomas, P.G., Price, D.A., Sewell, A.K., Chudakov, D.M., 2018. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res. 46, D419–D427. https://doi.org/10.1093/nar/gkx760

Sidhom, J.-W., Larman, H.B., Pardoll, D.M., Baras, A.S., 2021. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. Nat. Commun. 12, 1605. https://doi.org/10.1038/s41467-021-21879-w

Simpson, E.H., 1949. Measurement of Diversity. Nature 163, 688–688. https://doi.org/10.1038/163688a0

Simpson, N., Gatenby, P.A., Wilson, A., Malik, S., Fulcher, D.A., Tangye, S.G., Manku, H., Vyse, T.J., Roncador, G., Huttley, G.A., Goodnow, C.C., Vinuesa, C.G., Cook, M.C., 2010. Expansion of circulating T cells resembling follicular helper T cells is a fixed phenotype that identifies a subset of severe systemic lupus erythematosus. Arthritis Rheum. 62, 234–244. https://doi.org/10.1002/art.25032

Singer, A., 2002. New perspectives on a developmental dilemma: the kinetic signaling model and the importance of signal duration for the CD4/CD8 lineage decision. Curr. Opin. Immunol. 14, 207–215. https://doi.org/10.1016/s0952-7915(02)00323-0

Six, A., Mariotti-Ferrandiz, M.E., Chaara, W., Magadan, S., Pham, H.-P., Lefranc, M.-P., Mora, T., Thomas-Vaslin, V., Walczak, A.M., Boudinot, P., 2013. The Past, Present, and Future of Immune Repertoire Biology – The Rise of Next-Generation Repertoire Analysis. Front. Immunol. 4. https://doi.org/10.3389/fimmu.2013.00413

Snyder, T.M., Gittelman, R.M., Klinger, M., May, D.H., Osborne, E.J., Taniguchi, R., Zahid, H.J., Kaplan, I.M., Dines, J.N., Noakes, M.T., Pandya, R., Chen, X., Elasady, S., Svejnoha, E., Ebert, P., Pesesky, M.W., De Almeida, P., O'Donnell, H., DeGottardi, Q., Keitany, G., Lu, J., Vong, A., Elyanow, R., Fields, P., Greissl, J., Baldo, L., Semprini, S., Cerchione, C., Nicolini, F., Mazza, M., Delmonte, O.M., Dobbs, K., Laguna-Goya, R., Carreño-Tarragona, G., Barrio, S., Imberti, L., Sottini, A., Quiros-Roldan, E., Rossi, C., Biondi, A., Bettini, L.R., D'Angio, M., Bonfanti, P., Tompkins, M.F., Alba, C., Dalgard, C., Sambri, V., Martinelli, G., Goldman, J.D., Heath, J.R., Su, H.C., Notarangelo, L.D., Paz-Artal, E., Martinez-Lopez, J., Carlson, J.M., Robins, H.S., 2020. Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. MedRxiv Prepr. Serv. Health Sci. 2020.07.31.20165647. https://doi.org/10.1101/2020.07.31.20165647

Spolski, R., Li, P., Leonard, W.J., 2018. Biology and regulation of IL-2: from molecular mechanisms to human therapy. Nat. Rev. Immunol. 18, 648–659. https://doi.org/10.1038/s41577-018-0046-y

Spreafico, R., Rossetti, M., van Loosdregt, J., Wallace, C.A., Massa, M., Magni-Manzoni, S., Gattorno, M., Martini, A., Lovell, D.J., Albani, S., 2016. A circulating reservoir of pathogenic-like CD4+ T cells shares a genetic and phenotypic signature with the inflamed synovial micro-environment. Ann. Rheum. Dis. 75, 459–465. https://doi.org/10.1136/annrheumdis-2014-206226

Springer, I., Besser, H., Tickotsky-Moskovitz, N., Dvorkin, S., Louzoun, Y., 2020. Prediction of Specific TCR-Peptide Binding From Large Dictionaries of TCR-Peptide Pairs. Front. Immunol. 11, 1803. https://doi.org/10.3389/fimmu.2020.01803

Springer, I., Tickotsky, N., Louzoun, Y., 2021. Contribution of T Cell Receptor Alpha and Beta CDR3, MHC Typing, V and J Genes to Peptide Binding Prediction. Front. Immunol. 12, 664514. https://doi.org/10.3389/fimmu.2021.664514

Stadinski, B.D., Zhang, L., Crawford, F., Marrack, P., Eisenbarth, G.S., Kappler, J.W., 2010. Diabetogenic T cells recognize insulin bound to IA^{g7} in an unexpected, weakly binding register. Proc. Natl. Acad. Sci. 107, 10978–10983. https://doi.org/10.1073/pnas.1006545107

Steinman, R.M., Cohn, Z.A., 1973. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. J. Exp. Med. 137, 1142–1162. https://doi.org/10.1084/jem.137.5.1142

Stephan, V., Wahn, V., Le Deist, F., Dirksen, U., Bröker, B., Müller-Fleckenstein, I., Horneff, G., Schroten, H., Fischer, A., de Saint Basile, G., 1996. Atypical X-Linked Severe Combined Immunodeficiency Due to Possible Spontaneous Reversion of the Genetic Defect in T Cells. N. Engl. J. Med. 335, 1563–1567. https://doi.org/10.1056/NEJM199611213352104

Stephens, L.A., Malpass, K.H., Anderton, S.M., 2009. Curing CNS autoimmune disease with myelin-reactive Foxp3+ Treg. Eur. J. Immunol. 39, 1108–1117. https://doi.org/10.1002/eji.200839073

Su, L.F., Kidd, B.A., Han, A., Kotzin, J.J., Davis, M.M., 2013. Virus-Specific CD4+ Memory-Phenotype T Cells Are Abundant in Unexposed Adults. Immunity 38, 373–383. https://doi.org/10.1016/j.immuni.2012.10.021

Suen, J.-L., Li, H.-T., Jong, Y.-J., Chiang, B.-L., Yen, J.-H., 2009. Altered homeostasis of CD4 + FoxP3 + regulatory T-cell subpopulations in systemic lupus erythematosus. Immunology 127, 196–205. https://doi.org/10.1111/j.1365-2567.2008.02937.x

Sui, W., Hou, X., Zou, G., Che, W., Yang, M., Zheng, C., Liu, F., Chen, P., Wei, X., Lai, L., Dai, Y., 2015. Composition and variation analysis of the TCR β-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. Mol. Immunol. 67, 455–464. https://doi.org/10.1016/j.molimm.2015.07.012

Takahashi, T., Kuniyasu, Y., Toda, M., Sakaguchi, N., Itoh, M., Iwata, M., Shimizu, J., Sakaguchi, S., 1998. Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. Int. Immunol. 10, 1969–1980. https://doi.org/10.1093/intimm/10.12.1969

Tamzalit, F., Wang, M.S., Jin, W., Tello-Lafoz, M., Boyko, V., Heddleston, J.M., Black, C.T., Kam, L.C., Huse, M., 2019. Interfacial actin protrusions mechanically enhance killing by cytotoxic T cells. Sci. Immunol. 4, eaav5445. https://doi.org/10.1126/sciimmunol.aav5445

Tang, Q., Henriksen, K.J., Bi, M., Finger, E.B., Szot, G., Ye, J., Masteller, E.L., McDevitt, H., Bonyhadi, M., Bluestone, J.A., 2004a. In vitro-expanded antigen-specific regulatory T cells suppress autoimmune diabetes. J. Exp. Med. 199, 1455–1465. https://doi.org/10.1084/jem.20040139

Tang, Q., Henriksen, K.J., Bi, M., Finger, E.B., Szot, G., Ye, J., Masteller, E.L., McDevitt, H., Bonyhadi, M., Bluestone, J.A., 2004b. In vitro-expanded antigen-specific regulatory T cells suppress autoimmune diabetes. J. Exp. Med. 199, 1455–1465. https://doi.org/10.1084/jem.20040139

Tanno, H., Gould, T.M., McDaniel, J.R., Cao, W., Tanno, Y., Durrett, R.E., Park, D., Cate, S.J., Hildebrand, W.H., Dekker, C.L., Tian, L., Weyand, C.M., Georgiou, G., Goronzy, J.J., 2020. Determinants governing T cell receptor α/β-chain pairing in repertoire formation of identical twins. Proc. Natl. Acad. Sci. 117, 532–540. https://doi.org/10.1073/pnas.1915008117

Tarbell, K.V., Yamazaki, S., Olson, K., Toy, P., Steinman, R.M., 2004. CD25+ CD4+ T Cells, Expanded with Dendritic Cells Presenting a Single Autoantigenic Peptide, Suppress Autoimmune Diabetes. J. Exp. Med. 199, 1467–1477. https://doi.org/10.1084/jem.20040180

Taylor, P.A., Lees, C.J., Blazar, B.R., 2002. The infusion of ex vivo activated and expanded CD4(+)CD25(+) immune regulatory cells inhibits graft-versus-host disease lethality. Blood 99, 3493–3499. https://doi.org/10.1182/blood.v99.10.3493

Tennakoon, D.K., Mehta, R.S., Ortega, S.B., Bhoj, V., Racke, M.K., Karandikar, N.J., 2006. Therapeutic Induction of Regulatory, Cytotoxic CD8 + T Cells in Multiple Sclerosis. J. Immunol. 176, 7119–7129. https://doi.org/10.4049/jimmunol.176.11.7119

Teraguchi, S., Saputri, D.S., Llamas-Covarrubias, M.A., Davila, A., Diez, D., Nazlica, S.A., Rozewicki, J., Ismanto, H.S., Wilamowski, J., Xie, J., Xu, Z., Loza-Lopez, M. de J., van Eerden, F.J., Li, S., Standley, D.M., 2020. Methods for sequence and structural analysis of B and T cell receptor repertoires. Comput. Struct. Biotechnol. J. 18, 2000–2011. https://doi.org/10.1016/j.csbj.2020.07.008

Thomas, P.G., Crawford, J.C., 2019. Selected before selection: A case for inherent antigen bias in the T-cell receptor repertoire. Curr. Opin. Syst. Biol. 18, 36–43. https://doi.org/10.1016/j.coisb.2019.10.007

Thornton, A.M., Shevach, E.M., 1998. CD4+CD25+ immunoregulatory T cells suppress polyclonal T cell activation in vitro by inhibiting interleukin 2 production. J. Exp. Med. 188, 287–296. https://doi.org/10.1084/jem.188.2.287

Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E., Friedman, N., 2017. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinforma. Oxf. Engl. 33, 2924–2929. https://doi.org/10.1093/bioinformatics/btx286

Tomofuji, Y., Kishikawa, T., Maeda, Y., Ogawa, K., Nii, T., Okuno, T., Oguro-Igashira, E., Kinoshita, M., Yamamoto, K., Sonehara, K., Yagita, M., Hosokawa, A., Motooka, D., Matsumoto, Y., Matsuoka, H., Yoshimura, M., Ohshima, S., Nakamura, S., Inohara, H., Mochizuki, H., Takeda, K., Kumanogoh, A., Okada, Y., 2022. Whole gut virome analysis of 476 Japanese revealed a link between phage and autoimmune disease. Ann. Rheum. Dis. 81, 278. https://doi.org/10.1136/annrheumdis-2021-221267

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575–581. https://doi.org/10.1038/302575a0

Tong, Y., Li, Z., Zhang, H., Xia, L., Zhang, M., Xu, Y., Wang, Z., Deem, M.W., Sun, X., He, J., 2016. T Cell Repertoire Diversity Is Decreased in Type 1 Diabetes Patients. Genomics Proteomics Bioinformatics 14, 338–348. https://doi.org/10.1016/j.gpb.2016.10.003

Tsokos, G.C., 2011. Systemic Lupus Erythematosus. N. Engl. J. Med. 365, 2110–2121. https://doi.org/10.1056/NEJMra1100359

Umeshappa, C.S., Mbongue, J., Singha, S., Mohapatra, S., Yamanouchi, J., Lee, J.A., Nanjundappa, R.H., Shao, K., Christen, U., Yang, Y., Ellestad, K.K., Santamaria, P., 2020. Ubiquitous antigen-specific T regulatory type 1 cells variably suppress hepatic and extrahepatic autoimmunity. J. Clin. Invest. 130, 1823–1829. https://doi.org/10.1172/JCI130670

Vahl, J.C., Drees, C., Heger, K., Heink, S., Fischer, J.C., Nedjic, J., Ohkura, N., Morikawa, H., Poeck, H., Schallenberg, S., Rieß, D., Hein, M.Y., Buch, T., Polic, B., Schönle, A., Zeiser, R., Schmitt-Gräff, A., Kretschmer, K., Klein, L., Korn, T., Sakaguchi, S., Schmidt-Supprian, M., 2014. Continuous T Cell Receptor Signals Maintain a Functional Regulatory T Cell Pool. Immunity 41, 722–736. https://doi.org/10.1016/j.immuni.2014.10.012

Valkiers, S., de Vrij, N., Gielis, S., Verbandt, S., Ogunjimi, B., Laukens, K., Meysman, P., 2022. Recent advances in T-cell receptor repertoire analysis: Bridging the gap with multimodal single-cell RNA sequencing. ImmunoInformatics 5, 100009. https://doi.org/10.1016/j.immuno.2022.100009

Valkiers, S., Van Houcke, M., Laukens, K., Meysman, P., 2021. ClusTCR: a Python interface for rapid clustering of large sets of CDR3 sequences with unknown antigen specificity. Bioinforma. Oxf. Engl. btab446. https://doi.org/10.1093/bioinformatics/btab446

Van Boxel, J.A., Paget, S.A., 1975. Predominantly T-cell infiltrate in rheumatoid synovial membranes. N. Engl. J. Med. 293, 517–520. https://doi.org/10.1056/NEJM197509112931101

Van den Broek, M.F., Van de Langerijt, L.G., Van Bruggen, M.C., Billingham, M.E., Van den Berg, W.B., 1992. Treatment of rats with monoclonal anti-CD4 induces long-term resistance to streptococcal cell wall-induced arthritis. Eur. J. Immunol. 22, 57–61. https://doi.org/10.1002/eji.1830220110

van Gent, D.C., McBlane, J.F., Ramsden, D.A., Sadofsky, M.J., Hesse, J.E., Gellert, M., 1995. Initiation of V(D)J recombination in a cell-free system. Cell 81, 925–934. https://doi.org/10.1016/0092-8674(95)90012-8

van Lummel, M., van Veelen, P.A., Zaldumbide, A., de Ru, A., Janssen, G.M.C., Moustakas, A.K., Papadopoulos, G.K., Drijfhout, J.W., Roep, B.O., Koning, F., 2012. Type 1 diabetes-associated HLA-DQ8 transdimer accommodates a unique peptide repertoire. J. Biol. Chem. 287, 9514–9524. https://doi.org/10.1074/jbc.M111.313940

van Panhuys, N., 2016. TCR Signal Strength Alters T–DC Activation and Interaction Times and Directs the Outcome of Differentiation. Front. Immunol. 7. https://doi.org/10.3389/fimmu.2016.00006

Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., Peters, B., 2019. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 47, D339–D343. https://doi.org/10.1093/nar/gky1006

von Spee-Mayer, C., Siegert, E., Abdirama, D., Rose, A., Klaus, A., Alexander, T., Enghard, P., Sawitzki, B., Hiepe, F., Radbruch, A., Burmester, G.-R., Riemekasten, G., Humrich, J.Y., 2016. Low-dose interleukin-2 selectively corrects regulatory T cell defects in patients with systemic lupus erythematosus. Ann. Rheum. Dis. 75, 1407–1415. https://doi.org/10.1136/annrheumdis-2015-207776

Waase, I., Kayser, C., Carlson, P.J., Goronzy, J.J., Weyand, C.M., 1996. Oligoclonal T cell proliferation in patients with rheumatoid arthritis and their unaffected siblings. Arthritis Rheum. 39, 904–913. https://doi.org/10.1002/art.1780390606

Walker, L.S.K., Chodos, A., Eggena, M., Dooms, H., Abbas, A.K., 2003. Antigen-dependent proliferation of CD4+ CD25+ regulatory T cells in vivo. J. Exp. Med. 198, 249–258. https://doi.org/10.1084/jem.20030315

Wang, B., André, I., Gonzalez, A., Katz, J.D., Aguet, M., Benoist, C., Mathis, D., 1997. Interferon-gamma impacts at multiple points during the progression of autoimmune diabetes. Proc. Natl. Acad. Sci. U. S. A. 94, 13844–13849. https://doi.org/10.1073/pnas.94.25.13844

Wang, G., Mudgal, P., Wang, L., Shuen, T.W.H., Wu, H., Alexander, P.B., Wang, W.-W., Wan, Y., Toh, H.C., Wang, X.-F., Li, Q.-J., 2021. TCR repertoire characteristics predict clinical response to adoptive CTL therapy against nasopharyngeal carcinoma. OncoImmunology 10, 1955545. https://doi.org/10.1080/2162402X.2021.1955545

Wen, L., Ley, R.E., Volchkov, P.Yu., Stranges, P.B., Avanesyan, L., Stonebraker, A.C., Hu, C., Wong, F.S., Szot, G.L., Bluestone, J.A., Gordon, J.I., Chervonsky, A.V., 2008. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature 455, 1109–1113. https://doi.org/10.1038/nature07336

Whangbo, J.S., Kim, H.T., Nikiforow, S., Koreth, J., Kim, S., Alyea, E.P., Armand, P., Cutler, C., Ho, V.T., Antin, J.H., Soiffer, R.J., Ritz, J., 2018. Low-Dose IL-2 Therapy Preferentially Improves Diversity of the Regulatory T Cell Repertoire in Patients with Refractory Chronic Graft-Versus-Host Disease. Biol. Blood Marrow Transplant. 24, S70–S71. https://doi.org/10.1016/j.bbmt.2017.12.616

Wherry, E.J., Ahmed, R., 2004. Memory CD8 T-cell differentiation during viral infection. J. Virol. 78, 5535–5545. https://doi.org/10.1128/JVI.78.11.5535-5545.2004

Wing, J.B., Tanaka, A., Sakaguchi, S., 2019. Human FOXP3+ Regulatory T Cell Heterogeneity and Function in Autoimmunity and Cancer. Immunity 50, 302–316. https://doi.org/10.1016/j.immuni.2019.01.020

Wong, P.K.K., Quinn, J.M.W., Sims, N.A., van Nieuwenhuijze, A., Campbell, I.K., Wicks, I.P., 2006. Interleukin-6 modulates production of T lymphocyte-derived cytokines in antigen-induced arthritis and drives inflammation-induced osteoclastogenesis. Arthritis Rheum. 54, 158–168. https://doi.org/10.1002/art.21537

Woodsworth, D.J., Castellarin, M., Holt, R.A., 2013. Sequence analysis of T-cell repertoires in health and disease. Genome Med. 5, 98. https://doi.org/10.1186/gm502

Wooldridge, L., Ekeruche-Makinde, J., van den Berg, H.A., Skowera, A., Miles, J.J., Tan, M.P., Dolton, G., Clement, M., Llewellyn-Lacey, S., Price, D.A., Peakman, M., Sewell, A.K., 2012. A single autoimmune T cell receptor recognizes more than a million different peptides. J. Biol. Chem. 287, 1168–1177. https://doi.org/10.1074/jbc.M111.289488

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. https://doi.org/10.3389/fgene.2022.860510

Yagi, H., Nomura, T., Nakamura, K., Yamazaki, S., Kitawaki, T., Hori, S., Maeda, M., Onodera, M., Uchiyama, T., Fujii, S., Sakaguchi, S., 2004. Crucial role of FOXP3 in the development and function of human CD25+CD4+ regulatory T cells. Int. Immunol. 16, 1643–1656. https://doi.org/10.1093/intimm/dxh165

Ye, J., Lai, D., Cao, D., Tan, L., Hu, L., Zha, H., Yang, J., Shu, Q., 2021. Altered T-Cell Receptor β-Chain and Lactate Dehydrogenase Are Associated With the Immune Pathogenesis of Biliary Atresia. Front. Med. 8, 778500. https://doi.org/10.3389/fmed.2021.778500

Yee, C., Thompson, J.A., Byrd, D., Riddell, S.R., Roche, P., Celis, E., Greenberg, P.D., 2002. Adoptive T cell therapy using antigen-specific CD8+ T cell clones for the treatment of patients with metastatic melanoma: in vivo persistence, migration, and antitumor effect of transferred T cells. Proc. Natl. Acad. Sci. U. S. A. 99, 16168–16173. https://doi.org/10.1073/pnas.242600099

Yu, A., Zhu, L., Altman, N.H., Malek, T.R., 2009. A low interleukin-2 receptor signaling threshold supports the development and homeostasis of T regulatory cells. Immunity 30, 204–217. https://doi.org/10.1016/j.immuni.2008.11.014

Yu, W., Jiang, N., Ebert, P.J.R., Kidd, B.A., Müller, S., Lund, P.J., Juang, J., Adachi, K., Tse, T., Birnbaum, M.E., Newell, E.W., Wilson, D.M., Grotenbreg, G.M., Valitutti, S., Quake, S.R., Davis, M.M., 2015. Clonal Deletion Prunes but Does Not Eliminate Self-Specific αβ CD8(+) T Lymphocytes. Immunity 42, 929–941. https://doi.org/10.1016/j.immuni.2015.05.001

Zhang, N., Bevan, M.J., 2011. CD8(+) T cells: foot soldiers of the immune system. Immunity 35, 161–168. https://doi.org/10.1016/j.immuni.2011.07.010

Zhang, S.-X., Ma, X.-W., Li, Y.-F., Lai, N.-L., Huang, Z.-H., Fan, K., Wang, C.-H., Li, X.-F., 2018. The Proportion of Regulatory T Cells in Patients with Systemic Lupus Erythematosus: A Meta-Analysis. J. Immunol. Res. 2018, 1–11. https://doi.org/10.1155/2018/7103219

Zhang, Yanfang, Yang, X., Zhang, Yanxia, Zhang, Yan, Wang, M., Ou, J.X., Zhu, Y., Zeng, H., Wu, J., Lan, C., Zhou, H.-W., Yang, W., Zhang, Z., 2020. Tools for fundamental analysis functions of TCR repertoires: a systematic comparison. Brief. Bioinform. 21, 1706–1716. https://doi.org/10.1093/bib/bbz092

Zhao, D.-M., Thornton, A.M., DiPaolo, R.J., Shevach, E.M., 2006. Activated CD4+CD25+ T cells selectively kill B lymphocytes. Blood 107, 3925–3932. https://doi.org/10.1182/blood-2005-11-4502

Zheng, S.G., Wang, J., Wang, P., Gray, J.D., Horwitz, D.A., 2007. IL-2 Is Essential for TGF-β to Convert Naive CD4 $^{+}$ CD25 $^{-}$ Cells to CD25 $^{+}$ Foxp3 $^{+}$ Regulatory T Cells and for Expansion of These Cells. J. Immunol. 178, 2018–2027. https://doi.org/10.4049/jimmunol.178.4.2018

Zhou, J., Kong, C., Yu, J., Dong, H., Jin, C., Song, Q., 2014. Skewness of TCR Vβ of peripheral blood and synovial fluid of patients with rheumatoid arthritis. J. Immunoassay Immunochem. 35, 207–219. https://doi.org/10.1080/15321819.2013.841192

Zhu, J., Yamane, H., Paul, W.E., 2010. Differentiation of effector CD4 T cell populations (*). Annu. Rev. Immunol. 28, 445–489. https://doi.org/10.1146/annurev-immunol-030409-101212

Zinkernagel, R.M., Doherty, P.C., 1974. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. Nature 248, 701–702. https://doi.org/10.1038/248701a0

Annexes

Annexe 1.

Impaired Activated/Memory Regulatory T Cell Clonal Expansion Instigates Diabetes in NOD Mice

---

Vanessa Mhanna, Gwladys Fourcade, Pierre Barennes, Valentin Quiniou, Hang P. Pham, Paul-Gydeon Ritvo, Faustine Brimaud, Bruno Gouritin, Guillaume Churlaud, Adrien Six, Encarnita Mariotti-Ferrandiz, and David Klatzmann

---

DIABETES - 2021

Impaired Activated/Memory Regulatory T Cell Clonal Expansion Instigates Diabetes in NOD Mice

Vanessa Mhanna,[1] Gwladys Fourcade,[1] Pierre Barennes,[1] Valentin Quiniou,[1,2] Hang P. Pham,[3] Paul-Gydeon Ritvo,[1] Faustine Brimaud,[1] Bruno Gouritin,[1] Guillaume Churlaud,[1,2] Adrien Six,[1] Encarnita Mariotti-Ferrandiz,[1] and David Klatzmann[1,2]

Diabetes 2021;70:976–985 | https://doi.org/10.2337/db20-0896

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

PATHOPHYSIOLOGY

Regulatory T cell (Treg) insufficiency licenses the destruction of insulin-producing pancreatic β-cells by autoreactive effector T cells (Teffs), causing spontaneous autoimmune diabetes in NOD mice. We investigated the contribution to diabetes of the T-cell receptor (TCR) repertoires of naive regulatory T cells (nTregs), activated/memory Tregs (amTregs), and CD4$^+$ Teffs from prediabetic NOD mice and normal C57BL/6 (B6) mice. NOD mice amTreg and Teff repertoire diversity was unexpectedly higher than that of B6 mice. This was due to the presence of highly expanded clonotypes in B6 amTregs and Teffs that were largely lost in their NOD counterparts. Interleukin-2 (IL-2) administration to NOD mice restored such amTreg clonotype expansions and prevented diabetes development. In contrast, IL-2 administration only led to few or no clonotype expansions in nTregs and Teffs, respectively. Noteworthily, IL-2–expanded amTreg and nTreg clonotypes were markedly enriched in islet-antigen specific TCRs. Altogether, our results highlight the link between a reduced clonotype expansion within the activated Treg repertoire and the development of an autoimmune disease. They also indicate that the repertoire of amTregs is amenable to rejuvenation by IL-2.

NOD mice spontaneously develop autoimmune diabetes due to the destruction of insulin-producing β-cells by autoreactive CD4$^+$ and CD8$^+$ T cells (1). Diabetes susceptibility in NOD mice is attributed to genetic defects encoded by the insulin-dependent diabetes loci including >40 genes, among which the MHC class II and interleukin-2 (IL-2) genes have major contributions to the pathophysiological immune response leading to diabetes (2). The unique I-A^{g7} MHC class II molecule expressed by NOD mice has been reported to bind certain peptides with low affinity, thus impairing the thymocyte negative selection process (3,4). Reduced IL-2 production by activated effector T cells (Teffs) was found in NOD mice and linked to an impairment of regulatory T cell (Treg) numbers and functionality (5–7). An increased diabetes incidence was observed in NOD mice bearing an IL-2α receptor subunit with reduced affinity for IL-2, which was associated with a Treg but not a Teff defect (8). Likewise, the administration of low-dose IL-2 to NOD mice promotes Treg expansion and activation, prevents diabetes onset, and even cures diabetic mice (9,10). Thus, the altered tolerance promoting diabetes in NOD mice could be the result of: *1*) a defect in thymic selection of Treg and Teff T-cell receptor (TCR) repertoires, and/or *2*) an altered Treg fitness due to IL-2 deprivation.

Previous studies of the TCR repertoires of NOD mice and patients with type 1 diabetes yielded inconsistent results. Some studies showed a restricted TCR β-chain (TRB) V and TRBJ chain usage by CD4$^+$ T cells from pancreatic islets and pancreatic lymph nodes (PLN), indicating that overrepresented islet-specific TCRs preferentially express certain TRBVJ genes (11–13). Conversely, others showed a rather diverse gene usage among islet-infiltrating CD4$^+$ T cells in NOD mice (14), including islet-infiltrating memory CD4$^+$ T cells (15,16), and no common highly used TRBVBJ combinations in T cells from PLN of

[1]Sorbonne Universite, INSERM, UMRS959 Immunology-Immunopathology-Immunotherapy Laboratory, Paris, France
[2]Clinical Investigation Center in Biotherapy and Inflammation-Immunopathology-Biotherapy Department, Assistance Publique-Hôpitaux de Paris, Hôpital Universitaire Pitié-Salpêtrière, Paris, France
[3]Statistics Department, ILTOO Pharma, Paris, France

Corresponding author: David Klatzmann, david.klatzmann@sorbonne-universite.fr

Received 3 September 2020 and accepted 19 January 2021

This article contains supplementary material online at https://doi.org/10.2337/figshare.13611020.

A.S., E.M.-F., and D.K. are co-senior authors.

© 2021 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at https://www.diabetesjournals.org/content/license.

patients with type 1 diabetes (17). Moreover, NOD Tregs were found to have a restricted TCR α-chain (TRA) repertoire and limited overlap with thymic $CD4^+$ T cells (18,19), while we found similar TRB diversity between NOD peripheral Tregs and Teffs (20). These somehow conflicting results can be explained by the fact that: *1*) most of these studies were performed on unsorted $CD4^+$ T cells and thus did not distinguish between regulatory and effector populations, and/or *2*) most Treg repertoire analyses were done on a limited fraction of the repertoire by using TCR transgenic NOD mice or by focusing on few rearrangements (18,19). More importantly, most studies have overlooked the existence of Treg subsets.

Indeed, we previously showed that Tregs could be phenotypically divided into $CD4^+Foxp3^+CD44^{low}CD62L^{high}$ naive Tregs (nTregs) that have a low turnover and $CD4^+Foxp3^+CD44^{high}CD62L^{low}$ activated/memory Tregs (amTregs) that have a very rapid turnover due to their interaction with their cognate antigens (21). In the healthy C57BL/6 (B6) mouse background, we showed that amTregs are enriched in deep tissue–draining lymph nodes and are characterized by unique clonotype expansions (22). Moreover, we showed little overlap between amTreg and nTreg TRB repertoires (22). Thus, studying amTreg and nTreg TCR repertoires without separating them may be confounding.

In this study, we compared the TCR repertoire of splenic amTregs, nTregs, and Teffs of prediabetic NOD and normal B6 mice to investigate whether and how an altered TCR repertoire contributes to diabetes. We found that NOD amTregs have an unexpectedly more diverse repertoire than that of B6 mice, due to the absence of expanded clonotypes. Treatment of prediabetic NOD mice with IL-2, which prevents diabetes occurrence, restored the expansion of clonotypes that are preferentially diabetes-reactive. Thus, the TCR repertoire of amTregs is altered by an IL-2 insufficiency, which instigates diabetes development in NOD mice, and is amenable to therapeutic recovery.

RESEARCH DESIGN AND METHODS

Study Design

Given the rarity of some of the studied cell subsets, we pooled the spleens of six to eight young mice matched for genetic background, age, and sex. Six (three female and three male) nontreated C57BL/6 or NOD pools of mice were used to study the repertoire at homeostasis. Six young female NOD mice were treated with IL-2. Sample size was determined for statistical significance and experimental feasibility. Investigators were not blinded to the allocations during experiments and analyses.

Mice

Eight- to 10-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).

IL-2 Treatment

Recombinant AAV8 vectors were produced as described previously (9). Six female NOD-Foxp3-EGFP mice were injected once intraperitoneally with 10^{11} of recombinant AAVs diluted in PBS at 6 weeks of age. Serum was collected 2 weeks postinjection, and IL-2 levels were measured using a mouse IL-2 ELISA (eBioscience) according to the manufacturer's recommendations. Mice were sacrificed at 9 weeks of age.

Spleen Retrieval and Cell Sorting

Spleens were collected from euthanized mice and splenocytes were stained with anti–Ter-119-biotin (BD Biosciences), anti–B220-biotin (eBioscience), and anti–CD11c-biotin (eBioscience) antibodies for 20 min at 4°C and labeled with anti-biotin magnetic beads (Miltenyi Biotec) for 15 min at 4°C. B cells, erythrocytes, and monocytes were depleted on an AutoMACS separator (Miltenyi Biotec) following the manufacturer's procedure. Enriched T cells were stained for 20 min at 4°C with the following monoclonal antibodies at predetermined optimal dilutions: CD3-APC (BD Biosciences), CD4-Horizon-V500 (BD Biosciences), CD8-Alexa-700 (BD Biosciences), CD44-PE (BD Biosciences), and CD62L-eFluor-450 (eBioscience) and sorted on an FACSAria II cytometer (BD Biosciences) with a purity >95% into the following subsets: $CD4^+FoxP3^+CD62^{low}CD44^{high}$ (amTregs), $CD4^+FoxP3^+CD62^{high}CD44^{low}$ (nTregs), and $CD4^+FoxP3^-$ (Teffs). A total of 1×10^5–5×10^6 sorted cells were stored in RNAqueous kit lysis buffer (Invitrogen) at −80°C.

cDNA Library Preparation for TR Sequencing

RNA was extracted using the RNAqueous Total RNA Isolation Kit (Invitrogen). TRB libraries were prepared on 100 ng of RNA with the SMARTer Mouse TCR a/b Profiling Kit (Takara Bio) and sequenced with HiSeq 2,500 single read (300 bp) (Illumina) plus 10% PhiX at the LIGAN Genomics platform (23).

Raw Read Processing

Raw data in FASTQ format were processed for TRB sequence annotation using MiXCR software (v.3.0.3) (24), which ensures PCR and sequencing error corrections and provides for each sample a list of unique amino acid TRBs and their corresponding counts. Annotated sequences were then analyzed using an in-house workflow to eliminate long amino acid CDR3 sequences (length higher than the mean + 8). The resulting data sets are summarized in Supplementary Table 2.

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

Data Analysis

Statistical Comparison

Sample overlap was tested by computing the Jaccard distance (25) using the "vegan" R package and plotted into a heat map using the "pheatmap" package (26). Clustering analysis was performed using the "Ward" method. Principal component analysis was performed using the "stats" R package. Rényi entropy (27) was calculated using "vegan" and represents the distribution of clonal expansions as a function of the parameter α. Probability density plots were plotted using the "ggplot2" R package (28).

Network Analysis

For each cell population, the following method was adopted: the top most frequent 1,000 CDR3 amino acid sequences from each of the 6 B6 or NOD mice were pooled, and the mean frequency was calculated for every sequence. The list was then ordered decreasingly, and the first 1,000 sequences were selected. A matrix of pairwise Levenshtein distances (LDs) between the CDR3s was computed and used to construct the networks. In these networks, vertices (CDR3 sequences) are connected by edges (LD = 1) only if they differ by one amino acid (insertion/deletion/substitution). A connected cluster is defined as a set of two or more vertices connected by edges. The clustering coefficient of a node in a network quantifies how close its neighbors are to being a clique (a closed cluster). The maximum value is 1 if every neighbor connected to the node A is also connected to every other node within the neighborhood. The node size represents the mean frequency of each CDR3 in the network.

Analyses were performed using R packages: "stringdist" (29) was used to calculate LDs and "Igraph" (30) to calculate network properties. Graphics visualization was done using Cytoscape (31) and "ggplot2" for Figs. 3 and 5, respectively. In the latter, only clustered nodes are represented and edges are not shown.

Identification of Expanded Clonotypes

In each sample, the first (Q1) and third (Q3) quartiles and the interquartile range (IQR) were computed according to a negative binomial distribution fitted in the sequence read counts, while excluding private clonotypes with a count of 1. Expanded clonotypes are defined as those with counts greater than Q3 + 1.5 × IQR.

TCR Database

Mouse CDR3βs associated with diabetes ($n = 52$), experimental autoimmune encephalomyelitis (EAE; $n = 50$), systemic lupus erythematosus (SLE; $n = 49$), and cancer ($n = 52$) were collected from Friedman's published database (32). These sequences were derived from $CD4^+$ T cells and identified by different methods including reactivity assays, tetramer staining, selection of diabetes-specific TCRs following diabetes induction, and sequencing of tumor T-cell infiltrates in the context of cancer studies.

Statistical Analyses

Statistical analyses were performed using the nonparametric Wilcoxon test with Holm multiple-testing correction using "ggpubr" R package (33). A P value of ≤ 0.05 was considered statistically significant: $^*P \leq 0.05$; $^{**}P \leq 0.01$; $^{***}P \leq 0.001$; $^{****}P \leq 0.0001$; not significant is $P > 0.05$. Mean percentage comparisons of diabetes, EAE, SLE, and cancer CDR3s were done using the Wilcoxon paired test with diabetes as a reference group.

Data and Resource Availability

Source data are provided in the manuscript and the Supplementary Material. Fastq data were deposited in the National Center for Biotechnology Information Sequence Read Archive repository under the BioProject identification number PRJNA635928.

RESULTS

Differences in Repertoire Composition Between B6 and NOD

We analyzed and compared the TRB repertoire of splenic $CD4^+Foxp3^+CD44^{low}CD62L^{high}$ nTregs, $CD4^+Foxp3^+CD44^{high}CD62L^{low}$ amTregs, and $Foxp3^-CD4^+$ Teffs in B6 and NOD mice.

Compared with B6 mice, flow cytometry showed a significant decrease in the NOD Treg population ($P \leq 0.05$) associated with a decrease of amTreg proportions ($P \leq 0.01$) (Supplementary Table 1). Cell numbers for each sorted population and their corresponding number of TRB sequences and unique amino acid clonotypes are summarized in Supplementary Table 2.

Principal component analysis of TRBVBJ usage (i.e., the frequency of each combination in the repertoire irrespective of the clonotype abundance) clearly separated the genetic backgrounds and the cell populations (Fig. 1*A*). The first two components showed different TRBVBJ usage between B6 and NOD mice, with the first component (PC1) that separates the two genetic backgrounds explaining >68% of the variability. There was also a clear difference between the cell populations from the same background, with the two Treg subsets being close, but quite apart from Teffs. This reflects differences in the combined generation frequency and survival through the thymic selection and peripheral expansion of each generated TRBVBJ recombination.

To further investigate these differences, we assessed the clonotype (i.e., unique amino acid TRBV-CDR3-TRBJ sequence) overlap between samples using the Jaccard distance in which the higher the value, the more dissimilar the pair of samples compared (Fig. 1*B*). Hierarchical clustering on the Jaccard scores accurately separated B6 from NOD mouse samples. Surprisingly, within the latter cluster, amTregs are intermingled with nTregs and Teffs, while, as expected from our previous study (22), amTregs clustered separately in the B6 background. Overall, NOD mice display a peculiar repertoire structure compared with B6 mice both at the TRB gene and clonotype level.

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

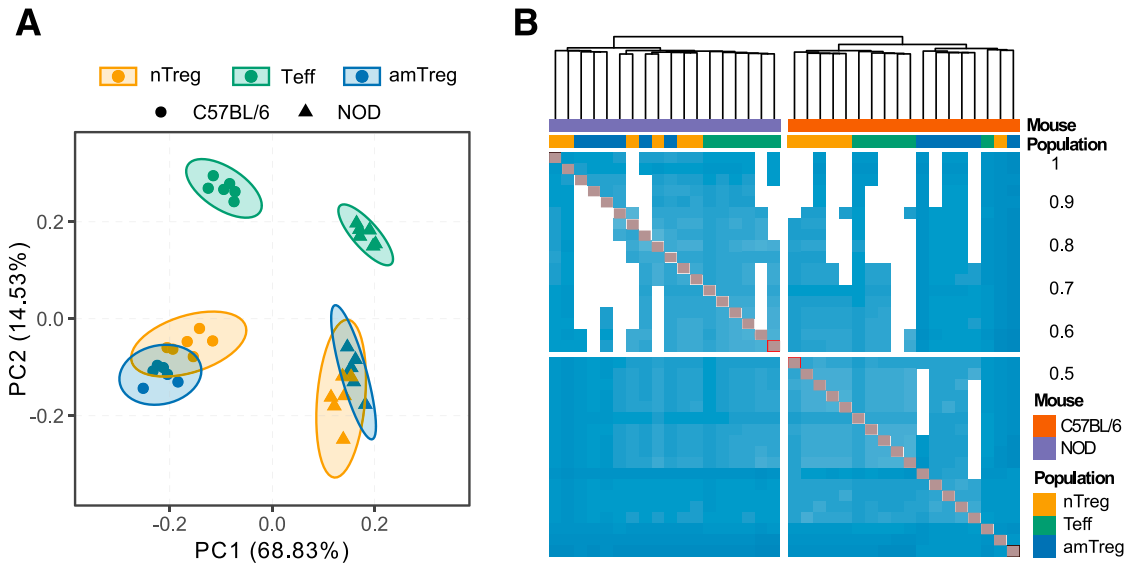


Figure 1—B6 and NOD mice express different TCR repertoires. *A*: Principal component analysis biplot for TRBVBJ usage of all B6 (dots) and NOD (triangles) samples. Cell populations are identified based on their colors. *B*: Heat map of the pairwise Jaccard distances calculated at the clonotype level. Scores range between 0 (complete similarity) and 1 (no similarity). The hierarchical clustering was performed using Ward's method.

Increased TCR Diversity of Teffs and amTregs in NOD Versus B6 Mice

We next explored the overall repertoire diversity at the clonotype level. We compared cell populations and mouse backgrounds by calculating the Rényi entropy, which evaluates diversity as a function of an α parameter (Fig. 2*A*). The higher the α, the higher is the weight put on frequent clonotypes. The higher the Rényi entropy, the more diverse is the repertoire.

In B6 mice, nTregs have the most diverse repertoire, while amTregs have the least diverse one, and Teffs fall in between. In contrast, while nTregs in NOD mice have a repertoire with a diversity similar to that in B6 mice, the diversities of the Teff and amTreg repertoires are increased compared with B6 mice (Fig. 2*A*). At $\alpha = 0$, for which the Rényi index represents clonotype richness, no differences were observed between the B6 and NOD backgrounds (Fig. 2*B*, left). This indicates that the overall richness of each cell subset is not affected in NOD mice. At $\alpha = 8$, which gives more weight to abundant clonotypes, Rényi indices were significantly higher in NOD amTregs and Teffs than in the corresponding subsets in B6 mice (Fig. 2*B*, right).

In view of these differences, we looked at the representativeness of the 100 most predominant clonotypes within each repertoire. The cumulative percentages of the 100 most predominant clonotypes in B6 amTregs and Teffs were significantly higher than their NOD counterparts (Fig. 2*C*). This is due to the presence, in the B6 cell populations, of highly frequent clonotypes depicted by blue lines, for which thickness is proportional to their percentage in the repertoire (Fig. 2*D*). In contrast, there were no such clonotypes in B6 nor in NOD nTregs.

Collectively, these results reveal that a healthy nTreg repertoire is diverse with no clonotype expansions. Conversely, the healthy amTreg repertoire, and to a lesser extent the Teff repertoire, is enriched in clonotypes with high frequency, likely resulting from antigen-driven expansions. In diabetes-prone NOD mice, such expansions are severely reduced in amTregs and Teffs.

Altered Similarity Structure of NOD amTreg and Teff Repertoires

Next, we examined the architecture of the repertoires. To this end, the 1,000 most frequent CDR3s per population were used to build similarity networks in which each CDR3 (node) has a size proportional to its frequency and is connected by edges to other CDR3s that differ to it by only one amino acid (i.e., having an LD = 1). This distance was shown to connect CDR3s that are likely to bind the same peptide (34).

The structure of the nTreg repertoire appeared similar for B6 and NOD mice, made of highly clustered CDR3s with low frequency (Fig. 3*A*). In contrast, B6 amTreg repertoire was mostly composed of individual or lightly connected with high frequencies, reflecting polyclonal self-antigen–driven expansions (22,35). In accordance with the previous observations (Figs. 1 and 2), these expansions are much reduced in amTregs from NOD mice. Similarly, node sizes of their Teffs appeared smaller than in B6 mice (Fig. 3*A*, middle).

Node frequency showed no differences for nTregs, a modest decrease for Teffs, and a marked decrease for amTregs in NOD mice compared with their counterparts in B6 mice (Fig. 3*B*). The clustering coefficient of each CDR3 within each network, which gives an indication of the

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

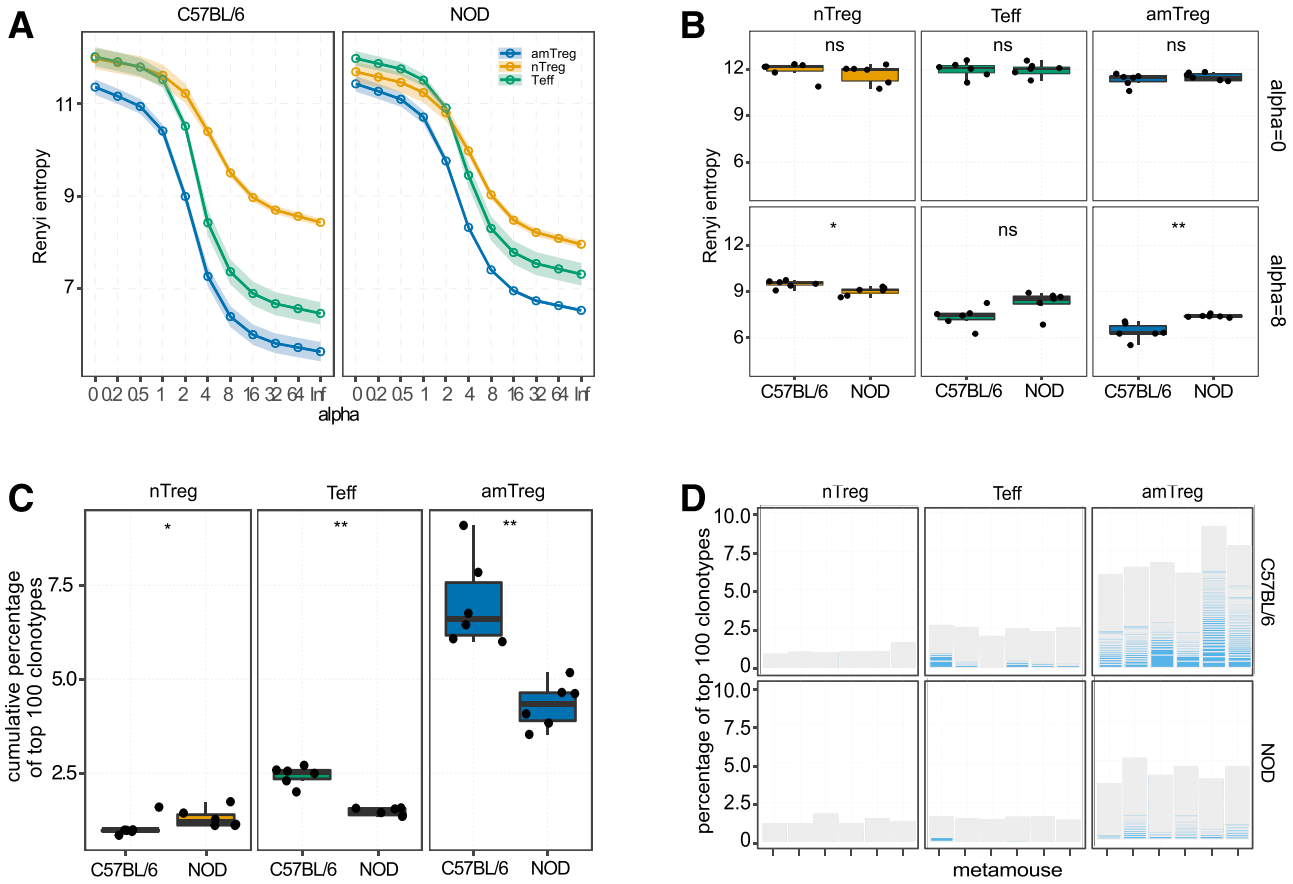


Figure 2— Increased diversity of NOD Teffs and amTregs compared with B6. *A*: Rényi profiles for nTregs (yellow), CD4 Teffs (green), and amTregs (blue) in B6 and NOD mice. Points represent the mean of all six mice, and shaded areas represent the SE. *B*: Rényi values were compared between B6 and NOD mice within each cell population at $\alpha = 0$ and $\alpha = 8$. *C*: Mean cumulative percentage of the 100 most predominant clonotypes of the six mice for each cell population. *D*: Percentage of the top 100 clonotypes within the three cell populations in B6 (top) and NOD (bottom) backgrounds. Each blue line and its thickness represent a unique clonotype and its occurrence; gray bars are the stacking of slim gray lines representing rare clonotypes. *B* and *C*: $n = 6$. $P > 0.05$ is NS. $^*P \leq 0.05$, $^{**}P \leq 0.01$, Wilcoxon test with Holm correction.

importance of these nodes in the global repertoire architecture, showed no significant difference between NOD and B6 nTregs (Fig. 3*C*). However, significantly higher clustering coefficients were observed in Teffs and amTregs from NOD compared with B6 mice.

Overall, the structural analysis of these networks highlights a higher sequence similarity of Teff and amTreg CDR3s in NOD mice. As IL-2 production is notoriously low in NOD mice and Tregs rely primarily on it for proliferation, we hypothesized that the reduced amTreg expansions may result from an IL-2 shortage.

Low-Dose IL-2 Normalizes NOD amTreg Repertoires

As administration of IL-2 to NOD mice specifically enables the expansion of Tregs (9), we investigated whether the abnormal repertoire diversity and structure of prediabetic NOD mice could be restored by an IL-2 treatment. NOD mice were injected with an adeno-associated virus coding for IL-2 (AAV-IL-2) that allows the continuous production of IL-2 at low dose and prevents the occurrence of diabetes (9). As expected, total Treg proportions increased significantly after IL-2 treatment (Fig. 4*A*). Importantly, within the Treg subsets, amTreg proportions showed a significant increase, at the expense of nTregs (Fig. 4*A* and Supplementary Table 3).

We next analyzed the expansions at the clonotype level after IL-2 treatment. To this end, we identified expanded clonotypes as the ones showing counts greater than the Q3 + 1.5 × IQR value, in which Q3 is the third quartile and IQR the IQR of sequence count distribution (see RESEARCH DESIGN AND METHODS and Supplementary Table 4). The abundance of the expanded clonotypes from each sample was plotted in Fig. 4*B* (left plots). Each circle represents a clonotype for which size is proportional to its frequency in the repertoire. amTregs and nTregs from AAV-IL-2–treated NOD mice were significantly enriched in expansions compared with amTregs and nTregs from B6 or NOD mice. This is confirmed by the density plots, which show that IL-2 treatment induces expansions of nTregs and amTregs compared with their B6 and NOD counterparts,

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

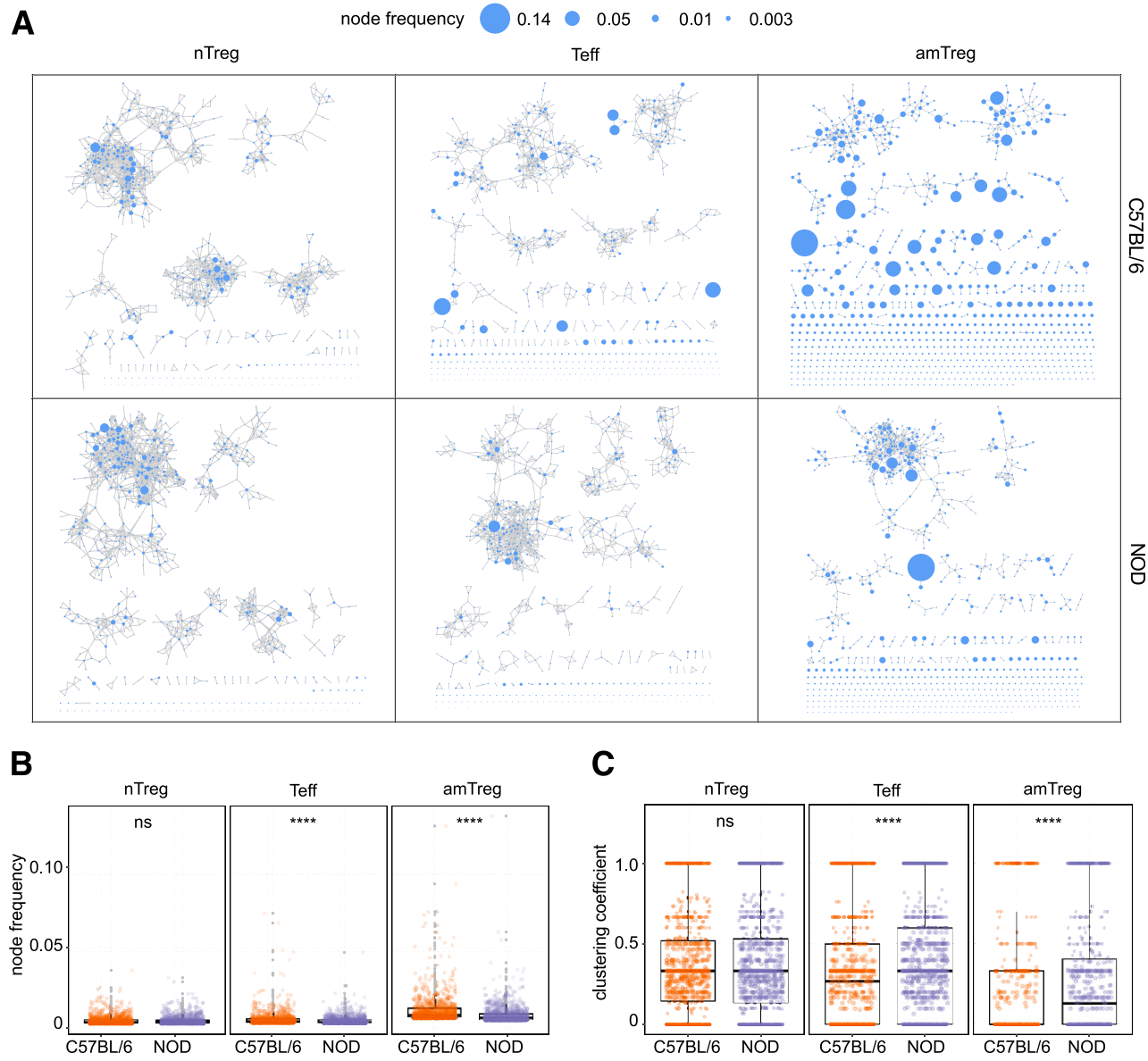


Figure 3—Similarity structure of nTreg, Teff, and amTreg CDR3 repertoires. *A*: For each cell population, the most frequent 1,000 CDR3s across all mice were identified based on their mean frequency and used to construct networks based on the LD: CDR3 sequences (nodes) with one amino acid difference (insertion/deletion/substitution) are connected by an edge. One network was constructed per cell population and mouse background. *B*: CDR3 frequency comparison between B6 and NOD mice for each cell subset. *C*: Comparison of the clustering coefficient, which quantifies the degree to which neighbor nodes tend to cluster together, between B6 and NOD mice in each cell subset. CDR3s with no connections are excluded. *B* and *C*: Each dot represents a CDR3, and box plots represent the median, first, and third quartiles. $P > 0.05$ is NS. ****$P \leq 0.0001$, Wilcoxon test with Holm correction.

but not of Teffs (Fig. 4*B*, right). Thus, IL-2 administration restores amTreg clonotype expansions that were absent in NOD mice.

IL-2–Expanded Tregs Are Enriched in Diabetes-Specific Clonotypes

We next examined whether the expanded clonotypes could be linked to diabetes. We queried a published catalog of TCRs with defined specificity (32). We identified a total of 203 mouse CDR3β amino acid sequences from $CD4^+$ T cells associated with diabetes ($n = 52$), EAE ($n = 50$), SLE ($n = 49$), and cancer ($n = 52$). We then searched for these sequences within the expanded clonotypes and calculated their mean enrichment in each cell subset (Fig. 5*A*). Interestingly, diabetes-specific sequences showed higher expansions compared with nondiabetes-specific sequences in nTregs and amTregs from IL-2–treated NOD mice. There was also a modest enrichment of these sequences in B6 amTregs. To further analyze the specificity pattern toward diabetes, we connected disease-specific

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

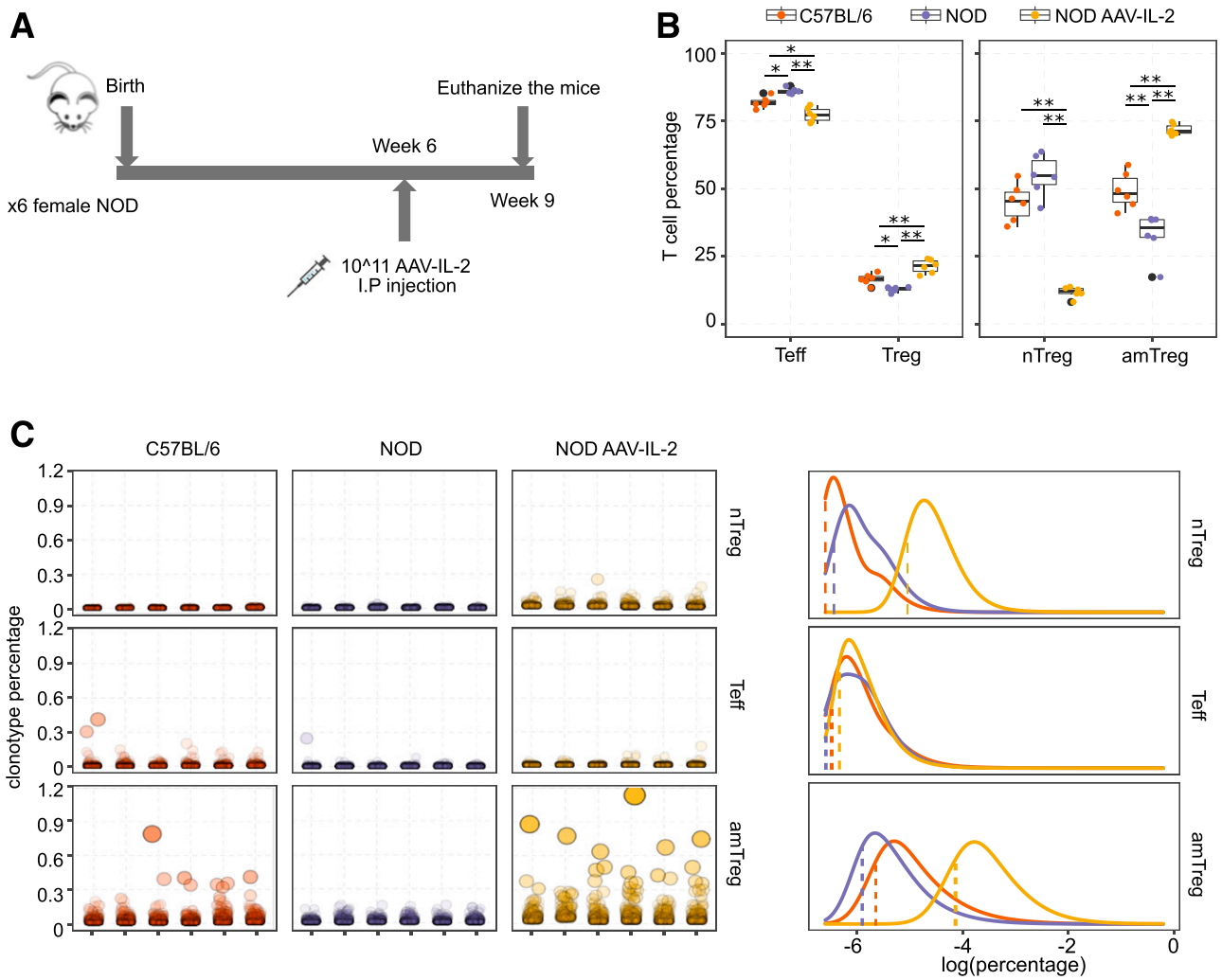


Figure 4— IL-2 induces Treg clonotype expansions. *A*: Experimental scheme. *B*: Box plots show the percentage of T cells in each B6 (orange), NOD (blue), and IL-2–treated NOD (yellow) mice (n = 6/background). *C*, left: The abundance (in percentage) of expanded clonotypes for each sample (each column represents a single sample) within the three populations and mouse backgrounds. Expanded clonotypes were identified as the ones with counts greater than Q3 + 1.5 × IQR according to a fitted negative binomial distribution. *C*, right: Probability density plot based on the mean expanded clonotype distribution between the three mouse conditions for each cell subset. Colored dashed lines represent the mode of each curve. *$P \leq 0.05$, **$P \leq 0.01$. I.P, intraperitoneal.

CDR3s by LD = 1 to the pool of expanded amTreg and nTreg CDR3s. The number of neighbors with LD = 1 to diabetes-related sequences showed a significant increase compared with other disease-specific CDR3s (Fig. 5*B*).

Noteworthily, clonotypes identified as specific for islet antigens such as insulin and GAD by tetramer staining were detected in large clusters of IL-2–expanded clonotypes (Fig. 5*C* and Supplementary Table 5).

DISCUSSION

In this study, we report an altered amTreg TCR repertoire in NOD mice that can be restored to normal by an IL-2 treatment. IL-2–expanded Tregs were enriched in islet-antigen–specific TCRs and were associated with protection from diabetes. These results have both heuristic and applied implications.

The Structure of the amTreg and nTreg Repertoires in a Healthy Mouse Background

In normal B6 mice, a control model that does not spontaneously develop any inflammatory or autoimmune diseases, nTreg repertoires are very diverse, even more than those of Teffs, with most clonotypes forming dense networks. These are the attributes of an unmodified and tightly connected postthymic selection repertoire (36). In contrast, amTreg repertoires have a reduced diversity and are enriched in frequent clonotypes with few connections. This illustrates the responses to distinct self-antigens, as expected from a cell population that is constantly activated to control autoimmunity in numerous tissues. Such similarity structures have also been described in the context of immunization and aging (34).

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

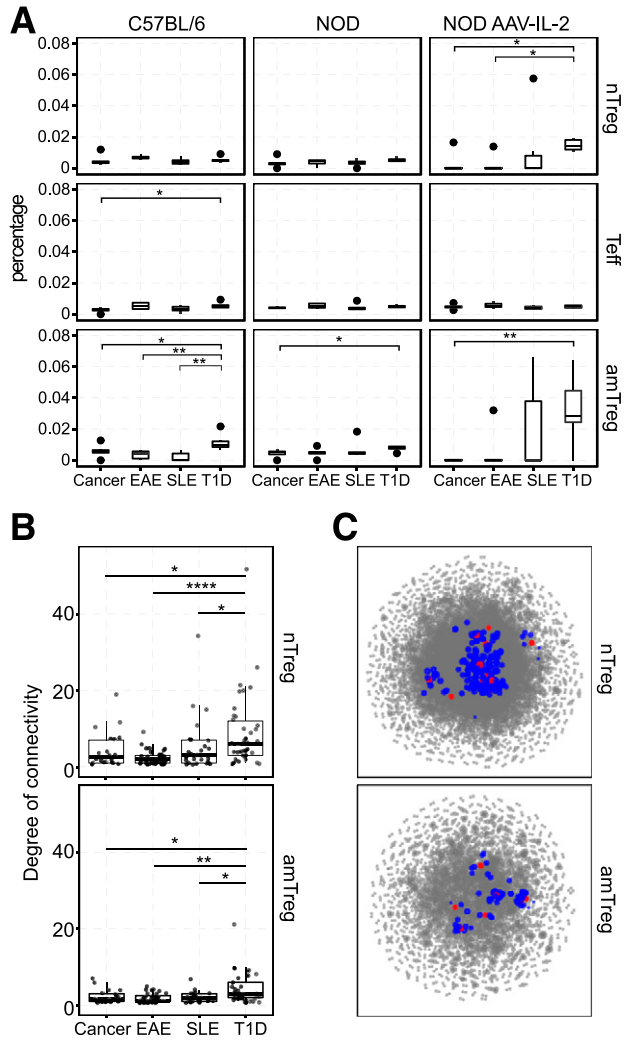


Figure 5—IL-2 preferentially expands clonotypes with a diabetes-related specificity. *A*: The proportion of CDR3s specific for cancer, EAE, SLE, and diabetes found within the expanded clonotypes of each population ($n = 6$). Box-and-whisker plots show the median and outliers. *B*: CDR3s with a disease-related specificity were connected by LD = 1 to the pool of amTreg and nTreg CDR3s identified across the six IL-2–treated NOD mice. The degree of connectivity (i.e., the number of LD = 1 edges connecting a given node) of disease-specific CDR3s is plotted ($n = 52$ in cancer, $n = 50$ in EAE, $n = 49$ in SLE, and $n = 52$ in type 1 diabetes [T1D]). *A* and *B*: Wilcoxon test was performed with the T1D group as a reference: $P > 0.05$ is NS; $^{*}P \leq 0.05$; $^{**}P \leq 0.01$; $^{****}P \leq 0.0001$. *C*: Graphical representation of the LD = 1 networks in NOD AAV-IL-2–treated nTregs and amTregs. Unconnected sequences are not shown. Identical sequences found in the database and samples from IL-2–treated mice are shown in red and their neighbors in blue as well as diabetes-related sequences from the database.

In addition, amTreg repertoires show a higher overlap with each other than with nTregs or Teffs, indicating that the self-antigens to which amTregs respond are shared between individuals. This is also in agreement with our previous report of a low sequence overlap between the two Treg subsets in the peripheral lymph nodes (22). The comparison with Treg TRB repertoires from the NOD autoimmune-prone mouse strain strengthens this conclusion. Reduction of amTreg major expansions is associated with disease susceptibility and their recovery with disease prevention.

Collectively, these results identify amTregs as the major subset involved in the pathogenesis of autoimmunity and warrant separate analysis of amTregs and nTregs when investigating their roles in health or disease.

amTreg Activation Depends on IL-2

The severe reduction in frequent clonotypes found in NOD mice, and their appearance under IL-2 provision, highlight a major role of IL-2 in supporting amTreg survival.

NOD mice have a genetically determined IL-2 insufficiency, which is also reported in human type 1 diabetes as well as in SLE and rheumatoid arthritis (37,38). It is thus likely that the observed defect in amTreg fitness contributes to the pathophysiology of these diseases. If so, the good news is that disease-relevant amTregs can be regenerated by IL-2 treatment. We show in this study that one injection of recombinant AAV-IL-2 in prediabetic mice was enough to expand NOD amTregs at the cell and repertoire levels. In our experiments, the repertoire was analyzed 21 days after the initiation of treatment. As the approximate duration of T-cell differentiation in the thymus is 28 days (39), newly generated thymocytes are unlikely to contribute to the expanded clonotypes in the amTreg population. Thus, the IL-2–induced amTregs most probably originate from peripheral amTregs and/or nTregs that received their activation signal in the presence of IL-2. The diabetes-related TCR expansions in both nTregs and amTregs further support that both populations contribute to the restoration of a more balanced and complete disease-relevant amTreg repertoire.

Treg Specificity to Diabetes-Related Antigens

The existence of public databases containing TCRs with known peptide specificities allowed us to investigate the presence of antigen-specific sequences in our repertoires. Diabetes-related CDR3s identified from the database were overrepresented in IL-2–treated NOD Tregs and most connected in LD = 1 networks, as shown by their high number of neighbors. Linking the CDR3 similarity networks to annotated TCR data sets was previously described by Madi et al. (34) and showed that CDR3s with similar specificity tend to cluster together. This indicates that diabetes-specific CDR3s could have core motifs engaged in pancreatic antigen recognition and shows the potential of such a tool in identifying novel pathogenic TCRs. Indeed, repertoire analysis of islet-infiltrating T cells revealed an accumulation of TCRs that were reactive in vitro to uncharacterized autoantigens, distinct from the well-known GAD and insulin antigens (13).

These observations highlight a relationship between the specific expansion of pancreatic antigen-related TCRs and the prevention of diabetes post–IL-2 treatment.

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

Contribution of I-A^{g7} to NOD Pathophysiology

It was hypothesized that diabetes development in NOD mice is associated with the I-A^{g7} molecule that they express. Coexpression with a different type of MHC or mutations within the I-A^{g7} molecule was shown to prevent insulitis and diabetes (4,40). In contrast, substitution of non-MHC regions from the C57BL/10 into the NOD strain was sufficient to prevent insulitis or diabetes, despite the expression of the I-A^{g7} molecule (41). Our studies highlight the importance of non-MHC genes in diabetes onset in the NOD mouse model. Indeed, we report in this study that IL-2 administration to prediabetic NOD mice, previously shown to prevent the disease outcome, improves amTreg fitness. This would argue that NOD Treg repertoires have the full potential to control autoimmune responses, but fail to do so because an IL-2 deficiency impairs the amTreg repertoire that can control islet autoreactivity.

However, our results do not exclude a potential enrichment of autoreactive T cells in the NOD mice repertoire that would take the lead in the periphery where Tregs fail to expand. In this regard, NOD mice have an overrepresentation of CDR3β sequences carrying hydrophobic amino acid doublets at positions 6 and 7 that have been shown to confer high self-reactivity (42). Further studies of the TCR repertoire selection process in the thymus of NOD mice should help address this important issue.

Altogether, our results reveal a link between a reduced expansion of clonotypes within the activated Treg TCR repertoire and the development of an autoimmune disease. IL-2 treatment not only activates Tregs, but also restores a repertoire of TCRs with relevant specificities. Thus, our work indicates that fit amTregs are key to protection from autoimmune diseases and highlights the value of studying these cells separately from the whole Treg population. Importantly, amTreg repertoire abnormalities and restoration are detected in the spleen, which affords a readout that is more representative of the circulation than PLN and thus has a translational value. Our results should guide further studies to elucidate the dynamics of pancreatic antigen-specific amTregs in diabetic versus nondiabetic NOD mice, including in mice outside of a pathogen-free environment, and identify TCR-based biomarkers of diabetes.

Acknowledgments. The authors thank F. Jebbawi and L. Barientos for the help in sample collection and cell sorting, M. Mendez and N. Papaya for the help in RNA extraction and library preparations, and K. El Soufi for the help in network analyses. The authors also thank the staff of the animal care facility for the excellent handling of mice and the 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, for performing the sequencing of our libraries.

Funding. This work was funded by the Training & Research Involving Polymer Optical Devices European Research Council Advanced EU (322856) and Recherche Hospitalo-Universitaire iMAP (ANR-16-RHUS-0001) grants to D.K., with additional support from the Sorbonne Université, INSERM, and Assistance Publique-Hôpitaux de Paris.

Duality of Interest. H.P.P was an employee of ILTOO Pharma. D.K. is an investor in a patent related to the use of IL-2 in the treatment of autoimmune diseases, owned by his public institution and licensed to ILTOO Pharma, in which D.K. has interests. No other potential conflicts of interest relevant to this article were reported.

Author Contributions. V.M., G.F., P.B., V.Q., P.-G.R. F.B., B.G., G.C., and E.M.-F. performed the experiments. V.M. analyzed the data, with contributions from all authors. V.M. and H.P.P. performed the statistical analyses. V.M., A.S., E.M.-F., and D.K. conceived the experiments. V.M., A.S., E.M.-F., and D.K. wrote the manuscript, with input from all authors. D.K. conceptualized and supervised the study. D.K. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Verdaguer J, Schmidt D, Amrani A, Anderson B, Averill N, Santamaria P. Spontaneous autoimmune diabetes in monoclonal T cell nonobese diabetic mice. J Exp Med 1997;186:1663–1676

2. Chen Y-G, Mathews CE, Driver JP. The role of NOD mice in type 1 diabetes research: lessons from the past and recommendations for the future. Front Endocrinol (Lausanne) 2018;9:51

3. Kanagawa O, Vaupel BA, Xu G, Unanue ER, Katz JD. Thymic positive selection and peripheral activation of islet antigen-specific T cells: separation of two diabetogenic steps by an I-A(g7) class II MHC beta-chain mutant. J Immunol 1998; 161:4489–4492

4. Lund T, O'Reilly L, Hutchings P, et al. Prevention of insulin-dependent diabetes mellitus in non-obese diabetic mice by transgenes encoding modified I-A β-chain or normal I-E α-chain. Nature 1990;345:727–729

5. Klatzmann D, Abbas AK. The promise of low-dose interleukin-2 therapy for autoimmune and inflammatory diseases. Nat Rev Immunol 2015;15:283–294

6. Long SA, Cerosaletti K, Bollyky PL, et al. Defects in IL-2R signaling contribute to diminished maintenance of FOXP3 expression in CD4(+)CD25(+) regulatory T-cells of type 1 diabetic subjects. Diabetes 2010;59:407–415

7. Sgouroudis E, Albanese A, Piccirillo CA. Impact of protective IL-2 allelic variants on CD4+ Foxp3+ regulatory T cell function in situ and resistance to autoimmune diabetes in NOD mice. J Immunol 2008;181:6283–6292

8. Dwyer CJ, Bayer AL, Fotino C, et al. Altered homeostasis and development of regulatory T cell subsets represent an IL-2R-dependent risk for diabetes in NOD mice. Sci Signal 2017;10:eaam9563

9. Churlaud G, Jimenez V, Ruberte J, et al. Sustained stimulation and expansion of Tregs by IL2 control autoimmunity without impairing immune responses to infection, vaccination and cancer. Clin Immunol 2014;151:114–126

10. Rosenzwajg M, Churlaud G, Mallone R, et al. Low-dose interleukin-2 fosters a dose-dependent regulatory T cell tuned milieu in T1D patients. J Autoimmun 2015;58:48–58

11. Marrero I, Vong A, Dai Y, Davies JD. T cell populations in the pancreatic lymph node naturally and consistently expand and contract in NOD mice as disease progresses. Mol Immunol 2012;52:9–18

12. Li L, He Q, Garland A, et al. beta cell-specific CD4+ T cell clonotypes in peripheral blood and the pancreatic islets are distinct. J Immunol 2009;183:7585–7591

13. Baker FJ, Lee M, Chien YH, Davis MM. Restricted islet-cell reactive T cell repertoire of early pancreatic islet infiltrates in NOD mice. Proc Natl Acad Sci U S A 2002;99:9374–9379

14. Nakano N, Kikutani H, Nishimoto H, Kishimoto T. T cell receptor V gene usage of islet beta cell-reactive T cells is not restricted in non-obese diabetic mice. J Exp Med 1991;173:1091–1097

15. Marrero I, Hamm DE, Davies JD. High-throughput sequencing of islet-infiltrating memory CD4+ T cells reveals a similar pattern of TCR Vβ usage in prediabetic and diabetic NOD mice. PLoS One 2013;8:e76546

16. Marrero I, Aguilera C, Hamm DE, Quinn A, Kumar V. High-throughput sequencing reveals restricted TCR Vβ usage and public TCRβ clonotypes among

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

pancreatic lymph node memory CD4(+) T cells and their involvement in autoimmune diabetes. Mol Immunol 2016;74:82–95
17. Seay HR, Yusko E, Rothweiler SJ, et al. Tissue distribution and clonal diversity of the T and B cell repertoire in type 1 diabetes. JCI Insight 2016;1:e88242
18. Ferreira C, Singh Y, Furmanski AL, Wong FS, Garden OA, Dyson J. Non-obese diabetic mice select a low-diversity repertoire of natural regulatory T cells. Proc Natl Acad Sci U S A 2009;106:8320–8325
19. Kern J, Drutel R, Leanhart S, Bogacz M, Pacholczyk R. Reduction of T cell receptor diversity in NOD mice prevents development of type 1 diabetes but not Sjögren's syndrome. PLoS One 2014;9:e112467
20. Ritvo P-G, Saadawi A, Barennes P, et al. High-resolution repertoire analysis reveals a major bystander activation of Tfh and Tfr cells. Proc Natl Acad Sci U S A 2018;115:9604–9609
21. Fisson S, Darrasse-Jèze G, Litvinova E, et al. Continuous activation of autoreactive CD4+ CD25+ regulatory T cells in the steady state. J Exp Med 2003; 198:737–746
22. Bergot A-S, Chaara W, Ruggiero E, et al. TCR sequences and tissue distribution discriminate the subsets of naïve and activated/memory Treg cells in mice. Eur J Immunol 2015;45:1524–1534
23. Barennes P, Quiniou V, Shugay M, et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. Nat Biotechnol 2021; 39:236–245
24. Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods 2015;12:380–381
25. Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bull Soc Vaudoise Sci Nat* 1901;37:547–579
26. Kolde R. pheatmap, 2019. Accessed 10 January 2021. Available from https://rdrr.io/cran/pheatmap/
27. Rényi A. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA*, 1961. Berkeley, CA, University of California Press, pp. 547–561
28. Wickham H, Navarro D, Lin Pedersen T. *ggplot2: Elegant Graphics for Data Analysis.* 2nd ed. New York, Springer, 2016
29. van der Loo MP. The stringdist package for approximate string matching. R J 2014;6:111–122
30. Csardi G, Nepusz T. The igraph software package for complex network research. Complex Syst 2006;1695:1–9
31. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13: 2498–2504
32. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. Bioinformatics 2017;33:2924–2929
33. Kassambara A. ggpubr R Package: ggplot2-based publication ready plots, 2020. Accessed 10 January 2021. Available from https://rpkgs.datanovia.com/ggpubr/
34. Madi A, Poran A, Shifrut E, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. eLife 2017;6:e22057
35. Darrasse-Jèze G, Bergot A-S, Durgeau A, et al. Tumor emergence is sensed by self-specific CD44hi memory Tregs that create a dominant tolerogenic environment for tumors in mice. J Clin Invest 2009;119:2648–2662
36. Elhanati Y, Murugan A, Callan CG Jr., Mora T, Walczak AM. Quantifying selection in immune receptor repertoires. Proc Natl Acad Sci U S A 2014;111: 9875–9880
37. Kitas GD, Salmon M, Farr M, Gaston JS, Bacon PA. Deficient interleukin 2 production in rheumatoid arthritis: association with active disease and systemic complications. Clin Exp Immunol 1988;73:242–249
38. Zier KS, Leo MM, Spielman RS, Baker L. Decreased synthesis of interleukin-2 (IL-2) in insulin-dependent diabetes mellitus. Diabetes 1984;33: 552–555
39. Thomas-Vaslin V, Altes HK, de Boer RJ, Klatzmann D. Comprehensive assessment and mathematical modeling of T cell population dynamics and homeostasis. J Immunol 2008;180:2240–2250
40. Gioia L, Holt M, Costanzo A, et al. Position β57 of I-A^{g7} controls early anti-insulin responses in NOD mice, linking an MHC susceptibility allele to type 1 diabetes onset. Sci Immunol 2019;4:eaaw6329
41. Serreze DV, Bridgett M, Chapman HD, Chen E, Richard SD, Leiter EH. Subcongenic analysis of the Idd13 locus in NOD/Lt mice: evidence for several susceptibility genes including a possible diabetogenic role for β 2-microglobulin. J Immunol 1998;160:1472–1478
42. Stadinski BD, Shekhar K, Gómez-Touriño I, et al. Hydrophobic CDR3 residues promote the development of self-reactive T cells. Nat Immunol 2016; 17:946–955

Downloaded from http://diabetesjournals.org/diabetes/article-pdf/70/4/976/439151/db200896.pdf by INSERM DISC IST user on 27 April 2022

Supplementary

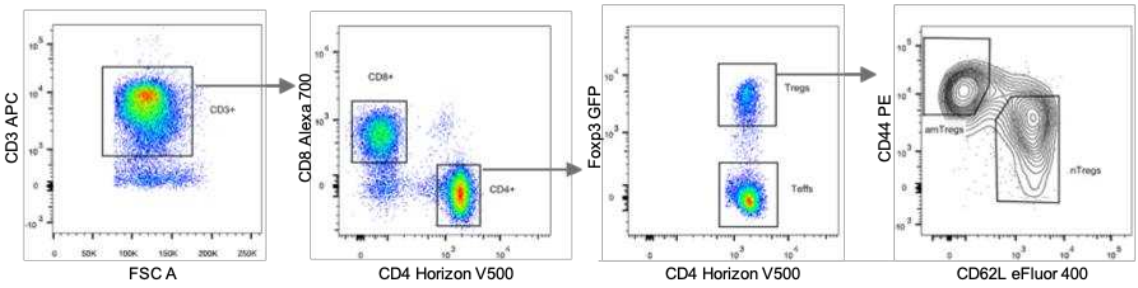


Figure S1. Representative flow cytometry gating

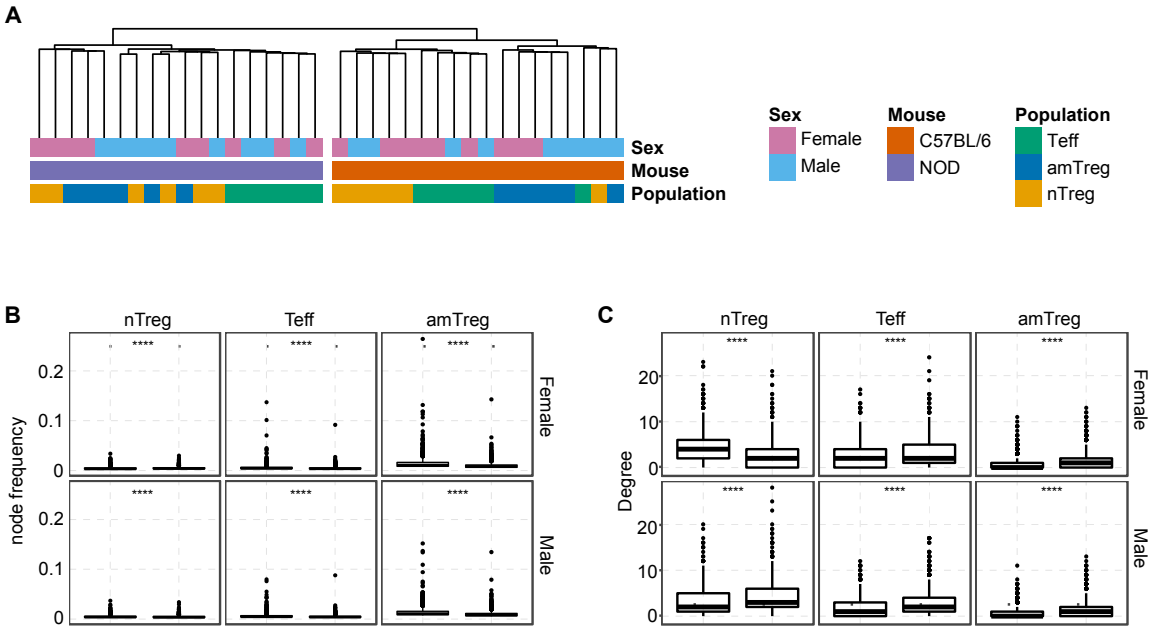


Figure S2: No sex-biased differences between B6 and NOD samples. (A) A hierarchical clustering was performed using Ward's method on the pairwise Jaccard distances calculated at the clonotype level. Colors represent the sex, mouse background and cell population. **(B)** For each cell population, the top most frequent 1000 CDR3s across same sex mice were identified based on their mean frequency and used to construct networks based on the Levenshtein distance. CDR3 frequency comparison between B6 and NOD

mice for each cell subset and sex. **(C)** Comparison of the degree of connectivity, which represent the number of connections of each node in a network, by sex within each cell subset. (B, C) Each dot represents a CDR3 and boxplots represent the median, first and third quartiles. Ns $P > 0.05$, *** $P \leq 0.001$, **** $P \leq 0.0001$, Wilcoxon test with Holm's correction.

		Proportion in CD4+ T cells (%)		Proportion in Tregs (%)	
Mouse background	Sample	Teffs	Tregs	nTregs	amTregs
C57BL/6	MM1	81.3	17.0	54.7	40.9
	MM2	79.2	19.3	44.6	49.4
	MM3	82.8	15.7	46.3	47.1
	MM4	81.0	17.7	38.4	55.3
	MM5	85.3	13.3	49.5	44.3
	MM6	82.1	16.4	36.0	58.8
NOD	MM7	86.0	12.8	62.1	17.3
	MM8	85.0	13.3	63.7	31.8
	MM9	88.0	11.1	50.6	32.6
	MM10	85.5	13.5	55.1	38.4
	MM11	86.5	12.4	42.8	38.6
	MM12	85.3	13.6	54.3	38.8

Table S1. Cell proportions in B6 and NOD mice

Mouse background	Sex	Cell population	Sample	No of sorted cells	No of raw reads	No of productive sequences	No of amino acid clonotypes
C57BL/6	Female	nTreg	MM1	1,323,341	2,127,673	1,015,715	233,660
			MM2	1,071,840	2,004,480	733,190	218,263
			MM3	1,246,936	2,817,720	446,115	135,535
		CD4 Teff	MM1	5,364,460	3,761,403	1,349,423	223,224
			MM2	4,234,206	2,581,178	714,488	120,433
			MM3	4,185,207	4,060,009	807,256	172,964
		amTreg	MM1	1,500,745	2,398,861	1,023,589	96,799
			MM2	1,358,814	1,631,017	589,358	101,868
			MM3	1,036,280	2,955,847	1,028,115	68,919
	Male	nTreg	MM4	1,131,312	1,279,418	500,408	199,423
			MM5	1,345,776	1,688,914	649,218	195,734
			MM6	1,071,282	1,533,625	451,775	54,575
		CD4 Teff	MM4	5,345,742	2,383,211	1,029,539	300,505
			MM5	3,975,052	2,946,289	794,384	192,373
			MM6	4,594,811	663,005	222,483	70,411
		amTreg	MM4	1,153,056	2,509,275	857,770	106,989
			MM5	1,201,394	2,551,643	823,937	125,671
			MM6	818,961	2,309,942	550,789	40,864
NOD	Female	nTreg	MM7	1,262,693	4,479,584	1,125,054	173,355
			MM8	663,051	2,429,643	592,815	71,037
			MM9	910,987	2,601,781	822,592	86,159
		CD4 Teff	MM7	3,560,531	5,999,296	1,212,278	207,209
			MM8	4,083,330	4,051,115	742,367	180,372
			MM9	4,794,056	3,286,016	734,086	112,917
		amTreg	MM7	1,015,255	3,069,939	685,274	97,993
			MM8	525,617	3,459,893	644,604	77,448
			MM9	1,044,756	4,664,421	912,418	134,700
	Male	nTreg	MM10	2,829,659	44,533,470	924,356	228,347
			MM11	1,187,673	4,842,671	1,152,406	171,911
			MM12	1,417,099	3,743,743	905,722	162,334
		CD4 Teff	MM10	1,828,793	8,402,032	1,440,109	293,362
			MM11	4,069,621	4,026,173	812,452	137,904
			MM12	5,129,815	5,104,745	1,242,181	172,096
		amTreg	MM10	1,439,570	4,483,668	1,118,172	140,069
			MM11	954,098	2,936,788	551,128	82,532
			MM12	1,073,035	5,172,199	926,015	116,134
NOD AAV-IL2	Female	nTreg	M1	260,496	561,066	180,251	24,024
			M2	298,114	1,722,615	578,756	34,327
			M3	279,845	1,903,257	563,135	30,487
			M4	221,511	2,582,384	817,599	45,441
			M5	160,748	3,020,522	884,592	45,018
			M6	178,671	3,628,737	943,936	51,778
		CD4 Teff	M1	4,202,771	3,634,559	706,155	165,175
			M2	4,683,268	2,885,576	689,779	165,703
			M3	3,922,300	4,688,586	1,170,572	182,343
			M4	4,226,220	3,575,702	1,003,248	115,356
			M5	3,814,764	4,931,386	1,054,540	181,844
			M6	3,763,450	3,252,391	644,635	139,630
		amTreg	M1	734,338	1,055,355	346,526	15,037
			M2	690,255	1,241,344	526,005	8,224
			M3	556,780	1,393,314	398,156	22,360
			M4	741,106	1,888,828	494,156	19,905
			M5	465,508	1,906,114	497,337	21,903
			M6	507,546	1,762,516	455,426	19,596

Table S2. Sample properties. MM: non-treated pools of mice, M: IL-2 treated NOD mice; Number of FACS-ARIA sorted cells, sequencing read counts, number of productive TR sequences, number of unique amino acid clonotypes.

		Proportion in CD4+ T cells (%)		Proportion in Tregs (%)	
Mouse background	Sample	Teffs	Tregs	nTregs	amTregs
NOD AAV-IL2	M1	74.2	24.1	8.1	74.7
	M2	76.8	22.0	11.3	70.4
	M3	74.8	23.7	11.2	73.8
	M4	77.8	21.0	13.7	71.5
	M5	81.0	17.8	13.2	69.6
	M6	79.8	18.9	12.7	70.9

Table S3. Cell proportions in AAV-IL-2–treated NOD mice

			Expanded clontoypes	
Mouse background	Cell population	Sample	Number	Percentage
C57BL/6	nTreg	MM1	8,480	3.63
		MM2	4,863	2.23
		MM3	2,878	2.12
	CD4 Teff	MM1	7,015	3.14
		MM2	3,324	2.76
		MM3	5,921	3.42
	amTreg	MM1	3,883	4.01
		MM2	3,352	3.29
		MM3	2,490	3.61
	nTreg	MM4	3,961	1.99
		MM5	4,230	2.16
		MM6	1,799	3.30
	CD4 Teff	MM4	6,393	2.13
		MM5	3,910	2.03
		MM6	1,336	1.90
	amTreg	MM4	4,269	3.99
		MM5	4,001	3.18
		MM6	1,625	3.98
NOD	nTreg	MM7	5,980	3.45
		MM8	3,148	4.43
		MM9	3,460	4.02
	CD4 Teff	MM7	8,085	3.90
		MM8	6,061	3.36
		MM9	4,921	4.36
	amTreg	MM7	3,821	3.90
		MM8	2,950	3.81
		MM9	4,797	3.56
	nTreg	MM10	8,023	3.51
		MM11	6,924	4.03
		MM12	6,410	3.95
	CD4 Teff	MM10	8,617	2.94
		MM11	5,445	3.95
		MM12	6,082	3.53
	amTreg	MM10	6,043	4.31
		MM11	3,007	3.64
		MM12	4,440	3.82
NOD AAV-IL2 Female	nTreg	M1	761	3.17
		M2	1,525	4.44
		M3	1,407	4.62
		M4	1,951	4.29
		M5	2,158	4.79
		M6	2,335	4.51
	CD4 Teff	M1	3,795	2.30
		M2	5,565	3.36
		M3	6,667	3.66
		M4	4,127	3.58
		M5	7,735	4.25
		M6	4,100	2.94
	amTreg	M1	765	5.09
		M2	365	4.44
		M3	1,043	4.66
		M4	925	4.65
		M5	1,121	5.12
		M6	937	4.78

Table S4. Expanded clonotypes. MM/M: mouse identifier; Number of expanded clonotypes, Percentage of expanded clonotypes: # of expanded clones divided by the total number of clonotypes

Population	Pathology	TRBV	CDR3	TRBJ	Sharing
amTreg	Type-1 diabetes	13-1	CASSDDNYEQYF	2-7	2/6
		5	CASSDRYEQYF	2-7	1/6
		16	CASSLGQQDTQYF	2-5	1/6
		5	CASSQRGANTEVFF	1-1	1/6
		16, 5	CASSRDSSYEQYF	2-7	3/6
		17	CASSRDWGYEQYF	2-7	1/6
		13-1	CASSDSQNTLYF	2-4	1/6
	Systemic lupus erythematosus	16	CASSPRLGGNQDTQYF	2-5	1/6
		1	CTCSGGLGGREQYF	2-7	1/6
nTreg	Type-1 diabetes	13-2	CASGGYEQYF	2-7	3/6
		13-3	CASSDDNYEQYF	2-7	1/6
		13-3	CASSDRYEQYF	2-7	1/6
		13-1	CASSDSQNTLYF	2-4	2/6
		15	CASSLGGNTEVFF	1-1	1/6
		16	CASSLGGYAEQFF	2-1	1/6
		14	CASSLGQQDTQYF	2-5	1/6
		12-1, 16	CASSLGTGGYAEQFF	2-1	2/6
		12-2	CASSLGWGDEQYF	2-7	1/6
		5	CASSQRGANTEVFF	1-1	1/6
		19, 29, 5	CASSRDSSYEQYF	2-7	4/6
		5	CASSRDWGYEQYF	2-7	5/6
	Systemic lupus erythematosus	19	CASSILGNYAEQFF	2-1	1/6
		5	CASSPRLGGNQDTQYF	2-5	1/6
		1	CTCSGGLGGREQYF	2-7	1/6
	Cancer	13-1	CASSGTGQDTQYF	2-5	1/6
	Experimental autoimmune encephalomyelitis	13-2	CASGDAGTGYEQYF	2-7	1/6

Table S5. List of sequences identified in both the annotated database and nTreg and amTreg samples from AAV-IL-2–treated NOD mice.

Annexe 2.

Human thymopoiesis produces polyspecific CD8+ a/b T cells responding to multiple viral antigens

---

Valentin Quiniou, Pierre Barennes, Vanessa Mhanna, Paul Stys, Hélène Vantomme, Zhicheng Zhou3, Federica Martina, Nicolas Coatnoan, Michèle Barbié-Sastre, Hang-Phuong Pham, Béatrice Clemenceau, Henri Vié, Mikhail Shugay, Adrien Six, Barbara Brandao, Roberto Mallone, Encarnita Mariotti-Ferrandiz, David Klatzmann

---

Human thymopoiesis produces polyspecific CD8$^+$ α/β T cells responding to multiple viral antigens

Authors:
Valentin Quiniou[1,2,8], Pierre Barennes[1,2], Vanessa Mhanna[1†], Paul Stys[1†], Hélène Vantomme[1,2†], Zhicheng Zhou[3†], Federica Martina[2], Nicolas Coatnoan[2], Michèle Barbié-Sastre[2], Hang-Phuong Pham[4,8], Béatrice Clemenceau[5], Henri Vié[5], Mikhail Shugay[6], Adrien Six[1], Barbara Brandao[3], Roberto Mallone[3,7], Encarnita Mariotti-Ferrandiz[1], David Klatzmann[1,2]*.

Affiliations:
[1]Sorbonne Université, INSERM, Immunology-Immunopathology-Immunotherapy (i3), 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]Université de Paris, Institut Cochin, CNRS, INSERM, Paris, France.
[4]ILTOO pharma, Statistical department, Paris, France.
[5]CRCINA, INSERM, CNRS, Université d'Angers, Université de Nantes, Nantes, France.
[6]Center of Life Sciences, Skoltech, Moscow, Russia.
[7]Assistance Publique Hôpitaux de Paris, Service de Diabétologie et Immunologie Clinique, Cochin Hospital, Paris, France.
[8]Present Institution: Parean biotechnologies, Saint-Malo, France.

† These authors, listed in alphabetical order, have contributed equally

*Corresponding author: David Klatzmann, Hôpital Pitié-Salpêtrière, 83 bd de l'Hôpital, 75651 Paris, France.
Phone: +33 1 42 17 74 61
Email: david.klatzmann@sorbonne-universite.fr

Abstract:

T cell receptors (TCRs) are formed by stochastic gene rearrangements, theoretically generating >10^{19} sequences. They are selected during thymopoiesis, which releases a repertoire of about 10^8 unique TCRs per individual. How evolution shaped a process that produces TCRs that can effectively handle a countless and evolving set of infectious agents is a central question of immunology. The paradigm is that a diverse enough repertoire of TCRs should always provide a proper, though rare, specificity for any given need. Expansion of such rare T cells would provide enough fighters for an effective immune response and enough antigen-experienced cells for memory. We show here that human thymopoiesis releases a large population of $CD8^+$ T cells harboring α/β paired TCRs that (i) have high generation probabilities and (ii) a preferential usage of some V and J genes, (iii) are shared between individuals and (iv) can each recognize and be activated by multiple unrelated viral peptides, notably from EBV, CMV and influenza. These polyspecific T cells may represent a first line of defense that is mobilized in response to infections before a more specific response subsequently ensures viral elimination. Our results support an evolutionary selection of polyspecific α/β TCRs for broad antiviral responses and heterologous immunity.

Main Text:

INTRODUCTION

Specificity is considered a hallmark of the adaptive immune response. For T cells, specificity is mediated by T cell receptors (TCRs), which interact with peptides presented by major histocompatibility complex (MHC) molecules through their complementary-determining-region-3 (CDR3). TCRs are formed by rearrangements between hundreds of gene segments at the alpha and beta loci, theoretically generating $>10^{60}$ possible sequences (*1–3*). Each rearrangement has its probability of generation, which has been modeled to vary by about 10 orders of magnitude for the beta chain alone (*4*). During their development within the thymus, T cells undergo a selection largely based on the strength of their activation by thymic antigen-presenting cells that eliminates >80% of them (*5–7*). For each individual, this process releases a pool of T cells expressing a repertoire of approximately 10^8 unique TCRs (*8*).

How evolution shaped a process that selects cells that would efficiently respond to antigens that are not yet present, i.e. from an infectious agent to come, is a central question of immunology. The paradigm is that the production and selection of a diverse enough repertoire of TCRs should always provide a proper, though rare, specificity for any given need. Expansion of such rare T cells would then provide enough fighters for an effective immune response and later enough antigen-experienced cells for immune memory (*9*). Thus, it is widely accepted that "*T cell-mediated immunity to infection is due to the proliferation and differentiation of rare clones in the preimmune repertoire that by chance express TCRs specific for peptide-MHC (pMHC) ligands derived from the microorganism*"(*10*).

However, this theory is challenged by overlooked data questioning the essence of specificity. First, a child with an X-linked severe combined immunodeficiency who had a reverse mutation in a single T cell progenitor was protected from infections despite having a very restricted repertoire of only about 1,000 different TCRs (*11*). How such a repertoire (10^{-5} of a normal repertoire size) could handle the numerous infections occurring during childhood awaits explanation. Second, robust experimental and epidemiological data have highlighted the importance of heterologous immunity, i.e. immunity to one pathogen afforded by the exposure to unrelated pathogens (*12–14*). Likewise, for example, (i) memory T cells that are specific for one virus can become activated during infection with an unrelated virus (*15*), (ii) there are numerous virus-specific memory phenotype T cells in unexposed individuals (*16*), (iii) CMV infection enhances the immune response to influenza (*17*) and (iv) tumor-resident memory HBV-specific T cell responses correlate with hepatocellular carcinoma outcome(18).

We hypothesized that high-resolution sequencing of the TCR repertoire of developing thymocytes and peripheral T cells would yield insights into repertoire selection and function towards pathogens. We focused on the development of $CD8^+$ T cells as they represent the main effectors against viral pathogens, those that have had most chances to influence the evolution of our immune system. We obtained samples from organ donors allowing access to thymocytes and peripheral T cells, as well as datasets from patients undergoing vaccinations or with infections; we also used TCR data available from multiple public repositories.

We identified a population of $CD8^+$ T cells harboring diversified and polyspecific α/β TCRs that each can bind to multiple unrelated viral peptides, notably those from commonly encountered viruses such as EBV, CMV or influenza. We hypothesize that these cells represent a

first line of defense that is promptly mobilized in response to infections, possibly containing them before a more specific response subsequently ensures their control. Our results support an evolutionary selection of polyspecific α/βTCRs for broad antiviral responses and heterologous immunity.

RESULTS

Thymopoiesis selects a large and diverse set of clustered CDR3s with high generation probabilities.

To investigate how TCR diversity influences immune response to infectious agents, we started by analyzing the TCR repertoire dynamics of developing thymocytes. We first focused on the hypervariable CDR3 region of the TCR that interacts with the antigenic peptide, while CDR1 and CDR2 usually interact with HLA molecules (*19*). CDR3 analyses can therefore be used to investigate the sharing of TCR specificities across individuals with distinct HLA molecules. We analyzed the repertoire of purified $CD4^+CD8^+CD3^-$ ($DPCD3^-$), $CD4^+CD8^+CD3^+$ ($DPCD3^+$) and $CD4^-CD8^+CD3^+$ ($CD8^+$) thymocytes (Fig. S1A). $DPCD3^-$ thymocytes represent the earliest stage of TCR β-chain gene recombination, and their repertoire embodies the unaltered outcome of the TCR generation process; $DPCD3^+$ thymocytes are at an early stage of the selection process and their repertoire should be minimally modified; $CD8^+$ thymocytes have passed the selection process and bear a fully selected repertoire. We analyzed and represented the structure of these repertoires by connecting CDR3s (nodes) differing by at most one single amino acid (AA) (Levenshtein distance less than or equal to one: LD≤1). LD≤1 connected CDR3s have been described to most often bind the same peptide (*20*–*26*) (Fig. S1B). In such networks, connected CDR3s are designated as *clustered nodes* and the others as *dispersed nodes*. For normalization, we represented the first 18,000 most expressed β or α CDR3s from each sample.

We observed a significant increase in the number of clustered βCDR3s from $DPCD3^-$ to $CD8^+$ thymocytes (Fig. 1A and B) ($p<0.0001$), which was remarkably consistent among all individuals studied, independently of their age, sex or HLA (Fig. S2). Node's degree, i.e., its number

of CDR3 neighbors based on LD≤1, was also significantly increased during T cell differentiation for clustered βCDR3s (Fig. 1C and D) ($p<0.0001$). Similar results were found for the αCDR3s (Fig. S3A-B). These observations suggest a positive selection during thymopoiesis of TCRs with close CDR3 sequences, and therefore with shared recognition properties.

The probability of generation (*Pgen*) of a given TCR varies enormously from one TCR to the other. The clustered TCRs from both $DPCD3^+$ and $CD8^+$ thymocytes have a significantly higher *Pgen* ($p<0.0001$) than the dispersed ones (Fig. 1E). *Pgen* also increased significantly ($p<0.0001$) from $DPCD3^+$ to $CD8^+$ thymocytes (Fig. 1E). Moreover, there is a robust and remarkable correlation between βCDR3 generation probability and their number of connections ($p<0.0001$) (Fig. 1F and Fig. S3C).

Clustered βCDR3s belong to TCRs that have a preferential usage of V and J genes, with a significant ($p<0.01$) overrepresentation of Vβ 12-3, 27, 5-1, and 7-9 and of JB 1-1 and 2-7 (Fig. S3D-E). This results in a clear separation of clustered versus dispersed TCRs in a principal component analysis of VJ usages (Fig. 1G).

Clustered CDR3s are enriched for publicness

Clustered βCDR3s are also enriched in public sequences, i.e. being shared between at least 2 individuals ($p<0.0001$) (Fig. 2A). The significant increase of public βCDR3s in $CD8^+$ versus $DPCD3^+$ thymocytes ($p<0.0001$) is mostly that of the clustered βCDR3s. For $CD8^+$ thymocytes, up to 31.7% of clustered βCDR3s are public compared to barely 1% of the dispersed ones (Fig. 2A, Fig. S4A). Public βCDR3 have a significantly higher generation probability than the private ones ($p<0.0001$) (Fig. 2B), which is robust across all the patients (Fig. S4B, Table S1). Noteworthily, the number of

public βCDR3 shared between individuals is independent of the number of shared HLA Class I alleles (Fig. 2C). Finally, clustered βCDR3s of one individual can be connected to those of other individuals (up to twelve), and more frequently in $CD8^+$ versus $DPCD3^+$ thymocytes (Fig. 2D and E), indicating a convergence of specificities between individuals' clustered repertoires. Altogether, these results indicate that the mechanisms for TCR generation and for their further thymic selection are biased to shape a public repertoire of connected βCDR3s with shared recognition properties.

Clustered public CDR3s are enriched in viral specificities

The preferential selection of clustered public CDR3s that could represent over 8% of the sampled repertoire (Fig. 2A) raises the question of their specificities. As the main function of $CD8^+$ T cells is cytotoxicity towards virally infected cells, we investigated whether the clustered CDR3s could be associated with virus recognition. We curated two databases of βCDR3s from TCRs specific for human infectious pathogens (*27*, *28*) to retain only a set of 5,437 βCDR3 that had been identified by the binding of soluble multimeric MHC/peptide complexes i.e. tetramers or dextramers (*29*). We detected an enrichment of these virus-specific βCDR3s in clustered versus dispersed $CD8^+$ thymocytes ($p<0.0001$) and in $CD8^+$ versus $DPCD3^+$ thymocytes ($p<0.0001$) (Fig. 3A). Moreover, these virus-specific βCDR3s were significantly enriched in βCDR3s with the highest *Pgen* and node degree (Fig. 3B and C, Fig. S5A, Table S2). They were also highly shared between individuals (Fig. 3D) and they contained a high representation of CMV, EBV and influenza specificities (Fig. S5B). To better represent the contribution of virus-specific CDR3s to the global repertoire of $CD8^+$ cells, we connected them to all $CD8^+$ cell CDR3s of the 12 individuals using

either a perfect matching (LD=0) (Fig. 3E left panel) or an LD≤1 (Fig. 3E right panel). In the latter representation, the 13,557 connected CDR3s (i) covered most of the clustered CDR3s of all individuals (Fig. 3E, Fig. S6A), (ii) amounted to up to 7.4% and 12.3% of the DPCD3$^+$ and CD8$^+$ cells CDR3 repertoires, respectively (Fig. 3F), and (iii) were highly shared between individuals (Fig. S6B).

Altogether, these results indicate that the selection of clustered CDR3s with high generation probabilities and high inter-individual sharing corresponds, at least in part, to the selection of virus-specific TCRs whose CDR3s are remarkably conserved between individuals independently of their HLA restriction and form what could be referred to as a "paratopic network".

Identification of polyspecific TCRs

After thymic selection, an *in-silico* estimation indicated up to 12% of the TCR repertoire as specific for the small set of 30 tested peptides derived from 7 viruses that was used for the analysis (Fig. 3F). This appears hardly compatible with the fact that the TCR repertoire should contain TCRs that react specifically to countless antigens. Actually, we observed that a single cluster of βCDR3s from CD8 thymocytes could comprise unique βCDR3s from TCRs assigned to different viral specificities in independent experiments, for example to CMV, influenza, EBV and YFV (Yellow Fever Virus). This observation is puzzling, as clustered CDR3s are likely to respond to similar antigens (Fig. S1B and (*20*–*26*)). It suggested that some virus-associated TCRs could have a "fuzzy" specificity that could allow the recognition of peptides from different viruses. We explored the public databases of virus-specific TCRs (*27*, *28*) in more detail and discovered that numerous single CDR3s had been assigned to multiple specificities, for example, single CDR3s binding both CMV

and influenza tetramers (Fig. S7A). In addition, among the 13,557 virus-associated βCDR3s that we previously estimated in $CD8^+$ thymocytes (Fig. 3E, "LD≤1"), almost 7% (958/13557) are *in silico* assigned (LD≤1) as specific for at least two distinct viruses (Fig. S7B).

Binding properties of polyspecific TCRs.

We further investigated the potential T cell specificities of CD8+ thymocytes by using the GLIPH2 algorithm to analyze a combined dataset of 216,000 CDR3β sequences from our $CD8^+$ thymocytes and from 32,496 tetramer-specific sequences from the curated VDJdb database. This algorithm groups together CDR3s that have structural characteristics that make them likely to recognize the same antigens, hence called "specificity groups" (*24*). We identified 93,182 such specificity groups. We then analyzed the presence among these groups of TCRs from the VDJdb database with a known viral specificity. Noteworthy, 31.6% of the specificity groups contained at least one CDR3 with a known specificity, an already surprising observation given that the VDJdb database relates to a quite limited sample of the universe of specificities. Even more surprisingly, 6.8% of the clusters identified by GLIPH2 contained CDR3s with more than one specificity, and up to 9 different ones (Fig. 4A).

We further analyzed the CDR3β repertoires identified with GLIPH2 as specific for at least one of 16 EBV peptides (Fig. 4B). The analysis of these CDR3s based on the Morisita-Horn (MH) dissimilarity index identifies peptides recognized by overlapping repertoires. Noteworthy, the presence of shared CDR3β EBV specific repertoire is observed in both a match/mismatch HLA context. The MDS representation of MH dissimilarity index for EBV repertoire (Fig. S7C) show the convergence of these shared CDR3β EBV specific repertoire across patients.

As these observations are made only from studying unpaired CDR3s, we aimed to confirm them with paired α and β virus-specific CDR3s obtained from single-cell TCR sequencing. These sequences were obtained from 160,914 blood $CD8^+$ T cells isolated from four healthy donors and incubated simultaneously with dextramers complexed with peptides from CMV, EBV, HIV, HPV, HTLV and influenza. We could identify numerous single cells harboring TCRs that could bind to multiple dextramers. As an example, single-cell TCRs that bind dextramers matching the HLA of a given individual loaded with EBV or CMV peptides are shown (Fig. 4C). Projecting the binding of EBV dextramers onto a t-SNE representation based on single-cell specificity identified a few distinct regions that contained positive cells (Fig. 4C, left t-SNE). Repeating the same analysis with CMV dextramers identified binding in these same regions, and often on the same cells (Fig. 4C, right t-SNE). Thus, single TCRs appear to bind to both EBV and CMV HLA-matched dextramers.

Actually, unique TCRs could recognize multiple unrelated peptides from the same virus as well as peptides from different viruses. We represented on chord diagram the binding of TCRs to HLA-matched EBV dextramers loaded with different peptides (Fig. 4D left). For example, one half of the TCRs that bind to the RAKFKQLL peptide also bind the FLRGRAYGL peptide. Furthermore, some TCRs recognizing this RAKFKQLL peptide can also bind peptides from CMV (red), EBV (green) or HIV (dark blue) (Fig. 4D right).

Finally, we analyzed in more detail the binding scores of TCRs from 2 patients that all use the same CASSIRSSYEQYF βCDR3 and that were associated with different αCDR3s. This represented 664 cells harboring 55 unique TCRs and 1410 cells harboring 91 unique TCRs, respectively (altogether 2074 cells expressing 131 unique TCRs). We represented the binding score of these CDR3s for 49 different dextramers. Noteworthily, these 2074 TCRs all had the highest

binding score for a matched HLA-A*0201 dextramer bound to an influenza peptide (Fig. S8). However, they also had a high binding score for a series of peptides from CMV and EBV bound to HLA-A*0301 and A*1101 dextramers. Most importantly, there was no significant binding to most of the dextramers, indicating that the binding of these few dextramers was specific. These results suggest that polyspecificity is biased towards recognition of common viruses.

Polyspecific T cells are activated in vitro by multiple viral peptides

These peculiar properties of polyspecific T cell (psT) TCRs led us to assess their functional relevance. We first evaluated the in vitro cross-activation of psT cells with different peptides. Human effector memory $CD8^+$ T cells were purified according to their binding of CMV or EBV HLA-matched dextramers (Fig S9); the sorted cells were then stimulated by either the peptide that was used to purify them, or by different ones, and their activation was measured by their IFN-γ production (Fig. 5A). All T cells were efficiently non-specifically activated by PMA/ionomycin. T cells that did not bind any dextramers could not be stimulated by any peptide. In contrast, dextramer-sorted cells could be activated by their cognate peptide and almost as well by the other (Fig. 5B).

We also tested the response of bulk T cells developed as cell therapy products for the treatment of CMV infections. These cells had been selected through 18 days of stimulation in the presence of a peptide pool covering the complete sequence of the pp65 protein of CMV. Upon stimulation by their cognate peptides, over 85% of the CD8+ T cells produced IFNγ (Fig. S10A). Surprisingly, a small fraction of the same cells produced interferon upon stimulation with unrelated CMV-IE1 peptides, and a large fraction upon stimulation with a peptide pool which

covers the sequence of the BZLF-1 protein of EBV (Fig. S10A). There was no interferon production upon stimulation of the same cells by a peptide pool which covers the EBNA1 protein of EBV (Fig. S10A). The stimulation with EBV peptides requires a 100-fold higher but still physiological concentration of peptides (Fig. S10B).

Finally, we directly confirmed the polyspecificity of unique TCRs upon their cloning and re-expression. We used a 5KC murine T-hybridoma cell line devoid of endogenous TCR, and transduced with a ZsGreen fluorescent reporter under the control of nuclear factor of activated T cells (NFAT) and with human CD8 (*30*). These 5KC transductants were then stimulated with K562 cells transduced with either HLA-A*02:01 or HLA-A*03:01 as antigen presenting cells, in the presence of individual peptides at increasing concentrations. From the single cell sequencing dataset, we selected TCRs that could bind dextramers loaded with Flu, CMV and EBV peptides. TCR#35-13 responded strongly to HLA-A*02:01-transduced K562 cells loaded with Flu MP 58-66 (EC50: 2.02x10-9 M) and CMV IE1-184-192 (EC50: 7.99x10-7 M), and poorly to EBV BMLF1 280-288 (EC50: 9.37x10-3 M) (Fig. 5C upper left). TCR#36-150 responded strongly to HLA-A*02:01-transduced K562 cells loaded with EBV BMLF1 280-288 (EC50: 9.99x10-11 M) and Flu MP 58-66 (EC50: 7.57x10-7 M) and poorly to CMV IE1 184-192 (EC50: 2.58x10-3 M) (Fig. 5C bottom left). None of the TCRs responded to peptide-pulsed K562 expressing HLA-A*03:01. In contrast, a control D222D TCR (*31*) known as specific for a ZnT8 186-194 peptide responded to this peptide (EC50: 2.68x10-7 M), but not to the Flu, EBV or CMV peptides (Fig. 5C upper right). We then further analyzed TCRs derived from single-sorted CD8+ T cells stained with tetramers loaded with pancreatic self-peptides. TCRs with known specificities were also cloned and expressed in the reporter cell line. We analyzed their response to a set of peptides comprising their cognate peptide

and peptides with no significant structural commonalities, selected by testing combinatorial peptide libraries(*30*). We also observed a marked polyspecificity of these TCRs. As an example, a TCR identified as responding to the VMNILLQYV peptide from the Glutamic Acid Decarboxylase protein responded even better to the structurally unrelated SLYNPAIVYA peptide from human chondroitin sulfate N-acetylgalactosaminyltransferase 2 (Fig 5D).

DISCUSSION

Our work brings to attention an overlooked phenomenon hidden in the literature. In a simple curation of public datasets of antiviral TCRs that were assigned to respond to known viruses - retaining only those from cells that could bind or respond to a known peptide - we found a surprisingly frequent representation of CDR3s that had been assigned to different viral specificities from independent experiments. Within the 40,939 curated βCDR3, 24% were from TCRs that could bind at least two totally unrelated viral peptides. This observation is highly relevant, as it arises in large part from datasets of full TCRs from single-cell sequencing. Noteworthily, in these latter experiments, cells are co-incubated with multiple dextramers bearing a large set of unrelated peptides; although this dataset is widely available, researchers have missed (or dismissed) that a single cell could bind multiple dextramers (sometimes a dozen). In the methodology description, the technology provider recommends determination of the specificity of a given T cell by the identity of the dextramer that has the highest binding counts on that cell. This is a rather crude readout, as the highest count is often marginally so. Furthermore, it should be realized that, as there are limited numbers of HLA molecules available per cell, there is competition between dextramers for binding at the cell surface. That a single cell binds multiple dextramers bearing unrelated peptides is in our view highly meaningful. As the binding of multimeric HLA/peptide complexes is a paradigmatic method to define specificity, multiple binding indicates multiple specificities. We thus defined TCRs binding multiple HLA/peptide complexes as polyspecific, and by extension called the cells harboring these TCRs as polyspecific T cells (psT cells).

TCRs are notoriously "cross-reactive," and it is estimated that each TCR could potentially recognize >10^6 different peptide/MHC combinations (*32*–*35*). Indeed, the HLA/peptide/TCR *ménage à trois* leaves room for mimotopic cross-reactivity, i.e. the fact that two unrelated peptides can bind differently to a given HLA molecule, but altogether generate a globally similar structure that would be recognized by the same TCR. Also, it was recently shown that peptides need to share only five residues at specific positions to bind the same TCR (*35*, *36*). This latter observation has functional relevance as bacterial-derived peptides sharing such 5 residues with a tissue-restricted MOG self-antigen were shown (i) to activate a known MOG-specific TCR as well as a MOG peptide and (ii) to trigger an autoimmune attack against brain tissues upon mouse immunization (*36*, *37*). Thus, if many peptides could interact with a given TCR, it is because they share a structural conservation of their TCR interaction surface (*33*).The functional relevance of these observations were described as that "*TCR cross-reactivity enables effective surveillance of diverse self and foreign antigens without necessitating degenerate recognition of non-homologous peptides*"(*33*).

In contrast to this mimotopic cross-reactivity, we did not observe any specific amino acid sharing structure for the multiple peptides recognized by polyspecific TCRs. Thus, polyspecificity appears to be what was described as "*degenerate recognition of non-homologous peptides"* (*33*). We prefer to describe polyspecificity as defining a fuzziness (rather than "degeneration") in TCR recognition that makes it capable of interacting with multiple unrelated peptides. In this regard, it is noteworthy that B cells have a machinery for somatic mutations of their BCRs that ultimately allows them to generate antibodies with increased affinity (specificity) for antigens. While TCR generation and BCR generation share many common mechanisms, the fact that T cells did not

evolve to use such somatic mutations suggest that T cell recognition has been selected to be more fuzzy than stringent.

As TCRs have been essentially selected during thymocyte differentiation for their ability to bind HLA molecules presented by thymic antigen-presenting cells, it could have been that polyspecificity represents some loose remnant of that property. However, it should be emphasized that a single cell does not bind randomly to any homologous-HLA/peptide complexes (Supplementary Fig. 7). Indeed, each polyspecific TCR has its own binding pattern regarding a large set of dextramers, with most that do not bind. Moreover, when two polyspecific TCRs bind to a given dextramer, they have a similar binding score for others that they also bind to. Thus, this binding is not just "*non-specific*" and must obey some structural rules that remain to be defined.

Polyspecific TCR binding properties are distinct from those of innate-like MAIT and NKT cells (*38*, *39*). These have a restricted diversity, with an invariant TCRα chain and a constrained TCRβ repertoire, and are MR1- or CD1-restricted, respectively. In contrast, the repertoire of polyspecific TCRs is highly diverse for both the TCR α and β chains, although with a specific usage of VDJ/VJ recombination. Also, polyspecific T cells found in adults cannot be the remnant of fetal/neonatal T cells that have been described as having a promiscuous repertoire, as these cells have repertoires enriched in germline-encoded TCRs(*40*, *41*).

The polyspecific nature of these TCRs has a functional relevance, as we show here that cloned polyspecific TCRs indeed trigger cell activation in response to multiple peptides, although with different efficiencies. In agreement with our results, the analysis of the TCR repertoire of lung tumor infiltrating T cells using GLIPH2, as we did here, recently identified a key TCR that, after cloning and expression, could bind and respond to a tumor-specific self-antigen as well as to a

peptide from E. coli and one from EBV (*42*). In the same line, a recent study of multiple sclerosis pathogenesis identified autoreactive CD4+ T cell clones that can respond to self-peptides as well as peptides from EBV and from the bacteria *Akkermansis muciniphila*(*43*). This further establishes the existence of psT cells that respond to multiple antigens from different microbes. Also, in the development of the T-scan platform aimed at characterizing the specificity of TCRs(*44*), the authors performed a set-up experiment with CD8 cells activated by the immunodominant NLV peptide from the pp65 protein of CMV. While T-scan could identify two hits related to NLV overlapping peptides, they also detected two hits related to the unrelated CMV I1E CMV protein. Actually, tetramer staining revealed that while 25% of the T cells in the NLV-expanded population bind NLV tetramers, 2% of the T cells bind IE1 tetramers(*44*). Moreover, the fold enrichment of the IE1 specific cells was close to that of NLV specific cells (60 vs 80, respectively). Thus, activation with a single peptide results in a major expansion of cells with TCRs that can also respond to an unrelated peptide.

The fact that CMV, EBV and influenza specificities are highly represented in polyspecific binding patterns likely reflects the biased representation of these specificities in the databases (15.9%, 16.2%, 50.6%, respectively), but also highlights that the detection of the diverse viral specificities of psT cells is largely underestimated due to the very low number of CDR3s with known viral specificities for viruses other than these. Our results led us to estimate that polyspecific TCRs could represent at least 1/5 of the CD8 selected TCR repertoire. Altogether, our results define peculiar binding properties of polyspecific TCRs. This warrants to study their physicochemical characteristics, with a special interest in solving the structure of such TCRs bound to two unrelated peptides.

Ideas and speculations

Our findings have important implications for the study of the adaptive immune response in health, diseases and immunotherapies. In the field of infectious diseases and vaccination, they prompt reconsideration of the paradigm of highly diverse adaptive immune repertoires providing a highly antigen-specific antiviral immune response. Indeed, there is an intrinsic weakness in the view that a highly diverse TCR repertoire is ideal to protect us from a countless repertoire of constantly evolving viruses. As best illustrated by the deadly outcome of flu in the elderly or young infants, there is a very limited amount of time for the immune system to react against rapidly replicating deadly viruses. For life-threatening situations, the initial recruitment of frequent polyspecific effector T cells might be more efficient and rapid than having to rely on rare cells with stringent specificity that would need a period of expansion to provide enough fighters. In the race between virus replication and the mounting of the immune response, this early response of polyspecific T cells could provide some control of the viral spread that will allow time for the development of an immunologically fittest response to ensure the final control of the infection. This would explain how a very restricted repertoire of only about 1,000 different TCRs arising from a single T cell progenitor was sufficient to cope with viral infections in a child (*11*). They would also explain why children vaccinated against measles in underdeveloped countries have a better life expectancy than unvaccinated ones (when excluding measles-related events), which could be linked to an overall better response to infections with other pathogens (*48*, *49*).

While our findings do not challenge that there are specific TCRs nor that there are specific immune responses, they highlight another mechanism of preparedness of the immune system, reminiscent of the role of (i) other unconventional T cells like MAIT and NKT cells(*45*), (ii) TCR

activation by bacterial superantigen (*46*) and (iii) natural antibodies specific for microbial determinants (*47*).

Altogether, polyspecific TCRs form a paratopic network, made of TCRs that have a high probability of generation and are positively selected during thymocyte differentiation. These properties indicate that these TCRs have been positively selected during evolution, further supporting their overall beneficial effects. A fuzzy recognition by polyspecific TCRs would explain the so-called "heterologous immunity" (*50*, *51*) in which T cell responses to one pathogen can have a major impact on the course and outcome of a subsequent infection with an unrelated pathogen (*17*). We speculate that individual histories of fuzzy immune responses may thus create patient-specific "antigenic sins" that might be responsible for the diverse quality of immune responses to viruses, from inapparent infection to fulminant immunopathology (*52–54*). Further studies will have to evaluate the contribution of fuzzy immune responses to the efficacy but also the immunopathology of antimicrobial responses and to autoimmunity. We speculate that the origin of many immune diseases should be sought in the repeated activation of polyspecific T cells.

MATERIALS AND METHODS

Patients and samples

Thirteen human thymus samples were obtained from organ donors undergoing surgery (Department of Cardiac Surgery, Pitié-Salpêtrière Hospital, France) after approval by the *Agence de Biomédecine* and the *Ministry of Research*. Their age at the time of sampling ranged from 19 to 65 years old. The male-to-female sex ratio was 2.6.

For cross-activation experiments, six leukapheresis samples were freshly collected from healthy donors at EFS Paris Saint-Antoine-Crozatier (Etablissement Français du Sang, Paris, France) after informed consent and according to institutional guidelines. Donor selection was based on matching HLA-A2 class I allele.

Isolation of thymocytes and extraction of RNA

Single-cell suspensions were prepared from the thymus by mechanical disruption through nylon mesh (cell strainer). Single-cell suspensions from whole thymus were stained with antibodies anti-CD3 (AF700), anti-CD4 (APC), anti-CD8 (FITC). Cells were sorted by fluorescent activated cell sorting (Becton Dickinson™ FACSAria II) with purity >95% to collect populations based on the following labeling: DPCD3$^-$ were gated as CD3$^-$CD4$^+$CD8$^+$, DPCD3$^+$ were gated as CD3$^+$CD4$^+$CD8$^+$ and CD8$^+$ were gated as CD3$^+$CD4$^-$CD8$^+$. RNA was isolated from sorted populations by means of lysis buffer with the RNAqueous-Kit (Invitrogen®) extraction kit, according to the manufacturer's protocol.

TCR repertoire library preparation and sequencing

TCR repertoire library preparation and sequencing was performed has previously described(*55*). Briefly, T cell receptor (TCR) alfa and beta libraries were prepared on 100 ng of RNA from each sample with the SMARTer Human TCR a/b Profiling Kit (Takarabio®) following the provider's protocol. Briefly, the reverse transcription was performed using TRBC reverse primers and further extended with a template-switching oligonucleotide (SMART-Seq® v4). cDNAs were then amplified following two semi-nested PCRs: a first PCR with TRBC and TRAC reverse primers as well as a forward primer hybridizing to the SMART-Seqv4 sequence added by template-switching and a second PCR targeting the PCR1 amplicons with reverse and forward primers including Illumina Indexes allowing for sample barcoding. PCR2 were then purified using AMPure beads (Beckman-Coulter®). The cDNA samples were quantified and their integrity was checked using DNA electrophoresis performed on an Agilent 2100 Bioanalyzer System in combination with the Agilent DNA 1000 kit, according to the manufacturer's protocol. Sequencing was performed with Hiseq 2500 (Illumina®) SR-300 protocols using the LIGAN-PM Genomics platform (Lille, France).

TCR deep sequencing data processing

FASTQ raw data files were processed for TRA and TRB sequences annotation using MiXCR (*56*) software (v2.1.10) with RNA-Seq parameters. MiXCR extracts TRAs and TRBs and provides corrections of PCR and sequencing errors.

Network generation and representation

To construct a network, we computed a distance matrix of pairwise Levenshtein distances between CDR3s using the "stringdist"(*57*) R package. When two sequences were similar under the defined threshold, LD>1 (i.e., at most one amino acid difference), they were connected and designated as "clustered" nodes. CDR3s with more than one amino acid difference from any other sequences are not connected and were designated as "dispersed" nodes.
Layout of networks for fig. 1C, 2D were obtained by using the graphopt algorithm of the "Igraph"(*58*) R package and plotted in 2D with "ggplot2" to generate figures (*59*). Only clustered nodes are represented, edges are not shown and colors represent the node degree (log scale). Layouts of detailed networks in fig. 1A were done with Cytoscape(*60*).

Statistical analysis and visualization

Normalization was performed by sampling on the top α or β 18,000 CDR3s based on their frequency in each sample. The repertoires with less than 18,000 α or β CDR3s were not included in the statistical analysis. The numbers of samples included in the statistical analysis for the β repertoire were: two for $DPCD3^-$, ten for $DPCD3^+$ and twelve for $CD8^+$. The numbers of samples included in the statistical analysis of the α repertoire were: six for $DPCD3^+$ and ten for $CD8^+$. Statistical tests used to analyze data are included in the figure legends. Comparisons of two groups were done using the Mann-Whitney test (fig.1B-D-E, fig.2A-C, fig.3A-B-F) and multiple t-test (fig. 2E). The correlation coefficient was calculated using the Pearson correlation coefficient (fig. 2B & 2C, fig. 2C). Enrichment of public CDR3s or virus-associated CDR3s was done using the two-tailed Chi-square test with Yates correction (Tables S1, S2 & S3). Statistical comparisons and multivariate analyses were performed using Prism (GraphPad Software, La Jolla, CA) and using R software version 3.5.0 (www.r-project.org). PCA was performed on the frequency of VJ combination usage frequency within each donor using the factoextra R package. T-SNE were generated using the binding scores of each cell across all the antigens present in the dataset. The function Rtsne of the homonymous R package (*61*) was applied with the perplexity parameter set to 10.

Probability of generation calculation

The generation probability (*Pgen*) of a sequence is inferred using the Olga_ algorithm, which is inferred by IGoR (*63*), for fig. 1E and fig.2B. IGoR uses out-of-frame sequence information to infer patient-dependent models of VDJ recombination, effectively bypassing selection. From these models, the probability of a given recombination scenario can be computed. The generation probability of a sequence is then obtained by summing over all the scenarios that are compatible with it. We also used OLGA to generate a random repertoire of 500,000 sequences for each α or β repertoire and 3 down sampling (of unique sequences) to get a control repertoire equal to the size of COVID-19 BAL dataset used in fig. 7B & C. The control repertoire was parametrized by the predefined genomic templates provided with the package.

CDR3 connections between individuals

In fig. 2D, the top 1,500 βCDR3s were sampled from each of the 12 datasets of $DPCD3^+$ and $CD8^+$, then merged to obtain two datasets of 18,000 βCDR3s for $DPCD3^+$ and $CD8^+$. We generated and represented networks, as described above, to investigate the βCDR3 inter-individual network structure.

Virus-specific CDR3 tetramer public databases

The virus-associated CDR3 databases used for the search for specificity was compiled from the most complete previously published McPAS-TCR (*27*) and VDJdb (*28*) databases. Virus-associated βCDR3s were selected from the original datasets only when derived from a TCR of sorted CD8 T cells that were bound by a specific tetramer. A total of 5,437 such unique tetramer-associated βCDR3s were identified and used. Peptides used for tetramer sorting were from cytomegalovirus (CMV), Epstein-Barr virus (EBV), hepatitis C virus (HCV), herpes simplex virus 2 (HSV2), human immunodeficiency virus (HIV), influenza and yellow fever virus (YFV).

Virus-specific CDR3 single-cell dextramer public dataset

This dataset contains single-cell alpha/beta TCRs from 160,914 $CD8^+$ T cells isolated from peripheral blood mononuclear cells (PBMCs) from 4 healthy donors. Briefly, 30 dCODE™ Dextramer® reagents (Immudex®) with antigenic peptides derived from infectious diseases (9 from CMV, 12 from EBV, 1 for influenza, 1 for HTLV, 2 for HPV and 5 for HIV) were simultaneously used to mark cells. Each Dextramer® reagent included a distinct nucleic acid barcode. A panel of fluorescently labeled antibodies was used to sort pure Dextramer®-positive cells within the $CD8^+$ T cell population using an MA900 Multi-Application Cell Sorter (Sony Biotechnology) in a reaction mix containing RT Reagent Mix and Poly dT RT primers. The Chromium Single Cell V(D)J workflow generates single cell V(D)J and Dextramer® libraries from amplified DNA derived from Dextramer®-conjugated barcode oligonucleotides, which are bound to TCRs. Chromium Single Cell V(D)J enriched libraries and cell surface protein libraries were quantified, normalized, and sequenced according to the user guide for Chromium Single Cell V(D)J reagent kits with feature barcoding technology for cell surface protein. We used this dataset to study the presence of multiple specificities in TCR and CDR3. There were 139,378 unambiguous TCRs (with only one α and one β chain). We set the threshold defining positive binding at UMI counts greater than 10 for any given dextramer. This identified 15,195 unique virus-specific TCRs with at least one binding.

Cross-activation experiment

PBMCs were separated on Ficoll gradient. CD8+ T cells were isolated from PBMCs by positive isolation using the DYNABEADS® CD8 Positive Isolation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. emCD8 T cells were purified after staining with CD3-AF700, CD8-KO, CD45RA-PeCy7 according to the manufacturer's instructions. The samples were also stained either with CMV pp65 NLVPMVATV or with EBV BMLF-1 GLCTLVAML PE-conjugated Dextramers (Immudex®). $emCD8^+Dex^+$ cells were sorted by FACS (FACS Aria II®; BD Biosciences) with a purity >95%. Sorted cells were cultured at a maximum of 5×10^5 cells/mL in round-bottom 96-well plates in RPMI 1640 medium supplemented with 10% FCS, 1% penicillin/streptomycin and glutamate at 37°C with 5% CO_2. *In vitro* stimulation was performed 24 hours after cell sorting. Sorted cells were stimulated for 6 hours with either nothing or 1 µg/mL of SIINFEKL ovalbumin peptide (OVA), NLVPMVATV cytomegalovirus pp65 peptide (CMV) or GLCTLVAML Epstein-Barr virus BMLF-1 peptide (Ozyme®). The positive control (Ctl PMA/Iono) was performed with 50 ng/mL phorbol myristate acetate (PMA) and 1 mM ionomycin. Intracellular IFN-γ production with an IFN-γ-FITC antibody (BD Pharmingen) was detected in the presence of Golgi-Plug (BD Pharmingen®) after fixation and permeabilization (BD Cytofix/Cytoperm). Data were acquired using a Navios flow cytometer and analyzed with Kaluza analysis software (Beckman Coulter).

In vitro CMVpp65 CTL selection

Peripheral blood mononuclear cells (PBMCs) were obtained from blood donors at the Etablissement Français du Sang (EFS) with informed consent (Blood products transfer agreement relating to biomedical research protocol 97/5-B—DAF 03/4868). CMV+ donor PBMCs were resuspended in the serum-free culture medium TexMacs (Miltenyi-Biotech) at 10^7 cells/mL and at a density of 5 X 10^6 cells/cm^2. Then, 20 µl PepTivator-CMVpp65 (Miltenyi-Biotech (130-093-435), 30 nmol of each peptide and approximately 140 peptides were added per mL of cell suspension. The final concentration of PepTivator-CMVpp65 was 0.6 nmol (approximately 1 µg of each peptide/mL). With these stimulation conditions, PBMCs were incubated for 4 hours and then transferred to culture flasks at 10^6 cells/mL in RPMI medium with 8% human serum (EFS Pays de la Loire, France) and 50 IU IL-2/mL. CTL were maintained in culture for 25 days.

Detection of CMV pp65-specific T cells by intracellular staining with anti-INF-γ-PE

CMVpp65 selected CTL were stimulated for 6 hours in TexMacs medium (without human serum and IL-2) with 20 µL PepTivator-CMVpp65 per mL of cell suspension. Brefeldin-A was added at 10 µg/ml. Negative control (CTL without stimulation) and positive control (CTL stimulated with PMA-ionomycin) were also included. Stimulation with irrelevant peptides was also performed with PepTivator-CMV IE-1 (Miltenyi (130-093-494) and PepTivator-EBV BZLF1 (Miltenyi (130-093-612). Cells were then stained with anti-CD8-FITC, fixed, permeabilized and intracellularly stained with anti-INF-γ-PE.

Peptide dose-response of TCR transductants

TCR transductants were generated as described [Mann, Zhou, Front Immunol 2020]. Briefly, 5KC T-hybridoma cells (kindly provided by M. Nakayama, Barbara Davis Center, Aurora, CO) were transduced with the NFAT- driven fluorescent reporter ZsGreen-1 along with human CD8 by spinoculation with retroviral supernatant produced from phoenix-eco cells (ATCC CRL-3214). These cells were subsequently transduced with retroviral vectors encoding a chimeric TCR alpha gene followed by a porcine teschovirus-1 2A (P2A) peptide and a chimeric TCR beta gene synthesized by TwistBioscience. K562 cells (ATCC CCL-243) were transduced with lentiviral vectors (TakaraBio) encoding HLA- A*02:01 or HLA-A*03:01 by spinoculation with ultracentrifuge concentrated viral particles produced from 293T-HEK-TN cells (System biosciences), followed by sorting of cells stained with anti-human HLA-A, B, C antibody conjugated with PE-Cy7 (clone W6/32, BioLegend).

TCR transductants (20,000 cells per well) were stimulated for 18 h with individual peptides at different concentrations (9 10-fold serial dilutions from 100 µM) in the presence of HLA- A*02:01+ or HLA-A*03:01+ K562 cells (50,000 cells per well), followed by analysis of ZsGreen-1 expression. Peptides (from Synpeptide) tested for the response by TCRs are GILGFVFTL (Flu MP 58–66), GLCTLVAML (EBV BMLF1 280-288), KLGGALQAK (CMV IE1 280-288), VMNILLQYV (GAD65 114-122), VAANIVLTV (ZnT8 186- 194) and other sequences selected by testing a GAD 114-122-reactive TCR by combinatorial peptide libraries, as described(*30*) EC50 values were calculated using the nonlinear regression log (agonist) vs. response (three parameters) equation model in GraphPad Prism 9.

Supplementary Materials

Fig. S1. Thymocytes cell sorting dot plot.
Fig. S2. ßCDR3 network during thymopoiesis.
Fig. S3. Properties of thymocytes' TCRs.
Fig. S4. Publicness of thymocytes' CDR3s.
Fig. S5. Virus specificities of thymocytes' CDR3s.
Fig. S6. Sharing of virus specific thymocytes' CDR3s.
Fig. S7. Binding properties of single cell TCRs.
Fig. S8. βCDR3s drive polyspecificity.
Fig. S9. Gating strategy for sorting effector memory (emCD8) Dextramer positive cells.
Fig. S10. Polyspecific properties of anti-CMV cytotoxic cells.
Table S1. Enrichment of public βCDR3s in CD8+ thymocytes vs DPCD3+.
Table S2. Enrichment of virus-specific βCDR3s from databases[14,15] in clustered CD8+ thymocytes.
Table S3. List of peptides represented on the chord plot. Related to 4I.

References and Notes

1. T. Dupic, Q. Marcou, A. M. Walczak, T. Mora, Genesis of the αβ T-cell receptor. *bioRxiv* (2018), doi:10.1101/353128.

2. Murphy, K., & Weaver, C., Janeway's immunobiology. Garland Science.(2016). *The Quarterly Review of Biology*. **87**, 266–267 (2012).

3. A. Murugan, T. Mora, A. M. Walczak, C. G. Callan, Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proceedings of the National Academy of Sciences*. **109**, 16161–16166 (2012).

4. P. Bradley, P. G. Thomas, Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. *Annual Review of Immunology*. **37**, 547–570 (2019).

5. L. Klein, B. Kyewski, P. M. Allen, K. A. Hogquist, Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol*. **14**, 377–391 (2014).

6. C. Sinclair, I. Bains, A. J. Yates, B. Seddon, Asymmetric thymocyte death underlies the CD4:CD8 T-cell ratio in the adaptive immune system. *Proceedings of the National Academy of Sciences*. **110**, E2905–E2914 (2013).

7. N. Vrisekoop, J. P. Monteiro, J. N. Mandl, R. N. Germain, Revisiting Thymic Positive Selection and the Mature T Cell Repertoire for Antigen. *Immunity*. **41**, 181–190 (2014).

8. Q. Qi, Y. Liu, Y. Cheng, J. Glanville, D. Zhang, J.-Y. Lee, R. A. Olshen, C. M. Weyand, S. D. Boyd, J. J. Goronzy, Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences*. **111**, 13139–13144 (2014).

9. M. M. Davis, T-cell antigen receptor genes and T-cell recognition, 8 (1988).

10. M. K. Jenkins, H. H. Chu, J. B. McLachlan, J. J. Moon, On the Composition of the Preimmune Repertoire of T Cells Specific for Peptide–Major Histocompatibility Complex Ligands. *Annu. Rev. Immunol.* **28**, 275–294 (2010).

11. P. Bousso, V. Wahn, I. Douagi, G. Horneff, C. Pannetier, F. Le Deist, F. Zepp, T. Niehues, P. Kourilsky, A. Fischer, G. de Saint Basile, Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor. *Proceedings of the National Academy of Sciences*. **97**, 274–278 (2000).

12. A. Roth, P. Gustafson, A. Nhaga, Q. Djana, A. Poulsen, M.-L. Garly, H. Jensen, M. Sodemann, A. Rodriques, P. Aaby, BCG vaccination scar associated with better childhood survival in Guinea-Bissau. *International Journal of Epidemiology*. **34**, 540–547 (2005).

13. L. B. Watkin, R. Mishra, A. Gil, N. Aslan, D. Ghersi, K. Luzuriaga, L. K. Selin, Unique influenza A cross-reactive memory CD8 T-cell receptor repertoire has a potential to protect against EBV seroconversion. *Journal of Allergy and Clinical Immunology*. **140**, 1206–1210 (2017).

14. R. M. Welsh, L. K. Selin, No one is naive: the significance of heterologous T-cell immunity. *Nature Reviews Immunology*. **2**, 417–426 (2002).

15. D. Verhoeven, J. R. Teijaro, D. L. Farber, Heterogeneous Memory T Cells in Antiviral Immunity and Immunopathology. *Viral Immunology*. **21**, 99–114 (2008).

16. L. F. Su, B. A. Kidd, A. Han, J. J. Kotzin, M. M. Davis, Virus-Specific CD4+ Memory-Phenotype T Cells Are Abundant in Unexposed Adults. *Immunity*. **38**, 373–383 (2013).

17. D. Furman, V. Jojic, S. Sharma, S. S. Shen-Orr, C. J. L. Angel, S. Onengut-Gumuscu, B. A. Kidd, H. T. Maecker, P. Concannon, C. L. Dekker, P. G. Thomas, M. M. Davis, Cytomegalovirus infection enhances the immune response to influenza. *Science Translational Medicine*. **7**, 281ra43-281ra43 (2015).

18. Y. Cheng, B. Gunasegaran, H. D. Singh, C.-A. Dutertre, C. Y. Loh, J. Q. Lim, J. C. Crawford, H. K. Lee, X. Zhang, B. Lee, E. Becht, W. J. Lim, J. Yeong, C. Y. Chan, A. Chung, B. K. P. Goh, P. K. H. Chow, J. K. Y. Chan, F. Ginhoux, D. Tai, J. Chen, S. G. Lim, W. Zhai, S. P. Choo, E. W. Newell, Non-terminally exhausted tumor-resident memory HBV-specific T cell responses correlate with relapse-free survival in hepatocellular carcinoma. *Immunity*. **54**, 1825-1840.e7 (2021).

19. J. Rossjohn, S. Gras, J. J. Miles, S. J. Turner, D. I. Godfrey, J. McCluskey, T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules. *Annual Review of Immunology*. **33**, 169–200 (2015).

20. P. Dash, A. J. Fiore-Gartland, T. Hertz, G. C. Wang, S. Sharma, A. Souquette, J. C. Crawford, E. B. Clemens, T. H. O. Nguyen, K. Kedzierska, N. L. La Gruta, P. Bradley, P. G. Thomas, Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. **547**, 89–93 (2017).

21. M. Klinger, F. Pepin, J. Wilkins, T. Asbury, T. Wittkop, J. Zheng, M. Moorhead, M. Faham, Multiplex Identification of Antigen-Specific T Cell Receptors Using a Combination of Immune Assays and Immune Receptor Sequencing. *PLOS ONE*. **10**, e0141561 (2015).

22. A. Madi, A. Poran, E. Shifrut, S. Reich-Zeliger, E. Greenstein, I. Zaretsky, T. Arnon, F. V. Laethem, A. Singer, J. Lu, P. D. Sun, I. R. Cohen, N. Friedman, T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *eLife*. **6** (2017), doi:10.7554/eLife.22057.

23. G. Chen, X. Yang, A. Ko, X. Sun, M. Gao, Y. Zhang, A. Shi, R. A. Mariuzza, N. Weng, Sequence and Structural Analyses Reveal Distinct and Highly Diverse Human CD8 + TCR Repertoires to Immunodominant Viral Antigens. *Cell Reports*. **19**, 569–583 (2017).

24. J. Glanville, H. Huang, A. Nau, O. Hatton, L. E. Wagar, F. Rubelt, X. Ji, A. Han, S. M. Krams, C. Pettus, N. Haas, C. S. L. Arlehamn, A. Sette, S. D. Boyd, T. J. Scriba, O. M. Martinez, M. M. Davis, Identifying specificity groups in the T cell receptor repertoire. *Nature*. **547**, 94–98 (2017).

25. Q. Qi, M. M. Cavanagh, S. Le Saux, H. NamKoong, C. Kim, E. Turgano, Y. Liu, C. Wang, S. Mackey, G. E. Swan, Diversification of the antigen-specific T cell receptor repertoire after varicella zoster vaccination. *Science translational medicine*. **8**, 332ra46-332ra46 (2016).

26. P. Meysman, N. De Neuter, S. Gielis, D. Bui Thi, B. Ogunjimi, K. Laukens, On the viability of unsupervised T-cell receptor sequence clustering for epitope preference. *Bioinformatics*. **35**, 1461–1468 (2019).

27. N. Tickotsky, T. Sagiv, J. Prilusky, E. Shifrut, N. Friedman, McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics*. **33**, 2924–2929 (2017).

28. M. Shugay, D. V. Bagaev, I. V. Zvyagin, R. M. Vroomans, J. C. Crawford, G. Dolton, E. A. Komech, A. L. Sycheva, A. E. Koneva, E. S. Egorov, A. V. Eliseev, E. Van Dyk, P. Dash, M. Attaf, C. Rius, K. Ladell, J. E. McLaren, K. K. Matthews, E. B. Clemens, D. C. Douek, F. Luciani, D. van Baarle, K. Kedzierska, C. Kesmir, P. G. Thomas, D. A. Price, A. K. Sewell, D. M. Chudakov, VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Research*. **46**, D419–D427 (2018).

29. G. Dolton, A. Lissina, A. Skowera, K. Ladell, K. Tungatt, E. Jones, D. Kronenberg-Versteeg, H. Akpovwa, J. M. Pentier, C. J. Holland, A. J. Godkin, D. K. Cole, M. A. Neller, J. J. Miles, D. A. Price, M. Peakman, A. K. Sewell, Comparison of peptidemajor histocompatibility complex tetramers and dextramers for the identification of antigenspecific T cells. *Clinical and Experimental Immunology*, 17 (2014).

30. S. E. Mann, Z. Zhou, L. G. Landry, A. S. Anderson, A. K. Alkanani, J. Fischer, M. Peakman, R. Mallone, K. Campbell, A. W. Michels, M. Nakayama, Multiplex T Cell Stimulation Assay Utilizing a T Cell Activation Reporter-Based Detection System. *Frontiers in Immunology*. **11**, 17 (2020).

31. S. Culina, A. I. Lalanne, G. Afonso, K. Cerosaletti, S. Pinto, G. Sebastiani, K. Kuranda, L. Nigi, A. Eugster, T. Østerbye, A. Maugein, J. E. McLaren, K. Ladell, E. Larger, J.-P. Beressi, A. Lissina, V. Appay, H. W. Davidson, S. Buus, D. A. Price, M. Kuhn, E. Bonifacio, M. Battaglia, S. Caillat-Zucman, F. Dotta, R. Scharfmann, B. Kyewski, R. Mallone, ImMaDiab Study Group, Islet-reactive CD8 $^{+}$ T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors. *Sci. Immunol.* **3**, eaao4013 (2018).

32. J. J. Adams, S. Narayanan, M. E. Birnbaum, S. S. Sidhu, S. J. Blevins, M. H. Gee, L. V. Sibener, B. M. Baker, D. M. Kranz, K. C. Garcia, Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nature Immunology*. **17**, 87–94 (2016).

33. M. E. Birnbaum, J. L. Mendoza, D. K. Sethi, S. Dong, J. Glanville, J. Dobbins, E. Özkan, M. M. Davis, K. W. Wucherpfennig, K. C. Garcia, Deconstructing the Peptide-MHC Specificity of T Cell Recognition. *Cell*. **157**, 1073–1087 (2014).

34. D. Mason, A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today*. **19**, 395–404 (1998).

35. L. Wooldridge, J. Ekeruche-Makinde, H. A. van den Berg, A. Skowera, J. J. Miles, M. P. Tan, G. Dolton, M. Clement, S. Llewellyn-Lacey, D. A. Price, M. Peakman, A. K. Sewell, A Single Autoimmune T Cell Receptor Recognizes More Than a Million Different Peptides. *Journal of Biological Chemistry*. **287**, 1168–1177 (2012).

36. R. W. Nelson, D. Beisang, N. J. Tubo, T. Dileepan, D. L. Wiesner, K. Nielsen, M. Wüthrich, B. S. Klein, D. I. Kotov, J. A. Spanier, B. T. Fife, J. J. Moon, M. K. Jenkins, T Cell Receptor Cross-Reactivity between Similar Foreign and Self Peptides Influences Naive Cell Population Size and Autoimmunity. *Immunity*. **42**, 95–107 (2015).

37. K. W. Wucherpfennig, J. L. Strominger, Molecular mimicry in T cell-mediated autoimmunity: Viral peptides activate human T cell clones specific for myelin basic protein. *Cell*. **80**, 695–705 (1995).

38. A. Toubal, I. Nel, S. Lotersztajn, A. Lehuen, Mucosal-associated invariant T cells and disease. *Nat Rev Immunol*. **19**, 643–657 (2019).

39. L. Mori, M. Lepore, G. De Libero, The Immunology of CD1- and MR1-Restricted T Cells. *Annu. Rev. Immunol.* **34**, 479–510 (2016).

40. B. D. Rudd, Neonatal T Cells: A Reinterpretation. *Annu. Rev. Immunol.* **38**, 229–247 (2020).

41. M. P. Davenport, N. L. Smith, B. D. Rudd, Building a T cell compartment: how immune cell development shapes function. *Nat Rev Immunol*. **20**, 499–506 (2020).

42. S.-H. Chiou, D. Tseng, A. Reuben, V. Mallajosyula, I. S. Molina, S. Conley, J. Wilhelmy, A. M. McSween, X. Yang, D. Nishimiya, R. Sinha, B. Y. Nabet, C. Wang, J. B. Shrager, M. F. Berry, L. Backhus, N. S. Lui, H. A. Wakelee, J. W. Neal, S. K. Padda, G. J. Berry, A. Delaidelli, P. H. Sorensen, E. Sotillo, P. Tran, J. A. Benson, R. Richards, L. Labanieh, D. D. Klysz, D. M. Louis, S. A. Feldman, M. Diehn, I. L. Weissman, J. Zhang, I. I. Wistuba, P. A. Futreal, J. V. Heymach, K. C. Garcia, C. L. Mackall, M. M. Davis, 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 (2021).

43. J. Wang, I. Jelcic, L. Mühlenbruch, V. Haunerdinger, N. C. Toussaint, Y. Zhao, C. Cruciani, W. Faigle, R. Naghavian, M. Foege, T. M. C. Binder, T. Eiermann, L. Opitz, L. Fuentes-Font, R. Reynolds, W. W. Kwok, J. T. Nguyen, J.-H. Lee, A. Lutterotti, C. Münz, H.-G. Rammensee, M. Hauri-Hohl, M. Sospedra, S. Stevanovic, R. Martin, HLA-DR15 Molecules Jointly Shape an Autoreactive T Cell Repertoire in Multiple Sclerosis. *Cell*. **183**, 1264-1281.e20 (2020).

44. T. Kula, M. H. Dezfulian, C. I. Wang, N. S. Abdelfattah, Z. C. Hartman, K. W. Wucherpfennig, H. K. Lyerly, S. J. Elledge, T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes. *Cell*. **178**, 1016-1028.e13 (2019).

45. D. I. Godfrey, A. P. Uldrich, J. McCluskey, J. Rossjohn, D. B. Moody, The burgeoning family of unconventional T cells. *Nat Immunol*. **16**, 1114–1123 (2015).

46. A. C. Hayday, P. Vantourout, *Annu. Rev. Immunol.*, in press, doi:10.1146/annurev-immunol-102819-023144.

47. S. Panda, J. L. Ding, Natural Antibodies Bridge Innate and Adaptive Immunity. *J.I.* **194**, 13–20 (2015).

48. P. Aaby, A. Bhuiya, L. Nahar, K. Knudsen, A. de Francisco, M. Strong, The survival benefit of measles immunization may not be explained entirely by the prevention of measles disease: a community study from rural Bangladesh. *International Journal of Epidemiology*. **32**, 106–115 (2003).

49. P. Aaby, C. L. Martins, M.-L. Garly, C. Bale, A. Andersen, A. Rodrigues, H. Ravn, I. M. Lisse, C. S. Benn, H. C. Whittle, Non-specific effects of standard measles vaccine at 4.5 and 9 months of age on childhood mortality: randomised controlled trial. *BMJ*. **341**, c6495–c6495 (2010).

50. R. M. Welsh, L. K. Selin, No one is naive: the significance of heterologous T-cell immunity. *Nature Reviews Immunology*. **2**, 417–426 (2002).

51. A. K. Sewell, Why must T cells be cross-reactive? *Nature Reviews Immunology*. **12**, 669–677 (2012).

52. C. Peteranderl, S. Herold, C. Schmoldt, Human Influenza Virus Infections. *Semin Respir Crit Care Med*. **37**, 487–500 (2016).

53. A. Bertoletti, C. Ferrari, Adaptive immunity in HBV infection. *Journal of Hepatology*. **64**, S71–S83 (2016).

54. M. Z. Tay, C. M. Poh, L. Rénia, P. A. MacAry, L. F. P. Ng, The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol* (2020), doi:10.1038/s41577-020-0311-8.

55. P. Barennes, V. Quiniou, M. Shugay, E. S. Egorov, A. N. Davydov, D. M. Chudakov, I. Uddin, M. Ismail, T. Oakes, B. Chain, A. Eugster, K. Kashofer, P. P. Rainer, S. Darko, A. Ransier, D. C. Douek, D. Klatzmann, E. Mariotti-Ferrandiz, Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. *Nat Biotechnol* (2020), doi:10.1038/s41587-020-0656-3.

56. 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. *Nature Methods*. **12**, 380–381 (2015).

57. M. P. J. van der Loo, The stringdist package for approximate string matching. *The R Journal, 6(1), 111-122.*

58. G. Csardi, T. Nepusz, The igraph software package for complex network research.

59. H. Wickham, *ggplot2: elegant graphics for data analysis* (Springer, Cham, Second edition., 2016), *Use R!*

60. P. Shannon, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*. **13**, 2498–2504 (2003).

61. J. H. Krijthe, Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. *R package version 0.13, URL https://github. com/jkrijthe/Rtsne* (2015).

62. Z. Sethna, Y. Elhanati, C. G. C. Jr, T. Mora, OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs, 8.

63. Q. Marcou, T. Mora, A. M. Walczak, High-throughput immune repertoire analysis with IGoR. *Nature Communications*. **9** (2018), doi:10.1038/s41467-018-02832-w.

Acknowledgments:

The authors would like to express their gratitude to the organ donors and their families who allowed the collection of samples for research under sad circumstances. The authors would also like to thank Prof. Pascal Leprince, Dr. Guillaume Lebreton, and Dr. Marina Rigolet of the cardiac surgery team, Prof. Bruno Riou and the graft coordination team of the Pitié-Salpêtrière hospital for their invaluable contribution to sample collection. The authors thank the UMR 8199 LIGAN-PM Genomics platform (Lille, France) for efficient sequencing. We also thank Thierry Mora, and Aleksandra Walczak for helpful discussion.

Funding:

This work was primarily funded by:
TRiPoD ERC-Advanced EU grant 322856 (DK).
LabEx Transimmunom ANR-11-IDEX-0004-02 (DK).
RHU iMAP ANR-16-RHUS-0001 (DK).
Besides, ZZ was supported by the JDRF Postdoctoral Fellowship 3-PDF-2020-942-A-N and RM by the Agence Nationale de la Recherche ANR-19-CE15-0014-01 and Fondation pour la Recherche Médicale EQU20193007831.

Author contributions:

Conceptualization: DK.
Data curation: VQ, PB, PS, HPP.
Data Analyses: VQ, PB, FM, VM, PS, HV, ZZ, HPP.
Funding acquisition: DK
Investigation: VQ, PB, VM, PS, HV, FM, NC, MBS, HPP, BC, HV, ZZ, BB, MS, EMF.
Methodology: VQ, DK.
Project administration: VQ, DK.
Resources: VQ, PB.
Software: VQ, PB, PS, HPP.
Supervision: DK.
Validation: DK.
Visualization: VQ, DK.
Writing – original draft: VQ, DK.
Writing – review & editing: VQ, DK.

Competing interests:

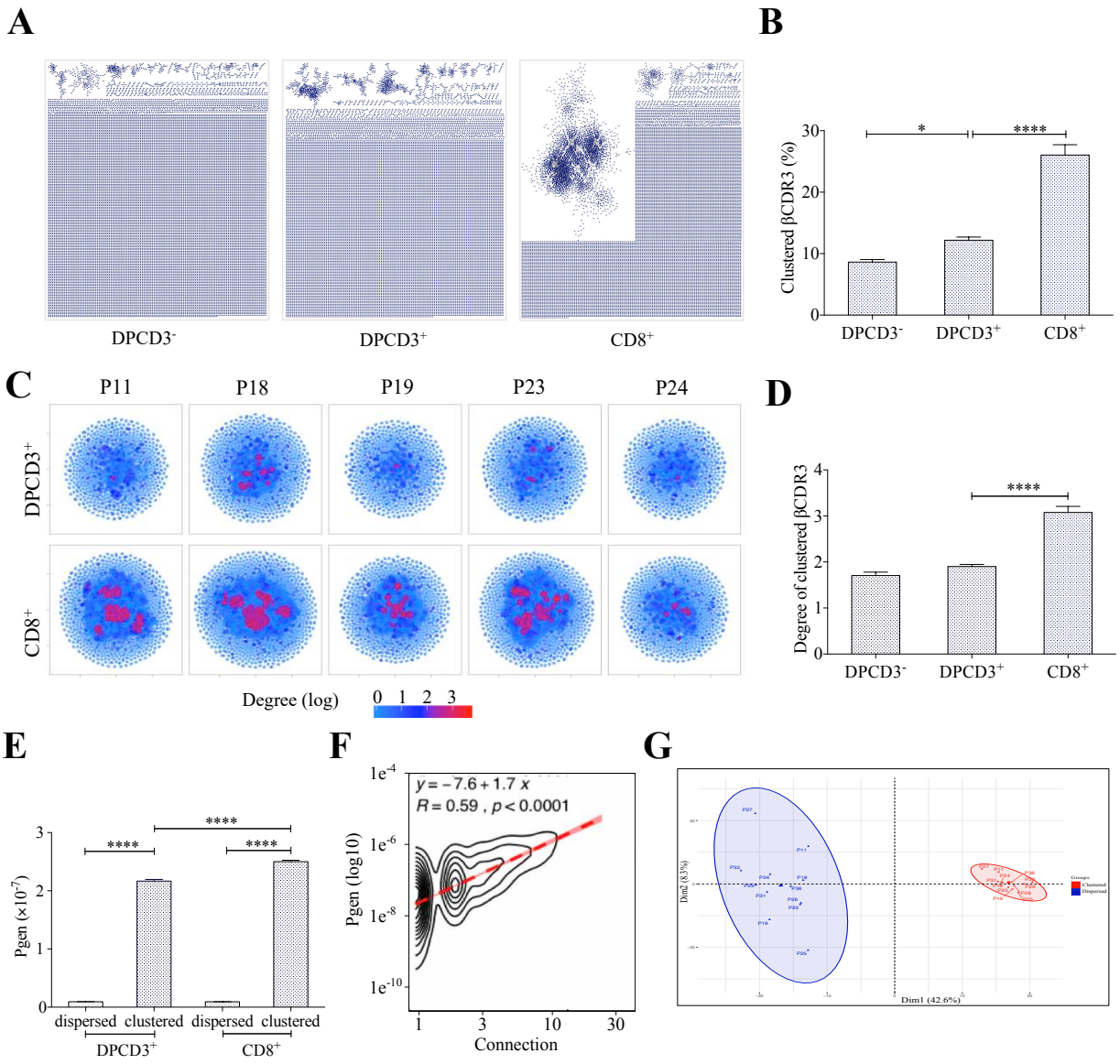
The authors declare no competing financial interests.
Hang-Phuong Pham is affiliated with ILTOO pharma and Parean biotechnologies. The author has no financial interests to declare.

Valentin Quiniou is affiliated with Parean biotechnologies. The author has no financial interests to declare.

Data and materials availability:

Datasets from VDJdb were downloaded from https://vdjdb.cdr3.net. Datasets from McPAS-TCR were downloaded from http://friedmanlab.weizmann.ac.il/McPAS-TCR/. We manually curated these datasets to be sure to use only βCDR3s from CD8 tetramer-specific cells.
Single-cell datasets from 10X genomics were downloaded from https://support.10xgenomics.com/single-cell-vdj/datasets ('Application Note - A New Way of Exploring Immunity' section, datasets 'CD8+ T cells of Healthy Donor' 1–4, available under the Creative Commons Attribution license).
Data from the donors are available on request to the authors.

Figures:



Fig; 1. Thymocyte differentiation produces clustered CDR3s with high generation probability and preferential TRB VJ gene combinations.

A. Representation of βCDR3aa networks from DPCD3$^-$, DPCD3$^+$ and CD8+. Each dot represents a single CDR3. Dot are connected (forming clusters) by edges defined by Levenshtein distance of ≤1 (one AA substitution/insertion/deletion).
B. Percentage of clustered βCDR3aa. (*p=0.0152 and ****p<0.0001, Mann-Whitney test).
C. βCDR3aa clustered from DPCD3$^+$ and ThyCD8. Each dot represents a single βCDR3aa. The colour scale represents the number of neighbours for each CDR3. Blue dots have only 1 connection while Red dots have more than 3 connections (up to 30).
D. Degree of clustered βCDR3aa. (****p<0.0001, Mann-Whitney test).
E. Generation probability of dispersed and clustered βCDR3aa in DPCD3+ or CD8+ cells. ****p<0.0001, Mann-Whitney test).

F. Correlation between *Pgen* and βCDR3 number of connections in the $CD8^+$ thymocyte repertoire. Contour plot represent the generation probability as a function of βCDR3 connections in the $CD8^+$ thymocytes for donor P29. Linear regression curves between *Pgen* and number of connections are represented as red dotted lines ("y" represent the regression curve's equation). Pearson correlation coefficient "R" and p-value "p" are calculated for each individual (Supplementary Figure 3C).
G. PCA analysis of TRB VJ gene combinations in CD8 thymocytes. Blue: dispersed nodes; Red: clustered nodes.

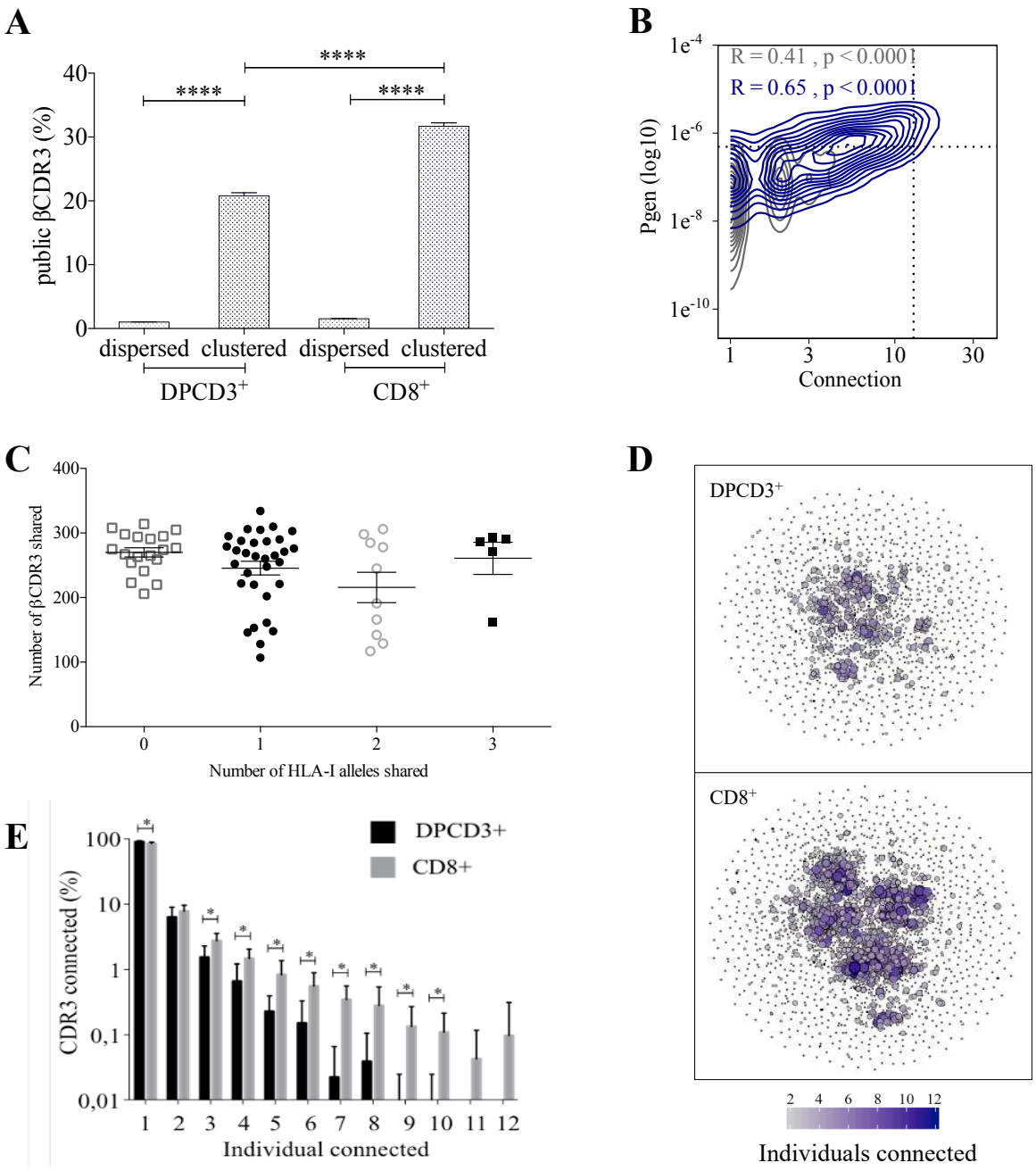


Figure 2. Thymocyte differentiation produces clustered CDR3s with high publicness.

A. Mean percentages of public (black) or private (grey) βCDR3s in all, dispersed or clustered nodes. (****p<0.0001, Mann-Whitney test).

B. Enrichment of public βCDR3s in the CD8⁺ thymocyte repertoires. Representation of the generation probability as a function of βCDR3 connections in individuals (Pn). The contour plots represent shared (blue) or private (grey) βCDR3s for a representative patient. The Pearson correlation coefficient "R" and p-value "p" are calculated for each group. The black dotted lines delimit the threshold for the 2.5% sequences with the higher *Pgen* and connection. βCDR3s with both the highest *Pgen* and connections are also the most

public for 12 out of 12 individuals ($p<0.0001$, Chi-square test, Supplementary table S1, Supplementary Figure 4B).

C. The number of public βCDR3s in $CD8^+$ thymocytes is independent of the number of HLA-I alleles shared. Each dot represents the number of βCDR3s shared between two donors in the first 18,000 $CD8^+$ thymocytes. There is no significant difference in the number of public βCDR3s according to the number of HLA-I alleles shared($p>0.1$, Mann-Whitney test). The number of public βCDR3s is independent of the number of HLA alleles shared.

D. Public convergence of βCDR3 recognition properties during thymopoiesis. CDR3 connections between individuals. The top 1,500 βCDR3s were sampled from $DPCD3^+$ (upper square) and $CD8^+$ (lower square) cells from each individual and pooled. The CDR3s are clustered based on $LD\leq1$ with colour and size both representing the level of sharing between individuals for each CDR3.

E. Convergence of public βCDR3 specificities during thymopoiesis. Bar plots representing the percentage of CDR3s from an individual that are connected to CDR3s of other individuals, for $DPCD3^+$ and $CD8^+$ thymocytes. The first two bars represent CDR3s that are not connected (n=1). The number of unconnected nodes in $DPCD3^+$ is higher than in $CD8^+$ (*$p=0.002$). The other bars represent the percentage of CDR3s connected between individuals. The number of nodes connected to 3 to 10 individuals is significantly higher in $CD8^+$ than in $DPCD3^+$ cells (*$p<0.01$, multiple t-test).

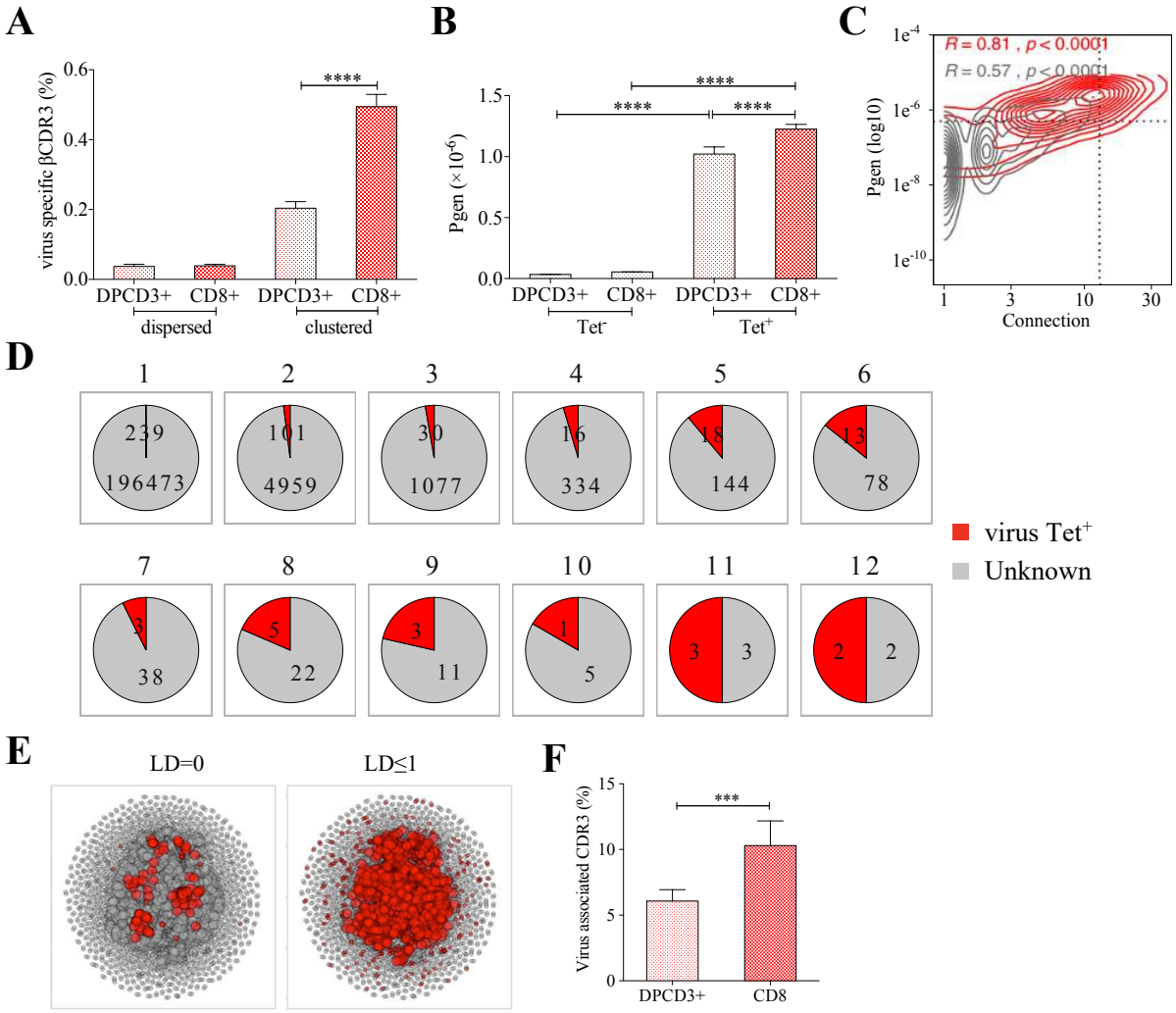


Figure 3. Clustered public TCRs are enriched for virus-specific TCRs.

A. Percentages of CDR3s associated with pathogens within DPCD3$^+$ and ThyCD8 cells and for dispersed or clustered CDR3s.

B. Mean generation probability of virus-specific βCDR3s, based on their identification by tetramer, in DPCD3$^+$ or CD8$^+$ thymocytes (****p<0.0001, Mann-Whitney test).

C. Enrichment of virus-specific βCDR3s in the CD8$^+$ thymocyte repertoire. Representation of the generation probability as a function of βCDR3 connections in a representative individual. The contour plot represents βCDR3s from TCRs identified as virus-specific based on tetramer identification[12,13] (red) or with unknown specificity (grey). Pearson correlation coefficient "R" and p-value "p" are calculated for each group. The black dashed lines delimit the threshold for the 2.5% sequences with both higher *Pgen* and degree of connection. (βCDR3s with both the highest *Pgen* and connections were also the most virus-specific for 11 out of 12 individuals, supplementary (p-value <0.0001, Chi-square test).

D. Virus-specific βCDR3s sharing in CD8$^+$ thymocytes. Pie charts represent the βCDR3s from private (1) to shared βCDR3s by all donors (12), in grey for βCDR3s with unknown specificity or in red for those with a virus specificity.

E. Network of virus-associated CDR3s. The Shared (identical) and Linked (LD≤ 1) virus-associated CDR3s within the CDR3 network of one individual are in red.

F. Virus antigen coverage. Barplot represents the mean percentage of LD≤ 1 to virus-associated CDR3s in all individuals (***p<0.001, Mann-Whitney test).

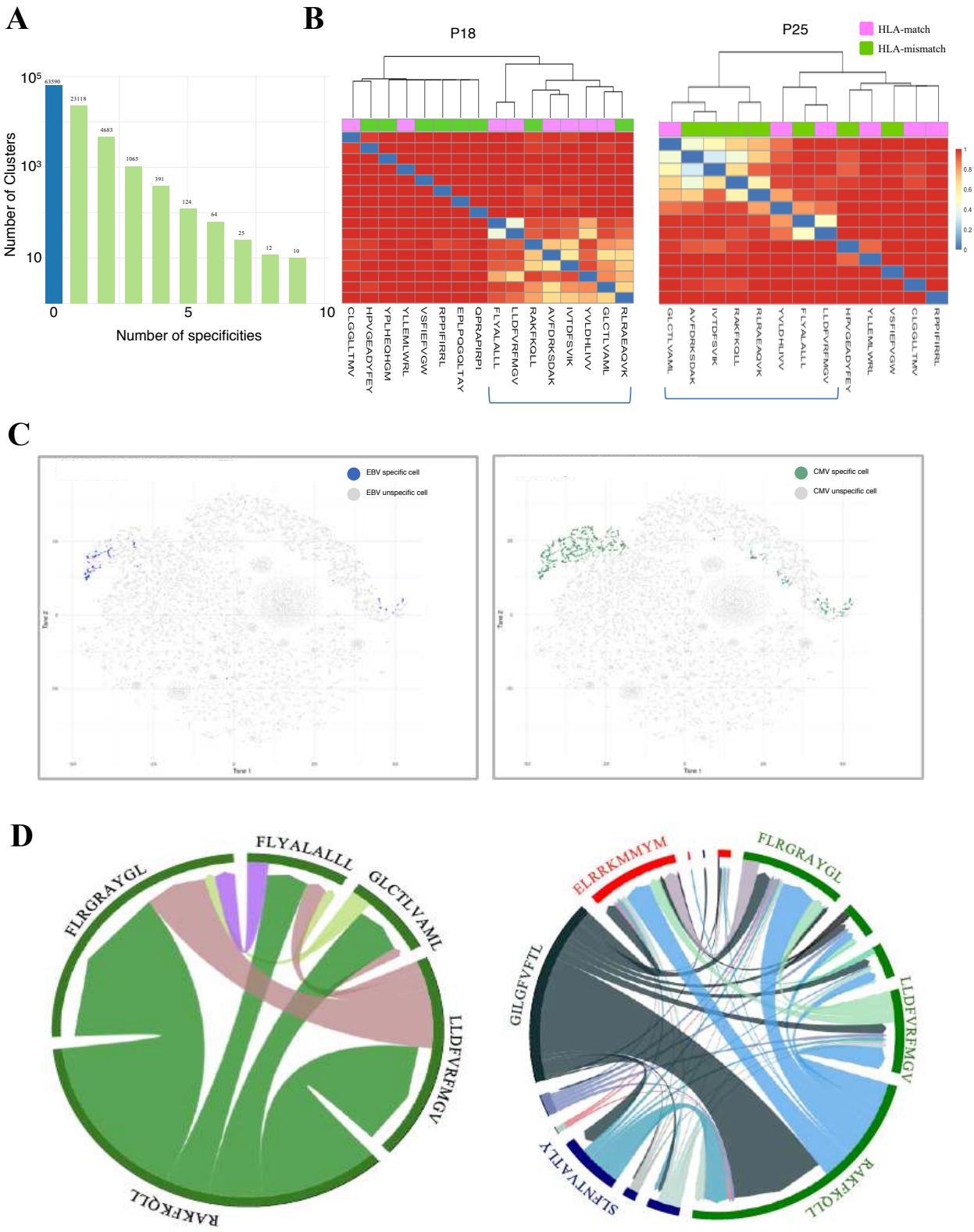


Figure 4. Identification of polyspecific TCRs.

A. Analysis of shared T cell specificities with the GLIPH2 algorithm. 216,000 CDR3β sequences from the CD8 thymocytes and 32,496 sequences from VDJdb that are tetramer specific were analysed to obtain 93,182 specificity groups. 31,6% of the groups contained at least one CDR3 with a known specificity for a virus. 6,8% of the clusters identified by GLIPH2 contained CDR3s with more than one specificity, and up to 9 different ones

B. In silico analysis of EBV specific CDR3β. Morisita-Horn dissimilarity analysis of CDR3β repertoire specific for EBV Tetramer identified by GLIPH Algorithm for P18 and P25 patients.

C. t-SNE representation of the single-cell EBV (left panel) and CMV (right panel) TCR specificities from one seronegative individual. $CD8^+$ T cells are able to bind both EBV and CMV HLA-matched dextramers. EBV Dex^+-specific cells are in blue and CMV Dex^+-specific cells are in green. Some of the cells appear specific for both.

D. Chord diagram showing TCR binding to multiple Dextramers. The chord diagram shows TCR binding to HLA-matched dextramers loaded with peptides from unrelated viruses. Each segment represents TCR binding to the peptides marked above. The size of the segments corresponds to the number of TCRs binding to these peptides. The link between segments identifies multiple TCR binding to different peptides. The colours of the segments represent the different viruses: CMV (red), EBV (green), HIV (dark blue), HPV (light green), HTLV (purple) and influenza (dark grey). The full list of the different peptides is in Supplementary table 3.

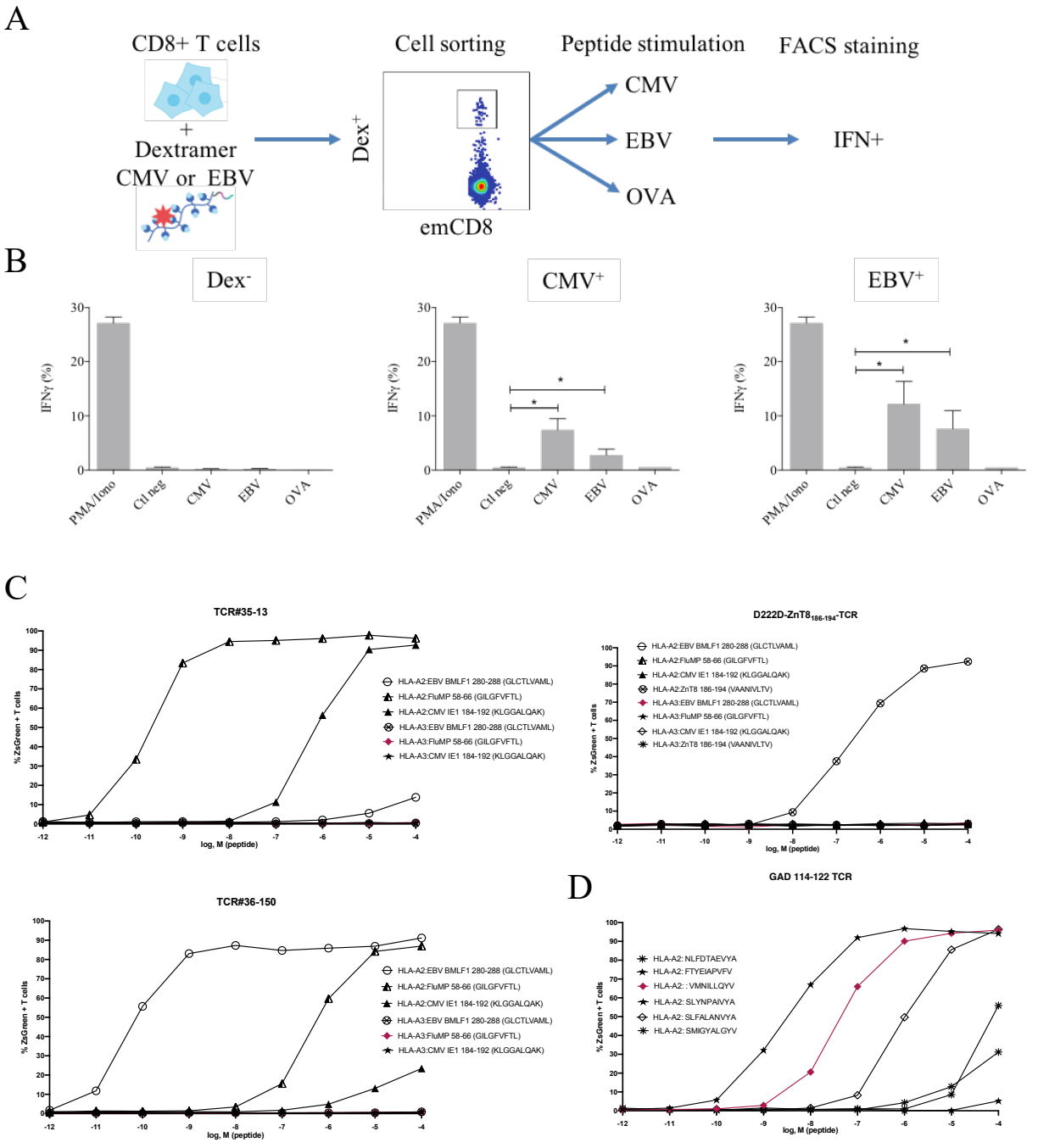


Figure 5. Polyreactivity of polyspecific TCRs.

A. Schematic representation of the in vitro cross-activation experiment.

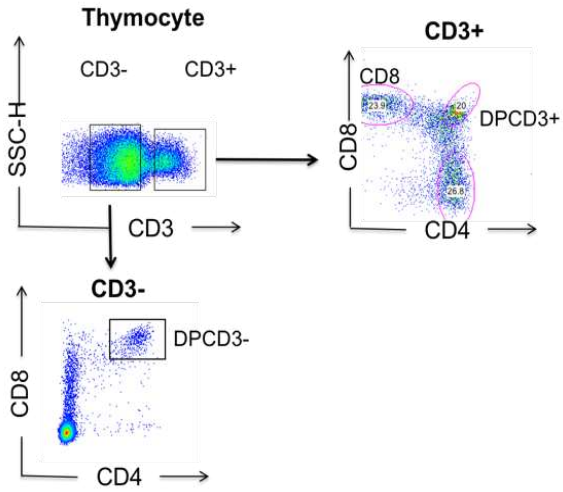
B. In vitro activation of polyspecific T cells. Percentage of IFNγ producing emCD8+ cells after activation with PMA/ionomycin (positive control), no peptide (Ctl neg) or CMV, EBV and OVA peptides (*p<0.05, Mann-Whitney test, mean ± s.e.m.).

C. Poly-reactivity of re-expressed viral epitope-reactive TCRs. 5KC cells transduced with TCR#35-13 (upper left), TCR#36-150 (bottom left) that responded to different peptides in the single cell screening, and D222D-ZnT8186-194 (upper right) that responded to its cognate peptide ZnT8186-

194 were stimulated with the indicated nonamer peptides at different concentrations, pulsed on K562 cells expressing either HLA-A*02:01 or HLA-A*03:01. The percent cells expressing the NFAT-driven ZsGreen reporter is shown as activation readout. A representative experiment out of two performed is shown.

D. TCRs from single-sorted CD8+ T cells stained with tetramers loaded with pancreatic self-peptides were cloned in reporter cell lines. We analyzed their response to various peptides: cognate peptide in red and peptides with no significant structural commonalities (grey).

A



B

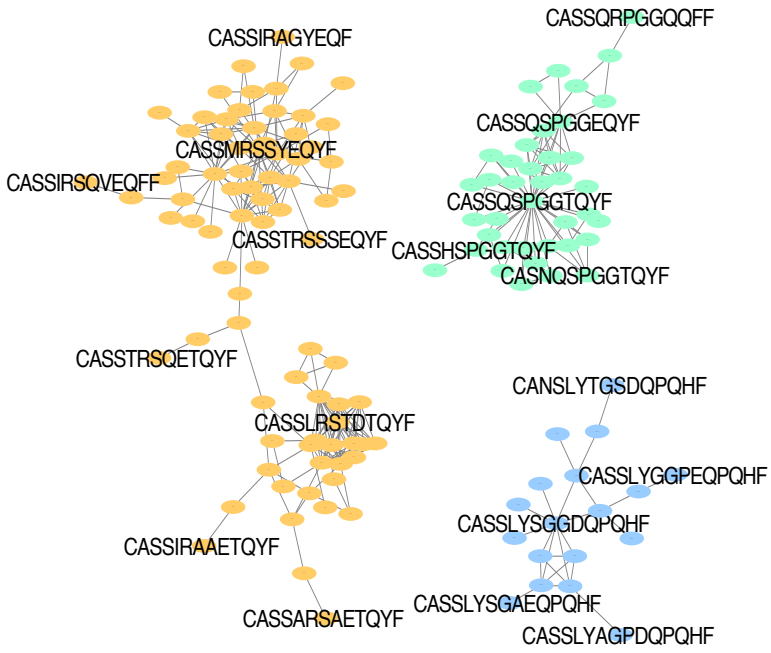


Figure S1: Dotplot of thymocytes sorting.

A. Dotplot of thymocytes sorting. Cells from the different patients were sorted according to their phenotypes: CD3+CD4-CD8+ for ThyCD8, CD3+CD4+CD8+ for DPCD3+ and CD3-CD4+CD8+ for DPCD3-.

B. Networks of βCDR3s specific for GILGFVFTL from influenza (orange), GLCTLVAML from Epstein-Barr virus (green) and FPRPWLHGL from human immunodeficiency virus (blue) are shown. These βCDR3s are from TCRs identified on CD8 T lymphocytes isolated with class I tetramer loaded with the indicated peptides(*27*, *28*). Each node represents a clonotype. Two different clonotypes are connected if their βCDR3s differ by at most one amino acid (LD≤1).

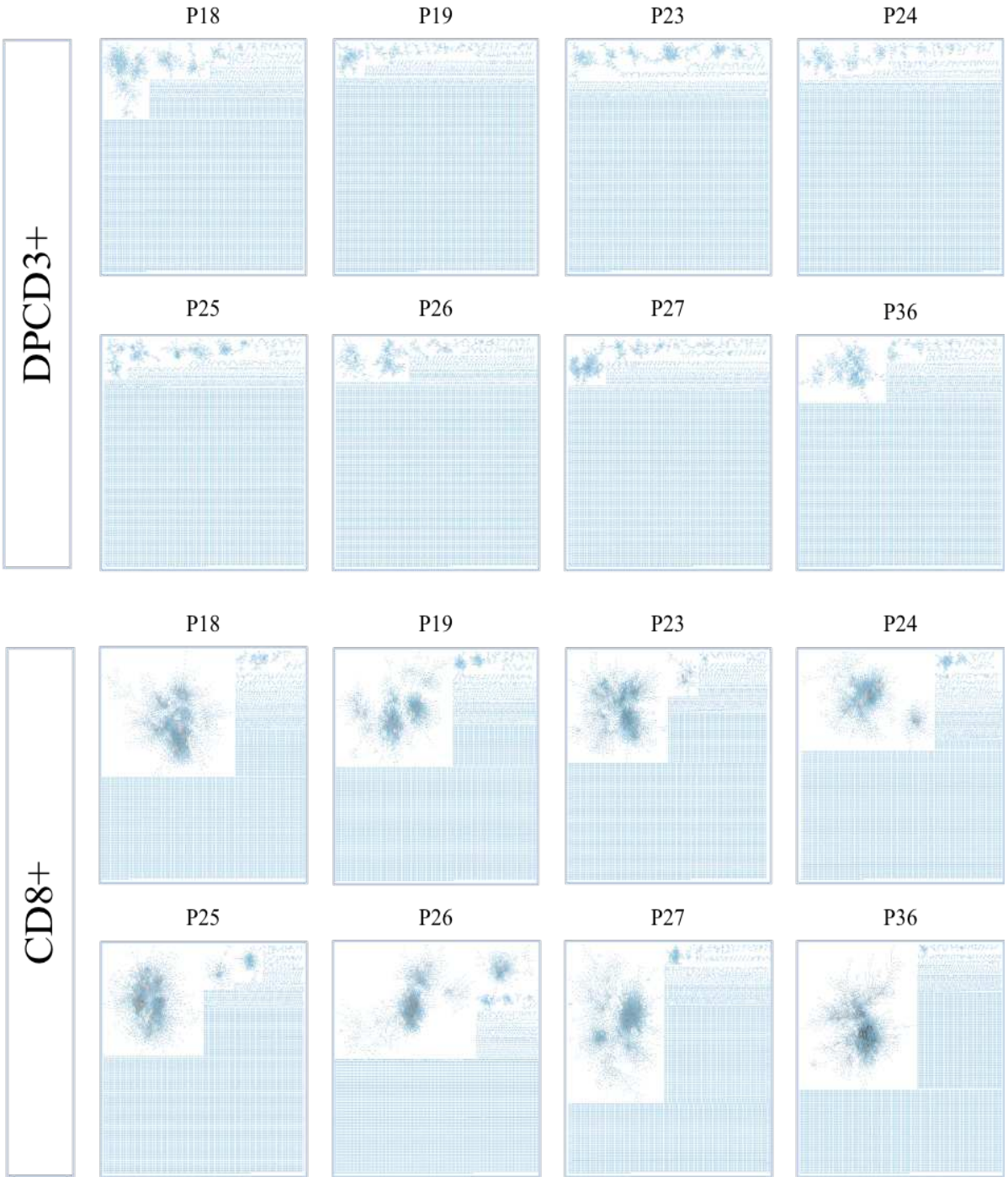
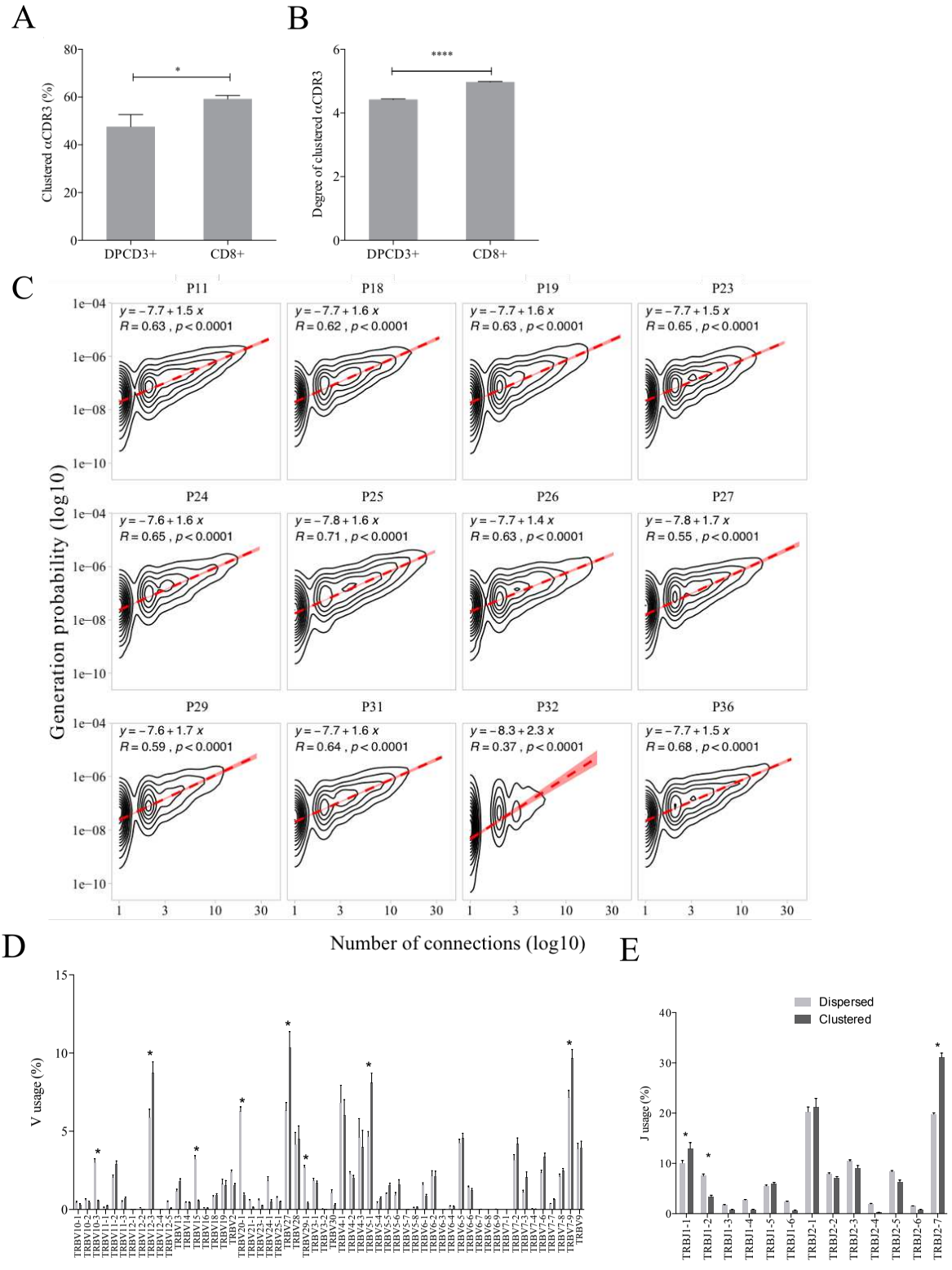


Figure S2: Properties of thymocytes' TCRs

βCDR3 network during thymopoiesis. Representation of the 18,000 most frequent βCDR3 networks from DPCD3^{+} and CD8^{+} thymocytes of eight donors (Pn).



Supplementary Figure 3.

(A-B) Clustered αCDR3s from DPCD3$^+$ and CD8$^+$ thymocytes. Analyses were performed on the first 18,000 most frequent αCDR3s per sample (n=6 for DPCD3$^+$ and n=10 for CD8$^+$ thymocytes). **(A)** Percentage of

clustered αCDR3s. (mean±s.e.m., *p=0.016, Mann-Whitney test). **(B)** Degree of clustered αCDR3s. (mean±s.e.m., ****p<0.0001, Mann-Whitney test).

(C) Correlation between *Pgen* and βCDR3 number of connections in the $CD8^+$ thymocyte repertoire. The contour plots represent the generation probability as a function of βCDR3 connections in the $CD8^+$ thymocytes for donors P11 to P36. Linear regression curves between *Pgen* and number of connections are represented as red dashed lines ("y" represents the regression curve's equation). The Pearson correlation coefficient "R" and p-value "p" are calculated for each individual.

(D-E) Clonogram representation of TCR Vβ **(E)** and Jβ **(F)** usage in clustered versus dispersed $CD8^+$ thymocytes. The bar plots represent the mean percentage of TCR Vβ (left) and Jβ (right) in dispersed (light grey) versus clustered (dark grey). (* p<0.01, multiple t-test).

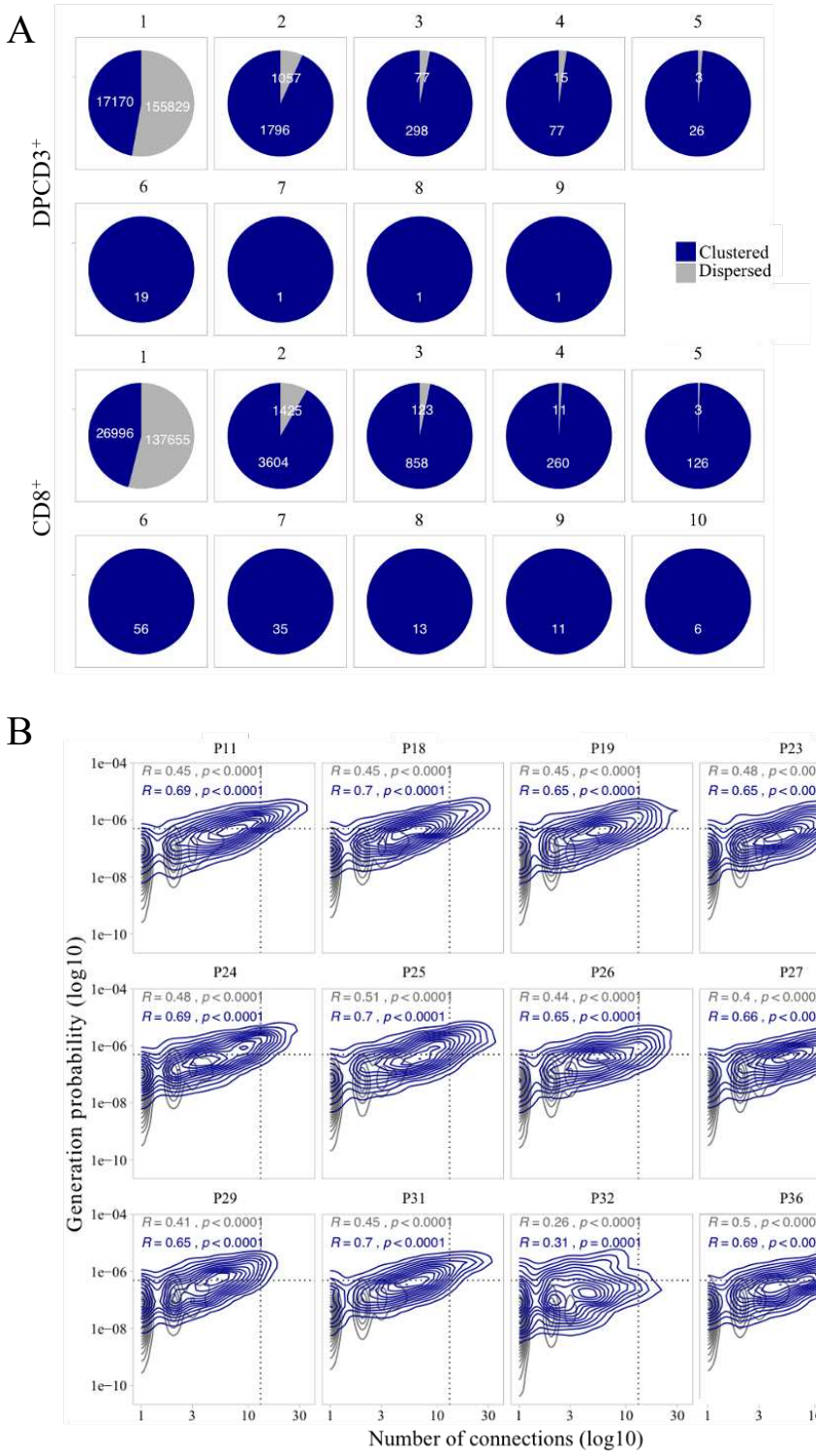


Figure S4. Publicness of thymocytes' CDR3s

(A) βCDR3 sharing between individuals. Pie charts represent the sharing between individuals before ($DPCD3^+$) and after thymic selection ($CD8^+$). Colours represent the dispersed (grey) or clustered (blue)

CDR3s. Sharing was analyzed within the 10 donors for which there were at least 18,000 βCDR3s in DPCD3^{+} and in CD8^{+} thymocytes.

(B) Enrichment of public CDR3s in the CD8+ thymocyte repertoires. Representation of the generation probability as a function of CDR3 connections in individuals (Pn). The contour plots represent shared (blue) or private (grey) CDR3s. The Pearson correlation coefficient "R" and p-value "p" are calculated for each group. The black dotted lines delimit the threshold for the 2.5% sequences with the higher Pgen and connection. CDR3s with both the highest Pgen and connections are also the most public for 12 out of 12 individuals ($p<0.0001$, two-tailed Fisher test).

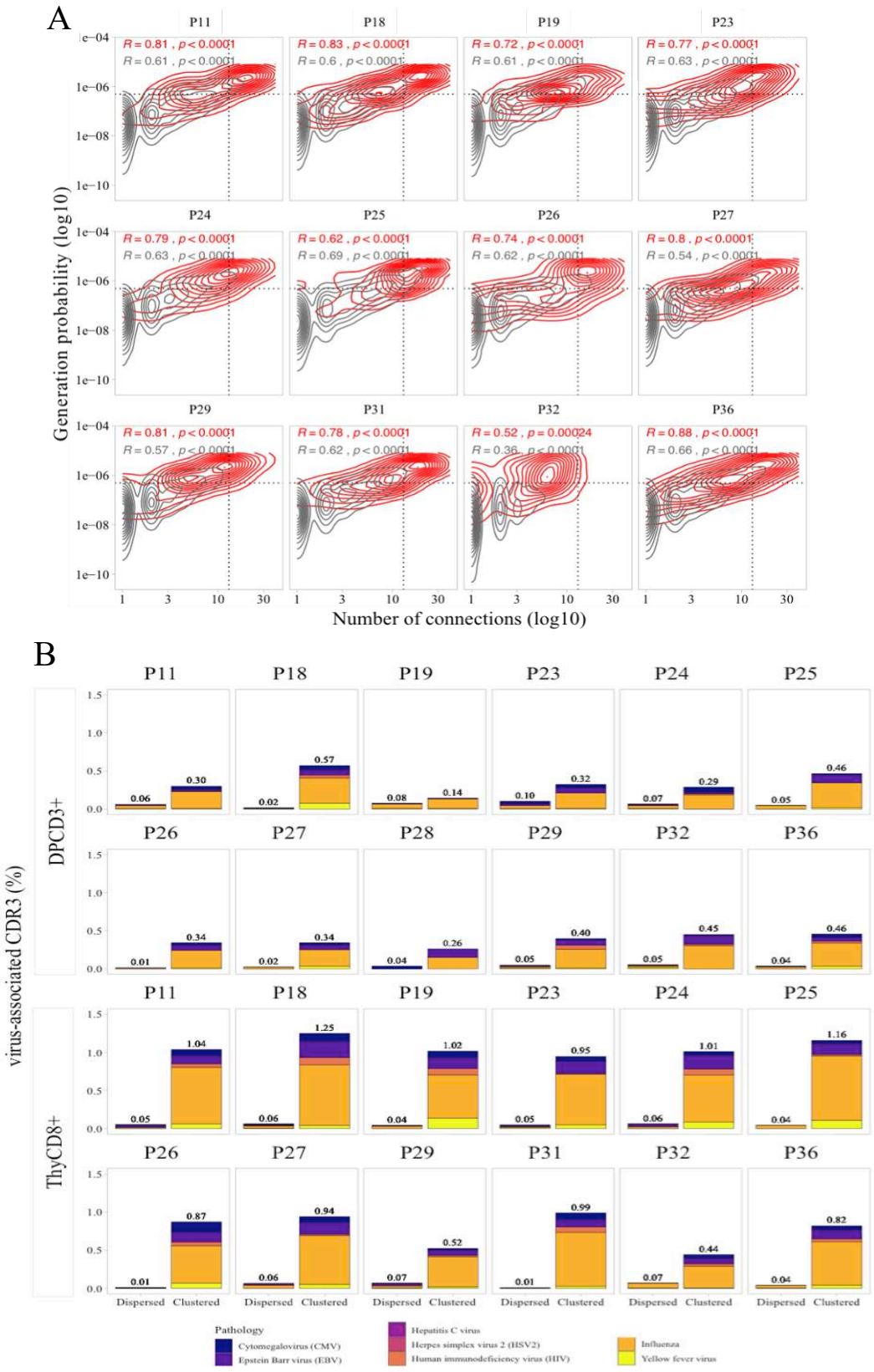

A
P11
P18
P19
P23
P24
P25
P26
P27
P29
P31
P32
P36
Generation probability (log10)
Number of connections (log10)
B
DPCD3+
ThyCD8+
virus-associated CDR3 (%)
Dispersed
Clustered
Pathology
Cytomegalovirus (CMV)
Epstein Barr virus (EBV)
Hepatitis C virus
Herpes simplex virus 2 (HSV2)
Human immunodeficiency virus (HIV)
Influenza
Yellow fever virus

Figure S5. Virus specificities of thymocytes' CDR3s

(A) Enrichment of virus-specific CDR3s in the CD8+ thymocyte repertoire. Representation of the generation probability as a function of CDR3 connections in individuals (Pn). The contour plots represent CDR3s from TCRs identified as virus-specific based on tetramer identification14,15 (red) or with unknown specificity (grey). The Pearson correlation coefficient "R" and p-value "p" are calculated for each group. The black dotted lines delimit the threshold for the 2.5% sequences with both higher Pgen and degree of connection. CDR3s with both the highest Pgen and connections were also the most virus-specific for 11 out of 12 individuals (p-value <0.0001, two-tailed Chi-square test).
(B) Virus-specific CDR3s among DPCD3^{+} and ThyCD8 cells. Barplots represent the percentage of CDR3s from TCR identified as virus-specific based on tetramer binding(*27, 28*). For each panel, calculation is done on dispersed (left boxplot) or clustered (right boxplot) CDR3s. Colours correspond to different viral specificities.

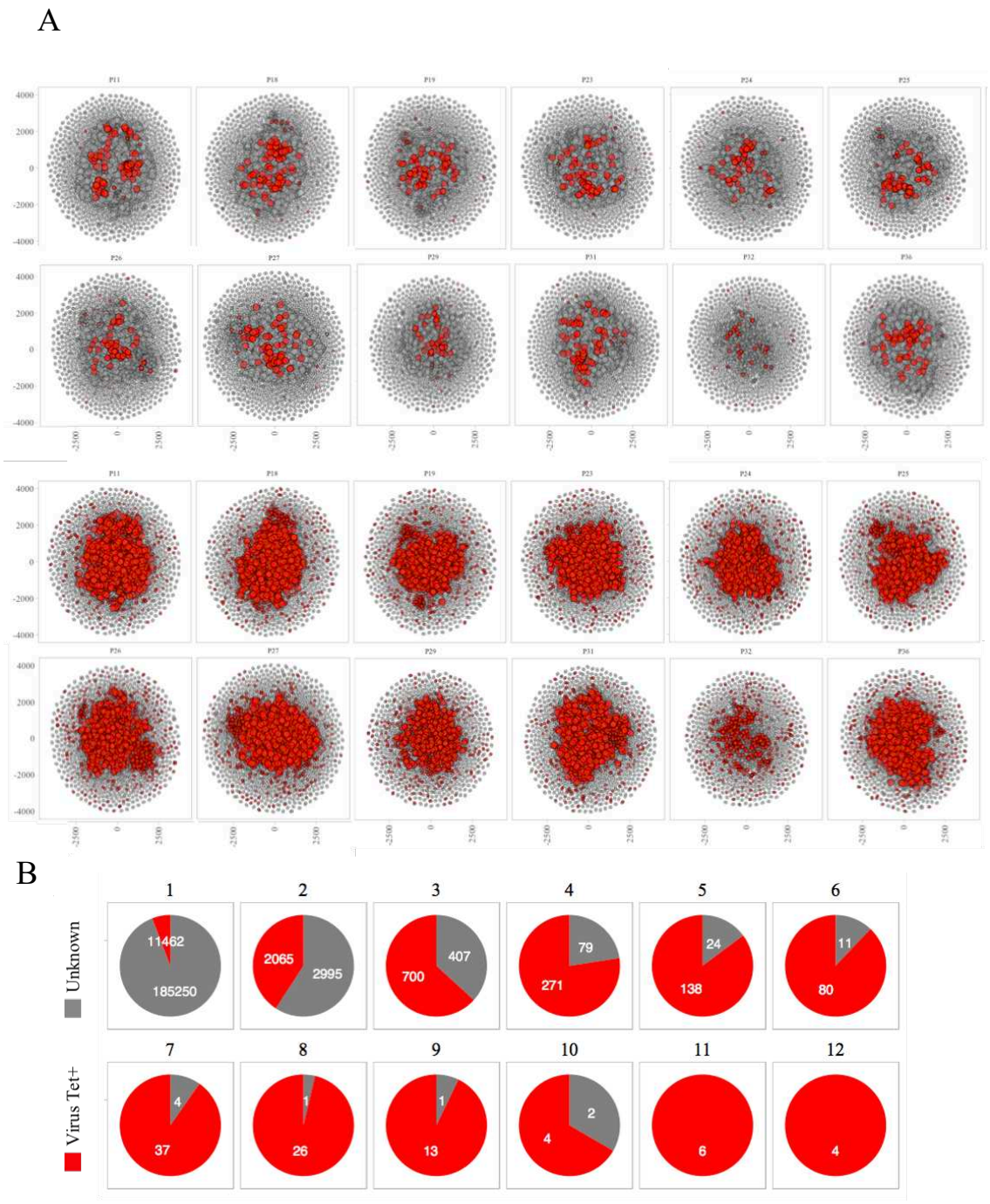


Figure S6. Sharing of virus specific thymocytes' CDR3s

(A) Network of clustered nodes from the first top ranking 18,000 CDR3s. For each individual, we plot CDR3s as viral-associated (red) either if they are in the public database (line 1 and 2) or if they have a LD≤ 1 with a CDR3 of the public database (line 3 and 4).
(B) Sharing of virus-associated CDR3s. The virus-associated CDR3s are highly shared between individuals.

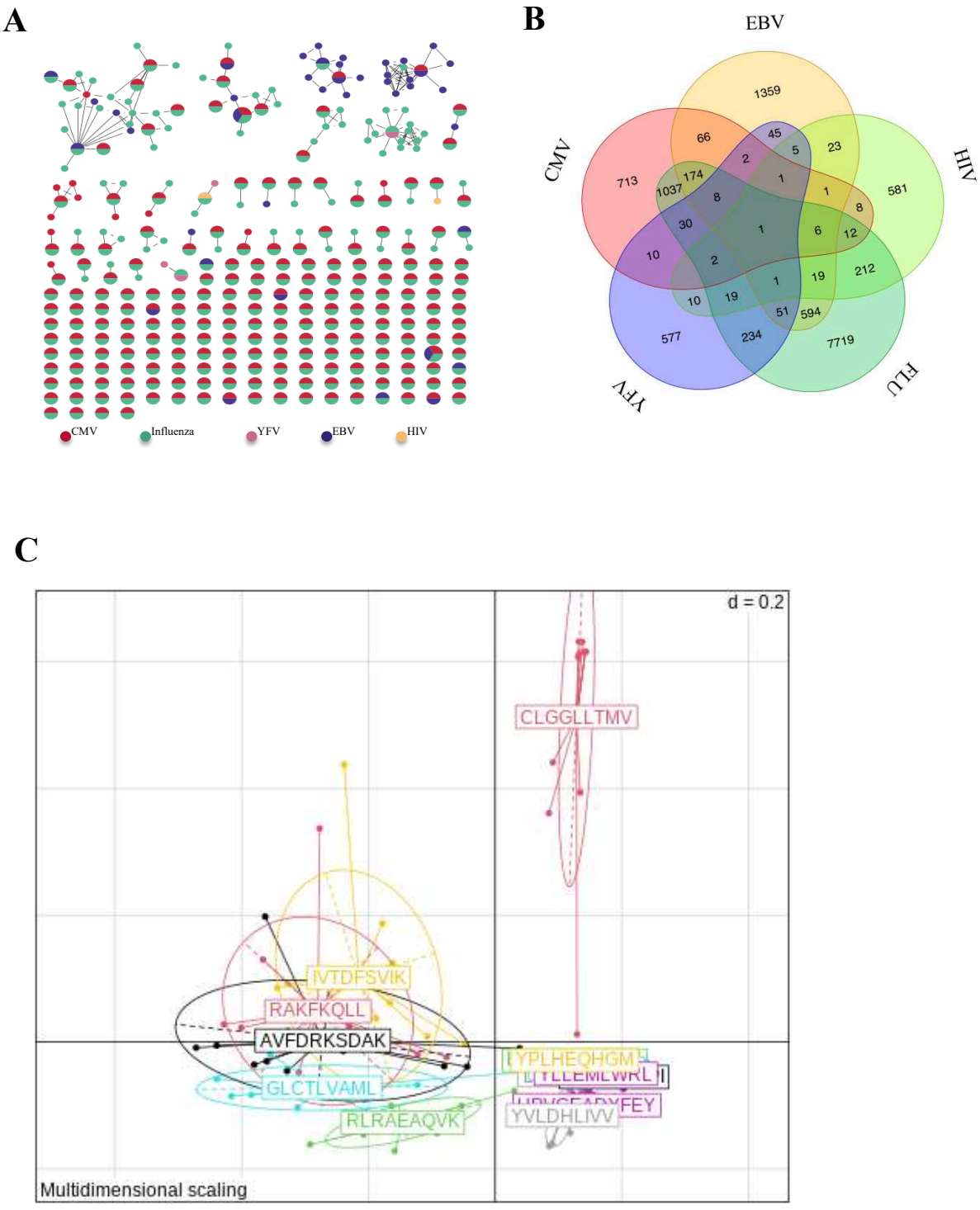


Figure S7. Binding properties of single cell TCRs

A. Network of tetramer-binding CDR3s from public database (VDJdb) identifying CDR3s with multiple viral specificities. Each dot represents a single CDR3. Dot are connected by edges defined by Levenshtein distance of ≤1 (one AA substitution/insertion/deletion). Each dot represents a CDR3 with at least 2

specificities for 2 different viruses. The colors are related to a specific virus. The specificities of these CDR3 were identified by tetramer staining.
B. Venn diagram representing specificities for 5 different viruses of the 13,557 unique virus-associated CDR3s (those from Figure 3e, LD≤1).
C. MDS representation of MH dissimilarity index for EBV repertoire. The MH index are calculated for each repertoire specific of EBV antigen (n=16). Each point represents a patient and each color an antigen.

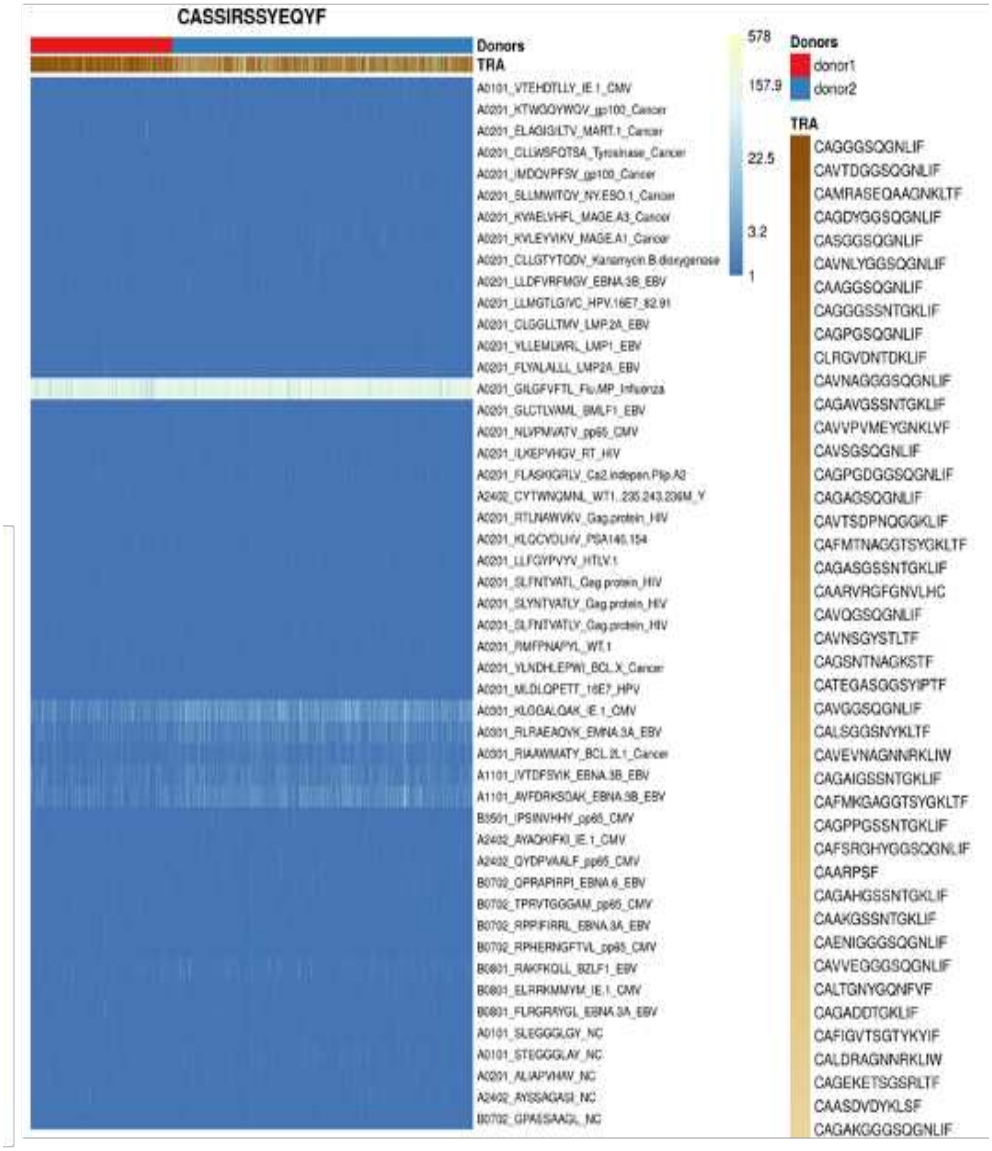


Figure S8**.** βCDR3s drive polyspecificity. Heatmap of the binding score for different dextramers of different TCRs using the same CASSIRSSYEQYF βCDR3, in two donors. The recognition properties are remarkably similar whatever the βCDR3. The polyspecificity is oriented towards the recognition of common viruses such as influenza, EBV, CMV.

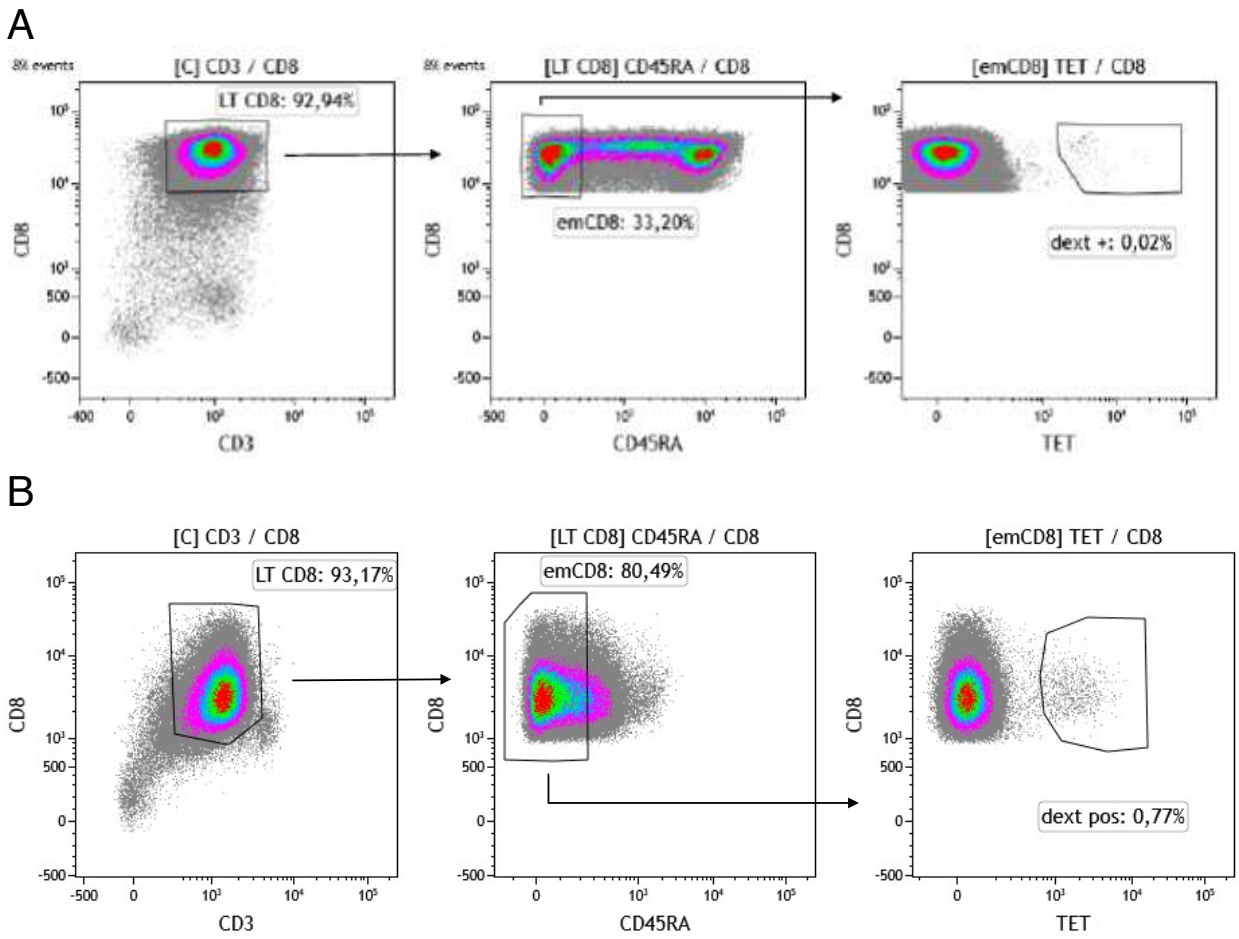


Figure S9. Gating strategy for sorting effector memory (emCD8) Dextramer positive cells.
CD8 were previously enriched from cytapheresis PBMC. Gating strategy of the sorting of Cytomegalovirus (A) or Epstein Barr Virus (B) dextramère positive effector memory CD8 (emCD8) from CD8 enriched fraction of PBMC's donor after excluding doublets. emCD8 cells are $CD3^+$ $CD8^+$ $CD45RA^-$ cells.

Figure S10. Polyspecific properties of anti-CMV cytotoxic cells.

(A) PBMC were stimulated in vitro using Peptivator CMV pp65 (that consists mainly of 15-mer peptides with 11-amino acid overlap, covering the complete sequence of the pp65 protein) and cultured for 25 days. Selected CTL were then restimulated in the presence of either Peptivator CMV pp65, Peptivator CMV IE-1, EBV-BZLF1 or Peptivator EBV EBNA-1.
(B) Polyspecificity is dependent of peptide level. We represent the % of cells secreting IFNγ as a function of peptides concentration. For pp65 CMV peptide (red) and BZLF-1 EBV peptide (blue), the level of IFN is different and depends of the concentration of the peptides, suggesting different affinity for the peptides. These observations were made with cells from 5 different donors.

		DP	CD8	p-value	Odds ratio
P11	Public	590	1630	p<0.0001	0.3403
	Private	17410	16370		
P18	Public	695	1567	p<0.0001	0.4212
	Private	17305	16433		
P19	Public	485	1489	p<0.0001	0.3071
	Private	17515	16511		
P23	Public	649	1489	p<0.0001	0.4148
	Private	17351	16511		
P24	Public	578	1416	p<0.0001	0.3886
	Private	17422	16584		
P25	Public	640	1536	p<0.0001	0.3952
	Private	17360	16464		
P26	Public	584	1470	p<0.0001	0.3771
	Private	17416	16530		
P27	Public	640	1583	p<0.0001	0.3823
	Private	17360	16417		
P29	Public	655	1583	p<0.0001	0.3916
	Private	17345	16417		
P36	Public	653	1468	p<0.0001	0.4239
	Private	17347	16532		

Supplementary table 1. Enrichment of public βCDR3s in $CD8^+$ thymocytes vs $DPCD3^+$.
Contingency table for the Chi-square analysis performed with Yates' correction to test the null hypothesis of independence between the sharing of βCDR3s (Public or Private) vs the cell phenotype ($DPCD3^+$ and $CD8^+$). We performed this test in the 10 donors for which we have 18,000 βCDR3s in both $DPCD3^+$ and $CD8^+$ thymocytes. A βCDR3 is defined as public if it is found at least once in the 18,000 βCDR3s of the same cell phenotype from other donors. The results (p-value < 0.0001) rejected the null hypothesis, thereby indicating the interdependency of the two variables.

		Clustered	Dispersed	p-value	Odds ratio
P11	Virus Tet+	106	8	p<0.0001	40.75
	Unknown	4389	13497		
P18	Virus Tet+	126	10	p<0.0001	39.93
	Unknown	4285	13579		
P19	Virus Tet+	102	8	p<0.0001	46.77
	Unknown	3832	14058		
P23	Virus Tet+	93	7	p<0.0001	47.82
	Unknown	3892	14008		
P24	Virus Tet+	96	9	p<0.0001	40.38
	Unknown	3739	14156		
P25	Virus Tet+	112	7	p<0.0001	63.60
	Unknown	3594	14287		
P26	Virus Tet+	83	2	p<0.0001	146.1
	Unknown	3963	13952		
P27	Virus Tet+	99	8	p<0.0001	34.60
	Unknown	4714	13179		
P29	Virus Tet+	55	9	p<0.0001	31.65
	Unknown	2903	15033		
P31	Virus Tet+	104	1	p<0.0001	360.8
	Unknown	4004	13891		
P32	Virus Tet+	44	9	p<0.0001	32.93
	Unknown	2320	15627		
P36	Virus Tet+	86	6	p<0.0001	54.12
	Unknown	3750	14158		

Supplementary table 2. Enrichment of virus-specific βCDR3s from databases14,15 in clustered CD8+ thymocytes.

Contingency table for the Chi-square analysis performed with Yates' correction to test the null hypothesis of independence between the specificity of βCDR3s (Virus Tet+ and Unknown specificities) vs the connection of βCDR3 ("clustered" and "dispersed") in all the CD8+ thymocytes from 12 donors. The results (p-value < 0.0001) rejected the null hypothesis, thereby indicating the interdependency of the two variables.

Peptide	Virus
FLRGRAYGL	EBV
FLYALALLL	EBV
GLCTLVAML	EBV
LLDFVRFMGV	EBV
RAKFKQLL	EBV
RTLNAWVKV	HIV
SLFNTVATL	HIV
SLFNTVATLY	HIV
MLDLQPETT	HPV
LLFGYPVYV	HTLV
GILGFVFTL	Influenza
ELRRKMMYM	CMV
VTEHDTLLY	CMV
SLYNTVATLY	HIV

Table S3. List of peptides represented on the chord plot. Related to 4G right.
The table is organized according to the clockwise order of the chord plot segments.

Annexe 3.

Biological controls for standardization and interpretation of adaptive immune receptor repertoire profiling

---

Johannes Trück, Anne Eugster, Pierre Barennes, Christopher M. Tipton, Eline T Luning Prak, Davide Bagnara, Cinque Soto, Jacob S. Sherkow, Aimee S. Payne, Marie-Paule Lefranc, Andrew Farmer, The AIRR Community, Magnolia Bostick*, Encarnita Mariotti-Ferrandiz*

ELIFE SCIENCES - 2021

REVIEW ARTICLE

Biological controls for standardization and interpretation of adaptive immune receptor repertoire profiling

Johannes Trück[1†], Anne Eugster[2†], Pierre Barennes[3,4†], Christopher M Tipton[5], Eline T Luning Prak[6], Davide Bagnara[7], Cinque Soto[8,9], Jacob S Sherkow[10,11,12], Aimee S Payne[6], Marie-Paule Lefranc[13,14,15], Andrew Farmer[16], The AIRR Community, Magnolia Bostick[16‡§*], Encarnita Mariotti-Ferrandiz[3‡*]

[1]University Children's Hospital and the Children's Research Center, University of Zurich, Zurich, Switzerland; [2]CRTD Center for Regenerative Therapies Dresden, Faculty of Medicine, Technische Universität Dresden, Dresden, Germany; [3]Sorbonne Université U959, Immunology-Immunopathology-Immunotherapy (i3), Paris, France; [4]AP-HP Hôpital Pitié-Salpêtrière, Biotherapy (CIC-BTi), Paris, France; [5]Lowance Center for Human Immunology, Emory University School of Medicine, Atlanta, United States; [6]Perelman School of Medicine, University of Pennsylvania, Philadelphia, United States; [7]University of Genoa, Department of Experimental Medicine, Genoa, Italy; [8]The Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, United States; [9]Department of Pediatrics, Vanderbilt University Medical Center, Nashville, United States; [10]College of Law, University of Illinois, Champaign, United States; [11]Center for Advanced Studies in Biomedical Innovation Law, University of Copenhagen Faculty of Law, Copenhagen, Denmark; [12]Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, Illinois, United States; [13]IMGT, The International ImMunoGeneTics Information System (IMGT), Laboratoire d'ImmunoGénétique Moléculaire (LIGM), Institut de Génétique Humaine (IGH), CNRS, University of Montpellier, Montpellier, France; [14] Laboratoire d'ImmunoGénétique Moléculaire (LIGM) CNRS, University of Montpellier, Montpellier, France; [15]Institut de Génétique Humaine (IGH), CNRS, University of Montpellier, Montpellier, France; [16]Takara Bio USA, Inc., Mountain View, United States

***For correspondence:** magnolia.bostick@gmail.com (MB); encarnita.mariotti@sorbonne-universite.fr (EM-F)

[†]These authors contributed equally to this work
[‡]These authors also contributed equally to this work

Present address: [§]Process Development, PACTPharma, Inc, South San Francisco, United States

Competing interest: See page 13

Funding: See page 13

Received: 13 January 2021
Accepted: 15 May 2021
Published: 26 May 2021

Reviewing editor: Lindsay Cowell, , United States

© Copyright Trück et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Abstract Use of adaptive immune receptor repertoire sequencing (AIRR-seq) has become widespread, providing new insights into the immune system with potential broad clinical and diagnostic applications. However, like many high-throughput technologies, it comes with several problems, and the AIRR Community was established to understand and help solve them. We, the AIRR Community's Biological Resources Working Group, have surveyed scientists about the need for standards and controls in generating and annotating AIRR-seq data. Here, we review the current status of AIRR-seq, provide the results of our survey, and based on them, offer recommendations for developing AIRR-seq standards and controls, including future work.

Introduction

Immunoglobulin chains (IG) and T-cell receptor chains (TR) are generated by DNA recombination, a process of somatic rearrangement of variable (V), diversity (D), and joining (J) genes

(*Tonegawa, 1983*). The diversity of the resulting rearranged genes (referred as V-J and V-D-J) is very high, due to not only the combination of different germline V, D, and J genes, but also to the deletion and addition of templated (P) nucleotides and the addition of non-templated (N) nucleotides at the junctions between the rearranged genes, and somatic hypermutation of expressed IG (*Papavasiliou and Schatz, 2002*; *Lefranc and Lefranc, 2020*). The total number of potential expressed rearranged IG and TR sequences in an individual is referred to as the adaptive immune receptor repertoire (AIRR). The adaptive immune repertoire is very diverse in a healthy individual, with the theoretically possible number of clonotypes reaching more than 10^{19} different TR (*Bradley and Thomas, 2019*) and 10^{11} IG (*Glanville et al., 2009*), far exceeds the number of B and T cells in a given individual (*Davis and Bjorkman, 1988*; *Freeman et al., 2009*; *Elhanati et al., 2015*). Thanks to next-generation sequencing (NGS), the AIRR can be sampled with sufficient depth for some of its complexity to be studied (*Weinstein et al., 2009*; *Six et al., 2013*). AIRR sequencing (AIRR-seq) provides insights into the immune status of an individual at steady-state or in altered conditions such as malignancy, autoimmune disease, immunodeficiency, infectious disease, or vaccination, and allows comparison of B- and T-cell populations between individuals and time points (*Benichou et al., 2012*; *Kirsch et al., 2015*; *Dziubianau et al., 2013*; *Hou et al., 2016*; *Ghraichy et al., 2018*). AIRR-seq permits the description and quantification of global diversity and characteristics of AIRR, the identification of clonal expansions, the tracking of particular clonotypes, and the prediction of their specificities (*Miho et al., 2018*; *Zvyagin et al., 2020*; *Sidhom et al., 2018*; *Glanville et al., 2017*; *Huang et al., 2020*; *Jokinen et al., 2021*; *Akbar et al., 2021*; *Hayashi et al., 2021*) as well as the antibody selection through phage display (*Rouet et al., 2018*; *Ravn et al., 2013*), thereby providing opportunities for new biomarker identification (*Gittelman, 2021*; *Dines, 2020*), therapeutic antibody discovery (*Akbar et al., 2021*; *Richardson et al., 2021*), CAR-T cell bioengineering (*Sheih et al., 2020*), vaccine development, cancer diagnostics and treatment (*Linette et al., 2019*; *Zhang et al., 2018*; *Lu et al., 2018*), including neoantigen discovery (*Chiou et al., 2021*; *Richters et al., 2019*) and immune intervention monitoring in diverse pathologies, such as stem cell transplantation (*Robinson, 2015*; *Fink, 2019*; *Jiang et al., 2019*; *Jacobsen et al., 2017*; *Theil, 2017*; *Link-Rachner et al., 2019*; *Rubelt et al., 2017*; *Parola et al., 2018*; *Georgiou et al., 2014*; *Arnaout et al., 2021*; *Anand et al., 2021*).

NGS-based approaches and methods have multiplied, now including high-throughput bulk sequencing of IG or TR starting from genomic DNA (gDNA) or mRNA (as cDNA), which typically provides information on one receptor chain only, and more recently to the sequencing of the two IG or TR chains expressed in a single cell, which provides information on the antigen-specific receptor. These approaches are increasingly applied, mostly to human AIRRs, but also to study AIRRs from other organisms (*Chaudhary and Wesemann, 2018*; *Minervina et al., 2019*). Molecular protocols to amplify IG or TR chains typically rely on Polymerase Chain Reaction (PCR), such as multiplex-PCR or RACE-PCR (*Robins et al., 2009*; *Wang et al., 2010*; *DeKosky et al., 2013*; *Eugster et al., 2013*; *Heather et al., 2015*; *Mamedov et al., 2013*).

To obtain reliable and comparable AIRR-seq data, the methods for performing AIRR-seq need to fulfill a number of requirements. First, the data generated by AIRR-seq must reflect the composition and diversity of the 'real cellular repertoire.' The cell subset(s), the cell sample size, and the sequencing depth used in an AIRR-seq experiment all influence the downstream data and should therefore be carefully adapted to the experimental question (*Rosenfeld et al., 2018*). PCR amplification of AIRR-seq libraries can introduce bias by preferentially targeting certain genes, by missing certain alleles (in the case of multiplex PCR with primers anchored in the V or J genes), and by overamplifying targets of certain genes and length (in multiplex PCR and RACE-PCR) (*Calis and Rosenberg, 2014*; *Alamyar et al., 2012*; *Gadala-Maria et al., 2015*; *Primi et al., 1986*; *Barennes et al., 2021*). All these parameters will influence how accurately AIRR profiling reflects the true abundance and diversity of clones in the immune repertoire. Second, AIRR V and J genes (and constant (C) genes for IG) must be identified in an unambiguous and unbiased manner as knowledge of the complementarity determining region 3 (CDR3; the site of V, (D), and J recombination in an IG or TR, and hence its greatest sequence diversity), but also of the CDR1 and CDR2 (and that of the C gene in IG) is critical to assess the physicochemical constraints that define specificity, affinity, and function of the

TR or the IG (**Li et al., 2013**; **Rossjohn et al., 2015**). Third, AIRR-seq data should be as free from sequencing error as possible. The CDR3, key for assigning a sequence to a clonotype, is by definition unknown, and in the case of IG sequences it is important to be able to distinguish *bona fide* somatic hypermutation from artifactual mutations introduced by PCR or by sequencing errors all along the rearranged molecule, as the latter can generate falsely elevated inter- and intra-clonal variation. Further complicating the matter is the fact that the germline V genes from the same subgroup (e.g., IGHV4-31 vs. IGHV4-30-4 for IG) and V alleles or polymorphic variants of a given gene (e.g., TRAV14/DV4 for TR) may have very few nucleotide differences (**Lefranc, 2014**; **Lefranc and Lefranc, 2001**). Therefore, distinguishing errors from true biological variants can be a major challenge. Finally, samples for AIRR-seq should be free from cross-sample contamination, to which AIRR-seq experiments are prone as multiple samples are often processed in parallel and sequenced on a single lane of a sequencing run.

In the past decade, multiple molecular biology protocols and approaches have been developed by academic and industrial investigators, rendering comparisons among studies difficult. Moreover, experimental and analytical protocols are highly complex and therefore prone to intra- and inter-experimental variability (**Barennes et al., 2021**; **Liu et al., 2016**; **Rosati et al., 2017**; **Bashford-Rogers et al., 2014**). AIRR-seq can be performed either on gDNA or cDNA from multiple T/B cell populations (bulk sequencing) or on individual cells (single-cell sequencing). The starting material and sequencing method used depend on the application, as each has advantages and disadvantages (**Table 1**). The availability of commercial kits can be helpful, since they are produced following standards and rigorous quality control, thus offering standardized reagents and protocols across laboratories. Currently, available commercial kits include gDNA-based methods (e.g., Adaptive Biotechnologies, iRepertoire) as well as mRNA-based methods that use cDNA (e.g., Illumina, Takara Bio, iRepertoire) for bulk sequencing. Mainly mRNA-based commercial methods (e.g., 10X Genomics, Takara Bio, HiFiBio) are used for single-cell analysis, which can provide sequences for full receptors or antibodies. This is an important consideration as the determination of the isotype for IG requires a full or partial sequence of the constant region. Single-cell approaches are further helping in the detection of clonotypes because they provide both paired chains and potentially allow full-length cDNA sequencing (**Stubbington et al., 2017**). The fidelity of sequence is an additional factor to consider. Unique molecular identifiers (UMI) consisting of random stretches of 8–12 nucleotides are incorporated into oligonucleotides that are used to generate cDNA from mRNA, such that statistically each cDNA molecule contains a unique sequence. Analysis of sequences that share the same UMI is used to generate a consensus sequence, greatly reducing sequencing errors (**Shugay et al., 2014**). In contrast, multiplex PCR approaches can be associated with artifacts arising from primer competition or off-target primer binding. Although this favors RACE-PCR when considering mRNA-based methods, gDNA-based multiplex PCR may offer higher fidelity since it does not rely on reverse transcription (reverse transcriptase enzymes have higher error rates than DNA polymerases [**Ellefson et al., 2016**; **Holland et al., 1982**]). Finally, cost may influence the choice of a particular protocol. There are many factors that contribute to the cost of AIRR-seq data generation. For example, the cost of sequencing, the sequencing depth, and the number of cells analyzed per sample are variable; also, the choice between commercial kits and 'homebrew' methods will influence costs. In general, gDNA analysis is the most cost-effective, because it requires the lowest sequencing depth with the largest representation of cells per sample, whereas single-cell analysis is on the opposite end of the scale, with bulk cDNA sequencing in the middle.

Several considerations should be taken into account when designing and planning an AIRR-seq experiment. In addition to the large number of different methods and protocols, other factors including budgetary constraints, timelines, sample types, and processing are also important. Given the diversity of AIRR-seq workflows, comparisons between different data sets are challenging or even impossible. Standards and controls are needed for optimal AIRR-seq data harmonization, interpretation, and sharing (**Rubelt et al., 2017**; **Breden et al., 2017**; **Vander Heiden et al., 2018**). This need led to the formation in 2015 of a grassroots community of scientists and other interested parties, known as the AIRR Community (https://www.antibodysociety.org/the-airr-community/). The objective of the Biological Resources Working Group (WG) within the AIRR Community is to coordinate the assessment and development of AIRR-seq controls, ultimately providing the scientific community with controls and standards for the generation, harmonization, and rigorous comparison and interpretation of AIRR-seq data. In order to recommend biological standards that are needed and to

Table 1. Current AIRR-seq methods and their typical use(s).

Bulk gDNA, bulk cDNA, and single-cell cDNA-based sequencing methods are compared with respect to their general features, uses, methods, and potential issues. Each is ranked using a semi-quantitative scale (from '+++" for best to '-" for worst or non-existent).

		Bulk gDNA sequencing	Bulk cDNA sequencing	Single-cell cDNA sequencing
General Features	PCR method	Multiplex	Multiplex and 5′ RACE	Multiplex and 5′ RACE
	Cell number	10^2–10^6	10^2–10^6	10^2–10^3
	Sample throughput	Low-high	Low-moderate	Low
	Length of receptor sequences	100–600 bp	150–600 bp	700–800 bp
	Availability of commercial kits and service providers	++	+++	+
Uses	Gene usage	++	++	+
	CDR3 length and properties	++	++	+
	Somatic hypermutation (for IG)	++	++	+
	Repertoire diversity	++	++	+/-
	Clonal expansion	+++	++	+
	Clonal evolution	++	+++	++
	Tracking of clonotypes	+++	++	+
	Clinical use (e.g., MRD detection)	++	+/-	-
	Unbiased detection of unproductive rearrangements	++	-	-
	Inference of germline	++	+	+/-
	Determination of constant gene	-	++	+
	Structural annotation	+/-	++	+
	Linkage of both antigen receptor chains	+/-	+/-	++
	Direct combination of AIRR-seq with single-cell immunophenotype (e.g., transcriptome or cell surface protein expression)	-	-	++
	Characterization of clonotype full antigen receptor/Functional testing	-	+/-	++
	Rare clonotype detection	++	++	+/-
Methods	Simplicity of workflow (library preparation)	+++	++	+
	Cost for library preparation commercial kits (per sample)	Low	Moderate	High
	Fidelity in sequences	Moderate	High	High
	Molecular barcoding (correcting PCR/sequencing error)	+/-	++	++
Potential Issues	V-gene amplification bias	++	+	+/-
	V-gene annotation issues	++	+	+
	PCR and sequencing error	++	+	+/-
	Difficulty with translation of copy number to cells	+/-	++	+/-
	Degradation of template	+	++	++

bp = base pairs; CDR3 = complementarity determining region 3; MRD = minimal residual disease; RACE = rapid amplification of cDNA ends; V = variable.

prioritize their development, a written survey was developed asking participants about the use of and need for controls for AIRR-seq experiments. The survey gathered information on participants' research interests, sample types, sequencing methodologies, currently available controls, and desired controls. In addition to the survey results, different AIRR-seq methods and controls were also gleaned from the literature, and finally, the WG also invited scientists with unpublished research on controls for their input. These three sources of data were then used to provide a comprehensive overview of sequencing methods, technical issues, current standards, and potential priorities for the development of future standards. Here we describe the progress of the Biological Resources WG of the AIRR Community, its information collection and proposed strategies to define, develop, and use AIRR-seq standards.

Biological controls in AIRR-seq experiments

AIRR Community survey: Overview and respondent demographics

To address the use, needs, and requirements for AIRR-seq controls and standards, we designed and disseminated a questionnaire to researchers in the AIRR Community, as well as to users of IMGT, the international ImMunoGeneTics information system (http://www.imgt.org) (***Lefranc, 2014***). The questionnaire was composed of 4 sections and included 28 questions, with tick-box predefined answers and free-text options allowing for participants' personalized answers (**Supplementary material**). After 6 months, 105 responses were recorded, including one incomplete response from a participant who neither produces nor analyzes AIRR-seq data. Three respondents participated twice, with consistent answers and same name and contact information, therefore only one completed form was considered for each. Answers from 101 remaining participants originated mainly from North America and Europe (***Figure 1A***) and were further analyzed. At the time of the survey, 96% of respondents were involved in AIRR-seq studies and 4% had plans to perform AIRR-seq studies in the future. Of the respondents, 92% were engaged in human studies, 48% in mouse studies, and 38% in the use of AIRR-seq to study other species or synthetic molecules (e.g., from phage-displayed antibody libraries; ***Figure 1B***). Approximately half of the respondents focused exclusively on IG while a quarter each studied TR and IG or TR alone (***Figure 1C***). Several respondents were interested in many different topics (***Figure 1D*** and ***Figure 1—figure supplement 1***), with their fields of interest dominated by 'immune system diseases' including infection, autoimmune disease, and cancer. Furthermore, the survey results clearly indicate an interest among the majority of respondents in developing bioinformatic tools for the analysis of AIRR-seq data, followed by major interests in other research areas such as vaccinology, immune repertoire homeostasis, immunotherapy, antibody engineering, hematology, and aging (***Figure 1D***). In addition, respondents with bioinformatic skills tended to be those who had broader research interests (***Figure 1—figure supplement 1***). Finally, 89/101 of survey respondents indicated an interest in using AIRR-seq to either track clonotypes over time or across samples, whereas 88/101 were interested in identifying highly expanded clonotypes, 87/101 on analyzing diversity, and 83/101 on studying clonal selection. In conclusion, participants in this survey came from diverse backgrounds, had wet and dry bench expertise, and had a breadth of research interests covering different aspects of AIRR-seq studies.

Survey results on sequencing methodologies used

Whereas 91% (92/101) questionnaire respondents commonly perform bulk sequencing experiments, 67% (68/101) combine bulk sequencing with the use of single-cell technologies. Only 8% of respondents focused exclusively on single-cell sequencing or phage display technologies only. With respect to the input biological material used in bulk sequencing, the majority of participants (83%) preferred to sequence long amplicons that covered the entire (or almost the entire) V-(D)-J region and part of the C region despite the associated higher sequencing cost per read and restrictions on the type of compatible sequencer. AIRR-seq researchers mainly used mRNA for cDNA sequencing (69%), while both mRNA and gDNA were used by 25%, and gDNA alone by 6% of respondents. ***Figure 2A*** shows that the majority of survey respondents performing bulk sequencing used multiplex PCR or the template-switching approach with a considerable number of AIRR-seq researchers using both methods. For those using either approach, mRNA was still the preferred starting material. In addition, UMIs were more commonly used with template switching than multiplex PCR approaches (***Figure 2B***). The association of UMIs with template switching methods is likely related to the

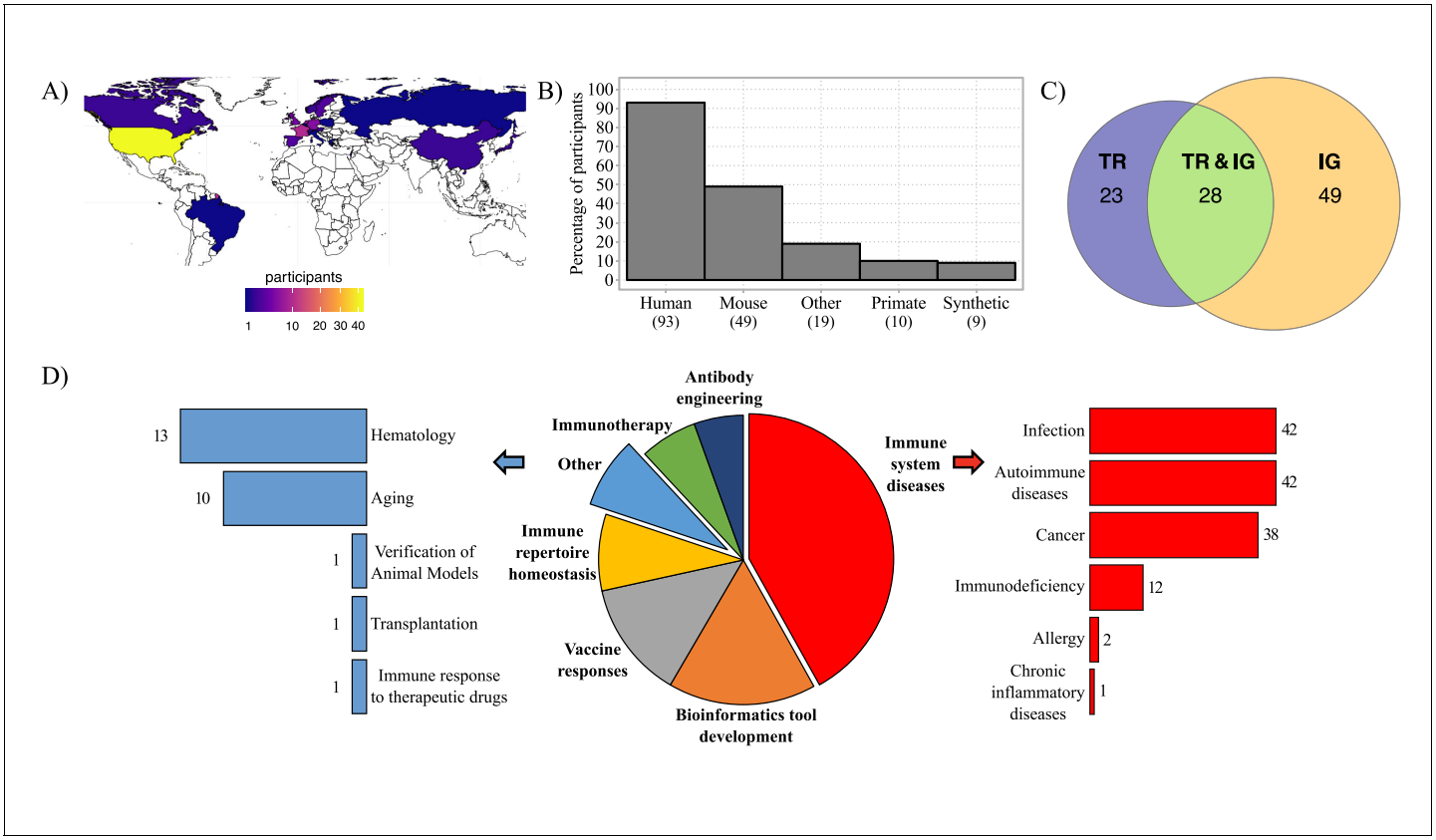


Figure 1. Geographic distribution of survey participants and their AIRR-seq research interests. (**A**) Map with geographic distribution of survey participants. (**B**) Histogram showing the principal studied organisms among the participants. The 'Other' category includes rat, ferret, rabbit, goat, pig, canine, bovis, cattle, chicken, fish, teleost, salmon, zebrafish, other fish species, transgenic animals. (**C**) Venn diagram representing the percentage of participants according to their interest in AIRR template type. (**D**) Pie-chart representing the distribution of survey participants according to their research interest(s). Immune system diseases and other categories are described in more detail in the bar plots (right and left). Numbers of respondents for each category are shown next to the bars.

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

Figure supplement 1. Heatmaps of the areas of study depending on the interest.

template choice, since UMIs are not easy to use with gDNA-based templates, due to the incorporation of the UMI into the amplified products; nevertheless, two participants in the survey reported using UMIs with gDNA (***Figure 2B***).

Altogether, these initial results suggest that among survey respondents bulk sequencing on mRNA and gDNA are the most frequently used methods. However, the literature reflects increasing use of single-cell approaches, including the computational construction of IG or TR from RNA-sequencing (RNA-seq) libraries (***Figure 2—figure supplement 1***) and the combined use of target capture plus single-cell RNA-seq (e.g., 10X Genomics' 5′ end kits, etc.).

To gain insight into which standards should be prioritized for development and sharing with the scientific community, we asked survey participants about the standards they currently use and about the types of standards they would like to use. Based on the results, we concentrate below on potential controls for bulk analysis, as this approach is being used by the majority of the respondents (91%) and because the development of standards for single-cell applications is somewhat distinct.

Survey results on controls used and desired controls

Most respondents (88%) were interested in using standards or controls in AIRR-seq experiments. The 47 respondents who already use controls (n = 47) did so for protocol development (12/47 = 26%), everyday use (7/47 = 15%), or for both (15/47 = 32%), with the remaining (13/47 = 28%) not indicating their specific application. Commercial controls were used by 11, and homebrew

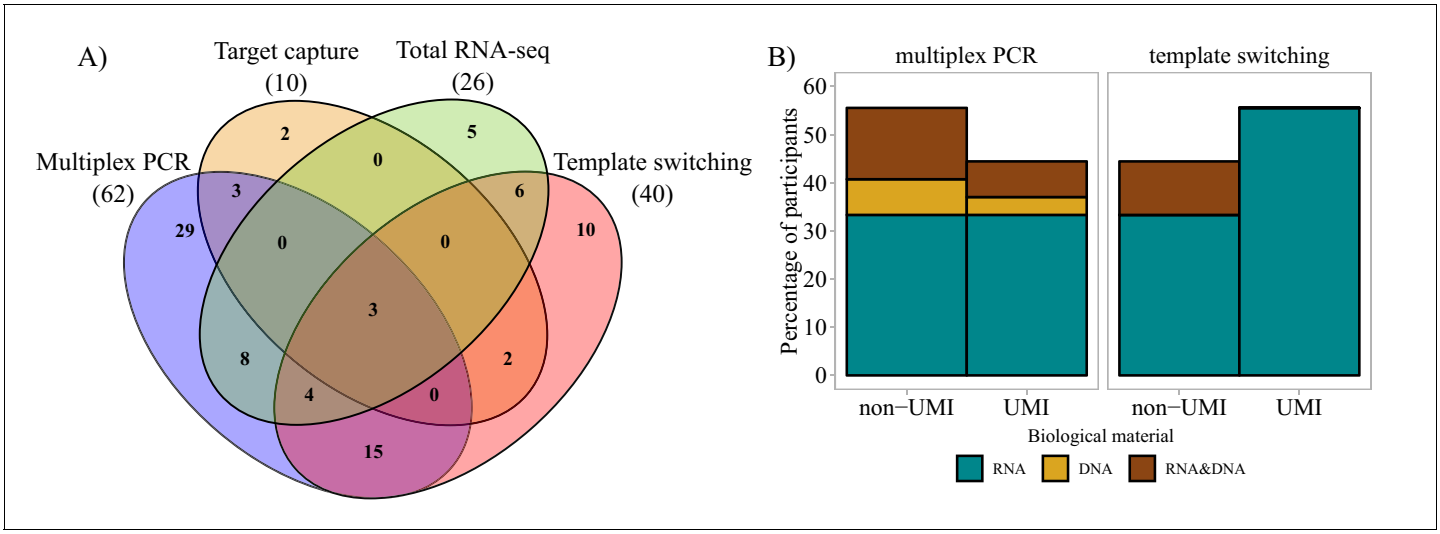


Figure 2. Molecular approaches used in bulk sequencing. (**A**) Venn diagram representing the most important molecular approaches used and their usage and sharing among participants. Numbers of respondents in each of the four main categories are shown in parentheses. (**B**) Bar plots representing biological material used and molecular barcoding proportion for the two major molecular biology approaches (multiplex PCR and template switching). Only the answers of respondents who used one technology exclusively are shown (Multiplex PCR: n = 29; Template switching: n = 10). UMI = unique molecular identifier.

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

Figure supplement 1. Yearly number of PubMed entries referring to single-cell AIRR sequencing (left panel) and bioinformatic AIRR reconstruction from RNA-seq studies (right panel) 1980–2020 (via https://www.ncbi.nlm.nih.gov/pubmed/, accessed on 16 January 2020).

controls were used by 23 participants, with 3 people using both and 16 not specifying their source. ***Figure 3A*** indicates that the most commonly used homebrew controls were cell lines or pooled-cell preparations. Respondents who did not use controls (black bars) were also asked how they might want to use controls. ***Figure 3B*** depicts the community survey responses to these questions.

In summary, the survey, as well as further discussions within the AIRR Biological Resources WG, identified major concerns arising from different steps in the AIRR-seq workflow (***Table 2***, left and middle columns). Additionally, based on these identified issues, the right column of ***Table 2*** describes potential controls to address them. These controls are described in more detail in the subsequent sections of the manuscript.

Current concepts in the use of biological standards for RNA-seq experiments

To prioritize the development of community-wide standards for AIRR data, we turned to examples of community-wide standards in the NGS space. Several such standards have been developed and are actively used, for example, the External RNA Control Consortium (ERCC) spike-in controls and the Microarray Quality Control (MAQC) RNA standards. Both of these standards were generated through broad collaborations of stakeholders including the United States government, industry, and academic researchers (***Reid and External RNA Controls Consortium, 2005***; ***Su et al., 2014***).

The ERCC spike-in controls were designed specifically to be used with RNA-seq experiments, for normalization of expression values during analysis. Two tubes with different compositions of the RNA sequences are commercially available from Thermo Fisher Scientific as Invitrogen ERCC RNA Spike-In Mix (https://perma.cc/WW9Z-D2NY) and ready for use with eukaryotic samples. Each tube contains 92 RNA species with each containing a predefined polyA tail with a different sequence, with their relative abundance covering a ~10^6 fold range, and a limited range of GC content and lengths. The external RNA mixture can be used to normalize relative quantities of transcripts across samples within a single experiment or project, and to optimize protocols for reproducibility and accuracy. The ERCC standards are widely accepted by the transcriptome community, but do suffer from a few limitations: (i) the GC content and range of lengths are not broad - representing the average, but not all possible mRNA moieties; (ii) the polyA region is shorter than endogenous mRNA,

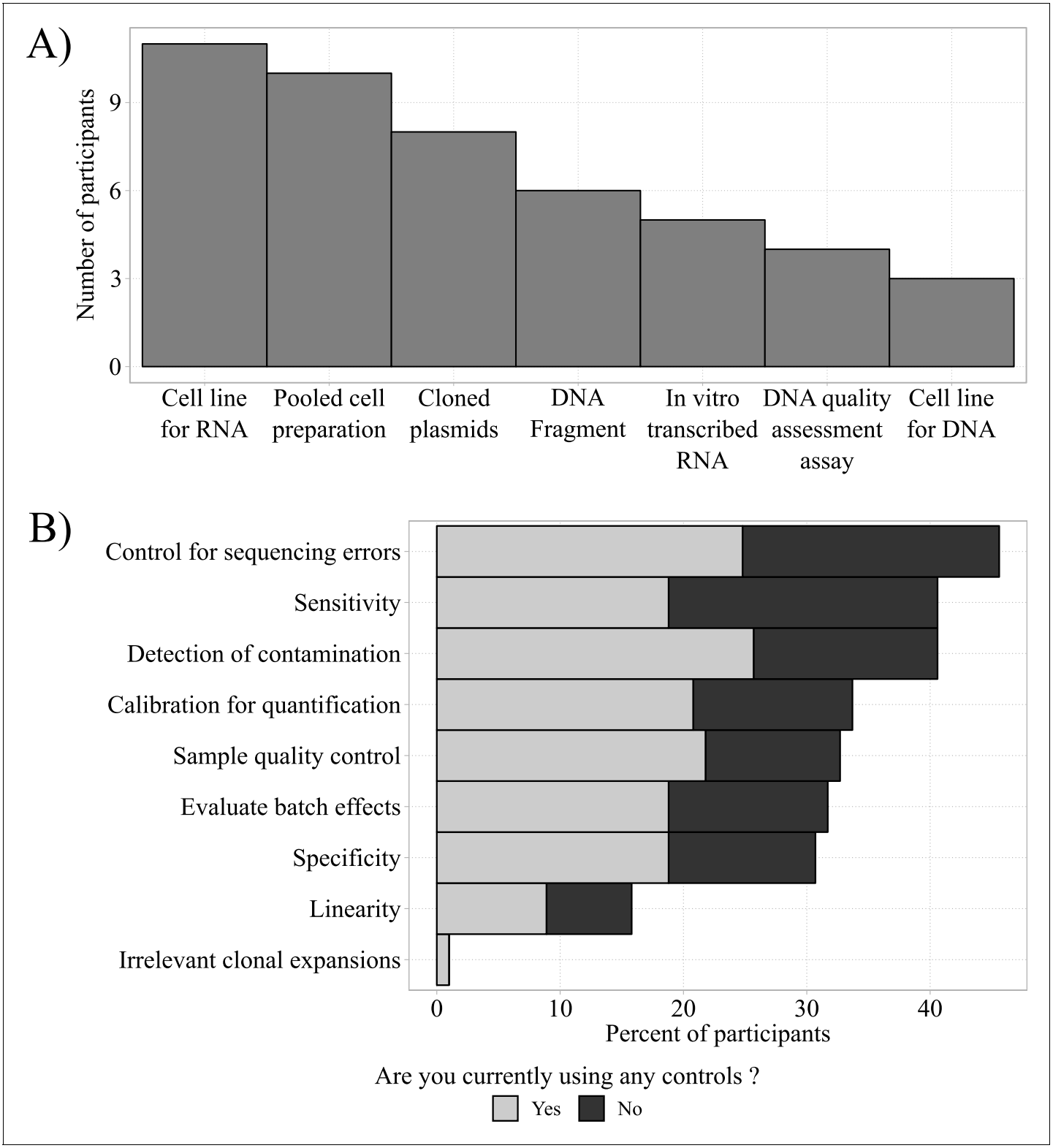


Figure 3. Homebrew controls and their desired applications. (**A**) Most frequently used homebrew controls (total n = 47). (**B**) Total frequencies of desired applications of homebrew controls for respondents currently using (gray bars; total n = 47) and currently not using controls (black bars; total n = 42).

Table 2. Concerns and expected errors introduced during AIRR-seq workflows and possible controls to detect them.
A typical workflow consists of 5 steps: Sample collection > Extraction > Amplification > Sequencing > Analysis.

Concern	Mechanism(s)	Example of potential controls
Sequence errors	Enzyme errors (RT, DNA polymerase); Sequencing errors	UMIs for bioinformatic error correction; Spike-in controls with defined sequences to evaluate error rates
Sensitivity	Enzymatic inefficiencies (RT or PCR conditions/polymerase); Sample collection size (e.g., cell input number, purity); Sequencing depth	Spike-in controls (synthetic or cellular) at known concentrations
Specificity	Enzyme bias (RT, DNA polymerase); Analysis pipelines (annotation, error correction)	Spike-in controls with defined sequences to identify overall V/D/J gene amplification bias
Detection of contamination	Bench-level cross contamination (sample mixing or PCR contamination) or barcode jumping during sequencing	Unique spike-in (synthetic) for each sample; UDIs for sequencing barcode crosstalk
Sample quality control	Sample collection or nucleic acid purification	Identified by spectroscopy or agarose electrophoresis
Evaluate batch effects	Subtle differences introduced at all stages of the workflow	Spike-in controls (synthetic or cellular); Parallel biological (clonal or complex) sample
Linearity/accuracy of clonotype quantification	Enzymatic inefficiencies (RT or PCR conditions); Analytical error correction	Spike-in controls (synthetic or cellular) at known concentrations
Reproducibility/ Batch effects	All stages	Spike-in controls (synthetic or cellular); Parallel biological (clonal or complex) sample; Comparison of replicate amplifications of the same sample; Comparison of sequences generated on the same sample in different sequencing runs
Data processing	Database/annotation limitations; filtering; error correction; collapsing/consensus algorithms	Spike-in controls (synthetic or cellular); Parallel biological (clonal or complex) sample

RT, reverse transcriptase; UMIs, unique molecular identifiers; UDIs, unique dual indices.

and (iii) no splicing variants are included (***Jiang et al., 2011***). These limitations result from the typical compromise accompanying any process for generating useful (and well validated) controls in a timely fashion.

The MAQC RNA standards were not generated as NGS controls initially, but instead were developed by the U.S. Food and Drug Administration (FDA)-led, community-wide consortium for the purpose of validating microarrays, instruments, and analysis methods. The MAQC project consists of four phases, with the first two focusing on microarray methods and the second two on NGS methods (SEQC; Sequencing QC). Initially, differential gene expression levels of nearly 1000 genes between two human reference RNA samples (Human Brain RNA and Human Universal RNA) were assessed by qRT-PCR and microarrays; these highly characterized RNA samples were subsequently used to validate different microarray or NGS library preparation methods, instruments, and data analysis methods (***Shi et al., 2006***). Phase 3 of the MAQC project applied similar concepts to NGS platforms and the comparison of results obtained by microarray or RNA sequencing (***Su et al., 2014***). Phase 4 of this project is ongoing, with the goal of developing robust analysis protocols and providing quality control metrics.

In addition, UMIs and unique dual indices (UDIs) have been proposed in the RNA-seq field (https://perma.cc/AMB4-WC86) (***Kircher et al., 2012***) to control and correct for sequencing errors, as well as sequencing index crosstalk.

Current practices for controls in AIRR-seq experiments

While the standards described above cannot be directly applied to AIRR-seq experimentation, they can serve as a blueprint for the development of standards. In ***Table 2*** (right column), we highlight different possible approaches to address the concerns of the AIRR Community. These approaches closely resemble the general standards described above, using spike-in controls or well-characterized biological samples, including UMIs or UDIs. As described below, AIRR-seq researchers have already initiated some studies to address the use of such controls.

Several groups have recently developed synthetic standards for use with AIRR-seq samples of mouse and human origin, all based on common principles, and generally available for academic

institutions via a material transfer agreement. These synthetic templates are either generated from plasmids (via in vitro transcription followed by RT-PCR) or directly produced as synthetic dsDNAs, with numbers of unique sequences produced ranging from dozens to over 1 million (**Khan et al., 2016**; **Friedensohn et al., 2018**; **Carlson et al., 2013**) (and unpublished work of J. Trück, University Children's Hospital Zurich and C. Tipton, Emory University).

The first use of a synthetic repertoire was reported by **Carlson et al., 2013**. They developed synthetic DNA templates combining 14 and 4 different V and J genes of the TR gamma (TRG) locus, resulting in a total of 56 templates of equal length. All sequences contained three barcodes to unambiguously identify individual V-J combinations. In addition, universal primer sites that were identical in all synthetic templates were added on both ends. This approach allowed identification of amplification bias and optimization of primer concentrations as well as informing the computational correction of residual bias. Work on both murine and human IG repertoire was performed by **Khan et al., 2016** and **Friedensohn et al., 2018** using a similar strategy in the same research laboratory. There, a total of 16 murine and 85 human synthetic sequences were used to assess primer bias from multiplex PCR library generation. In combination with incorporation of UMIs into amplicons and an error correction analysis pipeline, this approach increased workflow fidelity and produced more accurate data. A very important element in the strategy used in this process was the integration of UMI during initial reverse transcription, resulting in labeling of each cDNA on a single molecule level. In contrast to the study by Carlson et al., synthetic sequences used in both studies from Khan et al. and Friedensohn et al. contained different CDR3 sequences and were used in different relative concentrations within the spike-in pool. This approach allowed to not only assess primer-dependent amplification bias but also the impact of variable input concentrations of synthetic sequences on their relative abundance following sequencing.

In principle, synthetic templates are designed such that each mimics an individual recombined V-(D)-J region and a partial C region while also containing universal priming sites, barcodes for unique template identification. The universal priming sites allow for unbiased quantification. Furthermore, comparison of amplification efficiency with the universal vs. targeted primers (the latter usually binding to the leader or V and J or C regions) may be used to correct for target-specific differences in amplification efficiency. Through amplification of synthetic templates alone, multiplex primer sets can be tuned to individual concentrations that will more accurately amplify known targets within repertoires or they can be used to eliminate primers that perform poorly altogether. Through the use of known templated sequences of known abundance (e.g., cell lines spiked into other cells), quantification and amplification efficiency can be calculated. Experience from early testing has identified certain limitations of this approach (see below). Some standards additionally harbor mutations deviating from the germline (unmutated) IGHV sequences; these can be used to model the efficiency of amplification of somatically mutated templates (**Friedensohn et al., 2018**). In practical terms, synthetic standards can be used during method development (primer optimization, alteration of methods to account for amplification bias, etc.), as spike-ins, or can be run as separate positive controls alongside the samples of interest. Synthetic templates can also serve as spike-in controls for concurrent quantification measurements of run-to-run variability, amplification and sequencing efficiency, as a positive control, or as a measurement of sample-to-sample contamination and/or index of misidentification (using different synthetic library spike-ins for individual amplifications in a pooled sequencing run).

Although theoretically promising, the molecular design and bioinformatic analyses of synthetic sequences are challenging. Controls should mimic biological repertoires as closely as possible, and therefore are most effective when they contain a representative level of the biological diversity, which is tedious and expensive. In addition, they should also be distinguishable from other biological sequences so that even following nucleotide changes introduced through PCR or sequencing errors, such synthetic sequences can be unambiguously resolved from their biological counterparts. These challenges may explain the rare usage of spike-in controls except for initial method optimization or very specific applications.

Mixed-cell populations and cell lines can also be used as workflow controls, as has been documented recently by the Euroclonality Consortium (**Knecht et al., 2019**; **Brüggemann et al., 2019**). The first type of Euroclonality (**Knecht et al., 2019**) control monitors general primer and sequencing performance of a sequencing run (batch of samples) and consists of a poly-target control, comprised of gDNA isolated from healthy human thymus, tonsil, and peripheral blood mononuclear cells

(PBMCs), the latter derived from apheresis donors (in a 1:1:1 ratio). This control is included in the workflow alongside experimental samples in a separate tube. By bioinformatic identification of the primer sequences and comparison to stored reference sequencing profiles from the sample control mix, unusual amplification patterns or batch effects can be identified. One advantage of such a cell mixture control is that it fully models the immune repertoire complexity of a bulk cell population and provides quality monitoring for every step of the process including amplification of the template and its sequencing. A second advantage is that this type of control is very easy to generate and is therefore accessible to many laboratories. A disadvantage of using this type of complex control is that one does not fully know the identity of all of the rearrangements in the sample, which can be problematic if there is sample or PCR contamination. A second disadvantage is that it is generated in a finite amount - and once used up, the process of validating the control must be repeated.

The second type of control used by the Euroclonality Consortium (***Brüggemann et al., 2019***) is designed to evaluate assay sensitivity and linearity within each library. This in-parallel control consists of a gDNA sample obtained from 59 human B/T lymphoid cell lines with a total of 46 well-defined rearrangements mixed together in different ratios and added to each processed sample. An advantage of using cell lines as in-parallel controls is that their gene rearrangements are defined and thus easily identified; theoretically allowing for the conversion of reads into cell numbers and permitting relative quantification of template abundance. In practice, however, the use of in-parallel amplification controls can be very challenging, and requires careful interpretation. For example, in samples with poor gDNA quality or low template abundance, the control templates may outcompete the test sample. The depth of sequencing and relative amounts of sample input can affect the measured abundance (***Barennes et al., 2021***; ***Chaara et al., 2018***). An additional disadvantage is that cell lines do not model a fully diverse repertoire, only a fraction of V and J gene combinations are represented by the cell lines, and thus primer performance and bias, especially between samples, are not fully controlled.

Discussion and future work

AIRR-seq experiments are becoming increasingly commonplace, in both the research and clinical settings. In contrast, the development of controls and best practices for assay validation, interpretation and standardization have lagged behind. Here, the members of the AIRR Community Biological Resources WG have summarized the current practice regarding the use of standards and controls among its members as well as among other international AIRR-seq experts and in the literature. We also have identified differences in the types of standards and controls that are used among users. Some of these differences depend on the sample type (fresh vs. fixed cells), the starting material (single cell vs. bulk), the template (mRNA vs. gDNA), as well as the quality and quantity of the relevant cells and templates. In addition, the selection of controls is influenced by the amplification method, with single cell and mRNA-based methods relying more heavily on cellular and molecular barcoding approaches, for example. Last, but certainly not least, the downstream application of the assay can profoundly influence the choice and prioritization of controls. In some cases, assays need to be sensitive and specific (e.g., a clinical grade assay that detects minimal residual disease) whereas in others quantitation (e.g., clonal size analysis for monitoring clonal expansions) or unbiased amplification (e.g., assessment of repertoire skewing during an immune response) may be more important.

All AIRR-seq assays can clearly benefit from rigorous controls. There is broad agreement that controls and standards are desirable, with over half of AIRR-seq survey respondents currently already using controls (mostly of a homebrew variety) in their experiments. Furthermore, whether individuals used controls or did not, they appeared to agree on the types of issues in analyzing and interpreting AIRR-seq data that would benefit from the use of controls - most importantly measuring and controlling for sample quality, assay sensitivity and specificity, and calibration for the quantification of clonal size. Also, with the progress regarding antigen-specificity inference using computational tools (***Glanville et al., 2017***; ***Huang et al., 2020***; ***Jokinen et al., 2021***; ***Akbar et al., 2021***; ***Hayashi et al., 2021***; ***Richardson et al., 2021***; ***Galson et al., 2015***; ***Shomuradova et al., 2020***; ***Chronister et al., 2021***; ***Sidhom et al., 2021***; ***Pogorelyy et al., 2019***; ***Dash et al., 2017***) or more conventionally through technologically challenging using antigen-binding approaches, including single-cell (***Johnson et al., 2020***; ***Fuchs et al., 2019***), having controls would be of major interest to ensure the accuracy of the TR or IG identified. It is unlikely that a single control can fulfill all of these

needs across all methods and applications. There is at the moment also no obvious front-runner for a 'gold standard' that can be used to judge the adequacy of different types of controls.

Having identified a need, a diversity of methodologic approaches and the lack of a 'gold standard' for AIRR-seq controls, the AIRR Community Biological Resources WG is now coordinating the development of such controls. Although AIRR-seq researchers are aware of potential methodological problems, current solutions have not been systematically evaluated or compared. Based upon the current use of controls and needs identified by our survey respondents, the WG plans to focus on three forms of controls: one in-parallel (synthetic standards), one in-parallel and computational (UMIs), and one that is external (a complex cell mixture that is run in parallel to monitor amplification and sequencing run batch effects). To optimize these three types of standards, we first must determine how well they work. We therefore propose to carry out a multi-center analysis of three types of controls: (1) synthetic calibrators for bulk gDNA sequencing to measure clonal size and amplification bias; (2) UMIs for samples studied in parallel by bulk mRNA (with UMIs) and single cell sequencing (with cellular barcodes) for the analysis of amplification bias and sequencing error; and (3) a mixed cell population (either a human apheresis and/or pooled tissue product or murine spleen samples) for the evaluation of batch effects that compare between sequencing runs performed at the same site or between runs performed at different sites. These three types of controls can be run in parallel or in separate, dedicated experiments, allowing for greater participation of AIRR-seq investigators.

In order to perform these studies, the Biological Resources WG will first establish a framework suited to the analysis and quantitation of potential issues, depending on the type and amount of input material, the assay(s) used, and the analysis method(s). Since the method used for the production of AIRR-seq data can impact the results, as shown by the benchmarking of TR library preparation methods study (***Barennes et al., 2021***), we plan to evaluate different standards in the framework of a molecular biology method benchmarking study as well. For the TR repertoire, we will take advantage of the already evaluated methods to include more gDNA-based methods. For the IG repertoire analysis, we will launch a systematic study, leveraging high-volume pooled-cell collections and synthetic standards that can be shared by investigators at multiple sites. We plan four major experiments: (1) analysis of TR and IG rearrangements in PBMCs using bulk gDNA and RNA approaches; (2) analysis of TR on sorted naïve polyclonal T cells using bulk gDNA; (3) analysis of IG rearrangements on sorted naïve polyclonal B cells using bulk gDNA and RNA; and (4) analysis of IG rearrangements on spleen cells from organ donors using bulk gDNA and RNA approaches. Spleen cells are enriched for memory B-cell clones and are useful for modeling clonal expansion and somatic hypermutation (***Meng et al., 2017***). Synthetic controls and UMIs (in the case of RNA-based sequencing) will be added to triplicate samples of PBMCs and polyclonal splenocytes. Samples will be run with and without spike-in controls that will be included at different ratios. Ideally, the selected methods will all be handle by 2 to 3 labs, a compromise between feasibility and inter-lab validation. Protocols and workflows will be standardized and shared. To avoid sequencing batch effect, we will sequence all the replicates through the same facility. Based on the results and to determine the lowest possible cell input levels, we will then evaluate the impact of decreasing the quantity of cells and repeat the same schema, focusing on the most reproducible methods. Finally, we will work closely with other AIRR Community working groups and additional experts in the field to harmonize standardization efforts. Together with the AIRR Community Software WG, we will select a series of tools for data quality control, alignment, and annotation and identify the analysis pipelines required for the detection of contamination, amplification bias, and batch effects. By leveraging the diverse skills of AIRR Community investigators, the development, optimization, and dissemination of biological standards for AIRR-seq data should progress quickly. Such ambitious project will require financial support in order to help volunteer labs to handle the experiments, already under discussion at the level of the AIRR-community.

Using, testing, and comparing these standards is but the first step. Beyond that is their adoption by the wider scientific community. For widespread adoption, commercial and other partnerships are essential for high-quality production and dissemination of the standards. In addition, to ensure broad distribution, controls should either be free of significant intellectual property restrictions or, if proprietary, be well-documented. For this, the Biological Resources WG will reach out to academic research groups (e.g., IMGT), non-profit organizations (e.g., the Global Alliance for Genomics and Health), governmental organizations (e.g., the US Department of Commerce's National Institute of Standards and Technology), as well as commercial entities (e.g., the ATCC). With these efforts, we

will be able to improve the rigor, robustness and interchangeability of AIRR-seq studies, and to increase their utility for downstream applications, including clinical diagnostics.

Additional information

Competing interests

Eline T Luning Prak: is consulting or is an advisor for Roche Diagnostics Corporation, Enpicom, The Antibody Society, The American Autoimmune Related Diseases Association and IEDB. Andrew Farmer: works for Takara Bio USA, but has no ownership or stock in the company. Magnolia Bostick: During the writing of the manuscript, Magnolia Bostick was employed by Takara Bio USA, but has no ownership or stock in the company. The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Swiss National Science Foundation	PZ00P3_161147	Johannes Trück
Swiss National Science Foundation	PZ00P3_183777	Johannes Trück
Agence Nationale de la Recherche	ANR-16-rHUS-0001	Pierre Barennes
Agence Nationale de la Recherche	ANR-11-IDEX-0004-02	Encarnita Mariotti-Ferrandiz
FP7 Ideas: European Research Council	Fp7-IdEAS-ErC-322856	Encarnita Mariotti-Ferrandiz
Agence Nationale de la Recherche	ANR-18-ECVD-0001	Encarnita Mariotti-Ferrandiz
H2020 Fast Track to Innovation	825821	Encarnita Mariotti-Ferrandiz
DFG	FZ 111	Anne Eugster
National Institutes of Health	UM1 AI144288	Eline T Luning Prak
National Institutes of Health	P01 AI106697	Eline T Luning Prak
National Institutes of Health	P30-AI0450080	Eline T Luning Prak
National Institutes of Health	UC4 DK112217	Eline T Luning Prak
National Institutes of Health	P30-CA016520	Eline T Luning Prak

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Johannes Trück, Conceptualization, Formal analysis, Writing - original draft, Writing - review and editing; Anne Eugster, Conceptualization, Writing - original draft, Writing - review and editing; Pierre Barennes, Data curation, Formal analysis, Writing - original draft; Christopher M Tipton, Writing - original draft; Eline T Luning Prak, Conceptualization, Methodology, Writing - original draft, Writing - review and editing; Davide Bagnara, Investigation, Writing - review and editing; Cinque Soto, Writing - original draft, Writing - review and editing; Jacob S Sherkow, Investigation; Aimee S Payne, Marie-Paule Lefranc, Andrew Farmer, Writing - review and editing; The AIRR Community, Validation; Magnolia Bostick, Conceptualization, Supervision, Investigation, Writing - original draft, Project administration, Writing - review and editing; Encarnita Mariotti-Ferrandiz, Conceptualization, Supervision, Methodology, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs
Johannes Trück https://orcid.org/0000-0002-0418-7381
Anne Eugster http://orcid.org/0000-0001-8009-5959
Eline T Luning Prak https://orcid.org/0000-0002-9478-9211
Cinque Soto https://orcid.org/0000-0002-3997-6217
Jacob S Sherkow http://orcid.org/0000-0002-9724-9261
Marie-Paule Lefranc https://orcid.org/0000-0003-0116-9353
Encarnita Mariotti-Ferrandiz https://orcid.org/0000-0002-8770-0186

References

Akbar R, Robert PA, Pavlović M, Jeliazkov JR, Snapkov I, Slabodkin A, Weber CR, Scheffer L, Miho E, Haff IH, Haug DTT, Lund-Johansen F, Safonova Y, Sandve GK, Greiff V. 2021. A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. *Cell Reports* **34**:108856. DOI: https://doi.org/10.1016/j.celrep.2021.108856, PMID: 33730590

Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc M-P. 2012. IMGT/HIGHV-QUEST: the IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Research* **882**:569–604. DOI: https://doi.org/10.1007/978-1-61779-842-9_32

Anand T, Virmani N, Bera BC, Vaid RK, Vashisth M, Bardajatya P, Kumar A, Tripathi BN. 2021. Phage display technique as a tool for diagnosis and antibody selection for coronaviruses. *Current Microbiology* **78**:1124–1134. DOI: https://doi.org/10.1007/s00284-021-02398-9, PMID: 33687511

Arnaout RA, Prak ETL, Schwab N, Rubelt F, Adaptive Immune Receptor Repertoire Community. 2021. The future of blood testing is the immunome. *Frontiers in Immunology* **12**:626793. DOI: https://doi.org/10.3389/fimmu.2021.626793, PMID: 33790897

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. *Nature Biotechnology* **39**:236–245. DOI: https://doi.org/10.1038/s41587-020-0656-3, PMID: 32895550

Bashford-Rogers RJ, Palser AL, Idris SF, Carter L, Epstein M, Callard RE, Douek DC, Vassiliou GS, Follows GA, Hubank M, Kellam P. 2014. Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. *BMC Immunology* **15**:29. DOI: https://doi.org/10.1186/s12865-014-0029-0, PMID: 25189176

Benichou J, Ben-Hamo R, Louzoun Y, Efroni S. 2012. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing: rep-seq: ngs for the immunological repertoire. *Immunology* **135**:183–191. DOI: https://doi.org/10.1111/j.1365-2567.2011.03527.x

Bradley P, Thomas PG. 2019. Using T cell receptor repertoires to understand the principles of adaptive immune recognition. *Annual Review of Immunology* **37**:547–570. DOI: https://doi.org/10.1146/annurev-immunol-042718-041757, PMID: 30699000

Breden F, Luning Prak ET, Peters B, Rubelt F, Schramm CA, Busse CE, Vander Heiden JA, Christley S, Bukhari SAC, Thorogood A, Matsen Iv FA, Wine Y, Laserson U, Klatzmann D, Douek DC, Lefranc MP, Collins AM, Bubela T, Kleinstein SH, Watson CT, et al. 2017. Reproducibility and reuse of adaptive immune receptor repertoire data. *Frontiers in Immunology* **8**:1418. DOI: https://doi.org/10.3389/fimmu.2017.01418, PMID: 29163494

Brüggemann M, Kotrová M, Knecht H, Bartram J, Boudjogrha M, Bystry V, Fazio G, Froňková E, Giraud M, Grioni A, Hancock J, Herrmann D, Jiménez C, Krejci A, Moppett J, Reigl T, Salson M, Scheijen B, Schwarz M, Songia S, et al. 2019. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* **33**:2241–2253. DOI: https://doi.org/10.1038/s41375-019-0496-7, PMID: 31243313

Calis JJ, Rosenberg BR. 2014. Characterizing immune repertoires by high throughput sequencing: strategies and applications. *Trends in Immunology* **35**:581–590. DOI: https://doi.org/10.1016/j.it.2014.09.004, PMID: 25306219

Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, 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. *Nature Communications* **4**:2680. DOI: https://doi.org/10.1038/ncomms3680, PMID: 24157944

Chaara W, Gonzalez-Tort A, Florez LM, Klatzmann D, Mariotti-Ferrandiz E, Six A. 2018. RepSeq data representativeness and robustness assessment by shannon entropy. *Frontiers in Immunology* **9**:1038. DOI: https://doi.org/10.3389/fimmu.2018.01038, PMID: 29868003

Chaudhary N, Wesemann DR. 2018. Analyzing immunoglobulin repertoires. *Frontiers in Immunology* **9**:462. DOI: https://doi.org/10.3389/fimmu.2018.00462, PMID: 29593723

Chiou SH, 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, et al. 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. DOI: https://doi.org/10.1016/j.immuni.2021.02.014, PMID: 33691136

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. *Frontiers in Immunology* **12**:640725. DOI: https://doi.org/10.3389/fimmu.2021.640725, PMID: 33777034

Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, Crawford JC, Clemens EB, Nguyen THO, Kedzierska K, La Gruta NL, Bradley P, Thomas PG. 2017. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**:89–93. DOI: https://doi.org/10.1038/nature22383, PMID: 28636592

Davis MM, Bjorkman PJ. 1988. T-cell antigen receptor genes and T-cell recognition. *Nature* **334**:395–402. DOI: https://doi.org/10.1038/334395a0, PMID: 3043226

DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, Varadarajan N, Giesecke C, Dörner T, Andrews SF, Wilson PC, Hunicke-Smith SP, Willson CG, Ellington AD, Georgiou G. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature Biotechnology* **31**:166–169. DOI: https://doi.org/10.1038/nbt.2492, PMID: 23334449

Dines JN. 2020. The ImmuneRACE study: a prospective multicohort study of immune response action to COVID-19 events with the ImmuneCODETM open access database. *medRxiv*. DOI: https://doi.org/10.1101/2020.08.17.20175158

Dziubianau M, Hecht J, Kuchenbecker L, Sattler A, Stervbo U, Rödelsperger C, Nickel P, Neumann AU, Robinson PN, Mundlos S, Volk H-D, Thiel A, Reinke P, Babel N. 2013. TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T Cell-Related pathology. *American Journal of Transplantation* **13**:2842–2854. DOI: https://doi.org/10.1111/ajt.12431

Elhanati Y, Sethna Z, Marcou Q, Callan CG, Mora T, Walczak AM. 2015. Inferring processes underlying B-cell repertoire diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**:20140243. DOI: https://doi.org/10.1098/rstb.2014.0243

Ellefson JW, Gollihar J, Shroff R, Shivram H, Iyer VR, Ellington AD. 2016. Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science* **352**:1590–1593. DOI: https://doi.org/10.1126/science.aaf5409, PMID: 27339990

Eugster A, Lindner A, Heninger A-K, Wilhelm C, Dietz S, Catani M, Ziegler A-G, Bonifacio E. 2013. Measuring T cell receptor and T cell gene expression diversity in antigen-responsive human CD4+ T cells. *Journal of Immunological Methods* **400-401**:13–22. DOI: https://doi.org/10.1016/j.jim.2013.11.003

Fink K. 2019. Can we improve vaccine efficacy by targeting T and B cell repertoire convergence? *Frontiers in Immunology* **10**:110. DOI: https://doi.org/10.3389/fimmu.2019.00110, PMID: 30814993

Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA. 2009. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Research* **19**:1817–1824. DOI: https://doi.org/10.1101/gr.092924.109, PMID: 19541912

Friedensohn S, Lindner JM, Cornacchione V, Iazeolla M, Miho E, Zingg A, Meng S, Traggiai E, Reddy ST. 2018. 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. *Frontiers in Immunology* **9**:1401. DOI: https://doi.org/10.3389/fimmu.2018.01401, PMID: 29973938

Fuchs YF, Sharma V, Eugster A, Kraus G, Morgenstern R, Dahl A, Reinhardt S, Petzold A, Lindner A, Löbel D, Bonifacio E. 2019. Gene Expression-Based identification of Antigen-Responsive $CD8^+$ T cells on a Single-Cell level. *Frontiers in Immunology* **10**:2568. DOI: https://doi.org/10.3389/fimmu.2019.02568, PMID: 31781096

Gadala-Maria D, Yaari G, Uduman M, Kleinstein SH. 2015. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *PNAS* **112**:E862–E870. DOI: https://doi.org/10.1073/pnas.1417683112, PMID: 25675496

Galson JD, Kelly DF, Truck J. 2015. Identification of Antigen-Specific B-Cell receptor sequences from the total B-Cell repertoire. *Critical Reviews in Immunology* **35**:463–478. DOI: https://doi.org/10.1615/CritRevImmunol.2016016462, PMID: 27279044

Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. 2014. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology* **32**:158–168. DOI: https://doi.org/10.1038/nbt.2782, PMID: 24441474

Ghraichy M, Galson JD, Kelly DF, Trück J. 2018. B-cell receptor repertoire sequencing in patients with primary immunodeficiency: a review. *Immunology* **153**:145–160. DOI: https://doi.org/10.1111/imm.12865, PMID: 29140551

Gittelman RM. 2021. Diagnosis and tracking of SARS-CoV-2 infection by T-Cell receptor sequencing. *medRxiv*. DOI: https://doi.org/10.1101/2020.11.09.20228023

Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L, Sundar PD, Day GM, Cox D, Rajpal A, Pons J. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *PNAS* **106**:20216–20221. DOI: https://doi.org/10.1073/pnas.0909775106, PMID: 19875695

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: https://doi.org/10.1038/nature22976, PMID: 28636589

Hayashi F, Isobe N, Glanville J, Matsushita T, Maimaitijiang G, Fukumoto S, Watanabe M, Masaki K, Kira JI. 2021. A new clustering method identifies multiple sclerosis-specific T-cell receptors. *Annals of Clinical and Translational Neurology* **8**:163–176. DOI: https://doi.org/10.1002/acn3.51264, PMID: 33400858

Heather JM, Best K, Oakes T, Gray ER, Roe JK, Thomas N, Friedman N, Noursadeghi M, Chain B. 2015. Dynamic perturbations of the T-Cell receptor repertoire in chronic HIV infection and following antiretroviral therapy. *Frontiers in Immunology* **6**:644. DOI: https://doi.org/10.3389/fimmu.2015.00644, PMID: 26793190

Holland J, Spindler K, Horodyski F, Grabau E, Nichol S, VandePol S. 1982. Rapid evolution of RNA genomes. *Science* **215**:1577–1585. DOI: https://doi.org/10.1126/science.7041255, PMID: 7041255

Hou D, Chen C, Seely EJ, Chen S, Song Y. 2016. High-Throughput Sequencing-Based immune repertoire study during infectious disease. *Frontiers in Immunology* **7**:336. DOI: https://doi.org/10.3389/fimmu.2016.00336, PMID: 27630639

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: https://doi.org/10.1038/s41587-020-0505-4, PMID: 32341563

Jacobsen LM, Posgai A, Seay HR, Haller MJ, Brusko TM. 2017. T cell receptor profiling in type 1 diabetes. *Current Diabetes Reports* **17**:118. DOI: https://doi.org/10.1007/s11892-017-0946-4, PMID: 29022222

Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, Gingeras TR, Oliver B. 2011. Synthetic spike-in standards for RNA-seq experiments. *Genome Research* **21**:1543–1551. DOI: https://doi.org/10.1101/gr.121095.111, PMID: 21816910

Jiang N, Schonnesen AA, Ma KY. 2019. Ushering in integrated T cell repertoire profiling in Cancer. *Trends in Cancer* **5**:85–94. DOI: https://doi.org/10.1016/j.trecan.2018.11.005, PMID: 30755308

Johnson JL, Rosenthal RL, Knox JJ, Myles A, Naradikian MS, Madej J, Kostiv M, Rosenfeld AM, Meng W, Christensen SR, Hensley SE, Yewdell J, Canaday DH, Zhu J, McDermott AB, Dori Y, Itkin M, Wherry EJ, Pardi N, Weissman D, et al. 2020. The transcription factor T-bet resolves memory B cell subsets with distinct tissue distributions and antibody specificities in mice and humans. *Immunity* **52**:842–855. DOI: https://doi.org/10.1016/j.immuni.2020.03.020, PMID: 32353250

Jokinen E, Huuhtanen J, Mustjoki S, Heinonen M, Lähdesmäki H. 2021. Predicting recognition between T cell receptors and epitopes with TCRGP. *PLOS Computational Biology* **17**:e1008814. DOI: https://doi.org/10.1371/journal.pcbi.1008814, PMID: 33764977

Khan TA, Friedensohn S, Gorter de Vries AR, Straszewski J, Ruscheweyh HJ, Reddy ST. 2016. Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Science Advances* **2**:e1501371. DOI: https://doi.org/10.1126/sciadv.1501371, PMID: 26998518

Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* **40**:e3. DOI: https://doi.org/10.1093/nar/gkr771, PMID: 22021376

Kirsch IR, Watanabe R, O'Malley JT, Williamson DW, Scott L-L, Elco CP, Teague JE, Gehad A, Lowry EL, LeBoeuf NR, Krueger JG, Robins HS, Kupper TS, Clark RA. 2015. TCR sequencing facilitates diagnosis and identifies mature T cells as the cell of origin in CTCL. *Science Translational Medicine* **7**:308ra158. DOI: https://doi.org/10.1126/scitranslmed.aaa9122

Knecht H, Reigl T, Kotrová M, Appelt F, Stewart P, Bystry V, Krejci A, Grioni A, Pal K, Stranska K, Plevova K, Rijntjes J, Songia S, Svatoň M, Froňková E, Bartram J, Scheijen B, Herrmann D, García-Sanz R, Hancock J, et al. 2019. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. *Leukemia* **33**:2254–2265. DOI: https://doi.org/10.1038/s41375-019-0499-4, PMID: 31227779

Lefranc MP. 2014. Immunoglobulin and T cell receptor genes: IMGT and the birth and rise of immunoinformatics. *Frontiers in Immunology* **5**:22. DOI: https://doi.org/10.3389/fimmu.2014.00022, PMID: 24600447

Lefranc M-P, Lefranc G. 2001. *The Immunoglobulin FactsBook*. Elsevier.

Lefranc M-P, Lefranc G. 2020. Immunoglobulins or antibodies: IMGT bridging genes, structures and functions. *Biomedicines* **8**:319. DOI: https://doi.org/10.3390/biomedicines8090319

Li S, Lefranc MP, Miles JJ, Alamyar E, Giudicelli V, Duroux P, Freeman JD, Corbin VD, Scheerlinck JP, Frohman MA, Cameron PU, Plebanski M, Loveland B, Burrows SR, Papenfuss AT, Gowans EJ. 2013. IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. *Nature Communications* **4**:2333. DOI: https://doi.org/10.1038/ncomms3333, PMID: 23995877

Linette GP, Becker-Hapak M, Skidmore ZL, Baroja ML, Xu C, Hundal J, Spencer DH, Fu W, Cummins C, Robnett M, Kaabinejadian S, Hildebrand WH, Magrini V, Demeter R, Krupnick AS, Griffith OL, Griffith M, Mardis ER, Carreno BM. 2019. Immunological ignorance is an enabling feature of the oligo-clonal T cell response to melanoma neoantigens. *PNAS* **116**:23662–23670. DOI: https://doi.org/10.1073/pnas.1906026116, PMID: 31685621

Link-Rachner CS, Eugster A, Rücker-Braun E, Heidenreich F, Oelschlägel U, Dahl A, Klesse C, Kuhn M, Middeke JM, Bornhäuser M, Bonifacio E, Schetelig J. 2019. T-cell receptor-α repertoire of CD8+ T cells following allogeneic stem cell transplantation using next-generation sequencing. *Haematologica* **104**:622–631. DOI: https://doi.org/10.3324/haematol.2018.199802, PMID: 30262565

Liu X, Zhang W, Zeng X, Zhang R, Du Y, Hong X, Cao H, Su Z, Wang C, Wu J, Nie C, Xu X, Kristiansen K. 2016. Systematic comparative evaluation of methods for investigating the tcrβ repertoire. *PLOS ONE* **11**:e0152464. DOI: https://doi.org/10.1371/journal.pone.0152464, PMID: 27019362

Lu YC, Zheng Z, Robbins PF, Tran E, Prickett TD, Gartner JJ, Li YF, Ray S, Franco Z, Bliskovsky V, Fitzgerald PC, Rosenberg SA. 2018. An efficient Single-Cell RNA-Seq approach to identify Neoantigen-Specific T cell receptors. *Molecular Therapy : The Journal of the American Society of Gene Therapy* **26**:379–389. DOI: https://doi.org/10.1016/j.ymthe.2017.10.018, PMID: 29174843

Mamedov IZ, Britanova OV, Zvyagin IV, Turchaninova MA, Bolotin DA, Putintseva EV, Lebedev YB, Chudakov DM. 2013. Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling. *Frontiers in Immunology* **4**:456. DOI: https://doi.org/10.3389/fimmu.2013.00456, PMID: 24391640

Meng W, Zhang B, Schwartz GW, Rosenfeld AM, Ren D, Thome JJC, Carpenter DJ, Matsuoka N, Lerner H, Friedman AL, Granot T, Farber DL, Shlomchik MJ, Hershberg U, Luning Prak ET. 2017. An atlas of B-cell clonal distribution in the human body. *Nature Biotechnology* **35**:879–884. DOI: https://doi.org/10.1038/nbt.3942, PMID: 28829438

Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. 2018. Computational strategies for dissecting the High-Dimensional complexity of adaptive immune repertoires. *Frontiers in Immunology* **9**:224. DOI: https://doi.org/10.3389/fimmu.2018.00224, PMID: 29515569

Minervina A, Pogorelyy M, Mamedov I. 2019. T-cell receptor and B-cell receptor repertoire profiling in adaptive immunity. *Transplant International* **32**:1111–1123. DOI: https://doi.org/10.1111/tri.13475, PMID: 31250479

Papavasiliou FN, Schatz DG. 2002. Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. *Cell* **109 Suppl**:S35–S44. DOI: https://doi.org/10.1016/s0092-8674(02)00706-7, PMID: 11983151

Parola C, Neumeier D, Reddy ST. 2018. Integrating high-throughput screening and sequencing for monoclonal antibody discovery and engineering. *Immunology* **153**:31–41. DOI: https://doi.org/10.1111/imm.12838, PMID: 28898398

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

Primi D, Barbier E, Cazenave PA. 1986. Structural polymorphism of V kappa 21 E and V kappa 21 D gene products in laboratory mice. *European Journal of Immunology* **16**:292–296. DOI: https://doi.org/10.1002/eji.1830160315, PMID: 3082651

Ravn U, Didelot G, Venet S, Ng KT, Gueneau F, Rousseau F, Calloud S, Kosco-Vilbois M, Fischer N. 2013. Deep sequencing of phage display libraries to support antibody discovery. *Methods* **60**:99–110. DOI: https://doi.org/10.1016/j.ymeth.2013.03.001, PMID: 23500657

Reid L, External RNA Controls Consortium. 2005. Proposed methods for testing and selecting the ERCC external RNA controls. *BMC Genomics* **6**:150. DOI: https://doi.org/10.1186/1471-2164-6-150, PMID: 16266432

Richardson E, Galson JD, Kellam P, Kelly DF, Smith SE, Palser A, Watson S, Deane CM. 2021. A computational method for immune repertoire mining that identifies novel binders from different clonotypes, demonstrated by identifying anti-pertussis toxoid antibodies. *mAbs* **13**:1869406. DOI: https://doi.org/10.1080/19420862.2020.1869406, PMID: 33427589

Richters MM, Xia H, Campbell KM, Gillanders WE, Griffith OL, Griffith M. 2019. Best practices for bioinformatic characterization of neoantigens for clinical utility. *Genome Medicine* **11**:56. DOI: https://doi.org/10.1186/s13073-019-0666-2, PMID: 31462330

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 beta-chain diversity in Alphabeta T cells. *Blood* **114**:4099–4107. DOI: https://doi.org/10.1182/blood-2009-04-217604, PMID: 19706884

Robinson WH. 2015. Sequencing the functional antibody repertoire–diagnostic and therapeutic discovery. *Nature Reviews Rheumatology* **11**:171–182. DOI: https://doi.org/10.1038/nrrheum.2014.220, PMID: 25536486

Rosati E, Dowds CM, Liaskou E, Henriksen EKK, Karlsen TH, Franke A. 2017. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnology* **17**:61. DOI: https://doi.org/10.1186/s12896-017-0379-9, PMID: 28693542

Rosenfeld AM, Meng W, Chen DY, Zhang B, Granot T, Farber DL, Hershberg U, Luning Prak ET. 2018. Computational evaluation of B-Cell clone sizes in bulk populations. *Frontiers in Immunology* **9**:1472. DOI: https://doi.org/10.3389/fimmu.2018.01472, PMID: 30008715

Rossjohn J, Gras S, Miles JJ, Turner SJ, Godfrey DI, McCluskey J. 2015. T cell antigen receptor recognition of antigen-presenting molecules. *Annual Review of Immunology* **33**:169–200. DOI: https://doi.org/10.1146/annurev-immunol-032414-112334, PMID: 25493333

Rouet R, Jackson KJL, Langley DB, Christ D. 2018. Next-Generation sequencing of antibody display repertoires. *Frontiers in Immunology* **9**:118. DOI: https://doi.org/10.3389/fimmu.2018.00118, PMID: 29472918

Rubelt F, Busse CE, Bukhari SAC, Bürckert JP, Mariotti-Ferrandiz E, Cowell LG, Watson CT, Marthandan N, Faison WJ, Hershberg U, Laserson U, Corrie BD, Davis MM, Peters B, Lefranc MP, Scott JK, Breden F, Luning Prak ET, Kleinstein SH, AIRR Community. 2017. Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data. *Nature Immunology* **18**:1274–1278. DOI: https://doi.org/10.1038/ni.3873, PMID: 29144493

Sheih A, Voillet V, Hanafi LA, DeBerg HA, Yajima M, Hawkins R, Gersuk V, Riddell SR, Maloney DG, Wohlfahrt ME, Pande D, Enstrom MR, Kiem HP, Adair JE, Gottardo R, Linsley PS, Turtle CJ. 2020. Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy. *Nature Communications* **11**:219. DOI: https://doi.org/10.1038/s41467-019-13880-1, PMID: 31924795

Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, et al. 2006. The MicroArray quality control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology* **24**:1151–1161. DOI: https://doi.org/10.1038/nbt1239, PMID: 16964229

Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, Peshkova IO, Khmelevskaya A, Dianov DV, Malasheva M, Shmelev A, Serdyuk Y, Bagaev DV, Pivnyuk A, Shcherbinin DS, Maleeva AV, Shakirova NT, Pilunov A, Malko DB, Khamaganova EG, et al. 2020. SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. *Immunity* **53**:1245–1257. DOI: https://doi.org/10.1016/j.immuni.2020.11.004, PMID: 33326767

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

Sidhom J-W, Larman HB, Pardoll DM, Baras AS. 2018. DeepTCR: a deep learning framework for revealing structural concepts within TCR repertoire. *bioRxiv*. DOI: https://doi.org/10.1101/464107

Sidhom JW, Larman HB, Pardoll DM, Baras AS. 2021. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nature Communications* **12**:1605. DOI: https://doi.org/10.1038/s41467-021-21879-w, PMID: 33707415

Six A, Mariotti-Ferrandiz ME, Chaara W, Magadan S, Pham HP, Lefranc MP, 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. *Frontiers in Immunology* **4**:413. DOI: https://doi.org/10.3389/fimmu.2013.00413, PMID: 24348479

Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. 2017. Single-cell transcriptomics to explore the immune system in health and disease. *Science* **358**:58–63. DOI: https://doi.org/10.1126/science.aan6828, PMID: 28983043

Su Z, Labaj P, Li S, SEQC/MAQC-III Consortium. 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the sequencing quality control consortium. *Nature Biotechnology* **32**:903–914. DOI: https://doi.org/10.1038/nbt.2957, PMID: 25150838

Theil A. 2017. T cell receptor repertoires after adoptive transfer of expanded allogeneic regulatory T cells: t cell receptor repertoires post-T_{reg} cell therapy. *Clinical and Experimental Immunology* **187**:316–324. DOI: https://doi.org/10.1111/cei.12887

Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* **302**:575–581. DOI: https://doi.org/10.1038/302575a0, PMID: 6300689

Vander Heiden JA, Marquez S, Marthandan N, Bukhari SAC, Busse CE, Corrie B, Hershberg U, Kleinstein SH, Matsen Iv FA, Ralph DK, Rosenfeld AM, Schramm CA, Christley S, Laserson U, AIRR Community. 2018. AIRR community standardized representations for annotated immune repertoires. *Frontiers in Immunology* **9**:2206. DOI: https://doi.org/10.3389/fimmu.2018.02206, PMID: 30323809

Wang C, Sanders CM, Yang Q, Schroeder HW, Wang E, Babrzadeh F, Gharizadeh B, Myers RM, Hudson JR, Davis RW, Han J. 2010. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *PNAS* **107**:1518–1523. DOI: https://doi.org/10.1073/pnas.0913939107, PMID: 20080641

Weinstein JA, Jiang N, White RA, Fisher DS, Quake SR. 2009. High-throughput sequencing of the zebrafish antibody repertoire. *Science* **324**:807–810. DOI: https://doi.org/10.1126/science.1170020, PMID: 19423829

Zhang S-Q, Ma K-Y, Schonnesen AA, Zhang M, He C, Sun E, Williams CM, Jia W, Jiang N. 2018. High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nature Biotechnology* **36**:1156–1159. DOI: https://doi.org/10.1038/nbt.4282

Zvyagin IV, Tsvetkov VO, Chudakov DM, Shugay M. 2020. An overview of immunoinformatics approaches and databases linking T cell receptor repertoires to their antigen specificity. *Immunogenetics* **72**:77–84. DOI: https://doi.org/10.1007/s00251-019-01139-4, PMID: 31741011

Annexe 4.

Liste complète des travaux (par ordre chronologique)

Ritvo, P.-G., Saadawi, A., Barennes, P., et al. High-resolution repertoire analysis reveals a major bystander activation of Tfh and Tfr cells. (Proc. Natl. Acad. Sci., 2018)

Pouzolles, M., Machado, A.*, Guilbaud, M.*, Irla, M., Gailhac, S., Barennes, P., et al. Intrathymic adeno-associated virus gene transfer rapidly restores thymic function and long-term persistence of gene-corrected T cells. (J. Allergy Clin. Immunol., 2019)

Barennes, P., et al. Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases. (Nat. Biotechnol., 2021)

Mhanna, V., Fourcade, G., Barennes, P., et al. Impaired Activated/Memory Regulatory T Cell Clonal Expansion Instigates Diabetes in NOD Mice. (Diabetes, 2021)

Trück, J.*, Eugster, A.*, Barennes, P.* et al. Biological controls for standardization and interpretation of Adaptive Immune Receptor Repertoire profiling experiments: needs and challenges. (eLife science, 2021)

Quiniou, V., Barennes, P. et al. Human thymopoiesis produces polyspecific CD8+ α/β T cells responding to multiple viral antigens. (En révision dans eLife science)

Florez, L. M., Fourcade, G., Barennes, P.* et al. Uterine tissue-resident Tregs have trophic functions supporting fetus growth. (En préparation)

Barennes, P.*, Stys, P.* et al. Autoimmune disease-specific TCR repertoires in peripheral blood T cells. (Soumis dans Nature Immunology)

Mhanna V.*, Barennes, P.* et al. Pooling of cell subsets for TCR repertoire analysis reveals a highly private repertoire of activated regulatory T cells which has an impact on the interpretation of the results. (Soumis dans Communications biology)

Résumé

Signatures TCR des réponses immunes dans les pathologies auto-immunes et les infections

Les lymphocytes T représentent les cellules immunitaires centrale du système immunitaire adaptatif se distinguant en deux populations majeures : les cellules T effectrices (Teffs), auxillaires ou cytotoxiques, ayant un rôle pro-inflammatoire, contribuant à l'élimination des pathogènes en contexte infectieux et responsables des atteintes tissulaires dans le contexte des maladies auto-immunes (MAI) ; les cellules T régulatrices (Tregs) ayant un rôle anti-inflammatoire sur la réponse Teffs, permettant le contrôle des désordres auto-immuns. Teffs et Tregs se caractérisent par la reconnaissance spécifique d'antigène, nécessaire à l'activation cellulaire, et portée par le récepteur des lymphocytes T, autrement appelé T-cell receptor (TCR). Chaque TCR résulte d'un mécanisme de recombinaison génique générant une collection très diverse de TCR, communément appelé répertoire. Véritable reflet de l'histoire immunitaire de chaque individu, l'implication directe de la reconnaissance par le TCR dans les fonctions antagonistes des Tregs et des Teffs, à la fois dans la genèse mais aussi dans la progression des pathologies immunes, soulève la nécessité d'analyser distinctement leurs répertoires TCR. Dans le cadre de ma thèse, j'avais pour objectif d'évaluer le caractère prédictif du répertoire TCR au regard de l'état physiopathologique et de son évolution de patients atteints de MAI, pour lesquelles les outils diagnostiques et/ou pronostiques demeurent limités ainsi que de patients atteints de la COVID-19, maladie infectieuse émergente caractérisée par une forte mortalité des individus immunodéprimés et touchant gravement également des individus considérés sains. Pour cela, j'ai fait appel aux technologies de séquençage à haut-débit offrant aujourd'hui précision et profondeur nécessaire à l'étude de cet objet biologique complexe. Dans ce contexte, la fiabilité des protocoles de biologie moléculaire ainsi que les méthodes bio-informatiques et de modélisations représentent un défi majeur. J'ai dans un premier temps comparé neufs protocoles publiés et/ou commercialisés parmi les plus utilisés dans la littérature. A l'aide d'analyses de la diversité et de modélisation statistique, j'ai élaboré un score de performance, à partir duquel j'ai mis en évidence les forces et faiblesses de chaque protocole. J'ai ainsi identifié la méthode la plus fiable, robuste et reproductible. Avec ce protocole, j'ai ensuite analysé les répertoires TCR des Teffs auxillaires et des Tregs issus de prélèvements sanguins de patients atteints de diabète de type 1 et de polyarthrite rhumatoïde, deux MAI. A l'aide d'outils de modélisation mathématiques et d'apprentissage machine, j'ai développé une stratégie d'identification de TCR discriminants des groupes d'individus. Cette méthode innovante a permis d'identifier des signatures TCR des différents MAI étudiées ainsi que des Teffs et des Tregs, soutenant l'idée que la composition du répertoire TCR signe l'état physiopathologique, y compris dans des pathologies chroniques établies. Appliquée au répertoire TCR de patients atteints de lupus érythémateux systémique sous traitement à l'interleukine-2 à faible dose, j'ai identifié des signatures TCR des Teffs et des Tregs définissant les patients traités répondeurs, pouvant être utilisées comme potentiels biomarqueurs prédictifs de l'efficacité du traitement. Finalement, j'ai également appliqué cette approche sur des répertoires de patients atteints de Sars-CoV-2 pour lesquels l'issue clinique était connue. J'ai ainsi identifié une signature TCR pronostique de la guérison et discriminant les patients décédés. Ensemble, ces résultats renforcent l'intérêt de l'étude du répertoire TCR en montrant que celui-ci contient une information clinique fine, capable non seulement de séparer des patients atteints de différentes maladies mais aussi de distinguer, de façon précoce, la réponse au traitement ainsi que le niveau de sévérité face à une infection.

Mots-clés : Immunité adaptative, Répertoire TCR, Biomarqueurs, Maladies auto-immunes, Sars-CoV-2

Abstract

TCR signatures of immune responses in autoimmune diseases and infections

T cells represent the central immune cells of the adaptive immune system and are distinguished in two major populations. On one hand, effector T cells (Teffs), helper or cytotoxic, have a pro-inflammatory role and contribute to the elimination of pathogens in the context of infectious diseases as they can be responsible for tissue damage in the context of autoimmune diseases (AD). On the other hand, regulatory T cells (Tregs) have an anti-inflammatory role on the Teffs response, allowing the control of autoimmune disorders. Teffs and Tregs are characterized by antigen-specific recognition, a prerequisite for T cell activation, carried by the T-cell receptor (TCR). Each TCR results from a gene recombination mechanism generating a highly diverse collection of TCRs, commonly called repertoire. The direct implication of the TCR recognition in the antagonistic functions of Tregs and Teffs, both in the genesis and progression of immune diseases, raises the need to analyze their TCR repertoires separately. In the framework of my thesis, I aimed to evaluate the predictive character of the TCR repertoire with respect to the pathophysiological state and its evolution of patients. I focused on patients suffering from AD, for which diagnostic and/or prognostic tools remain limited, as well as of patients suffering from COVID-19, an emerging infectious disease characterized by a high mortality of immunocompromised individuals and also seriously affecting individuals considered healthy. For this purpose, I have used high-throughput sequencing technologies that now offer the precision and depth necessary to study this complex biological object. In this context, the reliability of molecular biology protocols as well as bioinformatics and modelling methods represent a major challenge. I first compared nine published and/or commercialized protocols among the most used in the literature. Using diversity analysis and statistical modeling, I developed a performance score, from which I highlighted the strengths and weaknesses of each protocol. I thus identified the most reliable, robust and reproducible method. With this protocol, I then analyzed the TCR repertoires of helper Teffs and Tregs from blood samples of patients with type 1 diabetes and rheumatoid arthritis, both ADs. Using mathematical modeling and machine learning tools, I developed a strategy to identify TCRs discriminating groups of individuals. This innovative method allowed identifying TCR signatures of the different AD studied as well as Teffs and Tregs, supporting the idea that the composition of the TCR repertoire signs the pathophysiological state, including in established chronic pathologies. Applied to the TCR repertoire of patients with systemic lupus erythematosus under low-dose interleukin-2 treatment, I identified TCR signatures of Teffs and Tregs defining responder treated patients, which can be used as potential predictive biomarkers of treatment efficacy. Finally, I also applied this approach on repertoires of Sars-CoV-2 patients for whom the clinical outcome was known. I thus identified a TCR signature that was prognostic of cure and discriminated against patients who died. Altogether, my work supports the interest of the TCR repertoire study by showing that it contains fine clinical information, capable not only of separating patients with different diseases but also of distinguishing, in an early stage, the response to treatment as well as the level of severity of an infection.

Keywords : Adaptive immunity, TCR repertoire, Biomarkers, Autoimmunes diseases, Sars-CoV-2