Université de Montréal

Développement et application de méthodes pour l'analyse virologique et immunologique de l'infection au virus de l'immunodéficience humaine

Par

Audrée Lemieux

Département de microbiologie, infectiologie et immunologie

Faculté de médecine

Mémoire présenté en vue de l'obtention du grade de maîtrise (M. Sc.) en microbiologie et immunologie

Avril 2024

Université de Montréal

Département de microbiologie, infectiologie et immunologie, Faculté de médecine

Ce mémoire intitulé

Développement et application de méthodes pour l'analyse virologique et immunologique de l'infection au virus de l'immunodéficience humaine

Présenté par

Audrée Lemieux

A été évaluée par un jury composé des personnes suivantes

Dre Nathalie Labrecque

Présidente-rapporteuse

Dr. Daniel Kaufmann

Directeur de recherche

Dr. Nicolas Chomont

Codirecteur

Dr. Matthieu Perreau

Membre du jury

Résumé

Dans les dernières années, il a été montré que les provirus intacts et défectueux du virus de l'immunodéficience humaine (VIH) codent pour des protéines virales qui contribuent à l'activation chronique des lymphocytes T spécifiques au VIH, et ce même sous traitement antirétroviral (ART). Nous avons développé deux méthodes pour étudier, d'un côté, l'intégrité et la clonalité de séquences provirales (virologie), et de l'autre, les réponses T spécifiques au VIH (immunologie). Tout d'abord, nous avons développé une bibliothèque R nommée « IntegrityAlgorithm », automatisant une séquence d'analyse utilisée précédemment pour déduire l'intégrité et la clonalité de séquences quasi-pleine longueur (NFL) de provirus. Chaque séquence provirale est analysée pour la présence, dans cet ordre, d'inversions, d'hypermutations, de grandes délétions internes, de codons stops, de défauts dans la région Ψ et de défauts dans les cadres de lectures ouverts (ORFs). Les provirus dépourvus de ces défauts sont identifiés comme « intacts ». Nous montrons que la hiérarchisation des défauts sous-estime certains défauts, comme les codons stops et les défauts dans le Ψ. L'algorithme produit des résultats en accord avec les données publiées et présente plusieurs avantages par rapport à d'autres algorithmes de prédiction.

Nous avons aussi multiplexé l'essai de marqueurs induits par l'activation (AIM), couramment utilisé pour la détection de cellules T spécifiques à un antigène, en mesurant simultanément la co-expression de quatre AIMs (CD69, 4-1BB, OX40 et CD40L) plutôt que deux. En utilisant une stratégie de *gating* booléen ET/OU, l'approche augmente la fréquence de détection des cellules T CD4$^+$ spécifiques à un antigène en comparaison aux paires d'AIMs standards. La détection des cellules T CD8$^+$ était plutôt dominée par la paire d'AIMs CD69$^+$4-1BB$^+$. Le multiplexage réduit les biais dans la qualification et la quantification des cellules T spécifiques à un antigène, tel qu'il a été montré avec des analyses de phénotypage et des comparaisons intra- et inter-cohorte. Nous avons ensuite modifié la stratégie de *gating* booléen ET/OU pour une stratégie de *gating* en « L », nous permettant de trier des cellules vivantes T CD4$^+$ et CD8$^+$ spécifiques au VIH et au cytomégalovirus (CMV).

En considérant l'étroite relation qu'entretiennent les provirus du VIH et le système immunitaire, améliorer nos méthodes pour la caractérisation à la fois des provirus, mais aussi des

réponses T spécifiques au VIH, demeure essentielle pour une meilleure compréhension de la persistance des réservoirs et pour élaborer des stratégies de vaccination et d'éradication du virus.

Mots-clés : VIH, cellules T CD4$^+$ spécifiques à un antigène, cellules T CD8$^+$ spécifiques à un antigène, réservoirs, séquençage quasi-pleine longueur, essai de marqueurs induits par l'activation, tri de cellules vivante, R, cytométrie en flux

Abstract

In the past years, intact and defective human immunodeficiency virus (HIV) proviruses have been shown to code for viral proteins that contribute to sustaining chronic HIV-specific T cell responses, even under antiretroviral therapy (ART). We have developed two methods to study, on one side, the intactness and clonality of proviral sequences (virology), and on the other, the HIV-specific T cell responses (immunology). We developed an R package named "IntegrityAlgorithm" that automates a previous analysis pipeline used to infer intactness and clonality of near full-length proviral sequences (NFL). Each proviral sequence undergoes analysis to identify, in this order, the presence of inversions, hypermutations, large internal deletions, stop codons, Ψ defects, and defects in the open reading frames (ORFs). Proviruses exhibiting none of these defects are inferred as "intact." We show that the hierarchization of defects underestimate some defects, such as stop codons or Ψ defects. The algorithm produces consistent results with the data published and presents several advantages over other prediction algorithms.

We also multiplexed the activation-induced marker (AIM) assay, commonly used to detect antigen (Ag)-specific T cells, by simultaneously measuring the co-expression of four AIMs (CD69, 4-1BB, OX40, and CD40L) instead of two. When combined in an AND/OR Boolean gating strategy, the approach increases the frequency of Ag-specific $CD4^+$ T cells detected in comparison to single AIM pairs. $CD8^+$ T cell detection is rather dominated by the AIM pair $CD69^+4\text{-}1BB^+$. Multiplexing reduces biases in the qualification and quantification of Ag-specific T cells, as shown with phenotyping analyses and intra- and inter-cohort comparisons. We then modified the Boolean AND/OR gating strategy to an "L gating" strategy for the purpose of sorting live Ag-specific $CD4^+$ and $CD8^+$ T cells.

Considering the interdependent relationship between HIV proviruses and the immune system, improving our methods to characterize both the intactness of proviruses and the HIV-specific T cell responses will remain essential for a better understanding of the reservoirs persistence and to elaborate vaccine and viral eradication strategies.

Table des matières

Liste des tableaux

Liste des figures

Section 1 – Introduction

Chapitre 3 - Développement et application d'un algorithme pour analyser l'intégrité et la clonalité des provirus quasi-pleine longueur du VIH

Chapitre 4 – Les bases moléculaires du dysfonctionnement persistant des lymphocytes T CD4$^+$ spécifiques au VIH

Liste des sigles et abréviations

AIM : marqueur induits par l'activation (*activation-induced marker*)

ARNm : ARN messager (*messenger RNA, mRNA)*

ARNv$^+$: ARN viral$^+$ (*vRNA$^+$*)

ART : thérapie antirétrovirale (*antiretroviral therapys*)

ATAC-Seq: *Assay for Transposase-Accessible Chromatin with highthroughput sequencing*

caRNA: ARN de VIH associé aux cellules (*cell-associated RNA*)

CCR : récepteur de chimiokine C-C (*C-C chemokine receptor*)

CMH : complexe majeur d'histocompatibilité

CMV : cytomégalovirus

COBRA : *Combined Bisulfite Restriction Analysis*

CP : progresseurs chroniques (*chronic progressors*)

CPA : cellule présentatrice d'antigènes

CTL : lymphocytes T cytotoxiques (*cytotoxic T lymphocytes*)

CTLA : l'antigène du lymphocyte T cytotoxique

CXCR : récepteur de chimiokine C-X-C (*C-X-C chemokine receptor*)

ddPCR : PCR digitale (*droplet digital PCR*)

EC : contrôleurs élites (*elite controllers*)

ELISpot : *enzyme-linked immunospot*

Env : *envelope*

FLIPS : essai FLIPS (*full-length individual proviral sequencing assay*)

Gag : *group-specific antigen*

GALT : tissus lymphatiques des intestins (*gut-associated lymphoid tissue*)

HLA : antigènes des leucocytes humains (*human leukocyte antigen*)

HXB2 : séquence de référence du VIH, sous-type B

ICI : inhibiteurs des points de contrôle immunitaire (*immune checkpoint inhibitors*)

ICS : marquage intracellulaire des cytokines (*intracellular cytokine staining*)

IL : interleukine

IFN : interféron

IPDA : *intact proviral DNA assay*

LRA : agents de réversion de la latence (*latency reversing agent*)

LTR : séquence terminale longue répétée (*long terminal repeat*)

MSD : site donneur principal d'épissage (*major splice donor*)

Nef : *negative factor*

NFL : quasi-pleine longueur (*near full-length*)

NGS: séquençage nouvelle generation (*next-generation sequencing*)

ORF : cadre de lecture ouvert (*open reading frame*)

PAMP : *pathogen-associated molecular patterns*

pb : paires de bases (*base pairs*, bp)

PCR : réaction en chaîne par polymérase (*Polymerase Chain Reaction*)

PD-1 : protéine 1 de la mort cellulaire programmée

Pol : *polymerase*

ProSeq-IT : *Proviral Sequence Annotation and Intactness Test*

PRR: *pathogen-recognition receptors*

PVIH : personnes avec le VIH (*people with HIV*, PWH)

qPCR : PCR quantitative ou PCR en temps réel (*quantitative PCR*)

REV : *regulator of virion expression*

RRBS: séquençage au bisulfite à représentation réduite (*Reduced representation bisulfite sequencing*)

RRE : *Rev response element*

SHIV : *Simian-Human immunodeficiency virus*

SIV : *Simian immunodeficiency virus*

Tat : *transactivator protein*

TAR : *tat responsive element*

T_{CM} : lymphocytes T mémoires centraux

TCR : récepteur de cellules T

T_{EM} : lymphocytes T mémoires effecteurs

Th : lymphocytes T auxiliaires (*T helper cells*)

TNF : facteurs de nécrose tumorale (*Tumour Necrosis Factor*)

T_{TM} : lymphocytes T mémoires transitionnels

SIDA : syndrome d'immunodéficience acquise (*acquired immunodeficiency syndrome, AIDS*)

Tfh : lymphocytes T auxiliaires folliculaires (*T follicular helper cells*)

UTR : région non-traduite (*untranslated region*)

VCML : Virus de la chorioméningite lymphocytaire (*lymphocytic choriomeningitis virus, LCMV*)

Vif : *viral infectivity factor*

VIH : Virus de l'immunodéficience humaine (*human immunodeficiency virus, HIV*)

Vpr : *viral protein R*

Vpu : *viral protein U*

Remerciements

Beaucoup de personnes m'ont aidée ou ont joué un rôle essentiel, tant au niveau professionnel que personnel, à la réalisation de cette maîtrise.

J'aimerais d'abord remercier mon co-directeur Nicolas Chomont, ainsi que l'associé de recherche du laboratoire Chomont, Rémi Fromentin. Merci à tous les deux pour la compréhension et la gentillesse que vous avez eue à mon égard, surtout vis-à-vis les défis qu'apportaient une co-direction. Merci de m'avoir intégré à toutes vos traditions, surtout celles de Noël, et toutes vos activités, que je garderai longtemps en mémoire. Merci également à tout le laboratoire Chomont – Hélène, César, Armando, Amélie, Caroline, Camille, Arthur, Thibaut et Julien – pour votre accueil chaleureux, vos précieux conseils et votre aide sur la réalisation de mes projets.

J'aimerais également remercier mon directeur de recherche, Daniel Kaufmann. Merci pour la confiance que tu m'as accordée dès le début de ce projet, pour ton encadrement, ta générosité et tes conseils. Travailler sous ta supervision a été extrêmement formateur et m'a permis de grandir en tant que jeune scientifique (et de réaliser de très belles figures).

Merci également à l'associé de recherche du laboratoire Kaufmann, Mathieu Dubé. Je suis très reconnaissante de ta supervision et de nos discussions sur le hockey et de la vie en générale. Tu as été un excellent mentor, en balançant efficacement les moments pédagogiques et les moments où j'avais besoin d'un coup de main pour débloquer. Merci aussi pour ta compréhension durant certains moments plus difficiles dans ma vie.

Merci à Philippe St-Onge et Gaël Dulude de la plateforme de cytométrie pour les 30 tris, souvent dynamiques, que nous avons faits ensemble.

Merci à mon groupe du cours de viro, Mehdi, Ketsia et Jonas, pour les bons moments passés ensemble, même si nous n'étions que 5 dans le cours. Merci particulièrement à Mehdi pour les petites discussions dans le corridor et pour t'être plaint avec moi à certains cours (tous les cours…) d'immuno et du séminaire, ou à chaque fois que nous devions nous rendre à des bâtiments qui prenaient plus que 5 minutes de marche.

Merci à Michael pour ta supervision en tant que post-doc, nos discussions sur le hockey, pour ton attitude « relaxe » qui avait le don de me faire déstresser et surtout, pour tous les conseils de vie que tu m'as donnés. Merci aussi à Gérémy pour toutes les discussions que nous avons eues, incluant tes connaissances qui m'ont aidées dans plusieurs de mes projets. Merci à Manon pour toutes les fois où tu m'as fait rire, le *tea* que tu nous partageais dans l'ascenseur et pour toutes les fois où tu m'as prêté ton chargeur de Mac (oups…)! Merci à Sabrina, que j'ai eu la chance de côtoyer pendant quelques mois lorsqu'elle était avec nous à Montréal, pour tes histoires, tes anecdotes et toutes les fois où tu me faisais rire en *lab meeting*. Merci à la meilleure voisine de bureau, Alina, pour toutes les fois où tu m'as dépanné dans le P2 ou que tu es allée chercher mes cellules dans le P3. J'ai adoré nos discussions sur tout et de rien, sur notre passion pour les tattoos et toutes les fois où nous sommes allées manger ou se chercher un *bubble tea* dans Chinatown. Merci à Rose, une des personnes les plus drôles que j'ai rencontrées dans ma vie, pour ton énergie, toutes les références québécoises que nous partagions et toutes les histoires que tu nous racontais aux dîners.

Merci à tous les autres membres du laboratoire Kaufmann avec qui je mangeais à presque chaque midi, mais plus spécifiquement :

- Merci à Gloria, la *lab manager*, mais aussi un peu la maman du laboratoire. Merci pour les leçons de vie (par exemple, les livres sur comment planifier ma retraite), nos discussions sur les films et émissions de Disney, les publications Instagram/Facebook qui te faisaient penser à nous, ou tout simplement pour l'aide sur mes projets.
- Merci à Tayma, ma *partner* de maîtrise. Merci pour toutes les fois où tu m'as rassurée pour quelconque raison par rapport à mes manips, pour les bons moments que nous avons passé à Keystone à écouter Netflix en soirée et pour ta gentillesse, ton calme et ta douceur.
- Merci à mon *Swiftie* préféré, Alex, pour avoir choisi *Speak Now* comme ton album préféré. Merci pour toutes les discussions très peu scientifiques que nous avions (incluant entre autres la musique, les Pokémons, Taylor Swift, les tattoos et surtout, LA bague).
- Merci énormément à une personne qui a joué un rôle clé à la réalisation de ce projet, Mélina. Merci de t'être levée à 6h00 du matin avec moi chaque mardi pour faire notre *staining*, pour ton support émotionnel quand le signal des cellules n'était pas aussi élevé que ce que j'avais prévu, pour tous les 10 000 dégels que tu as fait pour moi et pour toutes les journées entières

18

que nous passions au P3 à écouter toutes sortes de playlists de musique et à discuter de tous les sujets inimaginables.

Merci à toute ma famille – mes marraines, tantes, oncles, cousins, cousines et mes grands-parents.

Merci à mon chien Moose pour son soutien (très douteux la majorité du temps). Merci d'être le meilleur chien que j'ai eu dans ma vie, de me faire rire et de me changer les idées continuellement.

Je tiens à remercier toute la famille Lavictoire-Émond-Mercier – Noémie, Valérie, Pierre et Meghan – pour être une deuxième famille aussi formidable. Merci aussi à ma troisième paire de grands-parents, Mamie et Papou.

Merci à mes parents pour leur soutien inconditionnel, non seulement au cours de ces deux dernières années, mais aussi à travers tout mon parcours scolaire et de vie. Merci de m'avoir toujours poussée jusqu'aux plus loin de mes capacités, d'avoir toujours cru en moi et de m'avoir soutenue dans les décisions que j'ai prises. Je vous aime et j'ai hâte de vous revoir plus souvent.

Merci à ma sœur, Noémie, pour toutes les fois où nous avons joué à *Fort* ou NHL le soir après le travail, pour toutes les fois où tu es montée me voir à Montréal, même si ce n'était que pour une *game* d'hockey, et merci d'avance pour toutes les *rides* que tu vas me donner à Monfort les mardis.

Un énorme merci et probablement le plus précieux à Audrey, ma conjointe (et fiancée), une des personnes les plus importantes de ma vie, et sans qui ce mémoire n'aurait jamais été réalisé. Merci pour tes encouragements, les fous rires, les surprises que tu continues de me faire à chaque jour depuis les six dernières années. Merci pour ton amour, pour tout ce que tu fais pour moi et pour véritablement être ma supportrice numéro 1. C'est un nouveau chapitre de nos vies qui commence et je suis extrêmement chanceuse de pouvoir le partager à tes côtés.

Finalement, j'aimerais conclure cette section avec des remerciements et une pensée très spéciale pour ma grand-maman Andréa, la femme la plus forte que j'ai connue, à qui je dédie l'entièreté de ce mémoire.

Avant-propos

Pour donner un peu de contexte derrière l'origine de cette co-supervision, j'ai initialement débuté ma maîtrise dans le laboratoire de Dr. Daniel Kaufmann pour travailler sur le projet de la dysfonction des lymphocytes T CD4$^+$ spécifiques au VIH, qui sera présenté dans ce mémoire, avec Dr. Maaran Michael Rajah, un stagiaire postdoctoral.

Quelques mois après le début de ma maîtrise, considérant le déménagement du laboratoire Kaufmann à Lausanne, en Suisse, j'ai eu l'occasion d'entreprendre une co-supervision avec le Dr. Nicolas Chomont. Puisque j'ai une formation en bioinformatique et en programmation provenant de mon baccalauréat, on m'a proposé un projet qui mettrait en valeur ces compétences. Un projet, d'ailleurs, axé sur la virologie du VIH.

En parallèle, face au départ du Dr. Rajah quelques mois plus tard, on m'a donné la chance de prendre la tête du projet « dysfonction », un projet axé sur les réponses immunologiques face à l'infection au VIH.

Ainsi, dans ce mémoire, je présenterai deux projets réalisés dans deux laboratoires différents, chacun ayant son propre chapitre distinct: un projet virologique dans le laboratoire Chomont (**chapitre 3**) et un projet immunologique dans le laboratoire Kaufmann (**chapitre 4**). Ensemble, ces deux projets permettront une meilleure compréhension du VIH lui-même, de l'élaboration de stratégies vaccinales, de la persistance des réservoirs et de leur éradication.

Chapitre 1 – Introduction

1.1 Le virus de l'immunodéficience humaine

Le virus de l'immunodéficience humaine (VIH) est le rétrovirus responsable du syndrome d'immunodéficience acquise, communément appelé « SIDA ». Il existe deux types principaux de VIH : le VIH-1 et le VIH-2. Le VIH-1 est divisé en quatre groupes (M, N, O et P) et le groupe M, responsable de plus de 90% des cas mondiaux de VIH-1, est lui-même sous-divisé en sous-types A à L (1,2). Ce mémoire ne traitera que du VIH-1 et plus spécifiquement du sous-type B, le sous-type le plus répandu en Europe et dans les Amériques.

1.1.1 La pathogénèse

Le virus se transmet par échange de liquides corporels (sperme, sécrétions vaginales, sang et lait maternel) avec une personne avec le VIH (PVIH), principalement par contact sexuel, mais aussi par transmission verticale de mère à enfant ou par inoculation percutanée (3). Il y a trois phases à l'infection du VIH : la phase aigüe, la phrase chronique et la phase SIDA (**Figure 1.1.1**).
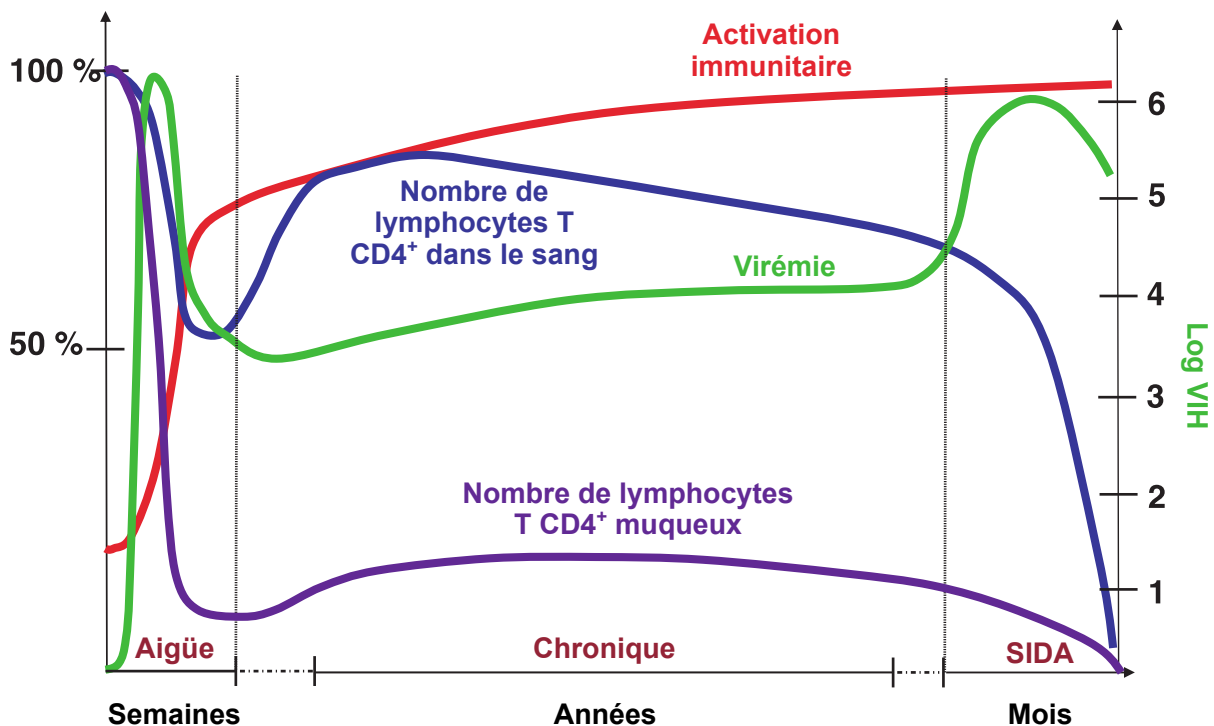
Figure 1.1.1 – Pathogénèse du VIH. Les courbes rouge, bleu, vert et mauve représentent, respectivement, l'activation immunitaire, le nombre de lymphocytes T CD4$^+$ dans le sang, la virémie et le nombre de lymphocytes T CD4$^+$ dans les muqueuses. Figure adaptée de Grossman et al. (4).

La phase aigüe dure de quelques jours à quelques semaines après l'événement de transmission. La première phase « d'éclipse » est asymptomatique, dure en moyenne 10 jours et précède la détection d'ARN viral dans le sang (5). Durant cette phase, le VIH infecte les lymphocytes T CD4$^+$, ses cellules cibles, des muqueuses du site d'infection. Le virus infecte préférentiellement les cellules qui expriment le récepteur de chimiokine C-C de type 5 (CCR5). L'infection déclenche une inflammation locale qui permet de recruter d'autres cellules immunitaires, comme des lymphocytes T CD4$^+$ additionnels, au site d'infection (6,7). À la fin de la phase d'éclipse, le VIH se rend jusqu'aux ganglions lymphatiques, rencontre et infecte d'autres cellules T CD4$^+$ activées (5,8) par dissémination systématique. Le VIH se réplique alors rapidement et se propage à d'autres tissus lymphatiques, en particulier à ceux des intestins (GALT). La charge virale (ou virémie) augmente exponentiellement jusqu'à atteindre des millions de copies d'ARN viral par mL de sang, trois à quatre semaines après le début de l'infection (5). En parallèle avec le pic de virémie, le système immunitaire est fortement activé, les lymphocytes T CD4$^+$ dans le sang sont en faible quantité et ceux des GALT sont largement diminués (4,9). Certains individus auront des symptômes aspécifiques similaires à ceux de la grippe. Vers la fin de la phase aigüe, la charge virale redescend, ce qui a été associé avec une augmentation de de la réponse cytotoxique des cellules T CD8$^+$ spécifiques au VIH (1). Le compte de cellules T CD4$^+$ est partiellement restauré dans le sang, mais pas dans les GALT.

La phase chronique est généralement asymptomatique et peut durer des années. Bien que la charge virale demeure relativement constante, le nombre de lymphocytes T CD4$^+$ diminue lentement, mais progressivement au cours du temps (4,10). L'activation immunitaire augmente aussi de manière constante. Sans traitement, le virus continue de se répliquer à bas niveaux. La perte de cellules T CD4$^+$ est dite linéaire et constante jusqu'au stade le plus avancé de l'infection au VIH, où lorsque le compte de cellules T CD4$^+$ atteint <200 cellules/ μL, le déclin se fait plus rapidement (10).

Dans la majorité des cas, la perte de lymphocytes T CD4$^+$ combinée avec l'augmentation graduelle de la virémie mène à la progression de l'infection en SIDA. À ce stade, le compte de cellules T CD4$^+$ est de <200 cellules/µL (1,11). Les individus deviennent alors susceptibles à toutes sortes d'infections opportunistes, devenant souvent mortelles dû à la diminution de leurs cellules T CD4$^+$, comme la pneumonie à *Pneumocystis jirovecii* ou la rétinite à cytomégalovirus (CMV), et à certains cancers tels que le sarcome de Kaposi (1,12).

En comparaison aux « progresseurs chroniques » (CP), l'infection au VIH ne progresse pas au SIDA chez un petit groupe de personnes (<1%). Ces individus, appelés « contrôleurs élites » (EC), contrôlent spontanément la virémie à des niveaux d'indétectabilité et maintiennent des comptes élevés et stables de cellules T CD4$^+$ pendant des années, et ce, sans traitement (13,14).

En 2022, environ 85,6 millions d'infections au VIH avaient été comptabilisées depuis sa découverte en 1981, avec toujours 1,3 millions de nouvelles infections à chaque année. Le virus a été responsable de plus de 40 millions de décès à travers le monde et, en moyenne, 630 000 personnes décédaient toujours en 2022 à la suite de maladies reliées au SIDA. Environ 39 millions de personnes vivent toujours avec le VIH (15).

L'introduction de la thérapie antirétrovirale (ART) a drastiquement changé le pronostic de l'infection par le VIH. Le traitement ART a permis de réduire la réplication virale à des niveaux d'indétectabilité et d'augmenter l'espérance de vie des PVIH (16). Sans traitement, dans la majorité des cas, l'infection au VIH progresse au SIDA en moyenne huit à dix ans après le début de l'infection (1).

1.1.2 Le cycle de réplication virale

Le VIH infecte préférentiellement les lymphocytes T CD4$^+$ activés, mais peut aussi infecter les cellules myéloïdes comme les macrophages, les cellules dendritiques et les cellules microgliales (1). Parmi les sous-populations de lymphocytes T CD4$^+$, les Th17, exprimant le récepteur de chimiokines et marqueur de migration vers les intestins CCR6, seraient plus susceptibles à l'infection (17–20).

Le cycle de réplication virale du VIH peut être divisé en sept étapes (**Figure 1.1.2**). Le VIH, avec la glycoprotéine gp120 de son enveloppe (Env), se lie à son récepteur CD4 et à un de ses corécepteurs, CCR5 ou le récepteur de chimiokine C-X-C de type 4 (CXCR4). L'attachement provoque un changement de conformation, entraînant la fusion des membranes virale et cellulaire de l'hôte, ce qui permet l'entrée du cœur viral dans le cytoplasme de la cellule (étape n°1). Grâce à l'enzyme virale de transcriptase inverse, l'ARN viral monocaténaire à polarité positive est rétrotranscrit en ADN viral double-brin (étape n°2). L'ADN viral néoformé est importé dans le noyau, où a lieu sa décapsidation et son relâchement à proximité des sites d'intégration (21). L'ADN viral s'intègre ensuite au génome de l'hôte à l'aide de l'intégrase virale (étape n°3). Le virus se trouve alors sous forme d'un « provirus », c'est-à-dire la séquence virale incorporée dans le génome de l'hôte. En utilisant la machinerie cellulaire, l'ADN viral est transcrit en ARN messagers (ARNm) viraux (étape n°4). Des copies du génome viral sous forme d'ARN monocaténaire sont aussi synthétisées. Après l'épissage des ARNm et leur exportation hors du noyau, les ARNm sont traduits en protéines et en enzymes virales nécessaires à la formation du cœur viral (étape n°5). L'assemblage se produit également dans le cytoplasme, où le génome et les protéines virales synthétisées sont encapsidés dans le cœur viral à l'aide de la polyprotéine Gag. Le bourgeonnement permet quant à lui la formation de l'enveloppe virale (étape n°6). Après le relâchement du virion immature hors de la cellule, l'enzyme de protéase virale induit sa maturation, le rendant ainsi compétent pour l'infection d'une nouvelle cellule (étape n°7) (1,12,22).

Les médicaments antirétroviraux actuels bloquent une ou plusieurs étapes du cycle viral, inhibant ainsi la réplication virale (**Figure 1.1.2**). Ils sont généralement regroupés en six grandes classes de médicaments : les inhibiteurs d'entrée (IEs), qui incluent les inhibiteurs d'attachement, les inhibiteurs de CCR5 et/ou CXCR4 et les inhibiteurs de fusion, les inhibiteurs nucléosidiques et nucléotidiques de la transcriptase inverse (INTIs), les inhibiteurs non nucléosidiques de la transcriptase inverse (INNTIs), les inhibiteurs de l'intégrase (INIs), les inhibiteurs de la protéase (IPs) et les inhibiteurs de la capside (ICs). Dans la majorité des cas, deux INTIs combinés à un INI, un IP ou un INNTI sont administrés à l'initiation du traitement (1).
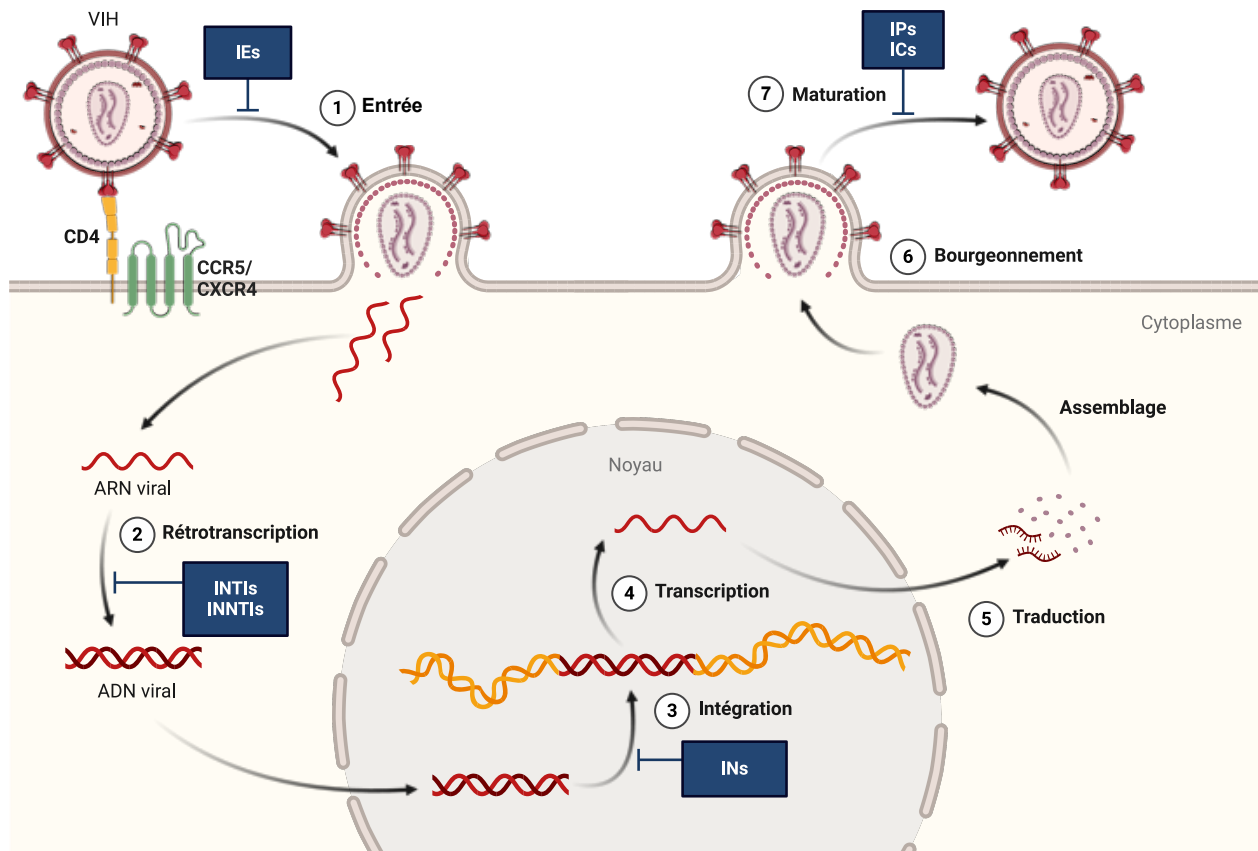
Figure 1.1.2 – Cycle viral du VIH. Figure adaptée de Deeks et al. (12) et Bekker et al. (1). Figure créée avec un modèle disponible sur BioRender.com. IEs : inhibiteurs d'entrée, incluant les inhibiteurs d'attachement, les inhibiteurs de CCR5 et/ou CXCR4 et les inhibiteurs de fusion; INTIs : inhibiteurs nucléosidiques/ nucléotidiques de la transcriptase inverse; INNTIs : inhibiteurs non nucléosidiques de la transcriptase inverse; INIs : inhibiteurs de l'intégrase; IPs : inhibiteurs de la protéase; ICs : inhibiteurs de la capside.

1.1.3 Les protéines virales

Le génome du VIH-1 a une taille approximative de 9 800 nucléotides (23) et code pour neufs cadres de lecture ouverts (ORFs) : trois protéines de structure (Gag, Pol et Env), deux protéines régulatrices essentielles (Tat et Rev) et quatre protéines régulatrices accessoires (Vif, Vpr, Vpu et Nef) principales (24) (**Figure 1.1.3**). En résumé, les rôles des protéines du VIH sont les suivants :

- La polyprotéine Gag (« *group-specific antigen* ») est composée des protéines de la matrice (p17), de la capside (p24) entourant le génome viral, de la nucléocapside (p7) et de p6, une protéine impliquée dans le bourgeonnement viral.

- La polyprotéine Pol (« *polymerase* ») est composée des enzymes virales de la protéase (p12), de la transcriptase inverse (p51) et de l'intégrase (p31).

- La polyprotéine Env (« *envelope* ») est formée de deux sous-unités : la glycoprotéine de surface (gp120) et la glycoprotéine fusogénique transmembranaire (gp41).

- La protéine régulatrice essentielle Tat (« *transactivator protein* ») se fixe à TAR (« *Tat responsive element* »), situé à l'extrémité 5' des régions séquences terminales longues répétées (LTR), pour favoriser l'élongation et augmenter l'efficacité de la transcription virale (24–26). Sans Tat, les transcrits ne dépassent habituellement pas quelques centaines de paires de bases.

- La protéine régulatrice essentielle Rev (« *regulator of virion expression* ») se fixe au RRE (« *Rev Responsive Element* ») pour exporter hors du noyau les ARNm partiellement et non épissés. Contrairement aux ARNm complètement épissés codant pour les protéines Tat, Rev et Nef, ceux partiellement et non épissés coderont pour les protéines structurales ou le génome (ARN génomiques) et nécessitent l'activité de Rev pour leur exportation (24,27).

- Le rôle de la protéine régulatrice accessoire Vif (« *viral infectivity factor* ») est d'inhiber l'activité du facteur de restriction APOBEC3G, connu pour induire l'hypermutation du génome viral, et ainsi limiter la production de protéines non-fonctionnelles (27,28).

- La protéine régulatrice accessoire Vpr (« *viral protein R* ») joue plusieurs rôles dans la réplication et la pathogénèse du virus, incluant un effet sur l'efficacité du processus de rétrotranscription, sur l'importation de l'ADN viral dans le noyau, sur la modulation de l'apoptose et sur l'induction de l'arrêt du cycle cellulaire en phase G2 (24,29,30).

- Le rôle de la protéine régulatrice accessoire Vpu (« *viral protein U* »), spécifique au VIH-1, est de dégrader le CD4 néo-synthétisé, de diminuer l'expression des molécules d'antigènes des leucocytes humains de type C (HLA-C) à la surface des cellules infectées, mais aussi de contrer l'action de la téthérine, un facteur de restriction qui bloque le relâchement des virions (24,31–33).

- La protéine régulatrice accessoire Nef (« *negative factor* ») améliore l'infectivité du virus et diminue l'expression de surface de CD4. Elle diminue aussi l'expression des molécules

du complexe majeur d'histocompatibilité (CMH) de classe I à la surface des cellules infectées, spécifiquement les HLA-A et les HLA-B, potentiellement pour les aider à évader la réponse cytotoxique (24,34,35).
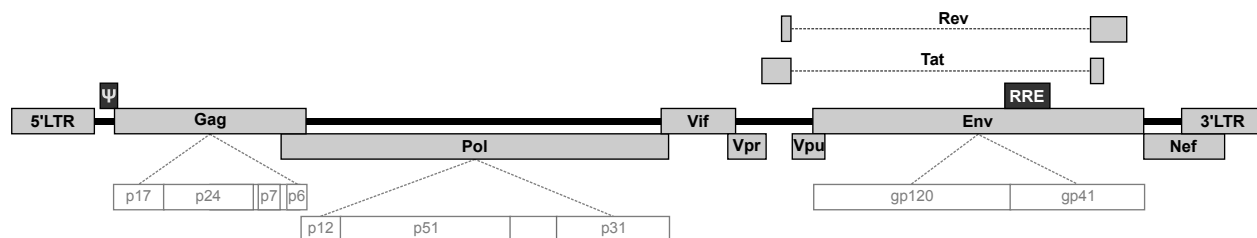
Figure 1.1.3 – Génome du VIH-1. Figure inspirée du *Los Alamos National Laboratory* (36).

À la suite de la rétrotranscription, l'ADN viral est flanqué de deux LTR identiques, l'une à l'extrémité 5' et l'autre, à l'extrémité 3' (**Figure 1.1.3**). Le 5'LTR contient le promoteur à la transcription virale, tandis que le 3'LTR joue un rôle dans la terminaison de la transcription et dans la polyadénylation des transcrits viraux (37). En aval du 5'LTR et en amont de Gag se situe le locus Psi (ou Ψ), qui contient le site donneur principal d'épissage (MSD) pour l'épissage des transcrits et les quatre boucles SL1, SL2, SL3 et SL4, qui ont un rôle dans la dimérisation et l'encapsidation de l'ARN viral (38).

1.1.4 Les réservoirs du VIH

Tel que mentionné plus haut, lors d'une infection par le VIH, l'ADN viral s'intègre dans le génome des cellules de l'hôte. On distingue l'infection dite « productive » d'une infection dite « latente ». L'infection productive est surtout observée chez les personnes qui ne sont pas sous traitement et est caractérisée par la production d'ARN, de protéines et de particules virales. Ces cellules transcriptionnellement actives composent le réservoir « actif » du VIH. Plusieurs études montrent cependant que le réservoir actif existe également chez les PVIH sous traitement, même si cette réplication résiduelle est à bas bruit (39–43). À l'opposé, dans l'infection latente, il n'y a pas ou très peu de transcription virale (1,17). L'ensemble de ces cellules infectées de manière latente constituent le réservoir « latent » du VIH, dans lesquels le génome viral persiste de façon prolongée

(44). Si les cellules infectées de manière latente portant des provirus compétents pour la réplication devaient se réactiver, ils pourraient causer une nouvelle infection productive.

L'infection latente est principalement observée dans les cellules T CD4$^+$ au repos qui ont un phénotype mémoire, principalement les lymphocytes T mémoires centraux (T$_{CM}$) et les lymphocytes T mémoires transitionnels (T$_{TM}$) (45,46), mais aussi les lymphocytes T effecteurs mémoires (T$_{EM}$) (44,47). Le VIH peut aussi persister dans différentes sous-populations fonctionnelles de cellules T mémoires, incluant les Th1, les Th17, les lymphocytes T folliculaires helper (Tfh) et les lymphocytes T régulateurs (Treg) (44). La latence serait établie soit post-activation des cellules T CD4$^+$, c'est-à-dire que des cellules activées et infectées retourneraient à un état de dormance, ou soit pré-activation, où le VIH infecterait directement des cellules T CD4$^+$ mémoires au repos ou des cellules activées dans le processus de se différentier en mémoire (1,17,44). Puisque les cellules infectées de manière latente n'expriment pas ou peu d'antigènes du VIH, elles échappent au système immunitaire.

Sous ART, la taille du réservoir demeure stable au cours du temps, avec une demi-vie approximative de 44 mois (48), malgré des décennies de thérapie sous ART (49). Plusieurs mécanismes permettraient au réservoir de persister. Parmi ceux-ci, il y a la prolifération homéostatique assurant la survie des cellules T mémoires pendant des années (45,50,51), mais aussi l'expansion clonale (50–55), où une cellule récemment activée prolifère en nombres par la création de cellules lui étant identiques. Ces clones possèdent alors des provirus identiques à la cellule initiale (56).

Dans la majorité des cas, un rebond de la charge virale est observé après l'interruption du traitement (57–59). Ainsi, bien que le traitement ART inhibe la réplication virale et permette de prévenir l'infection de nouvelles cellules, il ne parvient pas à éliminer les réservoirs du VIH, ces cellules où le génome viral est déjà intégré (12). Selon une étude utilisant un modèle simien, ces réservoirs s'établiraient aussi tôt que trois jours après le début de l'infection (60). Traiter l'infection rapidement limite la taille du réservoir, mais n'empêche pas son établissement (46,61–64).

La difficulté de réactiver, reconnaître et éliminer ces cellules portant des génomes compétents pour la réplication demeure un obstacle majeur à l'éradication du VIH. Même si le VIH semble avoir une préférence pour les lymphocytes T CD4$^+$ spécifiques au VIH (65), au CMV, à

l'influenza, au *Mycobacterium tuberculosis* (66) et au virus d'Epstein-Barr (EBV), des co-infections fréquentes avec le VIH, la spécificité des cellules T CD4$^+$ infectées continue d'être majoritairement inconnue (52). La stratégie thérapeutique du « *shock and kill* » utilise des agents de réversion de la latence (LRAs) pour induire la transcription virale, afin d'induire leur élimination par le système immunitaire (67).

1.1.5 Les mesures de l'intégrité du génome

Parmi les cellules du réservoir latent du VIH, celles portant des provirus inductibles et potentiellement compétents pour la réplication sont particulièrement étudiés, dû à leur potentiel d'alimenter le rebond de la virémie lors de l'arrêt du traitement. Une partie de la recherche sur le VIH se concentre ainsi à essayer de les identifier et de les éliminer. Toutefois, chez les individus sous traitement, ces provirus ne représentent que 2 à 5% de toutes les cellules infectées (47,68–70). Il existe de nombreuses méthodes pour caractériser quantitativement et qualitativement les réservoirs du VIH, que ce soit pour mesurer la taille du réservoir, pour informer sur leur phénotype, leur inductibilité, leur intégrité et/ou sur leur capacité à produire des virions compétents pour la réplication (71).

Parmi les méthodes plus communes, il y a la mesure de l'ADN total du VIH (72–74), qui quantifie toutes les formes du génome viral, et la mesure de l'ADN intégré de VIH (75), qui ne quantifie que les provirus intégrés (71,76,77) (**Figure 1.1.4**). Cela se fait par réaction en chaîne par polymérase (PCR), soit par PCR en temps réel (qPCR) ou par PCR digitale (ddPCR), en utilisant des amorces spécifiques au VIH (ADN total) ou en les combinant avec des amorces Alu, des courtes régions répétées spécifiques au génome humain (ADN intégré). Ce sont des approches simples, sensibles et qui requièrent peu de cellules. Cependant, elles mesurent majoritairement des génomes incompétents pour la réplication virale.

La mesure de l'ARN du VIH associé aux cellules (caRNA) (77–79), aussi basée sur la PCR, permet de quantifier le nombre de transcrits spontanés non-épissés (usRNA) ou multi-épissés (msRNA). Elle permet de mesurer la fréquence de cellules compétentes pour la transcription virale et d'estimer le réservoir actif du VIH.

Parmi les essais d'inductibilité du réservoir, le RNA-flow FISH (69–73) et le HIV-Flow (84) sont tous deux basés sur la cytométrie en flux. Le RNA-flow FISH permet de détecter différents transcrits d'ARNm (5'exon, Gag, Pol, Gag-Pol, Nef) à l'aide de sondes, indiquant les capacités transcriptionnelles du réservoir, ainsi que la protéine p24 pour ses capacités traductionnelles. Le HIV-Flow caractérise les réservoirs aux capacités traductionnelles en utilisant deux anticorps ciblant la p24. La combinaison de deux marqueurs (sondes d'ARNm et/ou p24) permet d'identifier les cellules infectées. Les événements détectés sont très rares, même en utilisant des agents de réactivation : chez les PVIH sous traitement, les médianes sont de 97 cellules compétentes pour la transcription par millions de lymphocytes T CD4$^+$ (RNA-flow FISH) (82), 3,6 cellules compétentes pour la transcription et la traduction (RNA-flow FISH) (80) et 4,3 cellules compétentes pour la traduction seulement (HIV-Flow) (84). Ce sont des essais assez coûteux qui requièrent un grand nombre de cellules, mais dont la résolution est au niveau de la cellule unique, permettant ainsi de phénotyper et d'obtenir la nature des cellules infectées.

Un autre essai d'inductibilité est le TILDA (« *Tat/rev Induced Limiting Dilution Assay* ») (85). Contrairement au caRNA, il mesure la transcription (ARN) induite. Spécifiquement, le TILDA mesure la fréquence de génomes pouvant produire des transcrits d'ARN multi-épissés (et plus spécifiquement, ceux de Tat et de Rev) en utilisant le principe de la dilution limite et de la RT-PCR. La dilution limite est une technique de dilution basée sur la loi de Poisson qui permet d'évaluer la fréquence de l'événement détecté, soit la présence d'ARN du VIH multi-épissés dans ce cas (86).

La seule méthode permettant de réellement quantifier les cellules portant des provirus compétents pour la réplication est le QVOA/VOA (« *quantitative viral outgrowth assay* » ou « *viral outgrowth assay* »). Il s'agit d'un autre essai d'inductibilité, mais cette fois-ci basé sur la culture cellulaire (58). Brièvement, les cellules de PVIH sont mises en culture en dilution limite, activées *in vitro* pour stimuler la réplication virale et la production de nouveaux virions infectieux, puis cultivées en présence de cellules cibles non-infectées. Les niveaux de protéines p24 dans le surnageant, indicatifs des capacités de production virale, sont mesurés avec la méthode immuno-enzymatique ELISA (71). En revanche, il s'agit d'un essai plus long et difficile (2-3 semaines).
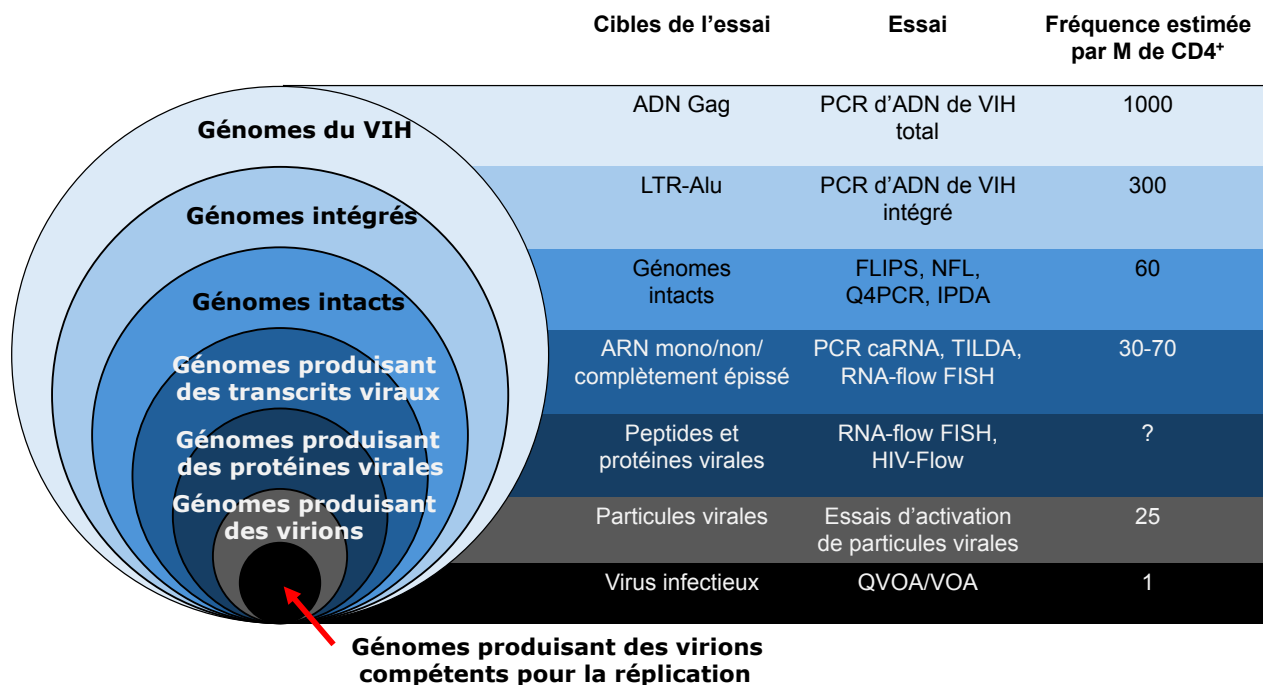
Figure 1.1.4 – **Différentes méthodes de caractérisation des réservoirs.** Figure adaptée de Deeks et al. (87).

1.1.5.1 Les approches de séquençage de provirus de quasi-pleine longueur ou NFL (« *near full-length* »)

Considérant les difficultés à identifier les provirus infectieux, plusieurs études se sont plutôt penchées à évaluer les génomes exempts de défauts, qualifiés de génétiquement « intacts », qui pourraient prédire leur compétence potentielle pour la réplication. Toutefois, aucune des techniques mentionnées ci-dessus ne permet de discriminer les génomes intacts et défectueux des provirus. Pour ce faire, des nouveaux essais basés sur le séquençage ont été développés pour estimer avec davantage de précision la proportion de génomes intacts.

Ho et al. ont développé une technique permettant d'obtenir la séquence quasi-pleine longueur (*near full-length* ou NFL) des provirus. Cette technique utilisait des PCR à deux étapes dites « nichées » (69). La première étape utilisait des amorces externes (de la région 5' à 3') pour amplifier la région d'intérêt en un seul fragment. Afin d'augmenter la spécificité du signal, ce produit d'amplification était subséquemment utilisé pour une deuxième PCR, cette fois-ci avec des

amorces internes. Les auteurs ont réalisé quatre PCR nichées avec des amorces internes se chevauchant, suivi par du séquençage Sanger, permettant de reconstruire la séquence quasi-pleine longueur du provirus.

Les équipes de Dr. Sarah Palmer (47) et Dr. Mathias Lichterfeld (88) ont ensuite modifié le protocole NFL pour amplifier, en une seule PCR nichée, un fragment d'environ 9 000 paires de bases (pb). Ce fragment couvrait approximativement 92% du génome du VIH (89). L'amplification a été suivie du séquençage nouvelle génération (NGS) Illumina MiSeq. Les fragments étaient ensuite assemblés *de novo*, puis alignés sur la séquence de référence du sous-type B du VIH-1, HXB2, et analysés pour la présence de plusieurs défauts. Hiener et al. ont estimé que 5% des provirus étaient intacts, trouvés en plus grande fréquence dans les cellules T effectrices mémoires (T_{EM}), et que plusieurs de ces génomes intacts étaient identiques, suggérant que la prolifération et l'expansion clonale jouent des rôles importants dans la persistance des réservoirs du VIH (47). Les résultats de Lee et al. concordaient avec ceux de Hiener et al. : Lee et al. ont trouvé que les génomes intacts clonaux représentaient une majeure partie des provirus trouvés dans les cellules mémoires, qu'ils étaient préférentiellement dans les cellules de polarisation Th1 et ont confirmé, à l'aide d'un VOA, que ces génomes intacts étaient réplicatifs et infectieux (88).

Une limite à ce type d'approches est la subjectivité des critères utilisés pour déterminer l'intégrité des provirus (47,69,88,90–92). Une hiérarchie de défauts est établie selon la probabilité que le défaut compromette la capacité de réplication du provirus, mais les catégories de défauts et leur hiérarchisation varient d'une séquence d'analyse à l'autre (**Figure 1.1.5**). Les grandes délétions internes semblent être un défaut important pour la discrimination des génomes intacts et défectueux. En revanche, les inversions ne sont trouvées que dans deux des quatre séquences d'analyse. Certains des défauts, comme les codons stop, les décalages de cadre de lecture et les indels (insertions et/ou délétions) étaient aussi analysés séparément ou regroupés dans des plus grandes catégories, ce qui fait grandement varier le portrait global des types de séquences provirales. Malgré tout, la proportion de génomes intacts demeure relativement constante entre les quatre séquences d'analyse (2,4% à 4,9%) (91).
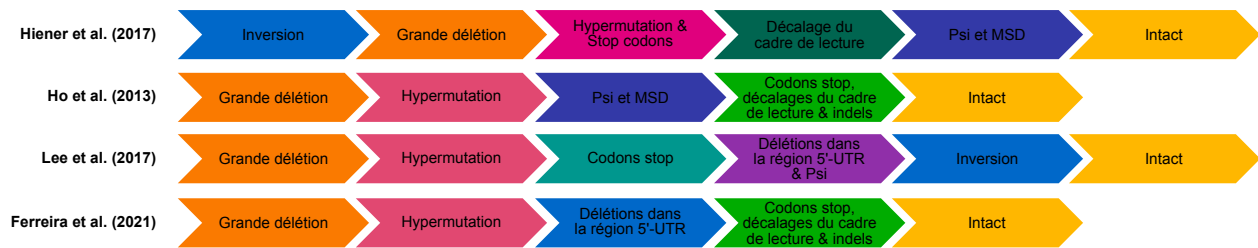
Figure 1.1.5 – Résumé des séquences d'analyse d'intégrité des provirus NFL. Figure adaptée de Ferreira et al. (91).

Considérant que le séquençage est l'étape la plus dispendieuse du protocole NFL, Bruner et al. ont développé une méthode ne requérant pas de séquençage, soit l'IPDA (« *intact proviral DNA assay* ») (68). En comparaison aux méthodes classiques de PCR, l'IPDA estime la proportion de génomes intacts avec plus de précision. Basée sur la ddPCR, cette technique permet de détecter les provirus intacts en utilisant des amorces dans deux régions fréquemment mutées et essentielles à l'intégrité du génome (locus Ψ et le RRE dans le gène Env). Initialement basée sur le sous-type B, elle a récemment été adaptée pour le sous-type C (93). En analyse bivariée (Env sur l'axe des abscisses, Ψ sur l'axe des ordonnées), quatre types de provirus peuvent alors être quantifiés selon le quadrant dans lequel il est détecté: (i) les provirus hypermutés et/ou délétés en 3' (simple positif pour Ψ), (ii) les provirus « intacts » (double positifs), (iii) l'absence de provirus du VIH ou les provirus défectueux pour les deux amplicons (double négatifs) et (iv) les provirus délétés en 5' (simple positif pour Env). Cependant, le nombre de provirus intacts est surestimé, car l'essai ne prend pas en considération les défauts à l'extérieur de ces deux régions (94,95). Dans tous les cas, la seule manière de vérifier l'intégrité des génomes proviraux, et donc les virus potentiellement compétents pour la réplication, est à l'aide du séquençage des provirus NFL (71).

Gaebler et al. ont quant à eux développé la NFL Q4PCR pour augmenter la fréquence de provirus intacts séquencés (96). Ils ont d'abord réalisé une PCR NFL avec des amorces externes, puis réalisé quatre qPCR ciblant les régions du Ψ, de Gag, de Pol et d'Env. Les puits positifs pour au moins deux des quatre sondes étaient soumises à une PCR NFL nichée et du séquençage Illumina MiSeq. La fréquence de génomes intacts était enrichie lorsque les puits étaient positifs à

trois (26%) ou quatre (51%) sondes, comparé à la condition en absence de préselection (12,9%) (96).

Des algorithmes pour prédire l'intégrité des génomes proviraux, comme ProSeq-IT (« *Proviral Sequence Annotation and Intactness Test* ») (97), HIVIntact (98) et HIVSeqinR (99,100), ont depuis été développés. Une description détaillée de ces algorithmes sera faite à la **section 3.3**.

1.1.5.2 Le protocole d'analyse de l'intégrité des provirus NFL développé dans le laboratoire Chomont

Le protocole NFL développé et publié par Dre Caroline Dufour (82,42,90,101), adapté du FLIPS (« *full-length individual proviral sequencing* ») de l'équipe de Dr. Sarah Palmer (47), utilise une PCR NFL pour amplifier les provirus de quasi-pleine longueur. En comparaison aux études de Hiener et al. (47) et Lee et al. (88), il utilise la technologie PacBio pour le séquençage. Tout comme le Nanopore, cette technologie permet de séquencer des longs fragments (*reads*) avec un faible taux d'erreurs (102,103). Dans le protocole d'analyse de l'intégrité des provirus NFL, chaque séquence provirale est analysée pour la présence de six catégories de défauts dans le génome. Une hiérarchie est établie en fonction de la probabilité qu'un défaut compromette la capacité de réplication du provirus. L'ordre des défauts, du plus probable au moins probable, est le suivant : l'inversion, puis l'hypermutation, la grande délétion interne, le codon stop et/ou le décalage de lecture, le défaut dans Ψ et finalement, la petite délétion interne dans les protéines du VIH (dans l'ordre : Gag, Pol, Vif, Vpr, Tat, Rev, incluant RRE, Vpu, Nef et Env). Pour cette dernière catégorie, la protéine est dite « intacte » si elle possède un codon start (sauf pour Pol), ne possède aucun codon stop, aucun décalage du cadre de lecture et que le nombre de nucléotides délétés ne dépasse pas 5% de sa longueur. Si la séquence provirale est dépourvue de toutes ces six catégories défauts, elle est désignée comme « intacte ».

Décrites précédemment (42,82,90,101), les étapes en amont de l'analyse se résument à l'amplification par PCR nichée des provirus quasi-pleine longueur du VIH, au séquençage par PacBio, au démultiplexage pour associer chaque séquence à son échantillon, à l'obtention des séquences consensus par l'assemblage *de novo* des petits fragments, au nettoyage des séquences et

à l'alignement des séquences avec la séquence de référence HXB2 (sous-type B), à l'aide de l'outil MAFFT (104,105). Les inversions sont identifiées et corrigées manuellement avec Geneious Prime v2023.2.1 (https ://www.geneious.com/), car seuls les compléments inverses des inversions s'alignent avec HXB2. Les hypermutations sont détectées avec le HIV Database QC Tool (https://www.hiv.lanl.gov/content/sequence/QC/index.html) (106), tandis que les codons stop le sont avec une combinaison de ProSeq-IT (https ://psd.cancer.gov/tools/pvs_annot.php) (97) et du HIV Database Gene Cutter (https ://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html) (107). Les grandes délétions internes, les décalages du cadre de lecture, les défauts dans Ψ et l'intégrité des protéines (incluant RRE) sont également identifiés avec ProSeq-IT. La clonalité, elle, est analysée avec le HIV Database ElimDupes (https ://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html) (108). QC Tool, Gene Cutter, ProSeq-IT et ElimDupes sont tous utilisés en ligne.

1.2 La réponse immunitaire contre le VIH

1.2.1 La réponse innée

La première ligne de défense contre le VIH est la réponse immunitaire innée, qui est aspécifique. Le virus est détecté par le système immunitaire via les récepteurs de reconnaissance de motifs moléculaires (PRR) des cellules hôtes, tels que les récepteurs de type Toll (TLR) et les récepteurs de type RIG-I (RLR), reconnaissant les motifs moléculaires associés aux pathogènes (PAMP) du virus (109–112). La signalisation via les PRR permet entre autres d'induire l'expression de plusieurs protéines antivirales, incluant APOBEC3, TRIM5a, SAMHD1 et la tétherine, permettant de limiter la réplication du VIH (109). Cependant, à travers l'activité de ses protéines accessoires, le VIH a appris à contrer certaines de ces mesures antivirales.

La reconnaissance des PAMP entraîne aussi l'activation du système immunitaire et déclenche une réponse antivirale caractérisée par la production de cytokines comme les interférons (IFN) de type I et III, ainsi que des chimiokines et des cytokines proinflammatoires recrutant des cellules de l'immunité innée (neutrophiles, macrophages, les cellules NK et les cellules dendritiques) au site de l'infection (109). La persistance du VIH dans la phase chronique crée toutefois un environnement proinflammatoire qui a des effets détrimentaires sur les réponses innées et adaptatives (113).

1.2.2 La réponse adaptative

La réponse adaptative est la deuxième ligne de défense du système immunitaire et, contrairement à la réponse innée, est spécifique à un pathogène. Elle se développe habituellement quelques jours après l'infection (114). Il existe deux types d'immunités : l'immunité humorale, médiée par les anticorps, et l'immunité cellulaire, médiée par les lymphocytes T.

Les lymphocytes B jouent un rôle central dans la réponse humorale contre le VIH. La génération d'anticorps neutralisants, qui permettent de bloquer l'infection des cellules, sont au centre des stratégies vaccinales contre le VIH. Les anticorps neutralisants sont générés plusieurs mois après l'infection dans la majorité des cas, mais le virus échappe rapidement à la pression

immunitaire (115). Parmi les individus infectés, uniquement 1% d'entre eux développeront des anticorps largement neutralisants, permettant de cibler et neutraliser des centaines de souches différentes du VIH (116). Les lymphocytes T CD4[+] fournissent une aide cruciale aux lymphocytes B pour leur activation et leur différentiation en plasmocytes, forme différentiée des cellules B produisant les anticorps.

Les lymphocytes T possèdent un récepteur de cellules T (TCR) spécifique à un complexe CMH/peptide. Lors de l'infection au VIH, les cellules présentatrices d'antigènes (CPA) capturent les antigènes, se rendent aux ganglions lymphatiques et les présentent aux lymphocytes T CD4[+] et CD8[+] naïfs à travers leurs CMH de classe I (CD8[+]) ou de classe II (CD4[+]). Selon le « modèle à deux signaux », les signaux requis pour l'activation des cellules T sont : (i) la reconnaissance de l'antigène par le TCR (signal spécifique et dépendant de l'antigène) et (ii) l'engagement des molécules de costimulation retrouvées sur la cellule T et sur la CPA (signal aspécifique et indépendant de l'antigène) (117,118). La plus importante est CD28, critique pour l'activation des cellules T, leur prolifération et la production de cytokines. L'expression d'autres molécules de co-stimulation, telles que ICOS, OX40, CD40L et 4-1BB, est induite spécifiquement lors de l'activation. L'expression de cytokines proinflammatoires est souvent considéré le troisième signal pour la prolifération et la différentiation des cellules T (119).

La différentiation est typiquement séparée en trois phases : la phase d'expansion, la phase de contraction et la phase mémoire (120,121). Dans la phase d'expansion, initiée dans les tissus lymphoïdes, les cellules T prolifèrent et se différentient en cellules T effectrices, soit des lymphocytes T CD4[+] auxiliaires (ou *helper*, Th) ou des lymphocytes T CD8[+] cytotoxiques (CTL), pour sécréter des cytokines inflammatoires et tuer les cellules infectées. Dans la phase de contraction, la grande majorité (>90%) des cellules T effectrices meurent par apoptose (121). Dans la dernière phase, la phase de maintenance, les cellules T restantes se stabilisent en nombre et persistent pendant des années en tant que cellules « mémoires ».

Les lymphocytes T CD4[+] spécifiques au virus jouent un rôle clé dans la défense contre les infections virales. Les cellules T CD4[+] auxiliaires produisent des cytokines modulant la fonction d'autres cellules immunitaires, comme les cellules T CD8[+] ou les lymphocytes B. Ces cellules se différentient et sont typiquement divisées en sous-populations fonctionnelles basées sur la sécrétion de leurs cytokines, leurs facteurs de transcription et/ou leurs récepteurs de chimiokines. Parmi les

plus communes, il y a les lymphocytes Th1 (interféron [IFN]-γ]/Tbet/CXCR3), importants dans la réponse antivirale et contre d'autres pathogènes intracellulaires; les lymphocytes Th17/Th22 (IL-17 et IL-22/RORγT/CCR6), jouant un rôle dans l'immunité des muqueuses, en particulier celles des intestins; les lymphocytes T folliculaires helper ou Tfh (IL-21/Bcl-6/CXCR5), qui interagissent avec les cellules B pour réguler la réponse humorale; et les lymphocytes T régulateurs ou Treg (IL-10 et IL-35/FoxP3), contribuant à réguler l'autoimmunité (122–126).

De leur côté, les cellules T CD8[+] jouent un rôle important dans la défense contre les pathogènes intracellulaires, incluant les virus. Les lymphocytes T CD8[+] cytotoxiques sécrètent des perforines et des granzymes pour induire la lyse directe des cellules infectées (120,126,127). La génération des lymphocytes T CD8[+] mémoires implique entre autres l'aide des lymphocytes T CD4[+] auxiliaires (120,126,128). Dans l'infection au VIH, les cellules T CD8[+] spécifiques au virus sécrètent des cytokines comme l'IFN-γ, qui inhibe la réplication du VIH, ainsi que les chimiokines MIP-1α, MIP-1β et RANTES, qui suppriment la réplication virale en compétitionnant et/ou diminuant l'expression de CCR5 (128). Les cellules T CD8[+] spécifiques au VIH produisent des niveaux significativement plus bas de perforine que les cellules spécifiques au CMV, suggérant une activité cytotoxique affaiblie (129).

1.2.3 L'activation immunitaire chronique et la dysfonction du système immunitaire

L'infection par le VIH est caractérisée par une activation chronique et systémique du système immunitaire, qui peut être à la fois aspécifique (translocation microbienne) ou spécifique aux réponses immunitaires du VIH. Elle est associée à une progression de la maladie vers le SIDA (130). De plus, les individus sous traitement ART suppressif sont plus à risque de comorbidités et de mortalité, ce qui est associé à des niveaux plus élevés d'activation immunitaire (131,132). Le traitement ART réduit efficacement la virémie, mais n'élimine pas complètement l'inflammation.

1.2.3.1 L'activation immunitaire chronique aspécifique

Chez les PVIH, les lymphocytes T CD4$^+$ des muqueuses sont largement éliminés et ne se rétablissent que très peu, même dans la phase chronique, affaiblissant ainsi l'intégrité de la barrière épithéliale de leur intestin (4). Des microbes ou produits microbiens de la flore gastrointestinale peuvent alors traverser la barrière muqueuse intestinale pour atteindre la lamina propria, les ganglions lymphatiques mésentériques et éventuellement, la circulation systémique (133). C'est ce qu'on appelle la « translocation microbienne ». Chez des individus non-infectés, ces microbes seront phagocytés dans la lamina propria ou dans les nœuds lymphatiques mésentériques (134). La translocation microbienne est un facteur important de l'activation immunitaire persistante (l'inflammation) au cours de l'infection chronique au VIH (133,135–137). Il a d'ailleurs été montré que les niveaux élevés de lipopolysaccharides, des molécules constituant la membrane externe des bactéries à Gram négatif et un marqueur important de la translocation microbienne, étaient positivement associées avec l'activation du système immunitaire inné et adaptatif chez les PVIH (138). Malgré la thérapie ART, l'intégrité de l'intestin n'est que partiellement restaurée.

1.2.3.2 L'activation immunitaire chronique spécifique au VIH

Tel que décrite précédemment, l'expression des molécules de co-stimulation consiste au deuxième signal d'activation des lymphocytes T. En plus des molécules de co-stimulation « positives », il y a des molécules co-inhibitrices (aussi appelées « inhibiteurs des points de contrôle immunitaire », ICIs) comme l'antigène 4 du lymphocyte T cytotoxique (CTLA-4) ou la protéine 1 de la mort cellulaire programmée (PD-1) (118). Quelques heures à quelques jours après l'activation de la cellule T, dans des conditions normales, les ICIs sont transitoirement exprimés sur les cellules T effectrices pour réduire l'activation et les fonctions immunitaires (139). Lorsque le pathogène est éliminé, les niveaux d'expression des ICIs retournent en principe à la normale. Cependant, dans des contextes d'infections chroniques comme le VIH, où le pathogène n'est pas complètement éliminé, la stimulation chronique des cellules T avec des antigènes engendre l'épuisement des réponses spécifiques, menant à leur dysfonction.

Dans les lymphocytes T CD4$^+$ et CD8$^+$ spécifiques au VIH, l'activation immunitaire et l'exposition chronique aux antigènes est associée à une perte des fonctions, un état appelé « épuisement immunitaire ». Cela inclut une perte du potentiel de prolifération, une perte des fonctions effectrices, une faible production de cytokines et des niveaux plus élevés d'expression des ICIs, aussi considérés comme marqueurs d'épuisement (139–142). Les lymphocytes T CD4$^+$ et CD8$^+$ spécifiques au VIH expriment également des niveaux plus élevés de PD-1 et d'autres molécules co-inhibitrices comme CTLA-4 (143,144). Le blocage de PD-1 avec un anticorps PD-L1 bloquant a augmenté la prolifération des cellules (143–145) et la fréquence des cellules ayant des fonctions Th1, Th17 et Th22, mais pas celle des cellules Tfh (146).

Le concept d'épuisement immunitaire a davantage été décrit dans les lymphocytes T CD8$^+$ et dans la perte des fonctions cytotoxiques que dans les lymphocytes T CD4$^+$ spécifiques au VIH (139). Chez les cellules T CD4$^+$ spécifiques, l'activation chronique mènerait possiblement à une altération persistante de leurs fonctions (146–149). Des études transcriptomiques réalisées par notre groupe ont montré que les ECs, contrôlant spontanément la virémie, expriment fortement des gènes associés avec des fonctions Th1, Th17 et Th22, tandis que les CPs, virémiques, avaient un profil immunologique dévié vers des fonctions Tfh, avec de faibles expressions de gènes associés aux Th17 et Th22 (146,147). Une analyse longitudinale a permis de mettre en évidence que les niveaux d'expression des gènes associés aux fonctions Tfh étaient restaurés avec le traitement ART, contrairement à ceux associés aux fonctions Th1, Th17 et Th22 (147,148). D'autres études ont montré une altération dans la différentiation des cellules T spécifiques lors d'une virémie persistante, qui serait déviée vers un phénotype effecteur plutôt que mémoire (150). Ainsi, l'altération dans la différentiation des lymphocytes T CD4$^+$ spécifiques au VIH serait au centre de sa dysfonction et inclurait à la fois la perte, mais aussi le gain, de fonctions.

Plusieurs hypothèses pourraient expliquer la dysfonction persistante des lymphocytes T CD4$^+$ spécifiques au VIH, même sous traitement ART. Une étude récemment publiée par notre groupe montre l'existence d'expression spontanée d'ARN viraux et de protéines virales chez les individus traités depuis au moins trois ans, ce qui maintiendrait l'activation des réponses T CD4$^+$ et CD8$^+$ spécifiques au VIH (42). Une collaboration passée a permis de montrer la présence de cicatrices épigénétiques dans les cellules T CD8$^+$ spécifiques au VIH, spécifiquement sur le gène codant pour PD-1, répresseur des réponses immunitaires. La méthylation a été associée avec une

répression de la transcription. Youngblood et al. ont montré qu'avant le traitement ART, le gène codant pour PD-1 était déméthylé, mais qu'il demeurait déméthylé malgré plus de deux ans de traitement (151). La molécule co-inhibitrice PD-1 était donc prête à être réexprimée rapidement, et ce malgré une diminution de son expression. De manière similaire, en utilisant des échantillons longitudinaux, Martin et al. ont découvert que les profils d'accessibilité à la chromatine de cellules T CD8+ spécifiques étaient similaires entre l'infection primaire (6 mois après le début de l'infection) et la phase chronique de l'infection au VIH (un an sans traitement) (152). Ces résultats suggèrent que les mécanismes épigénétiques sont établis tôt au début de l'infection. Après un an de traitement ART, un grand nombre de régions accessibles à la chromatine étaient plus accessibles ou moins accessibles comparé au début de l'infection, indiquant ainsi que le traitement ART modifie partiellement l'épigénome des cellules T CD8+ spécifiques au VIH. L'existence de changements épigénétiques importants a aussi été trouvée dans les cellules T CD8+ spécifiques au virus de la choriomeningite lymphocytaire (VCML), dans le modèle murin (153,154). Cela pourrait suggérer que des mécanismes épigénétiques sont aussi impliqués dans la dysfonction des cellules T CD4+ spécifiques.

1.2.3.3 Les méthodes de détection des lymphocytes T CD4+ et CD8+ spécifiques à un antigène pour le tri de cellules vivantes

Pour accéder au transcriptome et à l'épigénome des cellules, il est nécessaire de les trier et de les séquencer. Toutefois, la détection des lymphocytes T CD4+ et CD8+ *ex vivo* est limitée par leur faible fréquence et l'hétérogénéité de leurs différents sous-types fonctionnels et mémoires (155,156). Quatre techniques sont couramment utilisées pour détecter ce type de réponses: le marquage intracellulaire de cytokines (« *intracellular cytokine staining* », ICS), le test d'ELISpot (« *enzyme-linked immunosorbent spot* »), l'essai de marqueurs induits par l'activation (« *activation-induced marker assay* », AIM) et l'essai de tétramère.

L'ICS et l'ELISpot identifient tous deux les lymphocytes T spécifiques à un antigène en se basant sur la détection de cytokines. Cependant, certains types de cellules T CD4+, tels que les Tfh, ont des cytokines plus difficiles à détecter, si bien que ce type de cellules est souvent non détecté ou sous-représenté dans un ICS ou un ELISpot (155,157,158). L'ICS est limité par l'utilisation

41

d'inhibiteurs de sécrétion de cytokines comme la monensine ou la Brefeldin A (BFA), utilisés pour permettre l'accumulation intracellulaire des cytokines normalement sécrétées (159–161), mais qui peuvent changer l'expression des gènes. De plus, l'utilisation de réactifs pour fixer et perméabiliser les cellules entraîne leur mort et empêche l'utilisation de l'ICS pour des analyses subséquentes comme la transcriptomique (156). De son côté, l'ELISpot n'est pas un essai caractérisant les cellules uniques et ne permet pas de phénotyper les cellules.

L'essai AIM a été développé en partie pour surmonter les défis reliés à la production de cytokines. L'essai AIM quantifie les cellules T CD4$^+$ et CD8$^+$ augmentant l'expression de leurs marqueurs de surface (communément appelés « marqueurs induits par l'activation » ou « AIMs ») après une stimulation TCR avec les antigènes d'intérêts. Les cellules qui ont rapidement exprimé ces marqueurs induits par l'activation sont identifiées comme « spécifiques » à l'antigène. Toutefois, l'essai AIM demeure un essai fonctionnel : dans le contexte de l'infection chronique au VIH, une cellule épuisée qui ne produit pas de cytokines et/ou qui n'augmente pas l'expression de certains de ses marqueurs demeurera indétectable.

Contrairement à l'ICS, l'ELISpot et l'essai AIM, trois essais fonctionnels fondés sur la réponse des cellules T suite à leur stimulation (162), l'essai de tétramère est indépendant de la fonctionnalité. L'essai de peptide-CMH tétramère (simplement appelé « essai de tétramère » ou « marquage de tétramère ») mesure la fréquence de cellules T CD4$^+$ et CD8$^+$ qui possèdent un TCR spécifique à un complexe peptide-CMH donné. Cependant, les tétramères doivent être histocompatibles aux molécules de HLA de chaque individu et ne peuvent reconnaître que des épitopes spécifiques, nécessitant ainsi un tétramère par spécificité étudiée (163,164). Bien qu'ils soient communément utilisés pour détecter les cellules T CD8$^+$ spécifiques, marquer les cellules T CD4$^+$ est plus difficile, car les complexes peptide-CMH de classe II sont moins stables et plus difficiles à produire (165).

La réponse T CD4$^+$ étant très hétérogène, cibler une ou quelques cytokines (ICS ou ELISpot), ainsi qu'un seul épitope de la spécificité étudiée (essai de tétramère), sous-estime probablement l'ampleur de la réponse T spécifique à un antigène (155,166,167). En comparaison, l'essai AIM se concentre sur le marquage des molécules de surface et ne nécessite pas de traitement à la monensine/BFA et aux réactifs de fixation et de perméabilisation. L'avantage principal de cet essai est donc de garder les cellules en vie. Plusieurs marqueurs induits par l'activation ayant été

utilisés par notre groupe (42,146–148,155,168–171) et par d'autres (172–177) seront détaillés dans l'article « *Enhanced detection of antigen-specific T cells by a multiplexed AIM assay* » (178) de la **section 4.2**.

Chapitre 2 – Hypothèse et objectifs

Bien que la thérapie ART réprime la réplication du VIH, son arrêt mène dans la vaste majorité des cas à un rebond viral. Cela indique que, d'un côté, les traitements actuels ne peuvent complètement éliminer les réservoirs du VIH et, de l'autre, que les réponses immunitaires ne suffisent pas à contrôler ce regain de réplication virale. Le développement de méthodes et d'outils pour l'analyse virologique et immunologique de l'infection au VIH est donc primordial pour une meilleure compréhension de la persistance de ces réservoirs viraux et de la dysfonction des cellules T spécifiques au VIH.

Tel que discuté plus tôt dans l'**avant-propos**, ce mémoire présentera deux projets réalisés dans deux laboratoires différents, avec deux hypothèses distinctes. Du point de vue virologique, une limite à l'analyse de l'intégrité des provirus du VIH est la perte d'information due à la hiérarchisation des défauts (47,69,88,90–92). Puisque chaque séquence défectueuse est classifiée pour un seul défaut, les séquences qui possèdent plusieurs défauts ne sont pas intégralement représentées. Ainsi, des défauts analysés subséquemment dans la hiérarchie sont perdus et possiblement sous-représentés. Nous avons émis comme **première hypothèse** que l'algorithme développé dans ce projet fournirait des informations additionnelles comparée aux algorithmes actuels utilisés pour l'évaluation de l'intégrité des provirus du VIH.

Du point de vue immunologique, des résultats récents dans la dysfonction des cellules T CD8$^+$ pourrait suggérer que des mécanismes épigénétiques sont aussi impliqués dans la dysfonction des cellules T CD4$^+$ spécifiques au VIH. Nous avons émis comme **deuxième hypothèse** que le biais dans la différentiation des lignées des lymphocytes T CD4$^+$ spécifiques au VIH, causée par leur activation chronique, est uniquement partiellement réversible dû à des mécanismes épigénétiques.

Objectifs

Projet de l'analyse virologique de l'infection au VIH

1. Développer un algorithme R pour mesurer l'intégrité et la clonalité des provirus du VIH

2. Valider l'algorithme avec des données existantes

3. Comparer l'algorithme avec des outils existants

Projet de l'analyse immunologique de l'infection au VIH

1. Optimiser une stratégie exhaustive d'identification des cellules T CD4$^+$ et CD8$^+$ spécifiques au VIH

2. Optimiser une stratégie de tri cellulaire et de préparation des échantillons

3. Cartographier les mécanismes épigénétiques et transcriptionnels impliqués dans la dysfonction des cellules T spécifiques au VIH (en collaboration avec le laboratoire de Dr. Jeremy M. Boss, à l'Université d'Emory)

Chapitre 3 – Développement et application d'un algorithme pour analyser l'intégrité et la clonalité des provirus quasi-pleine longueur du VIH

3.1 Avant-propos

Dans le chapitre ci-dessous, je décrirai le projet réalisé dans le laboratoire Chomont, soit le développement d'un algorithme pour analyser l'intégrité et la clonalité des provirus quasi-pleine longueur du VIH. Cet algorithme, appelé « IntegrityAlgorithm », automatise un protocole d'analyse développé par Dre Caroline Dufour, tel que décrit plus tôt dans le chapitre 1. Cependant, ces analyses étaient longues et devaient être faites « manuellement » avec de nombreuses étapes de copier-coller dans des chiffriers Excel. Un des objectifs de cet algorithme était donc de faciliter et centraliser les analyses.

Dans la **section 3.2**, celle des méthodes, l'algorithme lui-même sera décrit. Dans la **section 3.3**, celle des résultats, nous avons comparé les résultats produits par l'algorithme avec (i) des résultats publiés par notre groupe et (ii) des résultats produits par d'autres algorithmes fréquemment utilisés (ProSeq-IT, HIVIntact, HIVSeqinR et ElimDupes). Les **sections 3.2** et **3.3** sont écrites en anglais, car elles seront utilisées dans des articles présentement en préparation, dont celui de Dufour et al. (voir l'**Annexe 9**).

3.2 Methods

3.2.1 The intactness of HIV proviruses

3.2.1.1 Description of the algorithm

We developed an R package to computerize the pipeline previously described (42,82,90,101). Each near full-length (NFL) provirus undergoes analysis to identify the presence of: (i) inversions, (ii) hypermutations, (iii) large internal deletions, (iv) stop codons, (v) Ψ defects, and (vi) small internal deletions in the Gag, Pol, Vif, Vpr, Tat, Rev (including the RRE), Vpu, Nef, and Env ORFs. If a proviral sequence does not exhibit any of these six types of defects, it is inferred as "intact" and thus potentially replication competent.

The pipeline for analyzing HIV proviral integrity is summarized in **Figure 3.2.1**. Users begin with a FASTA file containing their inversions-corrected, primers-free sequences, aligned to the reference sequence HXB2. Before running the algorithm, users submit this file to two online tools: HIV Database QC Tool and Gene Cutter. Both tools provide a hyperlink that is sent to their email address. Additionally, the QC Tool generates a summary text file than can be downloaded. Users should duplicate their FASTA file and remove the reference sequence, or rename it to "Reference_sequence," as ProSeq-IT does not allow the reference sequence to contain "HXB2" or its accession number (K03455). After submitting the duplicated FASTA file, ProSeq-IT generates an Excel file containing over 50 annotations for each sequence. These annotations include, but are not restricted to, the number of nucleotides inserted or deleted in the ORFs, and the presence or absence of frameshifts.

Since inversions can only be identified manually, we created an Excel template that can be read by the algorithm. This Excel template contains three tabs: "ProseqIT_criteria," "Manual_assessment," and "Hyperlinks." The "ProseqIT_criteria" tab is locked and solely used by the algorithm to analyze ProSeq-IT's results. In the "Manual_assessment" tab, users fill out the names of the sequences in the "Name" column and write "Y" for inversions or "N" otherwise in the "Inversions" column. Finally, users fill out the "Hyperlinks" tab with QC Tool and Gene Cutter's results hyperlinks.

After that, the results of QC Tool, Gene Cutter, and Geneious (for the inversions), through the Excel template, along with ProSeq-IT's, are input in the algorithm. To analyze the intactness of HIV proviruses, we developed an R (v4.3.2) function called `HIV_IntegrityAnalysis`. The algorithm first analyzes QC Tool's results to detect hypermutations (step #1, **Figure 3.2.1**). Then, it processes Gene Cutter's results by examining each of the nine ORFs to confirm the presence of start codons (AUG or methionine, "M") and any premature stop codons (UAA, UAG, or UGA) (step #2, **Figure 3.2.1**). Subsequently, it analyzes ProSeq-IT's results using criteria based on Dre Sarah Palmer's work (47,179) (step #3, **Figure 3.2.1**). Specifically, the algorithm identifies large internal deletions in sequences with lengths <8800 bp without primers. It also checks the Ψ region for deletions and mutations in the MSD. Users should manually confirm the Ψ defects by looking at the alignments with the reference HXB2, as previously stated (90). Additionally, it evaluates each of the nine ORFs for insertions, deletions (tolerated up to 5% of their length), and frameshifts. Start (except for Pol) and stop codons are included in the "small internal deletions," but have already been identified with Gene Cutter. However, while stop codons in Tat exon 2 and all defects in the Nef protein are documented, they are not considered critical for the intactness (34,179,180).

The R function uses up to six parameters, or "arguments," categorized either as "required" or "optional." Three of these arguments are required, meaning that they must be provided by users for the R function to work properly, while the remaining three are optional. They have default values and do not need require explicit user input. **Table 3.2.1** provides a summary of the required and optional arguments used in the `HIV_IntegrityAnalysis` function.

The algorithm creates two folders: one named "tmp" (temporary) and another named "FINAL_OUTPUT" (step #4, **Figure 3.2.1**). The temporary folder contains individual CSV file for each analyzed tool ("Analyzed_QCTool.csv," "Analyzed_GeneCutter.csv," and "Analyzed_ProseqIT.csv"), along with a detailed summary file ("intactness_detailedsummary.csv") that lists all types of defects found. In this file, defects are not hierarchized, meaning that one sequence can be classified for more than one type of defect. In contrast, the "FINAL_OUTPUT" folder contains the "intactness_summary.csv" file, which lists hierarchized defects. Both summary files infer the intactness ("intact" or "defective") of the HIV proviruses. We used a binary code where "0" and "1" represent the absence and presence of a defect, respectively. Examples of these files can be found in **Appendices 1** to **5**.
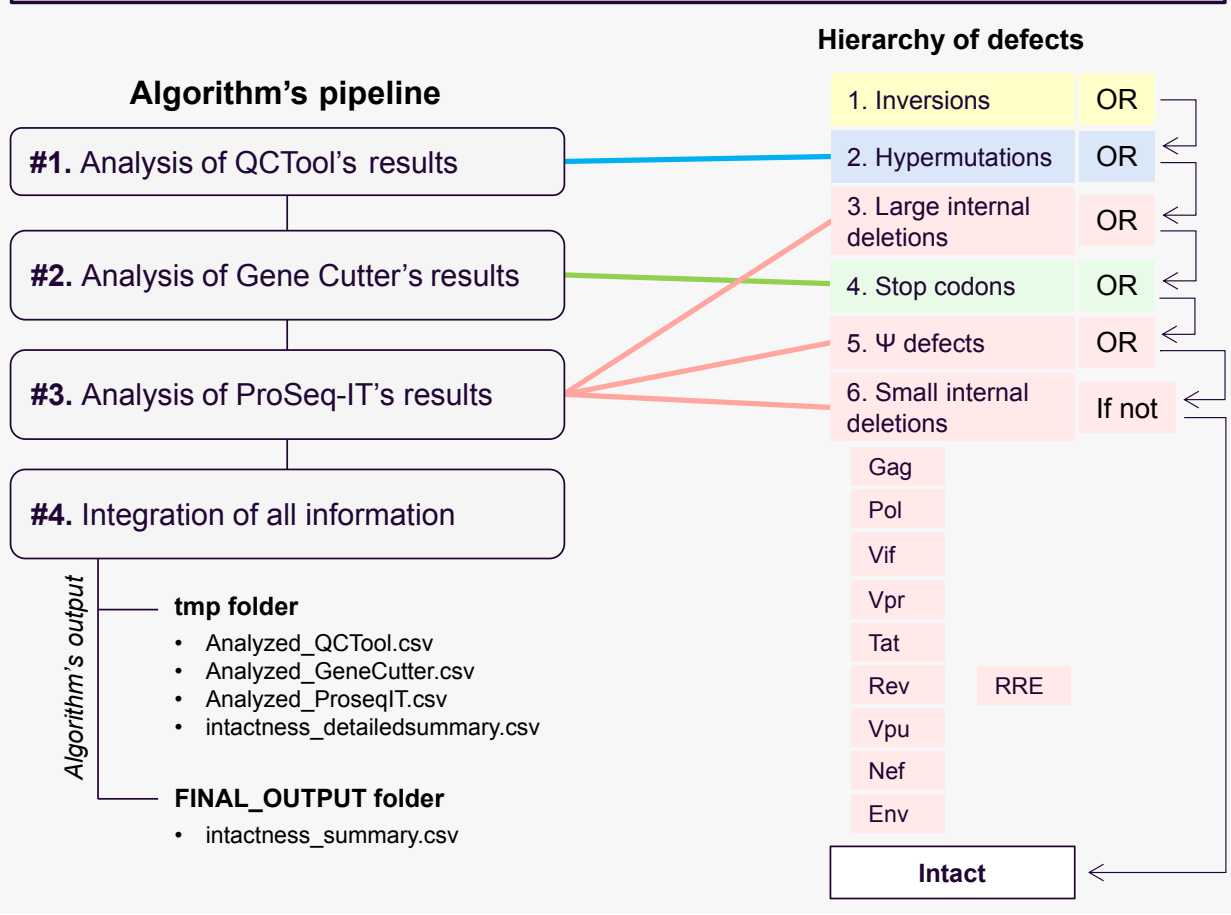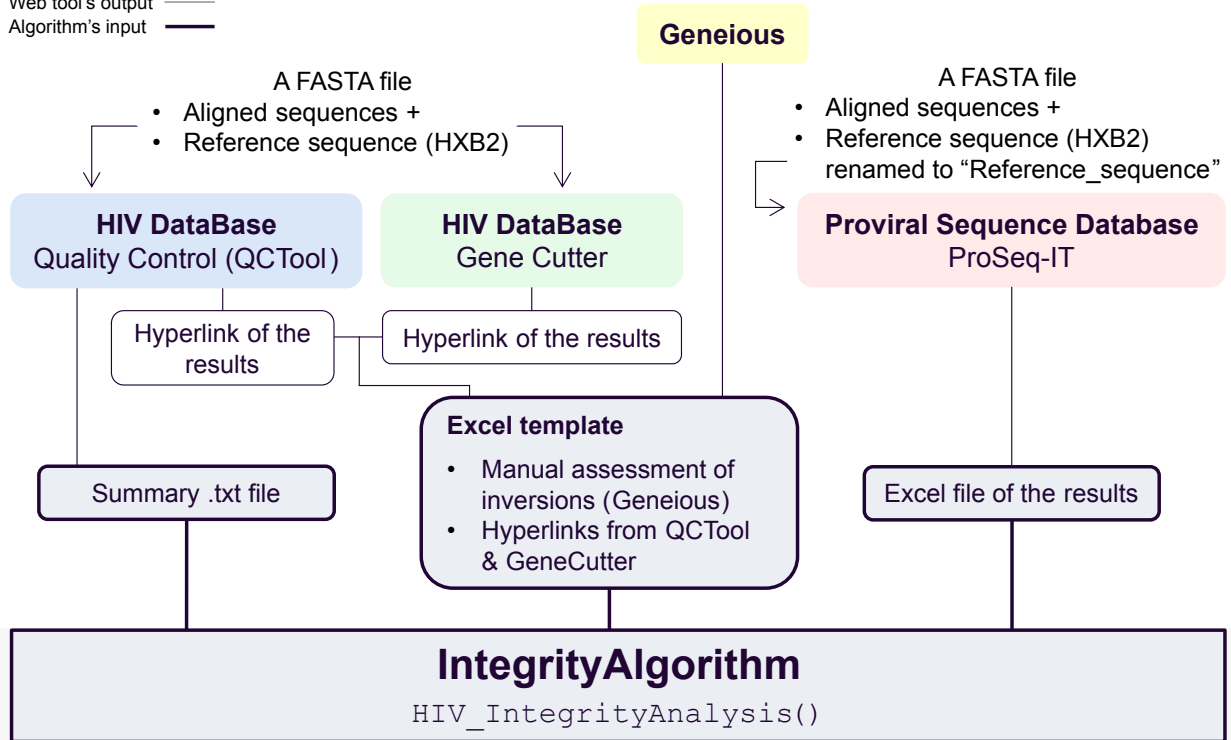
Figure 3.2.1 – Pipeline of the HIV proviral intactness analysis. The name of the R function to assess the HIV proviral intactness is called "`HIV_IntegrityAnalysis`." Different colours represent the tools used for defect identification: blue (QC Tool), green (Gene Cutter), yellow (Geneious Prime), and pink (ProSeq-IT). Dotted lines represent the outputs of the different web tools, while full lines represent inputs to the algorithm. The algorithm generates four files in a temporary ("tmp") folder and a single file in the "FINAL_OUTPUT" folder. The temporary folder contains individual CSV files for each tool's analysis output, and a detailed summary file listing all defects found. The "FINAL_OUTPUT" folder includes a summary file with hierarchized defect information, indicating the intactness ("intact" or "defective") of HIV proviruses.

Table 3.2.1 – Required and optional arguments of the `HIV_IntegrityAnalysis` function.

Argument	Required/Optional	Default value	Comment
template_filename	Required	-	Excel template containing the inversions assessment and QC Tool's and Gene Cutter's results Hyperlinks
QCTool_summary	Required	-	Text file containing the results from QC Tool
ProseqIT_rx	Required	-	Excel file containing the results from ProSeq-IT
ProseqIT_RefSeq	Optional	TRUE	If TRUE, the reference sequence (HXB2) is included in ProSeq-IT's results. It should be renamed to "Reference_sequence."
RefSeq	Optional	TRUE	If TRUE, the reference sequence (HXB2) is included in QC Tool and GeneCutter's results.
Analyzes	Optional	5	Specifies the analyzes to be done. 1: QC Tool only; 2: Gene Cutter only; 3: ProSeq-IT only; 4: IntegrateInfo only; 5: All.

3.2.1.2 Discrepancies between tools

While developing this algorithm, we observed discrepancies between the different tools commonly used. For stop codons, QC Tool, Gene Cutter, and ProSeq-IT produced slightly different results. Using Dubé et al.'s dataset (42) and the results from QC Tool, we calculated the number of ORFs containing stop codons for each sequence. We repeated this analysis for Gene Cutter and ProSeq-IT, then computed the pairwise differences between the algorithms. Out of 40 sequences,

discrepancies between QC Tool and Gene Cutter occurred in 21 sequences (52.5%) and in 25 sequences (62.5%) between QC Tool and ProSeq-IT (**Figure 3.2.2**). There were fewer differences (10 sequences, 25%) in the number of ORFs with stop codons between Gene Cutter and ProSeq-IT. QC Tool appears to underestimate the number of stop codons compared to Gene Cutter and ProSeq-IT.

A combination of Gene Cutter and ProSeq-IT was previously used to detect premature stop codons (42,82,90,101). However, since ProSeq-IT does not account for alignments, while Gene Cutter suggests using pre-aligned sequences with the reference sequence to yield better results, we decided to exclusively rely on Gene Cutter for identifying stop codons in the algorithm.
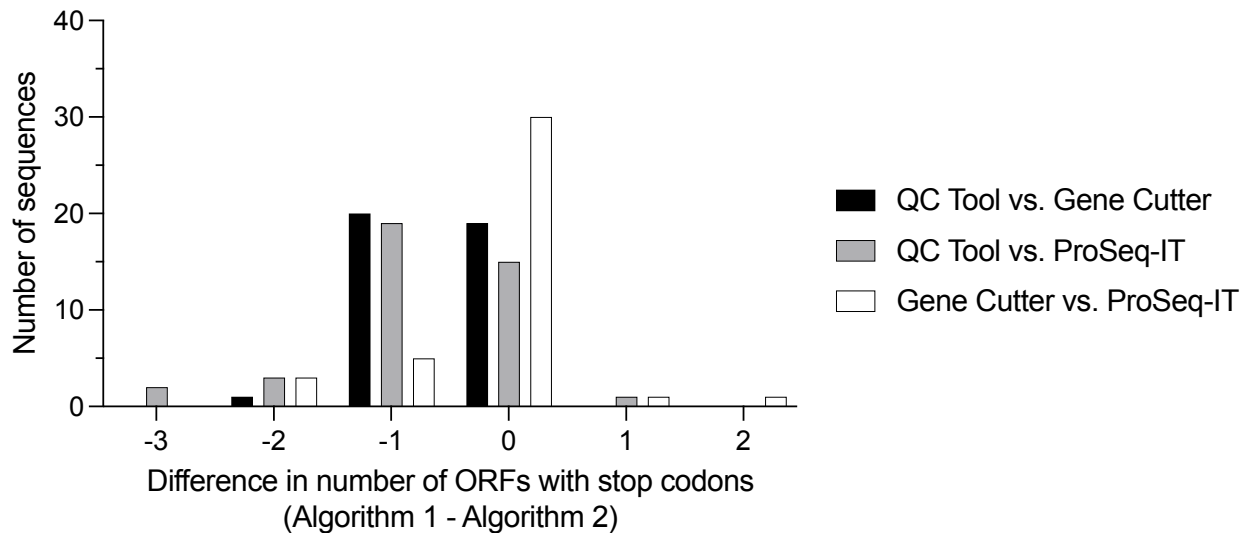
Figure 3.2.2 – Discrepancies in stop codon identification across algorithms. The number of open reading frames (ORFs) containing stop codons were calculated for each sequence based on the results of the different algorithms. Differences between the algorithms, calculated as Algorithm 1 – Algorithm 2, were then computed. Negative values indicate that Algorithm 2 found more stop codons than Algorithm 1, whereas positive values indicate that Algorithm 1 found more stop codons than Algorithm 2. A null value indicates that both algorithms identified the same number of stop codons.

We also found discrepancies between three tools commonly used to identify hypermutations: the HIV Database QC Tool, the HIV Database Hypermut (181), and the R package

HypermutR (https://github.com/philliplab/hypermutR) (182). These algorithms aim to detect hypermutations induced by APOBEC, using a one-sided Fisher's exact test. Both QC Tool and HypermutR are derived from Hypermut, with QC Tool using a simplified version of Hypermut and HypermutR being specifically developed for R implementation. Testing Dubé et al.'s dataset (42) on the three algorithms, we found that QC Tool and Hypermut identified only one and two possible hypermutants, respectively, out of 40 sequences. With HypermutR, however, eight sequences had been flagged as "potentially" hypermutated when consensus sequences are computed. When the algorithm considers HXB2 as the "ancestral" sequence, the differences become even more striking, with 25 sequences classified as possible hypermutants. Consequently, hypermutations might be overestimated in this R implementation.

Although QC Tool missed one possible hypermutant, it did not change the inference of the sequence's intactness, which was initially identified as inverted. Moreover, Hypermut cannot be automated without resorting to copying-and-pasting the entire results page. Due to these discrepancies and the fact that Hypermut and HypermutR have not been validated with the current NFL protocol (42,82,90,101), we chose QC Tool to detect hypermutations in the algorithm.

3.2.2 The clonality of HIV proviruses

3.2.2.1 Description of the algorithm

As previously described (42,82,90,101), the current pipeline defines two sequences as clonal if the number of different nucleotides equals 0. The pipeline for analyzing HIV proviral clonality is summarized in **Figure 3.2.3**. Similar to the integrity analysis, users start with a FASTA file containing their inversions-corrected, primers-free sequences aligned to the reference sequence HXB2. Since clonality is assessed within participants, the user also needs to provide a list of participant identifications (PID).

To analyze the clonality of HIV proviruses, we developed an R function called `Clonality_Analysis`. The FASTA file is first split according to the list of PIDs supplied (step #1**, Figure 3.2.3**). Then, for each participant, the algorithm generates a matrix of sequence length (step #2**, Figure 3.2.3**) and creates a list of "potential clones" with same-length sequences (step #3**,

Figure 3.2.3). Finally, clones and potential clones are confirmed by computing pairwise sequence alignments using MAFFT in the R package ips (v0.0.11) (183) (step #4**, Figure 3.2.3**).

To our knowledge, previous tools such as ElimDupes only considered the sequence coverage to determine whether two sequences are "duplicates" (or clones). In the `Clonality_Analysis` function, we added a threshold for the number of different nucleotides, which can be adjusted by the users. The default threshold value is 1, based on previous data indicating an extremely low error rate in PacBio sequencing, with 1 error every 34,025 nucleotides (90). Alongside the list of "confirmed clones," this threshold allows for the creation of a list of "confirmed potential clones," defined as two sequences with at least one different nucleotide but fewer than the specified threshold.

This R function takes up to three arguments (two required and one optional), which are summarized in **Table 3.2.2**. The algorithm creates the subfolder "Clonality" within the "FINAL_OUTPUT" folder. A file is generated for each participant (e.g., "PID_ClonalityAnalysis.csv"), containing the list of unique sequences, clones, and potential clones. This can be manually confirmed afterwards using Geneious Prime. An example of this file can be found in **Appendix 6**.

Table 3.2.2 – Required and optional arguments of the `Clonality_Analysis` function.

Argument	Required/Optional	Default value	Comment
FASTA_file	Required	-	FASTA file containing all primers-free, aligned sequences, including the reference sequence (HXB2).
Participants	Required	-	A string vector containing the list of participants. If more than one, use the `c()` function (for example, `c("PID1", "PID2")`.
Threshold	Optional	1	Threshold number of different nucleotides to consider two sequences as "potential clones."
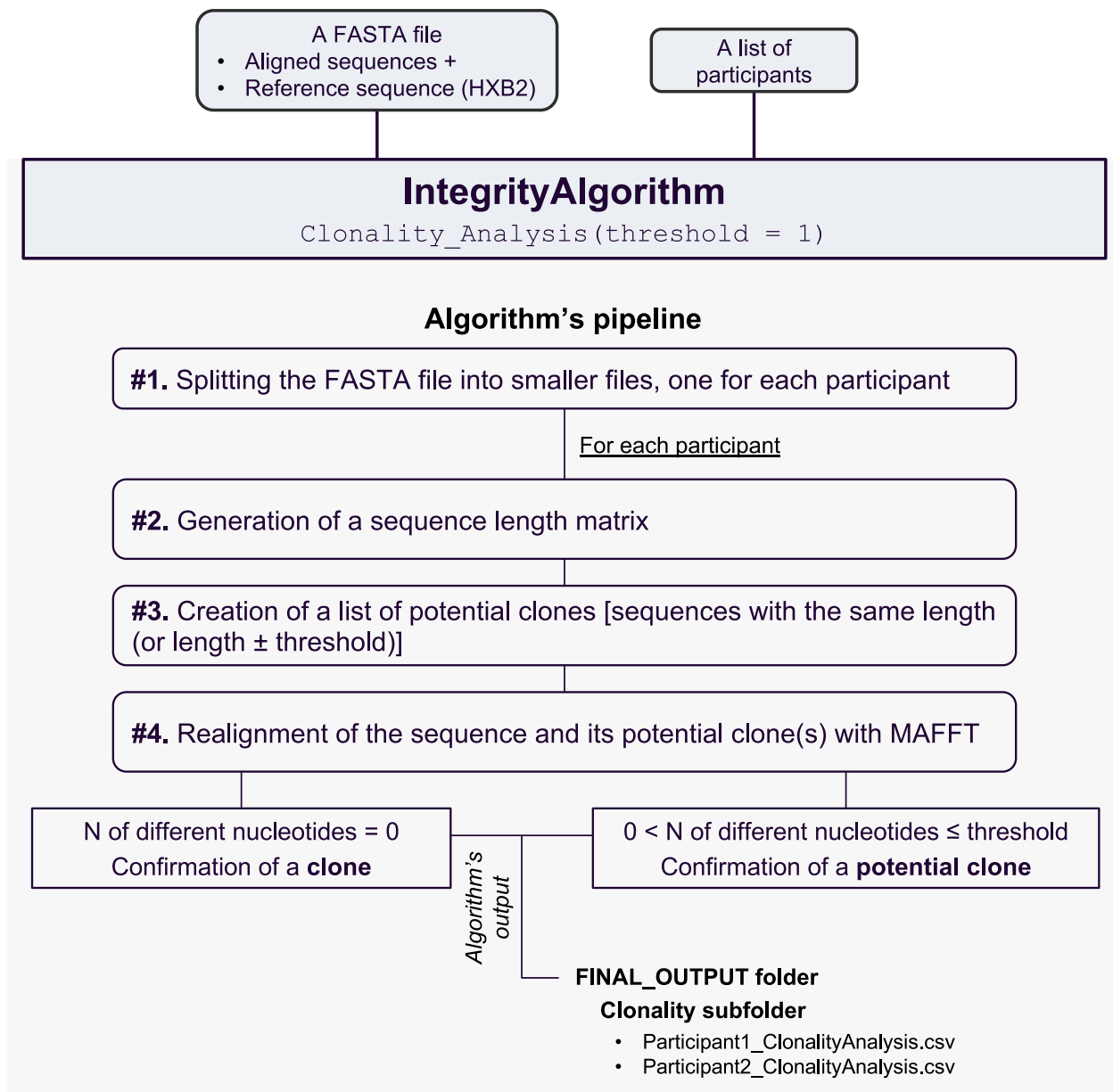
Figure 3.2.3 – Pipeline of the HIV proviral clonality analysis. The name of the R function to assess the HIV proviral clonality is called "Clonality_Analysis". The algorithm outputs, in the "FINAL_OUTPUT" folder, a subfolder called "Clonality" that contains one file for each participant identification provided.

3.2.3 Installation and availability

The R package "IntegrityAlgorithm" is available on Github at https://github.com/alemi055/IntegrityAlgorithm. It can be downloaded and updated directly through R, using the devtools R package (v2.4.5) (184) with the following command: `install_github("alemi055/IntegrityAlgorithm")`.

3.3 Results

The algorithm produces results that are consistent with the published datasets

To assess the algorithm's performance, we compared its results with four publications that used the same NFL protocol. We first conducted a comprehensive analysis of Dufour et al.'s HIV proviruses (90), which was the largest dataset available. In this paper, the authors characterized 635 HIV genomes retrieved from 309 inducible and translation-competent (p24$^+$) and 326 non-induced and/or translation-incompetent (p24$^-$) cells across six participants on suppressive ART. HIV reactivation was induced using PMA/Ionomycin. A sequence that was not included in the paper (308 genomes from p24$^+$ cells in the paper instead of 309) but published on GenBank was included in this analysis. Less than 4% of proviruses were found to be intact in p24$^+$ and p24$^-$ cells (**Figure 3.3.1A**), the majority of p24$^+$ cells being defective due to Ψ defects.

We downloaded the primers-free sequences of this dataset publicly available on GenBank (accession numbers ON816029 to ON816663), corrected the inversions, and re-aligned them with MAFFT, as previously described (42,82,90,101). The sequences were input into the IntegrityAlgorithm using the methods described in **section 3.2**. The algorithm identified the same number of intact genomes (**Figures 3.3.1A and 3.3.1B**), with only a few differences in the classification of defects for large internal deletions, stop codons, and Ψ defects. Six "largely deleted" sequences had a length of 8,799 bp, which is one nucleotide below the 8,800 bp threshold.

However, 117 and 105 Ψ defects in p24$^+$ and p24$^-$ cells were not counted due to the hierarchization of defects. Based on the algorithm's data, this strongly increases the proportion of sequences with Ψ defects from 58.3% to 96.1% (p24$^+$) and 5.8% to 38.0% (p24$^-$) (**Figure 3.3.1B**). To further assess the proportion of sequences with more than one defect, we computed all 64 ($2^6 =$ 64) theoretical combinations. We found that 100% of defective p24$^+$ proviruses had a Ψ defect, with a large portion of sequences (40.1%) being defective for the Ψ only. The remaining defective 56% all had at least one defective ORF (**Figure 3.3.1C**). In p24$^-$ cells, while 77.6% of sequences were classified as "largely deleted" in the hierarchy of defects, these sequences also harbored stop codons and/or defective Ψ, as well as small internal deletions.

A

Data published in Dufour et al., *Nat Commun.* (2023)
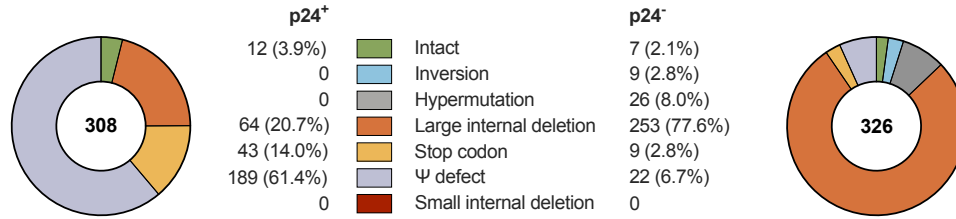
	p24+		p24-
Intact	12 (3.9%)		7 (2.1%)
Inversion	0		9 (2.8%)
Hypermutation	0		26 (8.0%)
Large internal deletion	64 (20.7%)		253 (77.6%)
Stop codon	43 (14.0%)		9 (2.8%)
Ψ defect	189 (61.4%)		22 (6.7%)
Small internal deletion	0		0

308 | 326

B

IntegrityAlgorithm: summary of all defects

p24	+	-	+	-	+	-	+	-	+	-	+	-	+	-
n hierarchized defects	12 3.9%	7 2.1%	0	9 2.8%	0	26 8.0%	71 23.0%	253 77.6%	46 14.9%	12 3.7%	180 58.3%	19 5.8%	0	0
n defects	12 3.9%	7 2.1%	0	9 2.8%	0	27 8.3%	71 23.0%	275 84.4%	63 20.4%	99 30.4%	297 96.1%	124 38.0%	173 56.0%	306 93.9%

C

IntegrityAlgorithm: sequences with more than one defect

57
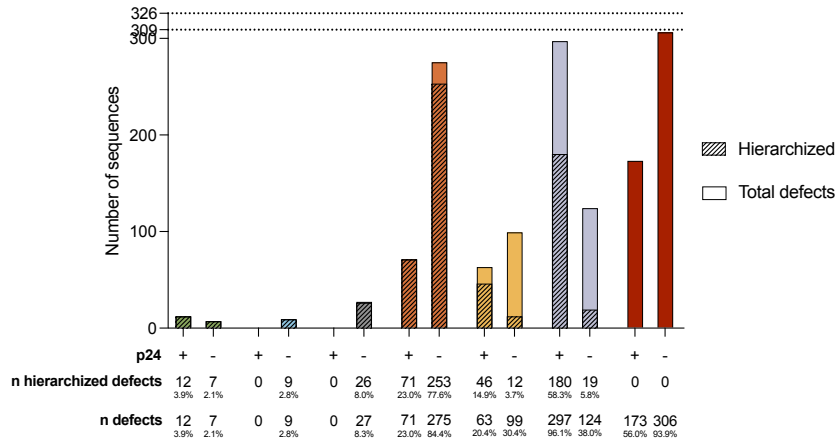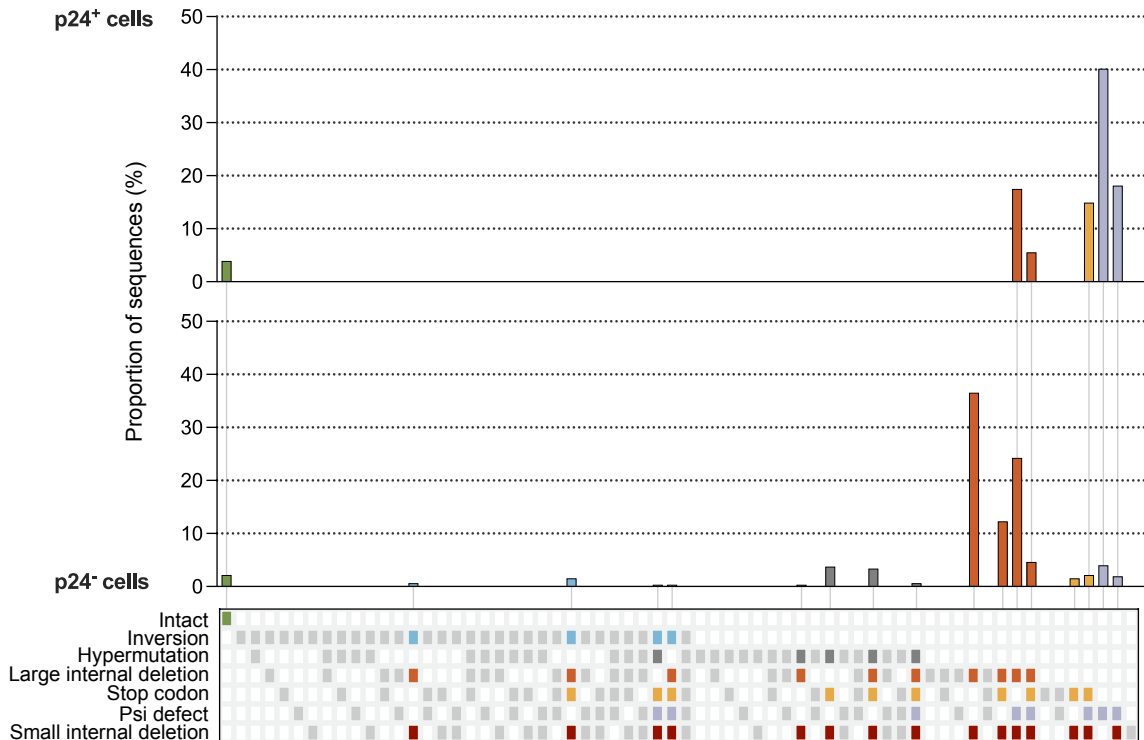
Figure 3.3.1 – Proviral landscape of p24$^+$ and p24$^-$ HIV genomes. (**A**) Data previously published in Dufour et al., *Nat. Commun.* (2023). An additional p24$^+$ sequence was found on GenBank compared to the results published in the paper. (**B**) Exhaustive list of defects provided by the IntegrityAlgorithm. The hatched bars represent the number of hierarchized defects, while the bars without pattern represent the total number of defects. The numbers and proportions of hierarchized and total defects are shown below. (**C**) Proportion of sequences being intact or having the combinations of defects. The combination of defects that each column refers to is indicated underneath the histograms. Coloured boxes indicate that at least one sequence has this combination, while grey boxes indicate that none of the sequences has this combination of defects. The bars are coloured by their main defect. n = 309 (or 308 for Dufour et al.) p24$^+$ and n = 326 p24$^-$ proviruses retrieved across six participants on suppressive ART.

Considering that a high proportion (93.9%) of p24$^-$ proviruses had small internal deletions, we further looked at the proportion of defective p24$^+$ and p24$^-$ proviral sequences that had intact regulatory regions (Ψ and RRE) and ORFs (Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Nef, and Env). Total defects were considered for this analysis. Most inverted and hypermutated p24$^-$ proviruses had an intact Ψ region, but their ORFs were almost all defective (**Figure 3.3.2**). Large internal deletions in translation-competent proviruses were located in the Ψ locus and the proteins Gag and Pol, as well as Nef, to a fewer extent. In p24$^+$ cells, few sequences with premature stop codons also had an intact Ψ or Gag protein. Interestingly, a high proportion of sequences with Ψ defects had intact ORFs. In comparison, almost all the ORFs of p24$^-$ sequences with large internal deletions and/or stop codons were defective. Defects were particularly located in the Gag and Pol ORFs for p24$^-$ proviruses with Ψ defects.

These results are a proof of concept of different levels of information that can be obtained with the IntegrityAlgorithm. Using this large dataset (n = 635 proviral sequences), we showed that a significant proportion of Ψ defects are missed due to the hierarchization of the defects.
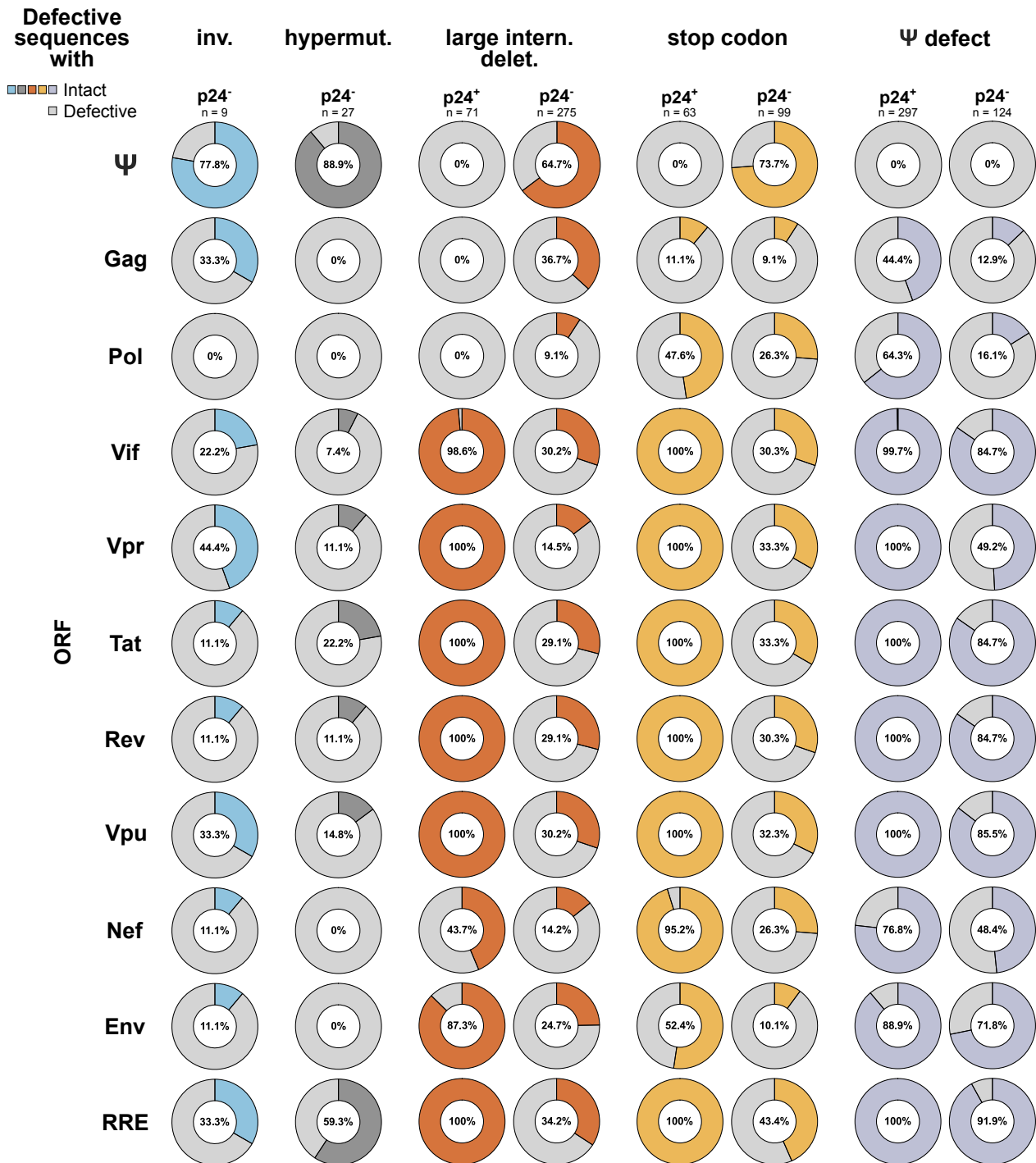
Figure 3.3.2 – Proportion of defective p24⁺ and p24⁻ proviruses having intact regulatory regions or ORFs. Number of intact regulatory regions (Ψ and RRE) and ORFs (Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Nef, and Env) in sequences that are defective for inversions, hypermutations, large internal deletions, stop codons, and/or a Ψ defects. Total defects provided by the IntegrityAlgorithm are considered for this analysis. Coloured and grey parts represent the proportion of sequences with intact and defective ORFs,

respectively. The proportion of sequences with intact ORFs is also computed in the middle of the pie charts. n = 309 p24$^+$ and n = 326 p24$^-$ proviruses retrieved across six participants on suppressive ART.

Additionally, we performed a comparative analysis on three other datasets to demonstrate the consistency and comparability of the developed tool. The second dataset consisted of proviral sequences obtained from postmortem collected tissues of two participants with HIV (101). This included the mediastinal, hilar, axillary, inguinal, and mesenteric lymph nodes, as well as the duodenum, jejunum, ileum, colon, rectum, liver, testes, lungs, and spleen. Consistent with the data published in the paper, the IntegrityAlgorithm detected the same number of intact, inverted, hypermutated, and largely deleted sequences (**Figure 3.3.3A**). Stop codons and small internal deletions were the defects most frequently missed due to the hierarchization (**Figure 3.3.3B**), with 100% of sequences that had inversions, hypermutations, large internal deletions, or stop codons also having small internal deletions (**Figure 3.3.3C**). Only 15.2% of sequences classified defective due to a Ψ defect had small internal deletions.

A Comparison: Dufour et al., *Cell Rep.* (2023) vs. IntegrityAlgorithm

Paper		IntegrityAlgorithm
41 (9.3%)	Intact	41 (9.3%)
4 (0.9%)	Inversion	4 (0.9%)
116 (26.3%)	Hypermutation	116 (26.3%)
215 (48.8%)	Large internal deletion	215 (48.8%)
20 (4.5%)	Stop codon	9 (2.0%)
42 (9.5%)	Ψ defect	46 (10.4%)
3 (0.7%)	Small internal deletion	10 (2.3%)

441 / 441

B IntegrityAlgorithm: summary of all defects

Hierarchized / Total defects

Net difference (total - hierarchized): 0 0 0 45 146 68 351

C IntegrityAlgorithm: sequences with more than one defect

Classification of sequence	int.	inv.	hypermut.	LID	stop	Ψ	SID
n	41	4	116	215	9	46	10

60

Figure 3.3.3 – Proviral landscape of HIV genomes retrieved from postmortem collected tissues. (A) Comparison between the data published in Dufour et al., *Cell Rep.* (2023) and the results produced by the IntegrityAlgorithm. **(B)** Exhaustive list of defects provided by the IntegrityAlgorithm. The hatched bars represent the number of hierarchized defects, while the bars without pattern represent the total number of defects. **(C)** Proportion of additional defects in sequences classified as intact (int.), inverted (inv.), hypermutated (hypermut.), largely deleted (LID), or harboring a stop codon (stop), a Ψ defect (Ψ), or a small internal deletion (SID). n = 441 proviruses retrieved from tissues collected postmortem of two participants on suppressive ART

The third and fourth datasets consisted of HIV genomes amplified from transcriptionally active viral reservoirs (vRNA$^+$) that were inducible (82), or spontaneously active (42). For the inducible dataset, the raw data of one sequence was not found, although the sequence itself was available on GenBank, and was hence removed from the following analyses (n = 195 instead of n = 196). Compared to the first datasets, the IntegrityAlgorithm found one more intact sequence than published in Sannier et al., *Cell Rep.* (2021) (**Figure 3.3.4A**). In this paper, the sequence was found to be defective for stop codons in Tat exons 1 and 2, while the IntegrityAlgorithm only found a single premature stop codon in Tat exon 2, which is not considered critical for the intactness (47,180). There was also a noticeable difference between the number of defective sequences with stop codons (48.2% vs. 3.1%). Nonetheless, the results published in Dubé et al., *Cell Host Microbe* (2023) and those produced by the IntegrityAlgorithm were almost identical (**Figure 3.3.4A**). In the inducible vRNA$^+$ reservoir dataset, the hierarchization hid 75 Ψ defects (**Figure 3.3.4B**). In both datasets, almost all defective sequences for large internal deletions and stop codons also had small internal deletions, although these proportions were below 20% for sequences classified as having Ψ defects (**Figure 3.3.4C**).

Overall, the results obtained with the IntegrityAlgorithm for these four datasets are consistent with the analysis done "manually" and published. In the majority of cases, hierarchizing underestimates defects lower in the hierarchy, such as stop codons, Ψ defects, and small internal deletions.
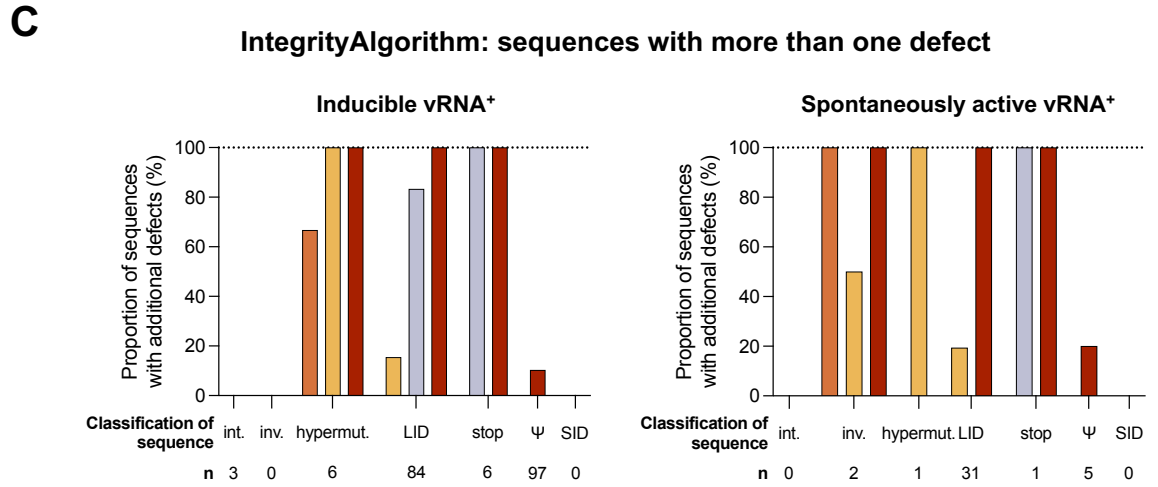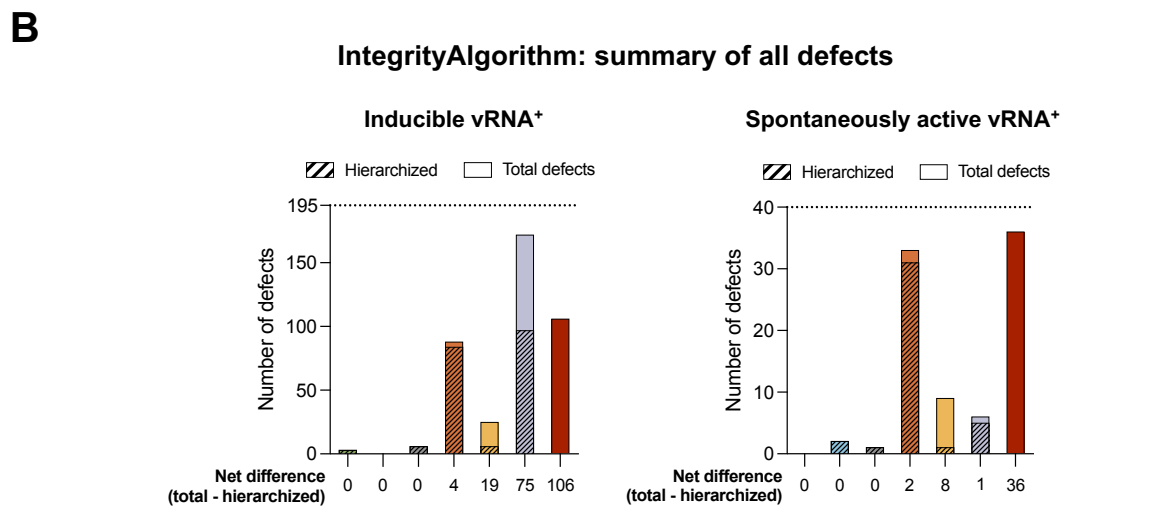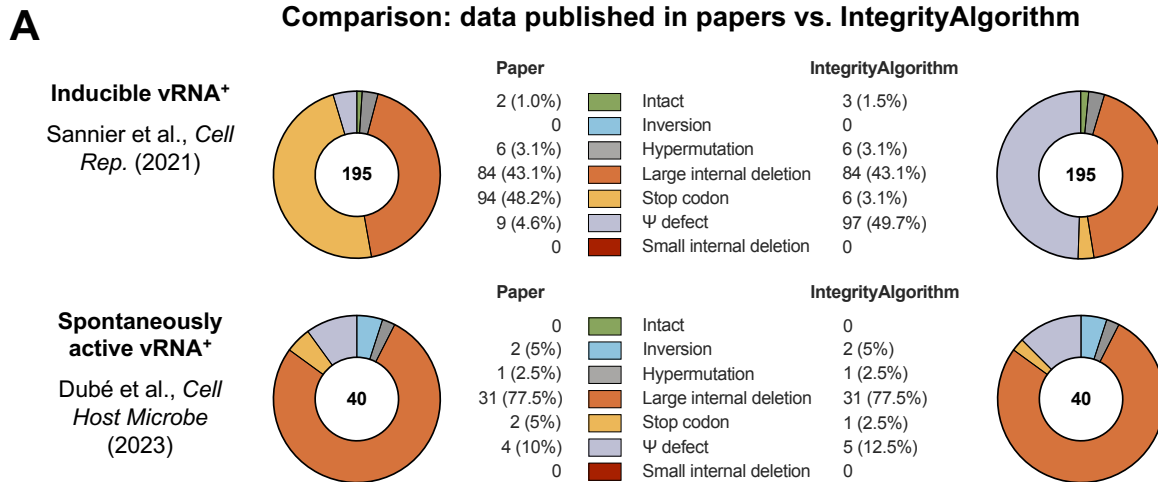
Figure 3.3.4 – Proviral landscape of HIV genomes retrieved from induced and spontaneously active vRNA⁺ cells. (**A**) Comparison between the data published in Sannier et al., *Cell Rep.* (2021), or in Dubé et al., *Cell Host Microbe* (2023), and the results produced by the IntegrityAlgorithm. (**B**) Exhaustive list of

defects provided by the IntegrityAlgorithm for inducible (Sannier et al.) and spontaneously active vRNA$^+$ reservoirs. The hatched bars represent the number of hierarchized defects, while the bars without pattern represent the total number of defects. (**C**) Proportion of additional defects in sequences classified as intact (int.), inverted (inv.), hypermutated (hypermut.), largely deleted (LID), or harboring a stop codon (stop), a Ψ defect (Ψ), or a small internal deletion (SID). n = 195 proviruses retrieved from inducible vRNA$^+$ reservoirs across three participants on suppressive ART. n = 40 proviruses retrieved from spontaneously active vRNA$^+$ reservoirs across three participants on suppressive ART.

The IntegrityAlgorithm presents several advantages over tools commonly used to assess proviral intactness and clonality

For the intactness, we compared the results produced by the IntegrityAlgorithm with other existing tools', namely ProSeq-IT (97), HIVIntact (98), and HIVSeqinR (99,100). HIVSeqinR requires to input sequences with their primers to work properly. We used Dubé et al.'s (42) dataset for this analysis, for which we had the original sequences, as we only had access to GenBank's publicly available primers-free sequences for the other three datasets.

All 40 sequences of this dataset were found to be defective (**Figures 3.3.4A and 3.3.5A**), mostly due to large internal deletions. Similar to the IntegrityAlgorithm, ProSeq-IT and HIVIntact classified all sequences as defective (**Figures 3.3.5B and 3.3.5C**). However, HIVSeqinR identified five of those sequences as intact, including one that was "intact but missing the Gag start codon" (**Figures 3.3.5D**). The other 35 sequences were classified as defective due to a large internal deletion, a premature stop codon, and a hypermutation.

Table 3.3.1 summarizes the characteristics of all four algorithms. ProSeq-IT has been used in multiple papers to infer HIV-1 proviral intactness (42,82,90,91,101,185–192). It is a fast and powerful tool that gives more than 50 annotations, including the number of deleted/inserted nucleotides, the number of stop codons, and the presence or absence of frameshifts in each of the nine ORFs. However, it only gives a conclusion ("defective" or "intact") and does not specify the defects found in the sequences, hence additional analyses are required. Currently, only defects in the structural proteins (Gag, Pol, and Env) are considered for the intactness. Except for Gag, the

algorithm does not verify for the presence of start codons in the ORFs, nor for inversions or hypermutations. It is also only applicable for HIV-1 subtypes B and C.

A
IntegrityAlgorithm

Intact	0
Inversion	2 (5%)
Hypermutation	1 (2.5%)
Large internal deletion	31 (77.5%)
Stop codon	1 (2.5%)
Ψ defect	5 (12.5%)
Small internal deletion	0

B
ProSeq-IT

Intact	0
Defective	40 (100%)

C
HIVIntact

Intact	0
Defective	40 (100%)

D
HIVSeqinR

Intact	4 (10%)
Intact, but missing the Gag start codon	1 (2.5%)
Hypermutation	1 (2.5%)
Large internal deletion	33 (82.5%)
Stop codon	1 (2.5%)

Figure 3.3.5 – Comparison between the four different algorithms used to predict HIV proviral intactness. (**A**) The IntegrityAlgorithm, (**B**) ProSeq-IT, (**C**) HIVIntact, and (**D**) HIVSeqinR were all tested on Dubé et al.'s (*Cell Host Microbe*) dataset. n = 40 proviruses retrieved from spontaneously active vRNA$^+$ reservoirs across three participants on suppressive ART.

HIVIntact is a Python-based algorithm that allows inference of proviral intactness using a single tool. It was used in another paper from the same group to develop a bioinformatics pipeline specifically for the Nanopore technology (193). Compared to ProSeq-IT, it can be used on HIV-1 subtypes A, B, C, D, F, G, and H. It does not require sequences to be aligned to its reference sequence. Sequences that pass the "initial check" (i.e., sequences with intact Gag, Pol, and Env ORFs) are further assessed for large internal deletions, frameshifts, smaller ORFs' intactness (Vif, Vpr, Tat, Rev, Vpu, and Nef), and intact Ψ region and RRE. HIVIntact uses its own implementation of Hypermut (181) to detect hypermutants, although hypermutations are not counted toward the overall intactness unless they introduce premature stop codons. The tool computes different files, including two sequence files (one FASTA for the inferred "intact" and one FASTA for the inferred

"defective" sequences) as well as a JavaScript Object Notation file detailing all defects within each sequence. However, this error file is hard to interpret without bioinformatics.

Similar to HIVIntact, HIVSeqinR, an R-based algorithm, is also a single tool that can infer proviral intactness. It was initially developed to analyze linear HIV genomes and is optimized for sequences *de novo* assembled with UltraCycle v1.0, an unpublished algorithm, spanning the 638 – 9632 region. Defects are initially assessed for non-HIV sequences. HIV sequences are then checked for large internal deletions (<8000 bp). Sequences with ≥8000 bp are subsequently analyzed for internal inversions, "scrambling without inversions," and hypermutations. Sequences lacking any of these defects are further analyzed for the absence of premature stop codons, deletions (tolerated up to 5%), and insertions (tolerated up to 20%) in the Gag, Pol, and Env ORFs. Finally, sequences with these intact ORFs were assessed for Ψ defects up to 15 deleted bp. The algorithm verifies for the presence of primers: if there are primers and the sequences do not harbor any of the defects mentioned above, the sequence is classified as "intact." Otherwise, it is classified as "inferred intact" (99). Similar to ProSeq-IT, defects in the regulatory regions and the accessory proteins are not considered for the intactness. HIVSeqinR also uses its own implementation of Hypermut. At the end, the algorithm generates a summary file, but due to the hierarchization of defects, only sequences that passed each step have further annotations. Nevertheless, each proviral sequences receives one of 31 possible verdicts detailing its intactnesss status.

In comparison, the IntegrityAlgorithm automates the NFL analysis protocol previously described (42,82,90,101). It generates both an exhaustive and a hierarchized list of defects that is easy to interpret. However, it is solely applicable for HIV-1 subtype B and it is dependent on multiple tools (QC Tool, Gene Cutter, ProSeq-IT, and Geneious), unlike ProSeq-IT itself, HIVIntact, or HIVSeqinR.

As done for the intactness, we thoroughly analyzed the clonality results obtained from the IntegrityAlgorithm for Dufour et al.'s dataset (90). In the current NFL protocol (42,82,90,101), and in others (194–196), clonality was assessed using the online tool ElimDupes (108). ElimDupes is not specific to HIV proviral sequences, as it has also been used for fungal (197,198), other viral (199,200), and other eukaryal (201,202) sequences. To our knowledge, this is the only publicly available tool that can identify duplicated sequences.

Table 3.3.1 – Characteristics of the four different algorithms used to predict HIV proviral intactness.

	IntegrityAlgorithm	ProSeq-IT	HIVIntact	HIVSeqinR
HIV-1 subtypes	B	B and C	A, B, C, D, F, G, and H	B and C
Single/multiple tool(s)	Multiple	Single	Single	Single
Pre-aligned sequences to the reference sequence required	Yes	No	No	No
Primers-free sequences required	Yes	Yes	Yes	No
Inversions detected	Yes	No	No	Yes
Hypermutations detected	Yes	No	Yes	Yes
Regulatory/accessory ORFs are considered critical for the intactness	Yes, except for Nef and stop codons in Tat exon 2	No	Yes	No
Defects are hierarchized	Yes	Not specified	Not specified	Yes
Exhaustive list of defects generated	Yes	Raw data only	Yes	Only for sequences with ≥ 8000 bp
Inferred intactness ("intact" or "defective")	Yes	Yes	Yes	Yes
Summary file generated	Yes	No	No	Yes
Additional analyses required to infer intactness status and/or defects	No	Yes	Yes, but bioinformatic skills are required to analyze the data	No
Reason(s) for defectiveness provided	Yes	No	No	Yes
Command line/web tool	Command line	Web	Command line	Command line

For the clonality, we consider two or more sequences as "clonal" if they are 100% identical. Compared to the results published in Dufour et al. (*Nat Commun.*), we obtained similar proportions of unique and clonal sequences as well as distributions of the sizes of the clones, the only difference being the additional sequence that was included in this analysis (**Figures 3.3.6A and 3.3.6B**). When compared with the other three datasets, four clones (three in PWH3 and one in PWH5) were

found in Dubé et al.'s dataset (42), compared to only three in the paper (**Figure 3.3.6C**). The algorithm identified the same number of clones for the datasets of Dufour et al. (*Cell Rep.*) and Sannier et al. (*Cell Rep.*).

A Comparison: Dufour et al., *Nat Commun.* (2023) vs. IntegrityAlgorithm
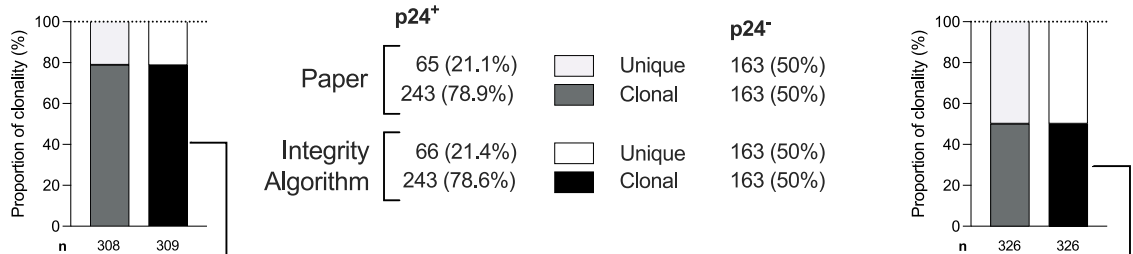
B IntegrityAlgorithm: size of clones
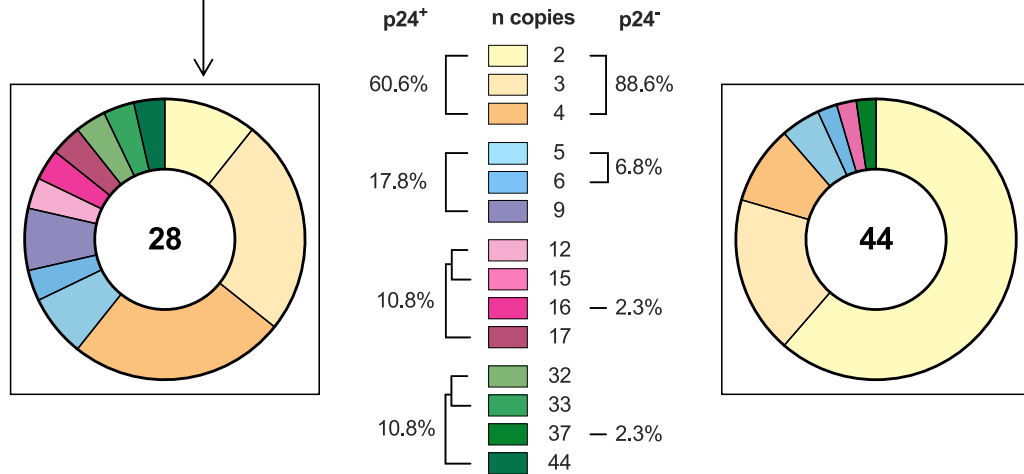
C Comparison: data published in papers vs. IntegrityAlgorithm

67
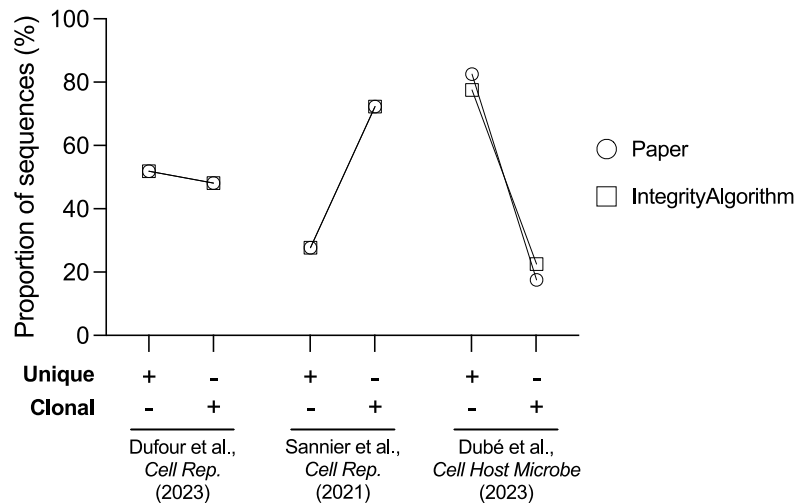
Figure 3.3.6 – Characterization of the clonality of HIV proviral sequences. (**A**) Comparison between the data published in Dufour et al., *Nat Commun.* (2023), computed by ElimDupes, and the results produced by the IntegrityAlgorithm. The proportion of unique and clonal sequences are represented in shades of white and black, respectively. (**B**) Distribution of the sizes of the clones according to the data provided by the IntegrityAlgorithm. (**C**) Proportion of unique and clonal sequences. Comparison between the data published and analyzed with ElimDupes (circles) and the results produced by the IntegrityAlgorithm (squares). n = 441 proviruses retrieved from tissues collected postmortem of two participants on suppressive ART (Dufour et al., *Cell Rep.*). n = 195 proviruses retrieved from inducible vRNA$^+$ reservoirs across three participants on suppressive ART (Sannier et al., *Cell Rep.*). n = 40 proviruses retrieved from spontaneously active vRNA$^+$ reservoirs across three participants on suppressive ART (Dubé et al., *Cell Host Micr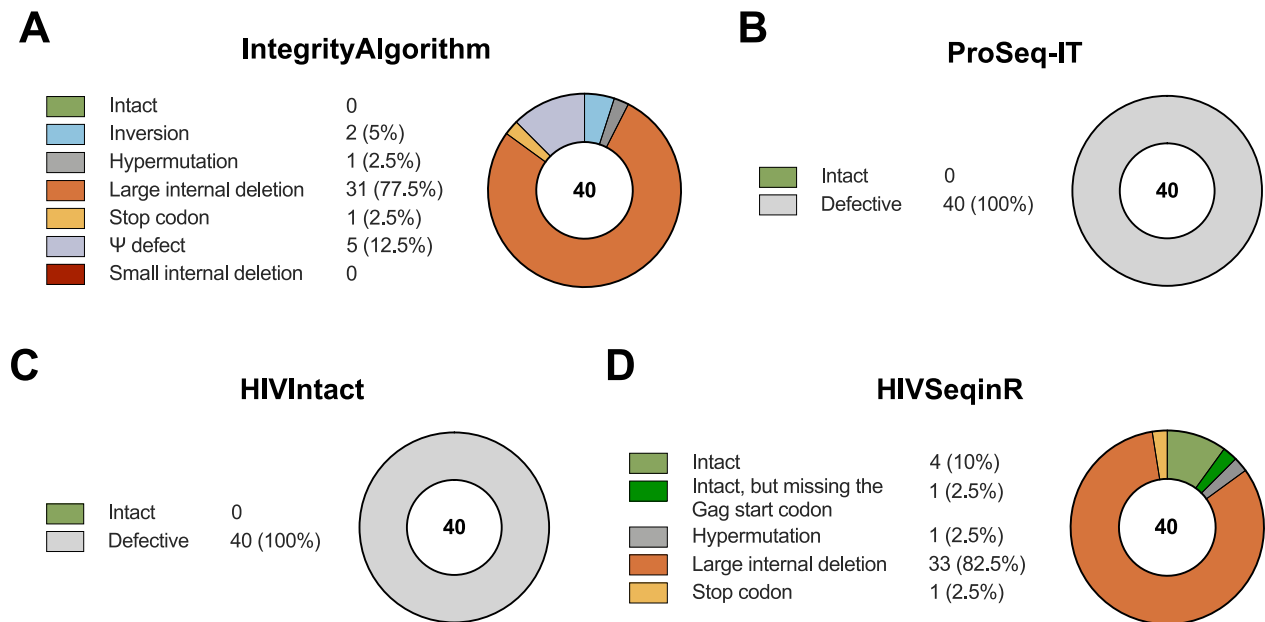obe*). (**A and B**) n = 309 (or 308 for Dufour et al.) p24$^+$ and n = 326 p24$^-$ proviruses retrieved across six participants on suppressive ART.

Compared to ElimDupes, the IntegrityAlgorithm can classify sequences with a single different nucleotide (or more, depending on the threshold provided by the user) as "potential clones." This nucleotide could either be mutated, deleted, or inserted. Using Geneious Prime, we manually aligned and confirmed each "clone" and "potential clone" in Dufour et al., *Nat. Commun.*'s dataset. Out of initially 286 unique sequences, and after accounting for duplicates (e.g., sequence 1 is a potential clone of sequence 2, but sequence 2 is also a potential clone of sequence 1), 39 sequences were confirmed to have a potential clone. In one participant, two large clones (n = 17 copies each) were classified as potential clones. We noticed that sequences aligned almost perfectly, except that 17 of these sequences were missing the first nucleotide of the amplicon, potentially due to an error when the primer sequences were removed. This also occurred for two smaller clones (n = 2 and 3 copies, respectively). Other differences between the sequences were most often caused by single mutations in the gp41 and gp120 subunits of the Env protein (33.3%), or in the retrotranscriptase- and protease-coding p51 and p31 genes of Pol (23.1%) (**Figure 3.3.7**).

Overall, the results obtained with the IntegrityAlgorithm for these four datasets are consistent with the analysis done with ElimDupes.
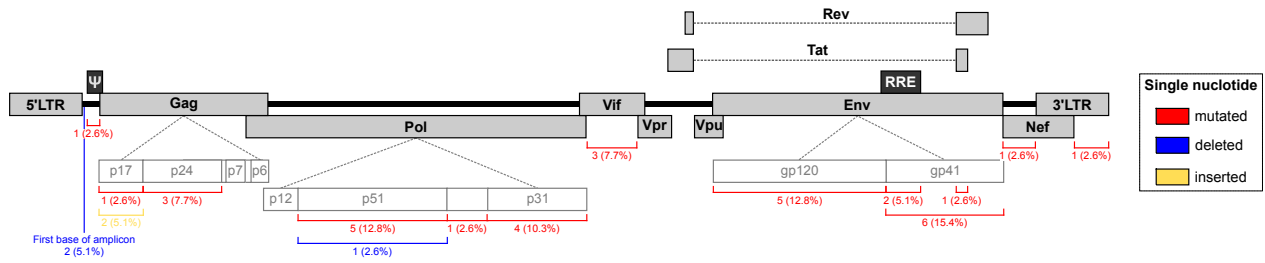
Figure 3.3.7 – Location in the HIV genome of the point mutations between potential clones. In this analysis, potential clones were defined as two or more sequences with a single different nucleotide, either mutated (red), deleted (blue), or inserted (yellow). The frequencies of point mutations are shown. n = 39 potential clones from the dataset of Dufour et al., *Nat Commun.* (2023).

Chapitre 4 – Les bases moléculaires du dysfonctionnement persistant des lymphocytes T CD4$^+$ spécifiques au VIH

4.1 Avant-propos

Dans le chapitre ci-dessous, je décrirai mon projet de maîtrise principal, réalisé dans le laboratoire Kaufmann. Le projet de la dysfonction des lymphocytes T CD4$^+$ spécifiques au VIH fait partie d'une collaboration avec le laboratoire de Dr. Jeremy M. Boss, à l'Université Emory, à Atlanta. Brièvement, des analyses de transcriptomiques, de méthylation de l'ADN et d'accessibilité de la chromatine seront réalisées sur des cellules T CD4$^+$ spécifiques au VIH. Dr. Maaran Michael Rajah, ancien étudiant postdoctoral du laboratoire Kaufmann, et moi avons optimisé le panel de cytométrie en flux utilisé pour le tri. À l'aide de Mélina Duchesne et Gloria Gabrielle Delgado, j'ai trié les cellules d'intérêt et extrait leur ARN ou ADN. Le laboratoire Boss se chargeait quant à lui du séquençage à l'échelle du génome et des analyses bioinformatiques. Une petite partie des échantillons a déjà été envoyée en septembre 2023 à nos collaborateurs dans le cadre d'un projet pilote. Dans les prochains mois, le laboratoire Boss traitera les échantillons restants pour compléter le jeu de données et débutera leurs analyses.

Dans la **section 4.2**, l'article « *Enhanced detection of antigen-specific T cells by a multiplexed AIM assay* », récemment publié dans *Cell Reports Methods*, sera présenté. Cet article montre une approche multiplexée de l'essai AIM, qui avait été utilisée par notre groupe dans des articles récemment publiés. La récolte des données avait initialement été faite par plusieurs des co-auteurs, moi y compris, mais ma plus grande contribution a été de faire son analyse complète et formelle ainsi que d'écrire le manuscrit. L'article nous a permis d'élaborer notre stratégie de *gating* et de tri. L'approche expérimentale de ce projet sera présentée dans la **section 4.3**. Puisque nous n'avons pas encore reçu les résultats du projet pilote, la section **section 4.4** se concentra sur les méthodes nous ayant permis d'obtenir les cellules triées et les extractions d'ADN, en plus d'un contrôle qualité réalisé par nos collaborateurs.

Les **sections 4.3** (approche expérimentale et méthodes) et **4.4** (résultats) sont écrites en anglais, car elles seront utilisées dans des articles subséquents, présentement en préparation (voir

l'**Annexe 9**). Les **Annexes 7** et **8** contiennent les figures et tableaux supplémentaires des **sections 4.3** et **4.4**.

4.2 Article: *Enhanced detection of antigen-specific T cells by a multiplexed AIM assay*

Statut :

Cet article a été publié dans *Cell Reports Methods*, le 15 janvier 2024.

https ://doi.org/10.1016/j.crmeth.2023.100690

Auteurs :

Audrée Lemieux,[1,2] Gérémy Sannier,[1,2] Alexandre Nicolas,[1,2] Manon Nayrac,[1,2] Gloria-Gabrielle Delgado,[1] Rose Cloutier,[1] Nathalie Brassard,[1] Mélanie Laporte,[1] Mélina Duchesne,[1] Alina Maria Sreng Flores,[1] Andrés Finzi,[1,2] Olivier Tastet,[1] Mathieu Dubé,[1] and Daniel E. Kaufmann[1,3,4,5]

Affiliations :

[1]Centre de Recherche du CHUM, Montréal, QC H2X 0A9, Canada

[2]Département de microbiologie, infectiologie et immunologie, Université de Montréal, Montréal, QC, H2X 0A9, Canada

[3]Consortium for HIV/AIDS Vaccine Development (CHAVD), La Jolla, CA, USA

[4]Département de Médecine, Université de Montréal, Montréal, QC, H2X 0A9, Canada

[5]Division of Infectious Diseases, Department of Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

Contribution des auteurs :

Conceptualisation : **A.L.**, M. Dubé et D.E.K.; AIM assays : **A.L**., G.S., A.N., M. Dubé, M.N., G.-G.D., R.C., N.B., M.L., M. Duchesne et A.M.S.F.; Analyse des données : **A.L.** et M. Dubé; Analyses non-supervisées : **A.L.** avec la contribution de O.T.; Sécurisation des échantillons : A.F. et D.E.K.; Rédaction du manuscrit original : **A.L.**, M. Dubé et D.E.K.; Lecture, édition et approbation du manuscrit final : **tous les auteurs**; Supervision de l'étude : M. Dubé et D.E.K.

Enhanced detection of antigen-specific T cells by a multiplexed AIM assay

Audrée Lemieux,[1,2] Gérémy Sannier,[1,2] Alexandre Nicolas,[1,2] Manon Nayrac,[1,2] Gloria-Gabrielle Delgado,[1] Rose Cloutier,[1] Nathalie Brassard,[1] Mélanie Laporte,[1] Mélina Duchesne,[1] Alina Maria Sreng Flores,[1] Andrés Finzi,[1,2] Olivier Tastet,[1] Mathieu Dubé,[1*] and Daniel E. Kaufmann[1,3,4,5*#]

[1]Centre de Recherche du CHUM, Montreal, QC H2X 0A9, Canada

[2]Département de Microbiologie, Infectiologie et Immunologie, Université de Montréal, Montreal, QC, H2X 0A9, Canada

[3]Consortium for HIV/AIDS Vaccine Development (CHAVD), La Jolla, CA, USA

[4]Département de Médecine, Université de Montréal, Montréal, QC, H2X 0A9, Canada

[5]Division of Infectious Diseases, Department of Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

*Correspondence: mathieu.dube.chum@ssss.gouv.qc.ca (M.D) and daniel.kaufmann@chuv.ch (D.E.K)

[#]Lead contact: daniel.kaufmann@chuv.ch (D.E.K)

In brief

Cytokine-independent AIM assays identify antigen-specific T cells. Lemieux et al. multiplex this approach by co-analyzing four AIMs instead of the usual two, enhancing detection of antigen-specific CD4$^+$ T cells while enabling simultaneous identification of antigen-specific CD8$^+$ T cells. This strategy reduces quantification and phenotyping biases caused by single AIM pairs.

Highlights

- The 6xAIM assay improves detection of antigen-specific CD4$^+$ T cells
- Four activation-induced markers generate six pairs that define T cell subsets
- A combination of CD69, 4-1BB, OX40, and CD40L detects CD4$^+$ and CD8$^+$ T cell responses
- This assay mitigates phenotyping biases arising through use of single AIM pairs

Summary

Broadly applicable methods to identify and characterize antigen-specific CD4$^+$ and CD8$^+$ T cells are key to immunology research, including studies of vaccine responses and immunity to infectious diseases. We developed a multiplexed activation-induced marker (AIM) assay that presents several advantages compared to single pairs of AIMs. The simultaneous measurement of four AIMs (CD69, 4-1BB, OX40, and CD40L) creates six AIM pairs that define CD4$^+$ T cell populations with partial and variable overlap. When combined in an AND/OR Boolean gating strategy for analysis, this approach enhances CD4$^+$ T cell detection compared to any single AIM pair, while CD8$^+$ T cells are dominated by CD69/4-1BB co-expression. Supervised and unsupervised clustering analyses show differential expression of the AIMs in defined T helper lineages and that multiplexing mitigates phenotypic biases. Paired and unpaired comparisons of responses to infections (HIV and cytomegalovirus [CMV]) and vaccination (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]) validate the robustness and versatility of the method.

Motivation

Detection of circulating antigen (Ag)-specific T cells *ex vivo* is hampered by their low frequency and heterogeneity. The use of conventional cytokine-based approaches limits identification of

some subsets such as T follicular helper (Tfh) cells, whose cytokines are difficult to detect. Flow cytometric activation-induced marker (AIM) assays, which work by measuring the upregulation of surface markers following T cell receptor (TCR) stimulation, are highly sensitive, cytokine-agnostic alternatives to detect Ag-specific T cells. However, they generally involve a single AIM pair. This absence of a consensus combination of molecules makes comparison between studies challenging. Many AIM pairs are also unable to detect Ag-specific CD4[+] and CD8[+] T cells simultaneously. Additionally, while expression of specific markers is expected to vary among T cell lineages and across antigens, side-by-side comparisons are lacking. We intend to develop an AIM assay that would mitigate these biases in the detection of infection- and vaccine-induced Ag-specific T cells.

Graphical abstract
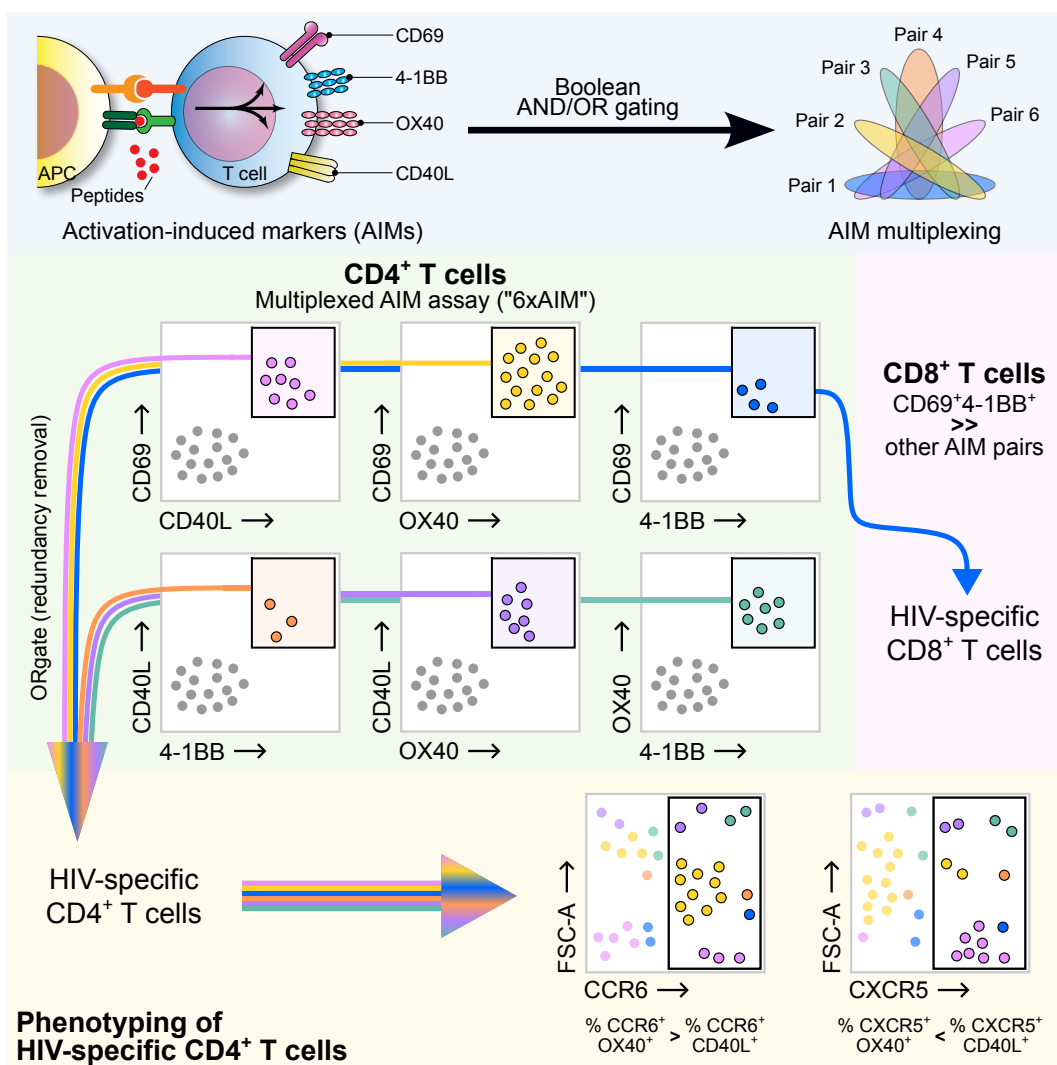
4.2.1 Introduction

Antigen (Ag)-specific CD4[+] and CD8[+] T cells play a critical role in immune responses against viral infections.[1–3] While CD8[+] T lymphocytes kill cells infected with intracellular pathogens via their cytotoxic activity, CD4[+] T cells produce cytokines that modulate the functions of other cells. CD4[+] T cells include subsets like T follicular helper (Tfh) cells, whose interaction with B cells is critical to regulate the antibody response; Th1, constituting important immune responses against viruses and other intracellular pathogens; and Th17 and Th22, implicated in mucosal immunity.[4–6] However, the direct detection of circulating Ag-specific T cells *ex vivo* is hampered by both their relatively low frequency and their heterogeneity.[7,8] Typically, pathogen-specific T cells are detected with conventional cytokine-based approaches such as intracellular cytokine staining (ICS) or enzyme-linked immunosorbent spot (ELISpot).[9] However, their use can bias the sampling of subsets like Tfh cells, whose cytokines are more difficult to detect.[7,10,11]

To overcome these limitations, the flow cytometric activation-induced marker (AIM) assay, which measures the upregulation of selected surface markers following T cell receptor (TCR) stimulation after encounter with the cognate Ag, was developed.[7,12–14] AIM assays rely on molecules that are quickly upregulated at the cell surface. In contrast to soluble cytokines, these membrane-associated markers do not need to be chemically trapped inside the cells, as done in ICS and related methods, thus allowing identification and, for specific downstream analyses, sorting of live cells. Moreover, while the AIM assay is a functional assay, as it requires the activation of the cell through its TCR, it is a robust and highly sensitive alternative for cytokine-independent detection of Ag-specific T cells.[7,12,13,15] In the past, single AIMs,[8,16–18] or single pairs of AIMs including, but not restricted to, CD40L[+]CD200[+],[18] OX40[+]CD25[+],[7,14,19,20] CD69[+]CD40L[+],[7,14,20–24] CD69[+]OX40[+],[14,25,26] or CD69[+]4-1BB[+],[14,24,26–29] have been used to detect T cells of different specificities. These molecules were specifically chosen for their known induction on activated T cells after TCR activation. However, there are limitations to using single AIM pairs: the absence of a consensus combination of molecules across the literature makes it difficult to compare between studies. The commonly used CD69[+]CD40L[+] assay, which works well on blood samples, cannot be used in lymphoid tissues, as CD69 is physiologically expressed at high levels on germinal center (GC) Tfh cells.[15] Many pairs of AIMs are also unable to detect Ag-specific CD4[+] and CD8[+] T cells simultaneously.

Here, we assessed in depth a multiplexed version of an AIM assay used in recent studies.[30–34] This assay integrates and co-analyzes four AIMs—CD69, 4-1BB (CD137), OX40 (CD134), and CD40L (CD154)—rather than two. We examined if this assay could overcome the possible biases related to the selection of a single pair of AIMs. We assessed the efficiency of the approach to capture HIV-specific CD4[+] and CD8[+] T cells and examined the additional coverage provided by each AIM pair. We evaluated whether the AIMs were preferentially expressed on CD4[+] T cells expressing defined lineage markers. We compared the observed profiles to T cell responses specific for another chronic virus (cytomegalovirus [CMV]), as well as those induced by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccination.

4.2.2 Results

The multiplexed AIM assay improves detection of HIV- and CMV-specific CD4[+] T cells

We first studied the ability of the multiplexed AIM assay to profile HIV-specific CD4[+] and CD8[+] T cell responses. We examined the co-expression patterns of CD69, 4-1BB, OX40, and CD40L on T cells after a 15 h stimulation with HIV Ags.[34] The 15 h stimulation was selected as a compromise for the different time frames used for single-pair assays relying on CD40L (9 h being optimal, as it is upregulated early)[7,35] and OX40 (up to 24 h being optimal).[7,36,37] We focused on a cohort of 16 people with HIV (PWH) on suppressive antiretroviral therapy (ART) with <40 plasma viral RNA (vRNA) copies/mL (**Table 4.2.S1**). The four activation markers can be combined to form six different AIM pairs post hoc to infer Ag specificity (**Figures 4.2.1A, 4.2.S1A, and 4.2.S1B**). AIMs, as functional molecules, may or may not be expressed by certain Ag-specific T cells. We hypothesized that combining multiple pairs would provide a more representative sampling of Ag-specific CD4[+] and CD8[+] T cells. On the other hand, CD4[+] and CD8[+] T cells are often polyfunctional[20,38–40] and can express more than one AIM pair, with sizable overlap between cell subpopulations. To measure the total Ag-specific responses, we therefore used an AND/OR Boolean combination gating strategy ensuring that AIM[+] cells are counted only once. In the example provided in **Figure 4.2.1B** about CD69[+]CD40L[+] and CD69[+]OX40[+] CD4[+] T cells, the combined signal would be 0.75% (Boolean OR gating) and not 1.05% (sum of frequencies), as there was a 0.30% redundancy between those two pairs.

We used the multiplexed AIM assay (henceforth termed "6xAIM") to broadly detect HIV-specific CD4[+] and CD8[+] T cells, as done previously.[21,22,41] We measured AIM[+] T cell responses following an *ex vivo* 15 h stimulation of cryopreserved peripheral blood mononuclear cells (PBMCs) with overlapping peptide pools spanning the HIV Gag, Pol, Envelope (Env), and Nef proteins (**Figure 4.2.1C**). To examine whether the patterns of AIM expression observed for HIV-specific T cell responses were specific to this pathogen or generalizable to other viral infections, we measured in parallel the responses against CMV, a life-long infection highly prevalent in PWH.[42] As CMV-specific CD4[+] T cells are known to be more terminally differentiated,[43] we hypothesized that the hierarchy of AIM pairs might differ from HIV-specific CD4[+] T cells. We studied the same cohort of PWH, excluding one donor who was negative for CMV (**Table 4.2.S1**). This allowed paired, intrahost comparisons of CD4[+] and CD8[+] T cells specific for two viruses (**Figure 4.2.1D**). As CMV is a large virus encoding numerous proteins, we focused on responses to the lower matrix protein pp65. The variations between technical replicates (**Figure 4.2.S1C**) and independent experiments (**Figure 4.2.S1D**) were low, indicating good reproducibility of the AIM assay.

78
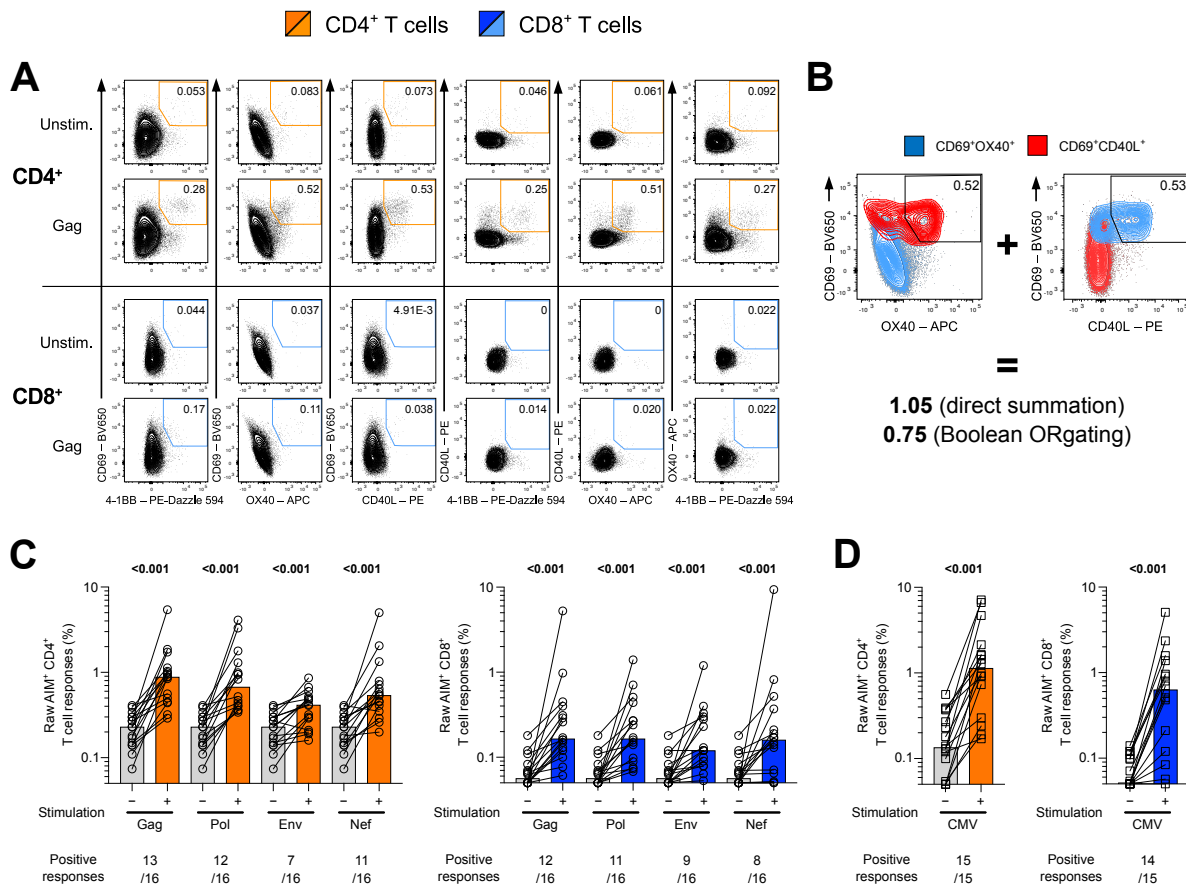
Figure 4.2.1 – Quantification of Ag-specific CD4$^+$ and CD8$^+$ T cell responses using Boolean OR gating. (**A**) Representative FACS plots depicting the multiplexed (6xAIM) strategy to identify Ag-specific CD4$^+$ (orange gate) and CD8$^+$ (blue gate) T cells. For simplicity, the example focuses on the HIV Gag stimulation. (**B**) FACS plots illustrating the AND/OR Boolean gating strategy. For simplicity, the example focuses on the AIM pairs CD69$^+$OX40$^+$ (blue population) and CD69$^+$CD40L$^+$ (red population). (**C and D**) Raw frequencies of (**C**) AIM$^+$ CD4$^+$ and CD8$^+$ T cell responses following an *ex vivo* 15 h stimulation of PBMCs with a pool of HIV Gag, Pol, Envelope (Env), or Nef peptides, or (**D**) following stimulation with HCMVA pp65 peptides. PBMCs were left unstimulated as a control (gray bars). Numbers of responders reaching >2X over the unstimulated condition are written below the histograms for each stimulation. The median 6xAIM values and Wilcoxon tests are shown. (**C**) n = 16 ART participants; (**D**) n = 15 CMV$^+$ ART participants.

The false positive rate is an inherent limitation of any functional assay. Bystander activation was found to be minimal in standard AIM assays relying on OX40$^+$CD25$^+$ and CD69$^+$CD40L$^+$.[7] We tested if the integration of multiple AIM pairs increased the false positive signal due to bystander activation. We collected supernatants from a primary AIM assay (AIM #1) conducted on Gag-stimulated PBMCs from ART participants and then performed a second AIM assay (AIM #2) using the conditioned supernatants to activate sentinel T cells from an HIV-naïve participant (**Figure 4.2.S1E**). The signal observed in AIM #2 remained similar to the background in AIM #1, indicating minimal bystander activation in this setting (**Figure 4.2.S1F**).

We subtracted the background signal (values of unstimulated conditions, used as a negative control)[7,13,17,19,21–23,41] to obtain net responses (**Figures 4.2.S2A and 4.2.S2B**). To assess the "total" HIV-specific responses in the ART-suppressed cohort, we then summed up the net Gag, Pol-, Env-, and Nef-specific responses (**Figure 4.2.S2A**). We compared the magnitude of HIV-specific T cell responses detected by each AIM pair as well as the 6xAIM multiplexed combination of the six AIM pairs (**Figure 4.2.2A**). Matched comparisons using the Friedman test with Dunn's correction revealed several statistically significant differences in these magnitudes (**Figure 4.2.2A**). For HIV-specific CD4$^+$ T cells, the CD69$^+$OX40$^+$ and OX40$^+$4-1BB$^+$ responses were of larger magnitudes, while the CD40L$^+$4-1BB$^+$ combination captured the least cells. Only subtle differences were observed for CMV responses when single AIM pairs were compared

(**Figure 4.2.2B**). The multiplexed 6xAIM quantification consistently provided higher detection of HIV- and CMV-specific CD4$^+$ T cell responses compared to single AIM pairs. In contrast to Ag-specific CD4$^+$ T cells, no such hierarchy existed for CD8$^+$ T cells. The multiplexed AIM$^+$ CD8$^+$ T cell response was strongly dominated by the CD69$^+$4-1BB$^+$ combination. The detection provided by CD69$^+$4-1BB$^+$ was equivalent to the 6xAIM strategy (**Figures 4.2.2A and 4.2.2B**).

Direct side-by-side comparisons with AIM pairs that have been extensively used in the past to detect Ag-specific CD4$^+$ T cells, i.e., CD69$^+$OX40[7,14,20–24] and CD69$^+$CD40L[+14,25,26] (**Figures 4.2.2C and 4.2.S2C**), further demonstrated the benefit of the 6xAIM in terms of the magnitude of responses detected. AIM$^+$ responses were, respectively, 1.9 and 3.2 times higher with the 6xAIM approach and reached significance with Wilcoxon signed-rank tests. Despite the different kinetics of AIM markers, the 6xAIM yielded the highest CD4$^+$ T cell responses throughout various stimulation durations commonly employed in standard AIM assays, such as 9,[7,21–23] 15,[30–34] and 24 h[18,19,25,27,28] (**Figure 4.2.S2D**). The 6xAIM analysis did not increase the false positive signal due to bystander activation (**Figure 4.2.S2E**).

Predictably, integrating multiple AIM pairs led to some increase in background signal (**Figure 4.2.S2C**). We next assessed if the enhanced detection of Ag-specific CD4$^+$ T cells achieved by the 6xAIM came at the expense of a proportionally higher background. We calculated the ratio of the signal in the unstimulated vs. Ag-stimulated conditions (**Figure 4.2.2D**). For individual pairs, the background signal constituted a median of 5%–11% of the stimulated signal. The median background level for 6xAIM was within that range, at 9%. Therefore, the 6xAIM approach generates a background signal proportionally equivalent to standard single AIM pairs.

Different pairs of AIMs can identify overlapping Ag-specific T cell populations. To map this redundancy, we generated combination gates, creating 64 ($2^6 = 64$) theoretical combinations (**Figure 4.2.S2F**). Each entry refers to a combination of pairs, for example a population identifiable by all six pairs of AIMs (indicated by "i" in **Figure 4.2.S2F**), or a population only identifiable by CD69$^+$4-1BB$^+$ (indicated by "ii"). Of all these combinations, only a handful represented sizable fractions of HIV-specific CD4$^+$ and CD8$^+$ T cells. For instance, none of the HIV-specific CD4$^+$ T cells exhibited exclusive positivity for CD40L$^+$4-1BB$^+$ and CD69$^+$OX40$^+$, whereas 20%

were positive for OX40⁺4-1BB⁺ exclusively, without any other pair being detected. Therefore, co-expression of AIMs followed constrained patterns.
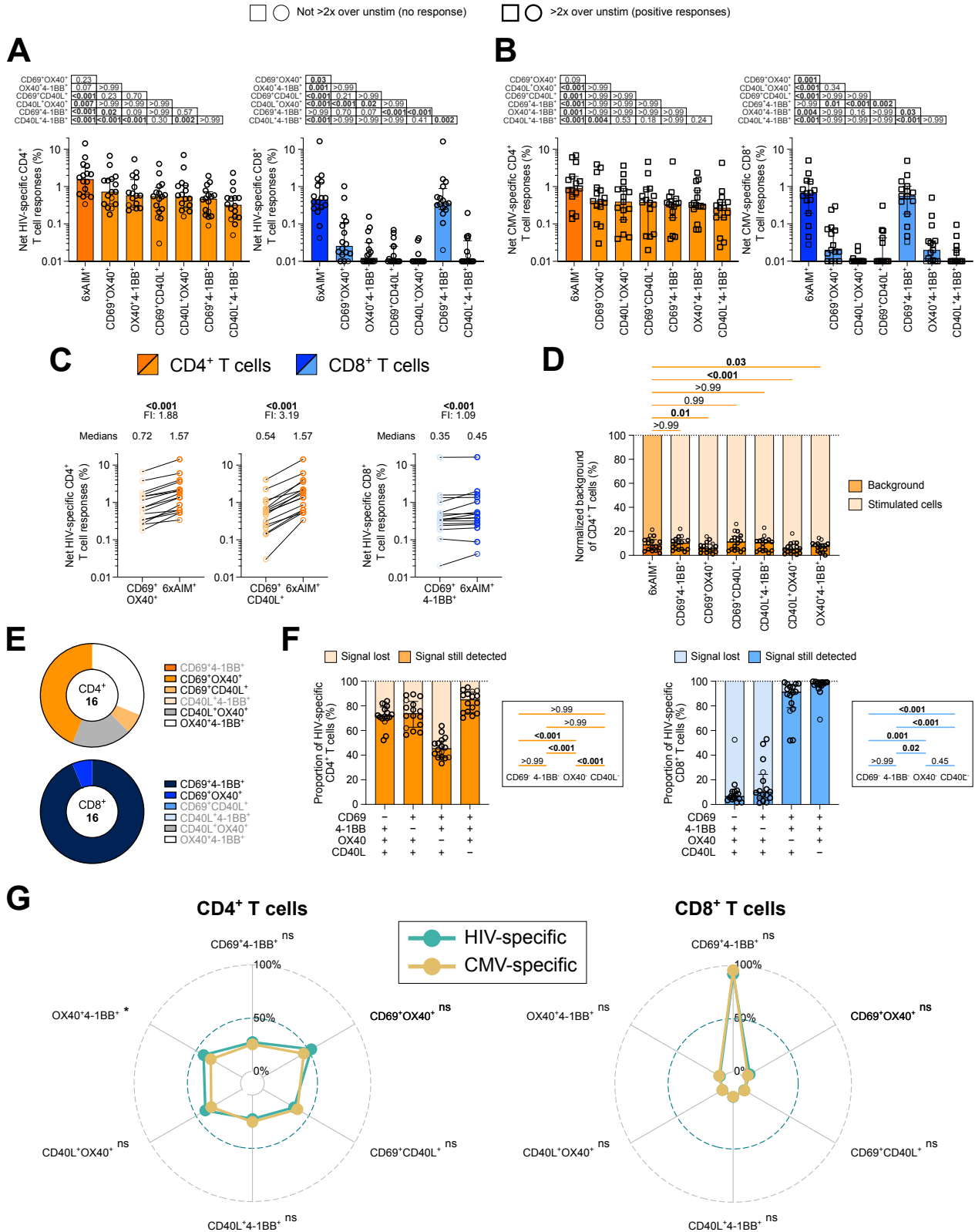
Figure 4.2.2 – The multiplexed 6xAIM assay improves detection of HIV- and CMV-specific CD4$^+$ T cells. (**A**) Net HIV-specific CD4$^+$ and CD8$^+$ T cell responses. The net HIV Gag, Pol, Env, and Nef responses were summed to assess the "total" HIV-specific responses. (**B**) Net CMV-specific CD4$^+$ and CD8$^+$ T cell responses. (**C**) Comparisons between the net responses of the 6xAIM strategy and benchmark AIM pairs to detect Ag-specific CD4$^+$ and CD8$^+$ T cell responses. Medians, median fold increases (Fis), and Wilcoxon tests are shown. (**D**) Proportion of background signal in HIV Ag-stimulated conditions. The Ag-stimulated conditions are normalized to 100% to better represent the autologous unstimulated signal. Light colors represent the total of stimulated cells; dark colors represent the background. (**E**) Donut charts depicting the proportion of participants for whom each AIM pair yielded the highest magnitude of HIV-specific CD4$^+$ and CD8$^+$ T cell responses. (**F**) Individual contributions of each AIM marker for the detection of HIV-specific CD4$^+$ and CD8$^+$ T cell responses. The detection levels were normalized to the value yielded by the 6xAIM. Light colors represent the signal lost when the indicated AIM marker was removed from the analysis; dark colors represent the residual signal still detected. (**G**) Spider charts depicting the comparisons between the normalized amplitudes of the net HIV- (teal) and CMV-specific (amber) CD4$^+$ and CD8$^+$ T cell responses. Within each cohort, responses were normalized to their 6xAIM values. Medians and Wilcoxon tests are shown for each pair (see also **Figure 4.2.S2H**). $*$ $p < 0.05$, $**$ $p < 0.01$, and $***$ $p < 0.001$. Matched donors who had at least one HIV stimulation and the CMV stimulation reaching >2X over the unstimulated conditions were kept. For CD4$^+$ and CD8$^+$ T cells, n = 13 CMV$^+$ ART-matched participants. (**A and B**) The 6xAIM and individual AIM pairs are ranked in decreasing order of the median, based on their respective CD4$^+$ T cells. The tables above the graphs represent the pairwise multiple comparisons performed using the Friedman test with Dunn's post hoc test. (**D and F**) Pairwise multiple comparisons were performed using the Friedman test with Dunn's post-hoc test. (**A, B, D, and F**) The bars represent medians ± interquartile ranges. Thicker borders represent donors who had at least one HIV stimulation, or had the CMV stimulation, reaching >2× over the unstimulated conditions. (**A, C, and D–F**) n = 16 ART participants; (**B**) n = 15 CMV$^+$ ART participants.

The AIM pair combinations detecting the largest fraction of the 6xAIM$^+$ HIV-specific CD4$^+$ T cell population varied from donor to donor (**Figure 4.2.2E**). Therefore, relying on a single AIM pair can lead to underestimation of the T helper responses. In contrast, the CD69$^+$4-1BB$^+$ combination dominated HIV-specific CD8$^+$ T cells, with a single exception.

We next assessed the contribution of each individual marker to the global detection of HIV- and CMV-specific T cells. To accomplish this, we iteratively removed them one by one from the

multiplexed analysis and calculated the proportion of the signal consequently lost (**Figures 4.2.2F and S2G**). Because of partial overlap and co-expression, we still achieved good detection of Ag-specific CD4$^+$ T cells in the combinations missing one marker. Removal of OX40 was significantly more detrimental than the removal of the other AIMs, showing that there is less redundancy and co-expression for this marker. In contrast, both CD69 and 4-1BB were required for the identification of HIV- and CMV-specific CD8$^+$ T cells.

As the 6xAIM represented the maximum responses obtained, we also normalized the virus-specific responses to their respective 6xAIM values and found that patterns of detection were relatively similar between HIV- and CMV-specific CD4$^+$ and CD8$^+$ T cells, respectively (**Figures 4.2.2G and 4.2.S2H**). These results indicate that the expression patterns of individual markers and marker pairs within the multiplexed AIM assay are quite consistent among T cell responses to two very different chronic viruses. These data also suggest that the risk of introducing biases by focusing on a single AIM pair to compare responses to different pathogens is probably small within the same group of participants.

These results indicate that a multiplexed AIM panel containing the markers CD69, 4-1BB, OX40, and CD40L allowed simultaneous detection of Ag-specific CD4$^+$ and CD8$^+$ T cells. The Boolean 6xAIM analysis increased the detection of HIV- and CMV-specific CD4$^+$ T cells compared to single-pair assays, while CD8$^+$ T cells were readily identified by the CD69$^+$4-1BB$^+$ combination.

CD69, 4-1BB, OX40, and CD40L are differentially expressed on HIV-specific CD4$^+$ T cell subsets

HIV-specific CD4$^+$ T cells defined by the 6xAIM strategy expressed different levels of AIMs, with larger fractions of CD69$^+$ and OX40$^+$ cells compared to 4-1BB and CD40L (**Figures 4.2.3A and 4.2.S3A**). AIMs are functional molecules that may be differentially expressed depending on the CD4$^+$ T cell subsets. Therefore, we assessed whether these differences in expression could impact the profile of the HIV-specific CD4$^+$ T cell subsets detected. We phenotyped AIM$^+$ HIV-specific CD4$^+$ T cells using a high-parameter fluorescence-activated cell sorting (FACS) panel (**Table 4.2.S2**) based on expression of chemokine receptors that are preferentially—but not exclusively—expressed by defined lineages: CXCR5 for Tfh, CXCR3 for

Th1, CCR6 for Th17/Th22, an inhibitory checkpoint (PD-1), and activation markers (CD38, HLA-DR). The uniform manifold approximation and projection (UMAP) algorithm[44] was used to illustrate the distribution of HIV-specific CD4$^+$ T cell populations (**Figure 4.2.3B**). These populations were clustered with the Phenograph algorithm,[45] which identified 12 distinct populations based on their relative marker expression (**Figures 4.2.3C and 4.2.S3B**). These populations could be further grouped in "superclusters" driven by their chemokine receptor expression (**Figure 4.2.3B**).

We next tested if individual AIMs were particularly associated with certain polarization markers. Univariate gatings were applied to identify, within the total 6xAIM$^+$ population, Ag-specific cells expressing CD69, 4-1BB, OX40, and/or CD40L (**Figure 4.2.S3A**). The phenotypes of AIM$^+$ CD69$^+$, AIM$^+$ 4-1BB$^+$, AIM$^+$ OX40$^+$, and AIM$^+$ CD40L$^+$ CD4$^+$ T cell subpopulations were compared to the global 6xAIM$^+$ CD4$^+$ T cells, which served as the reference parental population (**Figure 4.2.S3C**). These data are also summarized in **Figure 4.2.3D**, where relative enrichments were calculated for each phenotypic clusters in CD69$^+$, 4-1BB$^+$, OX40$^+$, and CD40L$^+$ Gag-specific CD4$^+$ T cells compared to the reference 6xAIM$^+$ Gag-specific population. The relative frequency of each of these populations varied depending on the AIM expressed. OX40$^+$ Gag-specific cells were the most comparable to the reference 6xAIM$^+$ CD4$^+$ T cell population. Only C1 and C12 strongly diverged, being underrepresented in OX40$^+$ Gag-specific cells. Yet, CCR6$^+$ clusters (C2, C4, C6, C7, and C9) were modestly but consistently enriched in this subpopulation. In contrast to OX40, the AIM$^+$ CD69$^+$ population tended to be underrepresented in CCR6$^+$ clusters (C2, C6, C7, and C9) but enriched in CXCR3$^+$ clusters (C3, C8, and C12). Differences were more profound for CD40L, as clusters with Tfh-like signatures (C3 and C5) were significantly overrepresented, to the detriment of Th17/Th22-like clusters (C2, C6, C7, and C9). Phenotypic biases were also observed in AIM$^+$ 4-1BB$^+$ cells, as C4 was largely decreased, whereas C6, C8, and 10 were overrepresented. Thus, individual AIMs preferentially identify different subsets of Ag-specific CD4$^+$ T cells.
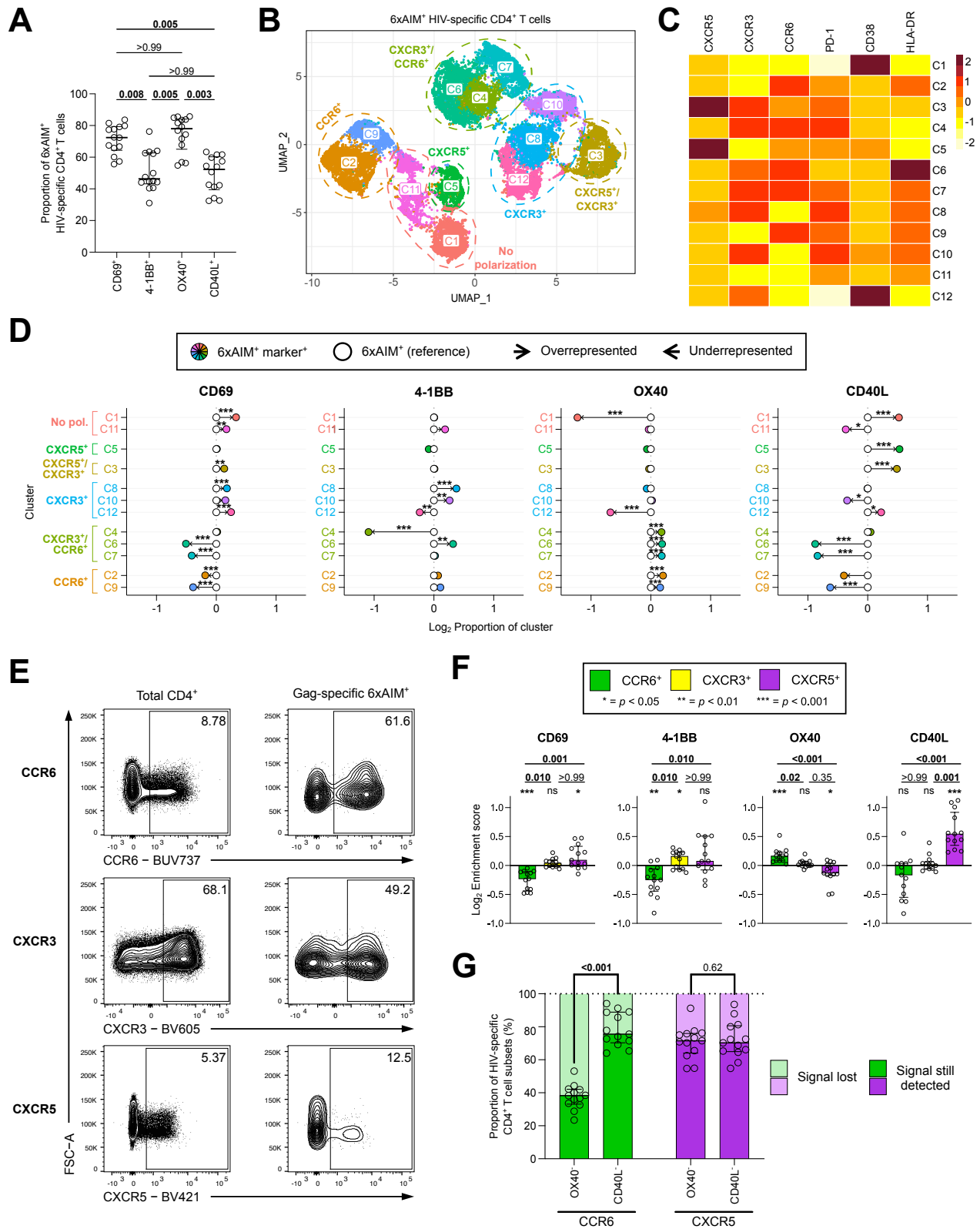
Figure 4.2.3 – CD69, 4-1BB, OX40, and CD40L are differentially expressed on HIV-specific CD4$^+$ T cell subsets. Phenotypic analysis of HIV-specific CD4$^+$ T cell responses. (**A**) Proportion of

6xAIM$^+$ HIV-specific CD4$^+$ T cells expressing CD69, 4-1BB, OX40, and CD40L in univariate analyses (see also **Figure 4.2.S3A** for gating strategy). (**B**) Multiparametric global UMAP representation of 6xAIM$^+$ HIV-specific CD4$^+$ T cells. The colors identify 12 populations clustered by unsupervised analyses and labeled on the UMAP. These populations could be further grouped in "superclusters" driven by their chemokine receptor expression. Clusters that were not polarized toward CCR6$^+$, CXCR3$^+$, or CXCR5$^+$ T cells were classified as "no polarization." (**C**) Heatmap summarizing the mean fluorescence intensity (MFI) of each loaded parameter. (**D**) Relative frequency of each identified cluster within 6xAIM$^+$ HIV-specific CD4$^+$ marker$^+$ T cell subpopulations (colored dots) compared to within the reference parental 6xAIM$^+$ cells (white dots). Arrows pointing to the right indicate that a population is overrepresented in the HIV-specific population compared to the reference CD4$^+$ T cells, while arrows pointing to the left indicate an underrepresentation. Medians and only significant results with the Wilcoxon tests (see also **Figure 4.2.S3C**) are shown. ∗ p <0.05, ∗∗ p <0.01, and ∗∗∗ p <0.001. (**E**) Example of univariate CCR6$^+$, CXCR3$^+$, and CXCR5$^+$ gatings on total and HIV Gag-specific CD4$^+$ T cell populations. (**F**) Enrichment scores of the AIMs in the Gag-specific CD4$^+$ CCR6$^+$ (green), CXCR3$^+$ (yellow), and CXCR5$^+$ (purple) T cell populations. The scores were calculated by dividing, for each individual AIM, the Gag-specific CD4$^+$ phenotype$^+$ population by the Gag-specific CD4$^+$ population, irrespective of the polarization (see also **Figure 4.2.S3D**). Wilcoxon tests (see also **Figure 4.2.S3E**) are shown. ∗ p <0.05, ∗∗ p <0.01, and ∗∗∗ p <0.001. (**G**) Proportion of HIV-specific CCR6$^+$ and CXCR5$^+$ CD4$^+$ T cells still detected after ad hoc removal of the indicated AIM molecules. The values were normalized to those of the 6xAIM. Wilcoxon tests are shown. (**A, F, and G**) Medians ± interquartile ranges are shown. (**A and F**) Pairwise multiple comparisons were performed using the Friedman test with Dunn's post-hoc test. (**D and F**) The graphs are on a log$_2$ scale. To avoid contaminating phenotype profiling with excessive background, donors who had at least one HIV stimulation (**A–D**) or the HIV Gag stimulation (**F and G**) reaching >2X over the unstimulated condition were kept. (**A–D**) n = 14 ART participants; (**F and G**) n = 13 participants.

We validated these observations with focused univariate analyses of chemokine receptor expression (**Figure 4.2.3E**). We examined Gag-specific responses, as we found strong AIM$^+$ CD4$^+$ T cell responses against this antigen (**Figure 4.2.1C**). We compared the expression of CD69, 4-1BB, OX40, and CD40L in Gag-specific CCR6$^+$, CXCR3$^+$, and CXCR5$^+$ CD4$^+$ T cells to their expression in the parental AIM$^+$ Gag-specific CD4$^+$ T cell population and calculated enrichment scores (**Figure 4.2.S3D**). The pattern of AIM expression in CXCR3$^+$ cells was similar

to the parental population (**Figures 4.2.3F and 4.2.S3E**). Consistent with the unsupervised analyses, we found that CD40L was preferentially expressed in the CXCR5$^+$ population to the detriment of CCR6$^+$ cells, while we saw the opposite for OX40, which was slightly enriched in Th17/Th22-like cells. We next tested whether the removal of OX40 or CD40L would affect the detection of HIV-specific CXCR5$^+$ and CCR6$^+$ cells (**Figure 4.2.3G**). Removal of CD40L resulted in a median 24% decrease in HIV-specific CCR6$^+$ CD4$^+$ T cells. The loss of detection was significantly exacerbated when OX40 was excluded instead, with a median 62% decrease. In contrast, removal of either OX40 or CD40L resulted in a similar median decrease (28% vs. 30%) of HIV-specific CXCR5$^+$ CD4$^+$ T cells, despite the enrichment of CD40L in Tfh cells.

We next verified if the 6xAIM could capture subtle Ag-specific phenotypic skewing by comparing Gag- and CMV-specific CD4$^+$ T cells (**Figure 4.2.S3F**). The 6xAIM approach captured the previously reported[22] higher proportions of CXCR3$^+$ and CXCR5$^+$ cells in the total Gag-specific CD4$^+$ T cell population as compared to CMV pp65-specific CD4 T$^+$ cells (**Figure 4.2.S3G**).

These results show that the sampling and phenotyping of HIV-specific CD4$^+$ T cells may vary depending on the pairs of AIMs selected. Integrating multiple AIMs mitigates these variations. However, a certain degree of redundancy exists between the AIMs, and the removal of a single AIM molecule still allows the detection of all unsupervised clusters.

Infection-specific (HIV) and vaccine-induced (SARS-CoV-2) responses are detected by the same AIM pairs

Finally, to test the versatility of our approach, we examined the multiplexed AIM responses in a cohort of 23 healthcare workers (HCWs) three weeks after the second dose of SARS-CoV-2 mRNA vaccines (**Figure 4.2.S4A**). These participants were vaccinated according to a long 16 week interval regimen as previously published by our group (**Table 4.2.S1**).[31–33] We next compared the features of the infection-induced HIV-specific T cells in PWH to those of vaccine-induced SARS-CoV-2 Spike-specific T cells in HCWs (**Figures 4.2.S4B and 4.2.S4C**). Consistent with our observations on HIV-specific T cell responses, we found that the 6xAIM approach provided more robust measures of Spike-specific CD4$^+$ T cell responses (**Figure 4.2.4A**). Moreover, OX40 (CD4$^+$) was still the most important AIM molecule for Spike-specific T cell detection

(**Figure 4.2.4B**). Both CD69 and 4-1BB were essential to detect Spike-specific CD8[+] T cells (**Figures 4.2.4A and 4.2.4B**).

We next tested whether individual AIM pairs provided similar estimations of Ag-specific CD4[+] and CD8[+] T cell responses in different cohorts. For this comparative purpose, we normalized the data based on the 6xAIM, which provided the highest detection levels in both cohorts (**Figures 4.2.4C and 4.2.S4D**). Despite some differences, the overall pattern of detection was similar between the two cohorts. In CD8[+] T cells, the CD69[+]4-1BB[+] pair was equivalently efficient at detecting Ag-specific cells. These data indicate that the use of a single AIM pair does not introduce major cohort-based biases in quantifications.
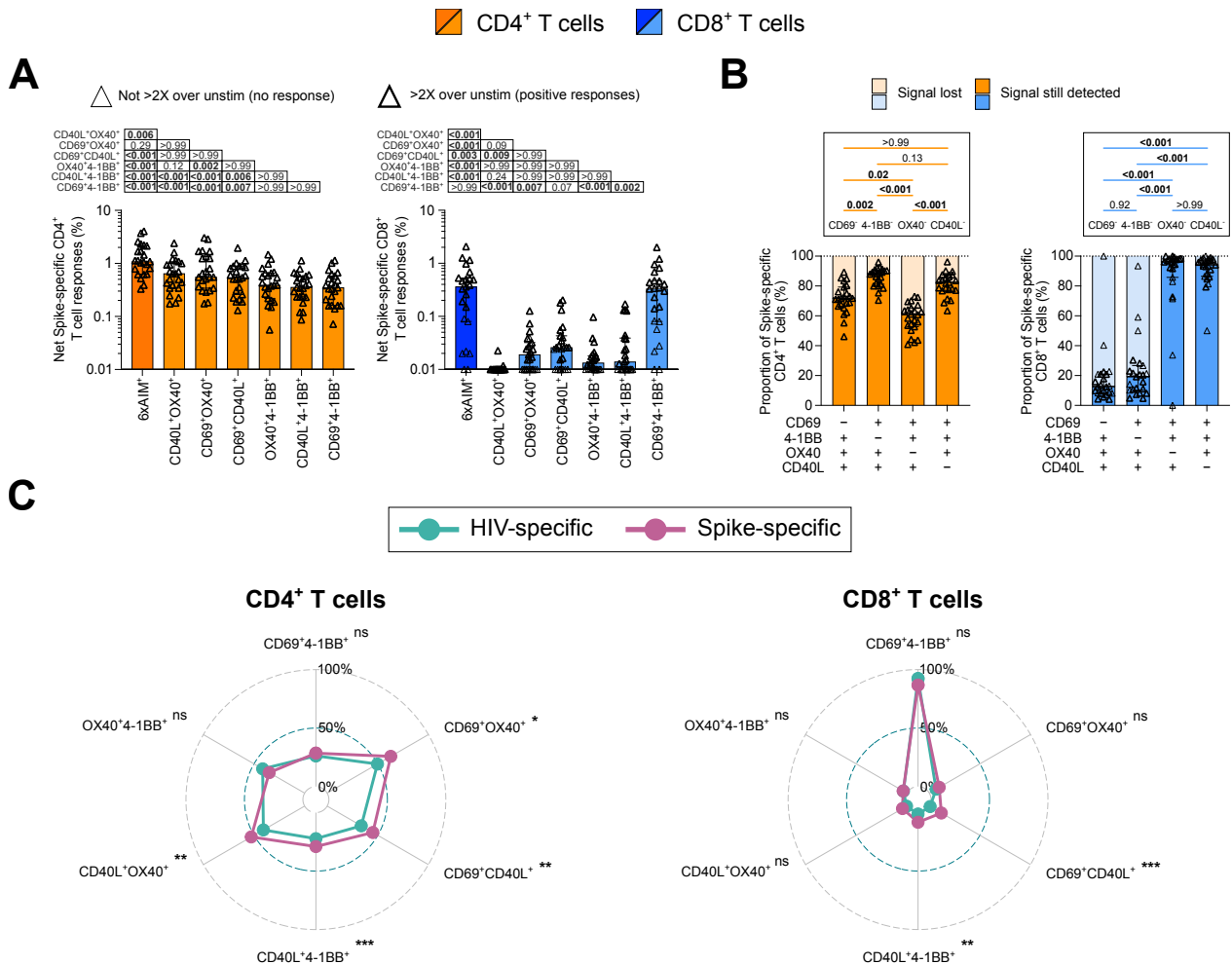
Figure 4.2.4 – Infection-specific (HIV) and vaccine-induced (SARS-CoV-2) responses are detected by the same AIM pairs. (A) Net SARS-CoV-2 Spike-specific CD4[+] (orange) and CD8[+] (blue) T cell responses. The 6xAIM and individual AIM pairs are ranked in decreasing order of the median, based on the

CD4$^+$ T cells. The tables above the graphs represent the pairwise multiple comparisons performed using the Friedman test with Dunn's post hoc test. (**B**) Individual contributions of each AIM marker for the detection of Spike-specific CD4$^+$ (left) and CD8$^+$ (right) T cell responses. The detection levels were normalized to the value yielded by the 6xAIM. Light colors represent the signal lost when the indicated AIM marker is removed from the analysis; dark colors represent the residual signal still detected. Pairwise multiple comparisons were performed using the Friedman test with Dunn's post-hoc test. (**C**) Spider charts depicting the comparisons between the normalized amplitudes of net HIV- (teal) and Spike-specific (lavender) CD4$^+$ and CD8$^+$ T cell responses. Within each cohort, responses were normalized to their 6xAIM values. Medians and Mann-Whitney tests are shown for each pair (see also **Figure 4.2.S4D**). $*$ p <0.05, $**$ p <0.01, and $***$ p <0.001. Donors who had at least one HIV stimulation reaching >2X over the unstimulated conditions were kept, as well as those whose Spike stimulation reached the same threshold. For CD4$^+$ T cells, n = 14 ART participants and n = 23 SARS-CoV-2-vaccinated participants; for CD8$^+$ T cells, n = 15 ART participants and n = 16 SARS-CoV-2-vaccinated participants. (**A and B**) The bars represent medians ± interquartile ranges. Thicker borders represent donors whose stimulation reached >2× over the unstimulated conditions. n = 23 SARS-CoV-2-vaccinated participants.

4.2.3 Discussion

In this study, we provide a new approach to enhance the sensitivity of the AIM assay by multiplexing AIM markers that have been validated in previous studies. We tested the robustness of a multiplexed AIM assay that combined CD69, 4-1BB, CD40L, and OX40 and comprehensively compared this integrated approach to individual AIM pairs. The multiplexed AIM approach consistently detected higher magnitudes of net Ag-specific CD4$^+$ T responses than any individual AIM pair while maintaining high specificity. We did not observe a clear benefit of multiplexing these markers for CD8$^+$ T cells, as most of the signal was provided by the CD69$^+$4-1BB$^+$ combination. These findings were consistent for immune responses to viral infections (HIV and CMV) and vaccines (SARS-CoV-2). We found CD40L to be enriched in CXCR5$^+$ CD4$^+$ T cells and OX40 in CCR6$^+$ CD4$^+$ T cells. Consequently, a slightly skewed phenotypic portrait of HIV-specific CD4$^+$ T cells emerged when relying on these markers in standard AIM assays. Multiplexing the AIM assay mitigated this issue.

Our data show that the multiplexed assay has several additional advantages. It allows co-assessment of CD4[+] and CD8[+] T cell responses in a single panel. In contrast, in previous studies, CD4[+] and CD8[+] T cells were often detected by distinct single pairs of AIMs.[14,17,18,24,25] This can be deconvoluted in more standard analyses (e.g., 6xAIM Boolean OR gating vs. single AIM pair), facilitating comparisons with other studies. For instance, two studies assessing vaccine responses to distinct vaccines reported apparently conflicting results: one, using the pairs OX40[+]4-1BB[+] for CD4[+] and CD69[+]4-1BB[+] for CD8[+] T cells, reported that AIM[+] responses were lower than those detected by ICS or ELISpot assays,[46] while the other one reported the opposite,[13] although it used different AIM pairs (OX40[+]CD25[+] and OX40[+]PDL1[+] for CD4[+] and OX40[+]CD25[+] and CD107a[+]CD25[+] for CD8[+] T cells). These discrepancies might be explained either by the antigens or by the different pairs of AIMs used, thus underlining the problematic absence of a consensual pair of AIMs. Multiplexing the assay also mitigates the donor-dependent variations observed in the specific AIM pair giving the strongest Ag-specific CD4[+] T cell response.

One unavoidable compromise of this approach is the slightly increased background, as the background of all pairs is summed up. This could be theoretically problematic when some pair combinations do not provide any specific signal per se like in CD8[+] T cells. Nevertheless, this had little impact on the signal-to-noise ratio that remained high in CD4[+] T cells. Our current iteration of the multiplexed AIM assay does not provide benefit for the detection of Ag-specific CD8[+] T cells. Alternatives are conceivable but would require the inclusion of (or substitution with) additional molecules upregulated in CD8[+] T cells upon activation, such as PD-L1 and CD107a.[13,41,47,48] Substitution for other AIM markers could also be required for more specific purposes, like identifying T cells from lymphoid tissue where CD69 is constitutively expressed at high levels.[15]

We observed a high degree of redundancy among the AIMs in CD4[+] T cells. For instance, when CD69, OX40, and 4-1BB markers are combined, CD40L appears to be dispensable for purely quantitative purposes—although it is a functionally important co-signaling molecule. Removing CD40L can be advantageous for some live-cell studies, as CD40L staining requires a preincubation with a CD40-blocking antibody, which might alter cellular interactions. OX40 was comparatively less redundant, further demonstrating its importance in core sets of AIM panels. Our results indicate that excluding OX40 could lead to underestimating the magnitude of responses. Therefore,

redundancy in AIM markers is desirable because it mitigates possible quantitative and qualitative biases associated with individual AIM pairs.

The inclusion of multiple AIMs, each with diverse biological functions, also offers the added benefit of providing a wealth of complementary information about the functional states of the Ag-specific cells. As AIMs are functional molecules, it is not surprising that they were differentially expressed in CD4[+] subsets. CD40L was more frequently expressed in CXCR5[+] (Tfh-like) cells, which is consistent with studies identifying CD40L as an important co-stimulatory molecule for Tfh differentiation and function, in particular the regulation of isotype switching in B cells.[49–51] Not much is known about the role of OX40 in Th22 cells,[52] and the studies on activated Th17 cells are conflicted: on one hand, OX40 may be involved in their maintenance and functionality,[52–54] but on the other hand, the OX40-OX40L pathway would inhibit their differentiation and the production of interleukin-17 (IL-17), one of the main cytokines produced by this subset.[52,53] While the 6xAIM allows for a broader phenotyping, it still captures pathogen-specific differences in T helper differentiation, such as the elevated proportion of CXCR5[+] and CXCR3[+] cells in Gag-specific CD4[+] T cells compared to CMV-specific CD4[+] T cells.[22] Of note, removing CD40L from our analysis had a limited impact on our capacity to detect HIV-specific CXCR5[+] CD4[+] T cells because of the frequent redundant detection of CD40L[+] cells by other AIM markers. In contrast, OX40 is less redundant, and its exclusion led to poor detection of HIV-specific CCR6[+] (Th17/Th22-like) CD4[+] T cells. Therefore, our data suggest that multiplexing the AIM assay mitigates the risk of underestimating certain CD4[+] T cell subsets over others.

We found good consistency in the detection provided by individual pairs of AIMs across different cohorts. This further demonstrates the robustness and flexibility of the AIM assay. Yet, several individual AIM[+] CD4[+] T cell pairs appeared to slightly underperform in the chronic HIV infection group compared to the vaccinal cohort. Group comparisons must be analyzed with caution, as these small differences may simply reflect heterogeneity between groups of people. Alternatively, T cell dysfunction caused by persistent activation during chronic HIV infection may affect certain activation markers more than others. The multiplexed strategy can represent a harmonizing approach, as it consistently provided the highest measures of Ag-specific CD4[+] T cells. Multiplexing the AIM can thus represent a prudent agnostic strategy when it is

unknown whether a particular pathological context can have an inherent effect on individual AIM expression.

Finally, the multiplexed AIM assay could be applied to numerous vaccine studies, especially for those against HIV, whose development remains challenging.[55] Specifically, detecting circulating Tfh (cTfh) cells in these types of studies is important, considering that they interact with B cells to help develop an efficient humoral immunity.[4] To name only a few, OX40$^+$4-1BB$^{+40,56}$ and CD40L$^+$OX40^{57} have been documented to do this in SARS-CoV-2 vaccine studies. These might not be applicable for HIV-vaccine trials, however, as we showed that the hierarchy of AIM pairs for CD4$^+$ T cell detection may slightly differ between ART-treated and SARS-CoV-2-vaccinated cohorts. Instead, multiplexing the AIM assay could be a pragmatic and promising approach to optimize both the magnitude and the diversity of vaccine-induced responses detected.

Limitations of the study

We found that the multiplexed strategy, using the AIMs CD69, 4-1BB, OX40, and CD40L, was a strong alternative to using single pairs of AIMs. However, these assays were solely done on peripheral blood and do not provide a full portrait of the tissues' immune cells.[7] Although it is a reliable alternative to conventional approaches based on cytokine secretion, the AIM assay remains a functional assay. Therefore, functionally exhausted cells that do not respond to TCR stimulation will persist in being undetected. We also acknowledge that false positive events are inherent limitations of any AIM assay. This is mitigated by calculating the net responses, which is now a standard practice to analyze AIM, as it is for ICS data, but also by reporting the proportion of background signal in the raw data. As in standard AIM assays, some bystander activation can also occur in the context 6xAIM assay, but this is mitigated by selecting AIM markers less susceptible to bystander activation after a 15 h incubation.[7] The setting of the bystander activation experiment maximized the possibility of detecting bystander activation. As soluble factors can accumulate over 15 h in the "conditioned" supernatants, they are given a full extra 15 h additional incubation time to exert their non-specific activation effect (total of 30 h) instead of the standard 15 h. Additionally, this 30 h incubation extends past the 24 h, where bystander activation has been reported to take place.[7] Finally, our panel was limited to four AIMs. While CD69$^+$4-1BB$^+$ proved to be a robust

AIM pair to detect Ag-specific CD8$^+$ T cells, the addition of other surface markers could help identify other or superior AIM combinations to characterize Ag-specific CD4$^+$ and CD8$^+$ T cells.

4.2.4 Acknowledgments

Author contributions

A.L., M. Dubé, and D.E.K. designed the study; A.L., G.S., A.N., M. Dubé, M.N., G.-G.D., R.C., N.B., M.L., M. Duchesne, and A.M.S.F. performed the AIM assays; A.L. and M. Dubé analyzed the data; A.L. performed the unsupervised clustering analyses with input from O.T.; A.F. and D.E.K. secured the samples; A.L., M. Dubé, and D.E.K. wrote the initial draft of the manuscript; every author read, edited, and approved the final manuscript; and M. Dubé and D.E.K. supervised the study.

Declaration of interests

The authors declare no competing interests.

4.2.5 STAR Methods

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Daniel E. Kaufmann (daniel.kaufmann@chuv.ch).

Materials availability

There are restrictions to the availability of clinical samples (PBMCs), in compliance with our IRB protocols and the informed consent obtained from each participant. This study did not generate new unique reagents.

Data and code availability

- The published article includes all of the processed datasets generated for this study. Further information and requests for resources and reagents should be directed to and fulfilled by the Lead contact Author (daniel.kaufmann@chuv.ch).

- We developed R codes scripted to create spider charts and perform unsupervised analyses on HIV-specific CD4$^+$ T cells. All original codes have been deposited at Github and are publicly available as of publication. DOIs are listed in the key resources table.

- Any additional information required to reanalyze the data reported in this paper is available from the Lead contact Author upon request (daniel.kaufmann@chuv.ch).

Experimental model and subject details

Ethics statement

All work was conducted in accordance with the Declaration of Helsinki in terms of informed consent and approval by an appropriate institutional board. Blood samples were obtained from donors who consented to participate in this research project at the CHUM (13.019 and 19.381). Plasma and PBMCs were isolated by centrifugation and Ficoll gradient, and samples were stored at −80°C and in liquid nitrogen, respectively, until use.

Participants and samples

We obtained the leukaphereses from study participants at the McGill University Health Center (Montréal, QC, Canada), and at the Center Hospitalier de l'Université de Montréal (CHUM; Montréal, QC, Canada). The study was approved by the respective IRBs, and written informed consent was obtained from all participants before enrollment. Our cohort of 16 ART-treated participants[34] with controlled viremia (<40 viral RNA copies/mL) was used to measure HIV and CMV-specific responses. One donor who was negative for CMV was removed from the CMV analyses. The SARS-CoV-2 vaccinated cohort included 23 healthcare workers (HCW) who had never been previously infected by the virus and received two doses of mRNA vaccine, with a 16-week interval. Blood samples were collected approximately three weeks after the second dose. The number of participants for this specific time point was the same as previously published.[31] The characteristics of all participants are summarized in **Table 4.2.S1**. PBMCs were isolated by the Ficoll density gradient method and stored in liquid nitrogen until use.

Method details

Activation-induced markers (AIM) assay

The multiplexed AIM assay[21–23,41] was adapted for HIV-,[34] CMV-, and SARS-CoV-2-specific[30–33] CD4$^+$ and CD8$^+$ T cells. PBMCs were thawed and rested for 3 h in 96-well flat-bottom plates in RPMI 1640 supplemented with HEPES, penicillin/streptomycin, and 10% FBS. For HIV, 2×10^6 PBMCs were stimulated with pools of 150 overlapping peptides of HIV-1 Gag, Pol, Env, or Nef proteins (0.5 µg/mL per peptide) (JPT) for 15 h (37°C, 5% CO$_2$). For CMV, 2×10^6 PBMCs were also stimulated with a pool of 138 overlapping HCMVA 65 kDa phosphoprotein (pp65) (0.5 µg/mL per peptide) (JPT) peptides. For SARS-CoV-2, the stimulation was done with 1.7×10^6 PBMCs and an S glycoprotein peptide pool (0.5 µg/mL per peptide), corresponding to the pool of 315 overlapping peptides (15-mers) spanning the complete amino acid sequence of the Spike Glycoprotein (JPT). CCR6, CXCR3, CXCR5, and CXCR6 antibodies, as well as a CD40 blocking antibody, were added in culture 15 min before stimulation. A DMSO-treated condition and a *Staphylococcus enterotoxin B* (SEB)-treated (0.5 µg/mL) condition served, respectively, as negative and positive controls. Cells were first stained for viability dye (Aquavivid, Thermofisher, 20 min, 4°C), surface markers (30 min, 4°C) (see **Table 4.2.S2** for the list of antibodies), and then

fixed using 1% paraformaldehyde (15 min, 4°C) before acquisition on the flow cytometer FACSymphony A5 Cell Analyzer (BD Biosciences). **Figure 4.2.S1A** describes the upstream gating strategy. Analyses were performed using FlowJo (v10.8.2).

Bystander activation experiment

A first AIM assay [AIM #1] was performed, as described above, on PBMCs of three ART participants thawed on day 1. On day 2, after 15 h of Gag stimulation, the conditioned supernatants from this AIM #1 were collected, clarified by two rounds of centrifugation (at 900 RPM and 1500 RPM, respectively, for 10 min each) in 1.5 mL microtubes, then transferred on freshly purified CD3$^+$ T lymphocytes (negative selection with the EasySep Human T cell Isolation Kit; StemCell) from an HIV-naïve participant (uninfected donor, UD) [AIM #2]. These "sentinel" T cells were incubated for 15 h in the conditioned supernatants and stained on day 3. Cells from AIM #1 and AIM #2 were acquired on the FACSymphony A5 Cell Analyzer on days 2 and 3, respectively.

Software scripts and visualization

Graphics were generated using GraphPad PRISM (v9.5.0) (GraphPad, San Diego, CA). For the unsupervised analyses, HIV-specific CD4$^+$ T cells were first downsampled to a comparable number of events (500 cells), and FCS files were loaded through the flowCore package (v2.10.0).[58] Scaling and logicle transformation of the flow cytometry data was applied using FlowSOM (v2.6.0),[59] as done previously.[31] The uniform manifold approximation and projection (UMAP) algorithm was performed using the R package M3C (v1.20.0),[60] while the clustering was achieved using Phenograph (v0.99.1)[45] with the *k* parameter (number of nearest-neighbors) set to 250, which we determined based on numerous iterations with varying *k* values until this plateau was reached. We obtained 12 clusters. These clusters were further grouped in "superclusters" based on their similar patterns of polarization markers expression. Heatmaps and spider charts were generated in R (v4.2.2) using the packages gglot2 (v3.4.0),[61] pheatmap (v1.0.12),[62] and ggradar (v0.2).[63] R codes scripted for this paper are available at https://github.com/alemi055/scripts-and-data/tree/master, in the archive Lemieuxetal_CellRepMet_2024.tar.bz2.

Quantification and statistical analysis

Antigen-specific T cell responses are expressed in percentages of CD4$^+$ and CD8$^+$ T cells. Symbols (circles for the HIV cohort, squares for the CMV cohort, and triangles for the SARS-CoV-2 cohort)

represent biologically independent samples. Lines connect data from the same donor. Median fold increases (Fis) were calculated by dividing the net T cell responses of the 6xAIM by the individual AIM pair's (net Fis), or the AIM$^+$ T cell responses of each pair by its respective unstimulated (background) values (raw Fis). For the intra- and inter-experiment reproducibility, coefficients of variation (CV) were calculated by dividing the standard deviation by the mean. Differences in responses for the same donor in the HIV and CMV cohorts were performed using Wilcoxon matched pair tests, while tests comparing different donors in the HIV and SARS-CoV-2 cohorts were performed using Mann-Whitney (unpaired) tests. Pairwise multiple comparisons were performed using the Friedman test with Dunn's post-hoc test. The Wilcoxon, Mann-Whitney, and Friedman tests were generated using GraphPad Prism. *P* values are indicated for each comparison assessed, and *p* <0.05 were considered significant. For graphical representation on a log scale (but not for statistical tests), null values were arbitrarily set at the minimal values for each assay.

4.2.6 References

1. Swain, S.L., McKinstry, K.K., and Strutt, T.M. (2012). Expanding roles for CD4+ T cells in immunity to viruses. Nat. Rev. Immunol. 12, 136–148. 10.1038/nri3152.

2. Sant, A.J., and McMichael, A. (2012). Revealing the role of CD4+ T cells in viral immunity. J. Exp. Med. 209, 1391–1395. 10.1084/jem.20121517.

3. Cox, M.A., Kahan, S.M., and Zajac, A.J. (2013). Anti-viral CD8 T cells and the cytokines that they love. 2013 Rev. Issue 435, 157–169. 10.1016/j.virol.2012.09.012.

4. Crotty, S. (2014). T Follicular Helper Cell Differentiation, Function, and Roles in Disease. Immunity 41, 529–542. 10.1016/j.immuni.2014.10.004.

5. Zhu, X., and Zhu, J. (2020). CD4 T Helper Cell Subsets and Related Human Immunological Disorders. Int. J. Mol. Sci. 21. 10.3390/ijms21218011.

6. Chatzileontiadou, D.S.M., Sloane, H., Nguyen, A.T., Gras, S., and Grant, E.J. (2021). The Many Faces of CD4+ T Cells: Immunological and Structural Characteristics. Int. J. Mol. Sci. 22. 10.3390/ijms22010073.

7. Reiss, S., Baxter, A.E., Cirelli, K.M., Dan, J.M., Morou, A., Daigneault, A., Brassard, N., Silvestri, G., Routy, J.-P., Havenar-Daughton, C., et al. (2017). Comparative analysis of activation induced marker (AIM) assays for sensitive identification of antigen-specific CD4 T cells. PLOS ONE 12, e0186998. 10.1371/journal.pone.0186998.

8. Chattopadhyay, P.K., Yu, J., and Roederer, M. (2005). A live-cell assay to detect antigen-specific CD4+ T cells with diverse cytokine profiles. Nat. Med. 11, 1113–1117. 10.1038/nm1293.

9. Lamoreaux, L., Roederer, M., and Koup, R. (2006). Intracellular cytokine optimization and standard operating procedure. Nat. Protoc. 1, 1507–1516. 10.1038/nprot.2006.268.

10. Yu, D., Rao, S., Tsai, L.M., Lee, S.K., He, Y., Sutcliffe, E.L., Srivastava, M., Linterman, M., Zheng, L., Simpson, N., et al. (2009). The Transcriptional Repressor Bcl-6 Directs T Follicular Helper Cell Lineage Commitment. Immunity 31, 457–468. 10.1016/j.immuni.2009.07.002.

11. Kroenke, M.A., Eto, D., Locci, M., Cho, M., Davidson, T., Haddad, E.K., and Crotty, S. (2012). Bcl6 and Maf Cooperate To Instruct Human Follicular Helper CD4 T Cell Differentiation. J. Immunol. 188, 3734–3744. 10.4049/jimmunol.1103246.

12. Bacher, P., and Scheffold, A. (2013). Flow-cytometric analysis of rare antigen-specific T cells. Cytometry A 83A, 692–701. 10.1002/cyto.a.22317.

13. Bowyer, G., Rampling, T., Powlson, J., Morter, R., Wright, D., Hill, A.V.S., and Ewer, K.J. (2018). Activation-induced Markers Detect Vaccine-Specific CD4+ T Cell Responses Not Measured by Assays Conventionally Used in Clinical Trials. Vaccines 6. 10.3390/vaccines6030050.

14. Ferragut, F., Cruz, K.M., Gallardo, J.P., Fernández, M., Hernández Vasquez, Y., and Gómez, K.A. (2022). Activation-induced marker assays for identification of Trypanosoma cruzi-specific CD4 or CD8 T cells in chronic Chagas disease patients. Immunology. 10.1111/imm.13622.

15. Havenar-Daughton, C., Reiss, S.M., Carnathan, D.G., Wu, J.E., Kendric, K., Torrents de la Peña, A., Kasturi, S.P., Dan, J.M., Bothwell, M., Sanders, R.W., et al. (2016). Cytokine-Independent Detection of Antigen-Specific Germinal Center T Follicular Helper Cells in Immunized Nonhuman Primates Using a Live Cell Activation-Induced Marker Technique. J. Immunol. 197, 994–1002. 10.4049/jimmunol.1600320.

16. Frentsch, M., Arbach, O., Kirchhoff, D., Moewes, B., Worm, M., Rothe, M., Scheffold, A., and Thiel, A. (2005). Direct access to CD4+ T cells specific for defined antigens according to CD154 expression. Nat. Med. 11, 1118–1124. 10.1038/nm1292.

17. Painter, M.M., Mathew, D., Goel, R.R., Apostolidis, S.A., Pattekar, A., Kuthuru, O., Baxter, A.E., Herati, R.S., Oldridge, D.A., Gouma, S., et al. (2021). Rapid induction of antigen-specific CD4+ T cells is associated with coordinated humoral and cellular immunity to SARS-CoV-2 mRNA vaccination. Immunity 54, 2133-2142.e3. 10.1016/j.immuni.2021.08.001.

18. Goel, R.R., Painter, M.M., Apostolidis, S.A., Mathew, D., Meng, W., Rosenfeld, A.M., Lundgreen, K.A., Reynaldi, A., Khoury, D.S., Pattekar, A., et al. (2021). mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. Science 374, abm0829. 10.1126/science.abm0829.

19. Dan, J.M., Lindestam Arlehamn, C.S., Weiskopf, D., da Silva Antunes, R., Havenar-Daughton, C., Reiss, S.M., Brigger, M., Bothwell, M., Sette, A., and Crotty, S. (2016). A Cytokine-Independent Approach To Identify Antigen-Specific Human Germinal Center T Follicular Helper Cells and Rare Antigen-Specific CD4+ T Cells in Blood. J. Immunol. 197, 983–993. 10.4049/jimmunol.1600318.

20. Barham, M.S., Whatney, W.E., Khayumbi, J., Ongalo, J., Sasser, L.E., Campbell, A., Franczek, M., Kabongo, M.M., Ouma, S.G., Hayara, F.O., et al. (2020). Activation-Induced Marker Expression Identifies Mycobacterium tuberculosis–Specific CD4 T Cells in a Cytokine-Independent Manner in HIV-Infected Individuals with Latent Tuberculosis. ImmunoHorizons 4, 573–584. 10.4049/immunohorizons.2000051.

21. Morou, A., Brunet-Ratnasingham, E., Dubé, M., Charlebois, R., Mercier, E., Darko, S., Brassard, N., Nganou-Makamdop, K., Arumugam, S., Gendron-Lepage, G., et al. (2019). Altered differentiation is central to HIV-specific CD4+ T cell dysfunction in progressive disease. Nat. Immunol. 20, 1059–1070. 10.1038/s41590-019-0418-x.

22. Niessl, J., Baxter, A.E., Morou, A., Brunet-Ratnasingham, E., Sannier, G., Gendron-Lepage, G., Richard, J., Delgado, G.-G., Brassard, N., Turcotte, I., et al. (2020). Persistent expansion and Th1-like skewing of HIV-specific circulating T follicular helper cells during antiretroviral therapy. EbioMedicine 54, 102727. 10.1016/j.ebiom.2020.102727.

23. Brunet-Ratnasingham, E., Morou, A., Dubé, M., Niessl, J., Baxter, A.E., Tastet, O., Brassard, N., Ortega-Delgado, G., Charlebois, R., Freeman, G.J., et al. (2022). Immune checkpoint expression on HIV-specific CD4+ T cells and response to their blockade are dependent on lineage and function. eBioMedicine 84, 104254. 10.1016/j.ebiom.2022.104254.

24. Busà, R., Sorrentino, M.C., Russelli, G., Amico, G., Miceli, V., Miele, M., Di Bella, M., Timoneri, F., Gallo, A., Zito, G., et al. (2022). Specific Anti-SARS-CoV-2 Humoral and Cellular Immune Responses After Booster Dose of BNT162b2 Pfizer-BioNTech mRNA-Based Vaccine: Integrated Study of Adaptive Immune System Components. Front. Immunol. 13. 10.3389/fimmu.2022.856657.

25. Pallikkuth, S., Williams, E., Pahwa, R., Hoffer, M., and Pahwa, S. (2021). Association of Flu specific and SARS-CoV-2 specific CD4 T cell responses in SARS-CoV-2 infected asymptomatic heath care workers. Vaccine 39, 6019–6024. 10.1016/j.vaccine.2021.08.092.

26. Naaber, P., Tserel, L., Kangro, K., Sepp, E., Jürjenson, V., Adamson, A., Haljasmägi, L., Rumm, A.P., Maruste, R., Kärner, J., et al. (2021). Dynamics of antibody response to BNT162b2 vaccine after six months: a longitudinal prospective study. Lancet Reg. Health – Eur. 10, 100208. 10.1016/j.lanepe.2021.100208.

27. Tarke, A., Coelho, C.H., Zhang, Z., Dan, J.M., Yu, E.D., Methot, N., Bloom, N.I., Goodwin, B., Phillips, E., Mallal, S., et al. (2022). SARS-CoV-2 vaccination induces immunological T cell memory able to cross-recognize variants from Alpha to Omicron. Cell 185, 847-859.e11. 10.1016/j.cell.2022.01.015.

28. Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S., et al. (2020). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. Cell 181, 1489-1501.e15. 10.1016/j.cell.2020.05.015.

29. Rydyznski Moderbacher, C., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, C., Choi, J., et al. (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. 10.1016/j.cell.2020.09.038.

30. Tauzin, A., Nayrac, M., Benlarbi, M., Gong, S.Y., Gasser, R., Beaudoin-Bussières, G., Brassard, N., Laumaea, A., Vézina, D., Prévost, J., et al. (2021). A single dose of the SARS-CoV-2 vaccine BNT162b2 elicits Fc-mediated antibody effector functions and T cell responses. Cell Host Microbe 29, 1137-1150.e6. 10.1016/j.chom.2021.06.001.

31. Nayrac, M., Dubé, M., Sannier, G., Nicolas, A., Marchitto, L., Tastet, O., Tauzin, A., Brassard, N., Lima-Barbosa, R., Beaudoin-Bussières, G., et al. (2022). Temporal associations of B and T cell immunity with robust vaccine responsiveness in a 16-week interval BNT162b2 regimen. Cell Rep. 39, 111013. 10.1016/j.celrep.2022.111013.

32. Nicolas, A., Sannier, G., Dubé, M., Nayrac, M., Tauzin, A., Painter, M.M., Goel, R.R., Laporte, M., Gendron-Lepage, G., Medjahed, H., et al. (2023). An extended SARS-CoV-2 mRNA

vaccine prime-boost interval enhances B cell immunity with limited impact on T cells. iScience 26, 105904. 10.1016/j.isci.2022.105904.

33. Sannier, G., Nicolas, A., Dubé, M., Marchitto, L., Nayrac, M., Tastet, O., Chatterjee, D., Tauzin, A., Lima-Barbosa, R., Laporte, M., et al. (2023). A third SARS-CoV-2 mRNA vaccine dose in people receiving hemodialysis overcomes B cell defects but elicits a skewed CD4+ T cell profile. Cell Rep. Med. 4, 100955. 10.1016/j.xcrm.2023.100955.

34. Dubé, M., Tastet, O., Dufour, C., Sannier, G., Brassard, N., Delgado, G.-G., Pagliuzza, A., Richard, C., Nayrac, M., Routy, J.-P., et al. (2023). Spontaneous HIV expression during suppressive ART is associated with the magnitude and function of HIV-specific CD4+ and CD8+ T cells. Cell Host Microbe 31, 1507-1522.e5. 10.1016/j.chom.2023.08.006.

35. Daoussis Dimitris, Andonopoulos Andrew P., and Liossis Stamatis-Nick C. (2004). Targeting CD40L: a Promising Therapeutic Approach. Clin. Vaccine Immunol. 11, 635–641. 10.1128/CDLI.11.4.635-641.2004.

36. Croft, M., So, T., Duan, W., and Soroosh, P. (2009). The significance of OX40 and OX40L to T-cell biology and immune disease. Immunol. Rev. 229, 173–191. 10.1111/j.1600-065X.2009.00766.x.

37. Fu, N., Xie, F., Sun, Z., and Wang, Q. (2021). The OX40/OX40L Axis Regulates T Follicular Helper Cell Differentiation: Implications for Autoimmune Diseases. Front. Immunol. 12.

38. Lachmann Raskit, Bajwa Martha, Vita Serena, Smith Helen, Cheek Elizabeth, Akbar Arne, and Kern Florian (2012). Polyfunctional T Cells Accumulate in Large Human Cytomegalovirus-Specific T Cell Responses. J. Virol. 86, 1001–1009. 10.1128/JVI.00873-11.

39. Sonnet, F., Namork, E., Stylianou, E., Gaare-Olstad, I., Huse, K., Andorf, S., Mjaaland, S., Dirven, H., and Nygaard, U. (2020). Reduced polyfunctional T cells and increased cellular activation markers in adult allergy patients reporting adverse reactions to food. BMC Immunol. 21, 43. 10.1186/s12865-020-00373-w.

40. Thieme, C.J., Anft, M., Paniskaki, K., Blazquez-Navarro, A., Doevelaar, A., Seibert, F.S., Hoelzer, B., Konik, M.J., Berger, M.M., Brenner, T., et al. (2020). Robust T Cell Response Toward Spike, Membrane, and Nucleocapsid SARS-CoV-2 Proteins Is Not Associated with Recovery in Critical COVID-19 Patients. Cell Rep. Med. 1, 100092. 10.1016/j.xcrm.2020.100092.

41. Niessl, J., Baxter, A.E., Mendoza, P., Jankovic, M., Cohen, Y.Z., Butler, A.L., Lu, C.-L., Dubé, M., Shimeliovich, I., Gruell, H., et al. (2020). Combination anti-HIV-1 antibody therapy is associated with increased virus-specific T cell immunity. Nat. Med. 26, 222–227. 10.1038/s41591-019-0747-1.

42. Lichtner, M., Cicconi, P., Vita, S., Cozzi-Lepri, A., Galli, M., Lo Caputo, S., Saracino, A., De Luca, A., Moioli, M., Maggiolo, F., et al. (2015). Cytomegalovirus Coinfection Is Associated With an Increased Risk of Severe Non–AIDS-Defining Events in a Large Cohort of HIV-Infected Patients. J. Infect. Dis. 211, 178–186. 10.1093/infdis/jiu417.

43. Booiman, T., Wit, F.W., Girigorie, A.F., Maurer, I., De Francesco, D., Sabin, C.A., Harskamp, A.M., Prins, M., Franceschi, C., Deeks, S.G., et al. (2017). Terminal differentiation of T cells is strongly associated with CMV infection and increased in HIV-positive individuals on ART and lifestyle matched controls. PLOS ONE 12, e0183357. 10.1371/journal.pone.0183357.

44. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. 37, 38–44. 10.1038/nbt.4314.

45. Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir, E.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 162, 184–197. 10.1016/j.cell.2015.05.047.

46. Tan, A.T., Lim, J.M.E., Le Bert, N., Kunasegaran, K., Chia, A., Qui, M.D.C., Tan, N., Chia, W.N., de Alwis, R., Ying, D., et al. (2021). Rapid measurement of SARS-CoV-2 spike T cells in whole blood from vaccinated and naturally infected individuals. J. Clin. Invest. 131. 10.1172/JCI152379.

47. Betts, M.R., Brenchley, J.M., Price, D.A., De Rosa, S.C., Douek, D.C., Roederer, M., and Koup, R.A. (2003). Sensitive and viable identification of antigen-specific CD8+ T cells by a flow cytometric assay for degranulation. J. Immunol. Methods 281, 65–78. 10.1016/S0022-1759(03)00265-5.

48. Wolint, P., Betts, M.R., Koup, R.A., and Oxenius, A. (2004). Immediate Cytotoxicity But Not Degranulation Distinguishes Effector and Memory Subsets of CD8+ T Cells. J. Exp. Med. 199, 925–936. 10.1084/jem.20031799.

49. Deenick, E.K., and Ma, C.S. (2011). The regulation and role of T follicular helper cells in immunity. Immunology 134, 361–367. 10.1111/j.1365-2567.2011.03487.x.

50. Awe, O., Hufford, M.M., Wu, H., Pham, D., Chang, H.-C., Jabeen, R., Dent, A.L., and Kaplan, M.H. (2015). PU.1 Expression in T Follicular Helper Cells Limits CD40L-Dependent Germinal Center B Cell Development. J. Immunol. 195, 3705–3715. 10.4049/jimmunol.1500780.

51. Cicalese, M.P., Gerosa, J., Baronio, M., Montin, D., Licciardi, F., Soresina, A., Dellepiane, R.M., Miano, M., Baselli, L.A., Volpi, S., et al. (2018). Circulating Follicular Helper and Follicular Regulatory T Cells Are Severely Compromised in Human CD40 Deficiency: A Case Report. Front. Immunol. 9.

52. Fu, Y., Lin, Q., Zhang, Z., and Zhang, L. (2020). Therapeutic strategies for the costimulatory molecule OX40 in T-cell-mediated immunity. Acta Pharm. Sin. B 10, 414–433. 10.1016/j.apsb.2019.08.010.

53. Furue, M., and Furue, M. (2021). OX40L–OX40 Signaling in Atopic Dermatitis. J. Clin. Med. 10. 10.3390/jcm10122578.

54. Zhang, Z., Zhong, W., Hinrichs, D., Wu, X., Weinberg, A., Hall, M., Spencer, D., Wegmann, K., and Rosenbaum, J.T. (2010). Activation of OX40 Augments Th17 Cytokine Expression and Antigen-Specific Uveitis. Am. J. Pathol. 177, 2912–2920. 10.2353/ajpath.2010.100353.

55. Haynes, B.F., Wiehe, K., Borrow, P., Saunders, K.O., Korber, B., Wagh, K., McMichael, A.J., Kelsoe, G., Hahn, B.H., Alt, F., et al. (2023). Strategies for HIV-1 vaccines that induce broadly neutralizing antibodies. Nat. Rev. Immunol. 23, 142–158. 10.1038/s41577-022-00753-w.

56. Mise-Omata, S., Ikeda, M., Takeshita, M., Uwamino, Y., Wakui, M., Arai, T., Yoshifuji, A., Murano, K., Siomi, H., Nakagawara, K., et al. (2022). Memory B Cells and Memory T Cells Induced by SARS-CoV-2 Booster Vaccination or Infection Show Different Dynamics and Responsiveness to the Omicron Variant. J. Immunol. 209, 2104–2113. 10.4049/jimmunol.2200525.

57. Rydyznski Moderbacher, C., Kim, C., Mateus, J., Plested, J., Zhu, M., Cloney-Clark, S., Weiskopf, D., Sette, A., Fries, L., Glenn, G., et al. (2022). NVX-CoV2373 vaccination induces functional SARS-CoV-2–specific CD4+ and CD8+ T cell responses. J. Clin. Invest. 132. 10.1172/JCI160898.

58. Ellis, B., Haaland, P., Florian, H., Le Meur, N., Gopalakrishnan, N., Spidlen, J., Jiang, M., Finak, G., and Grandjeaud, S. (2022). flowCore: Basic structures for flow cytometry data.

59. Quintelier, K., Couckuyt, A., Emmaneel, A., Aerts, J., Saeys, Y., and Van Gassen, S. (2021). Analyzing high-dimensional cytometry data using FlowSOM. Nat. Protoc. 16, 3775–3801. 10.1038/s41596-021-00550-0.

60. John, C.R., Watson, D., Russ, D., Goldmann, K., Ehrenstein, M., Pitzalis, C., Lewis, M., and Barnes, M. (2020). M3C: Monte Carlo reference-based consensus clustering. Sci. Rep. 10, 1816. 10.1038/s41598-020-58766-1.

61. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis.

62. Kolde, R. (2019). Pheatmap: Pretty Heatmaps.

63. Bion, R. ggradar: Create radar charts using ggplot2.

4.2.7 Supplemental Figures
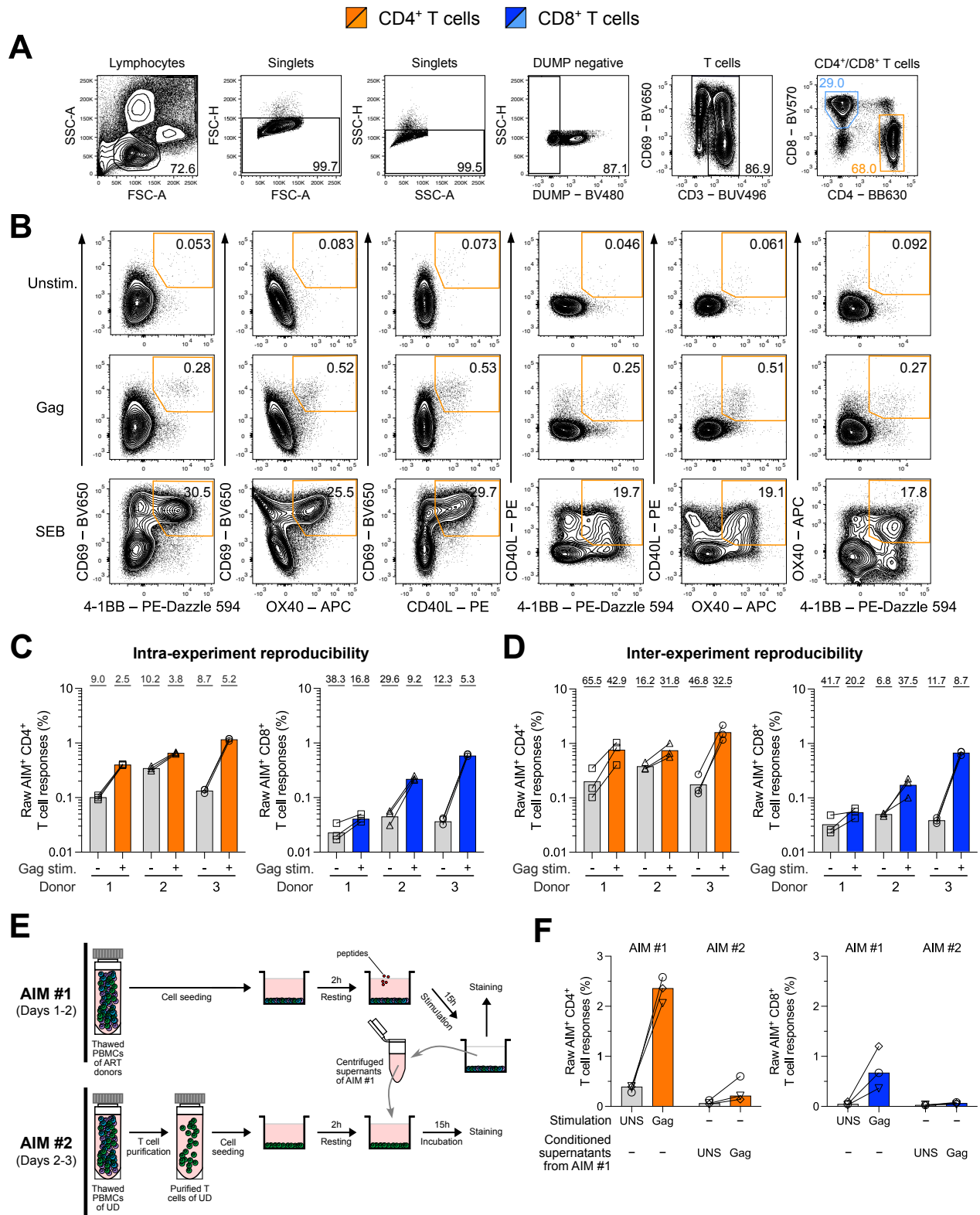
Figure 4.2.S1 – Quantification of Ag-specific CD4$^+$ and CD8$^+$ T cell responses using Boolean OR Gating. Related to Figure 4.2.1. (A) Representative upstream generic gating. **(B)** FACS plots depicting the multiplexed (6xAIM) strategy to identify HIV-specific CD4$^+$ and CD8$^+$ T cells. A DMSO-treated (unstimulated) condition and a *Staphylococcus enterotoxin B* (SEB)-treated condition served, respectively, as negative and positive controls. For simplicity, the example focuses on CD4$^+$ T cells and the HIV Gag stimulation. **(C and D)** Intra- **(C)** and inter-experiment **(D)** reproducibility. PBMCs were left unstimulated as a control (gray bars). Orange (CD4$^+$) and blue (CD8$^+$) bars represent cells stimulated with HIV Gag peptides. Means are represented. Coefficients of variation are shown above the bars. n = 3 ART participants. **(E)** Design of the bystander activation experiment, as described in the STAR Methods. **(F)** Side-by-side comparison of the primary AIM #1 and bystander AIM #2 results. Medians are shown. n = 3 ART participants and n = 1 UD. **(C, D, and F)** Raw AIM$^+$ T cell responses (6xAIM values) are represented. Distinct symbols represent individual participants.
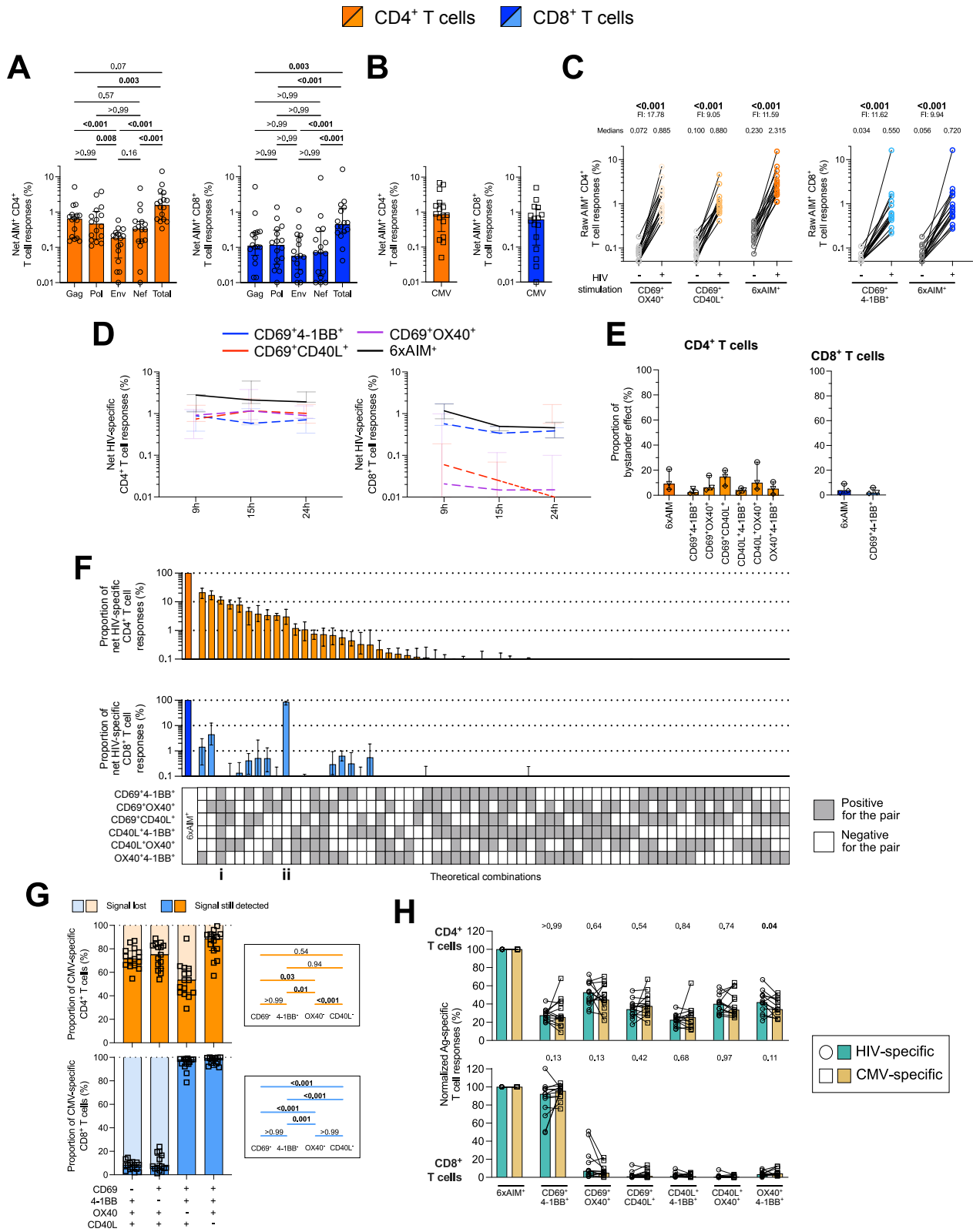
Figure 4.2.S2 – The multiplexed AIM assay improves detection of HIV- and CMV-specific CD4[+] T cells. Related to Figure 4.2.2. Net frequencies of (**A**) HIV-specific CD4[+] (orange) and CD8[+] (blue) T cell

responses following an *ex vivo* 15h stimulation of PBMCs with a pool of HIV Gag, Pol, Envelope (Env), or Nef peptides, or (**B**) CMV-specific CD4$^+$ and CD8$^+$ T cell responses following stimulation with HCMVA pp65 peptides. The 6xAIM values are represented. The net HIV Gag, Pol, Env, and Nef responses were summed to assess the "total" HIV-specific responses. (**C**) Comparisons between the raw responses of the 6xAIM strategy and benchmark AIM pairs to detect HIV-specific CD4$^+$ and CD8$^+$ T cell responses. Medians and median fold increases (Fis) are shown. (**D**) Comparisons between the net HIV-specific CD4$^+$ and CD8$^+$ T cell responses of the 6xAIM pairs strategy with those of individual pairs [CD69$^+$4-1BB$^+$ (blue), CD69$^+$OX40$^+$ (purple), and CD69$^+$CD40L$^+$ (red)], at different stimulation durations. The full and dashed lines represent, respectively, the 6xAIM strategy and the individual AIM pairs. n = 3 ART participants. (**E**) Relative proportion of bystander activation. This was estimated by dividing the net Gag-specific T cell responses of AIM #2 by those obtained in AIM #1 (see also **Figures 4.2.S1E and 4.2.S1F**). Distinct symbols represent individual ART participants. n = 1 UD. (**F**) Proportion of HIV-specific CD4$^+$ and CD8$^+$ T cell responses defined by the 6xAIM strategy. Values of each of the 64 theoretical combinations were normalized to those of the 6xAIM. Combinations are ranked in decreasing order based on the CD4$^+$ T cell responses. The AIM pair combination that each column refers to is indicated underneath the histograms: grey boxes indicate AIM pairs expressed, whereas white spaces indicate those that were not. (**G**) Individual contributions of each AIM marker for the detection of CMV-specific CD4$^+$ and CD8$^+$ T cell responses. The detection levels were normalized to the value yielded by the 6xAIM. Light colors represent the signal lost when the indicated AIM marker is removed from the analysis; dark colors represent the residual signal still detected. (**H**) Comparisons between the normalized magnitudes of net HIV- (teal) and CMV-specific (amber) CD4$^+$ and CD8$^+$ T cell responses. Within each cohort, responses were normalized to their 6xAIM pairs values. Medians are shown. Matched donors who had at least one HIV stimulation and the CMV stimulation reaching >2x over unstimulated condition were kept. n = 13 CMV$^+$ ART matched participants. (**A, B, and D−G**) Medians ± interquartile ranges are shown. (**C and G**) Thicker borders represent donors who had at least one HIV stimulation, or the CMV stimulation, reaching >2x over unstimulated conditions. (**A and G**) Pairwise multiple comparisons were performed using the Friedman test with the Dunn's post-hoc test. (**C and H**) Wilcoxon tests are shown. (**A−C, G, and H**) Circles (HIV) and squares (CMV) represent biologically independent samples. (**A, C, and F**) n = 16 ART participants. (**B and G**) n = 15 CMV$^+$ ART participants.
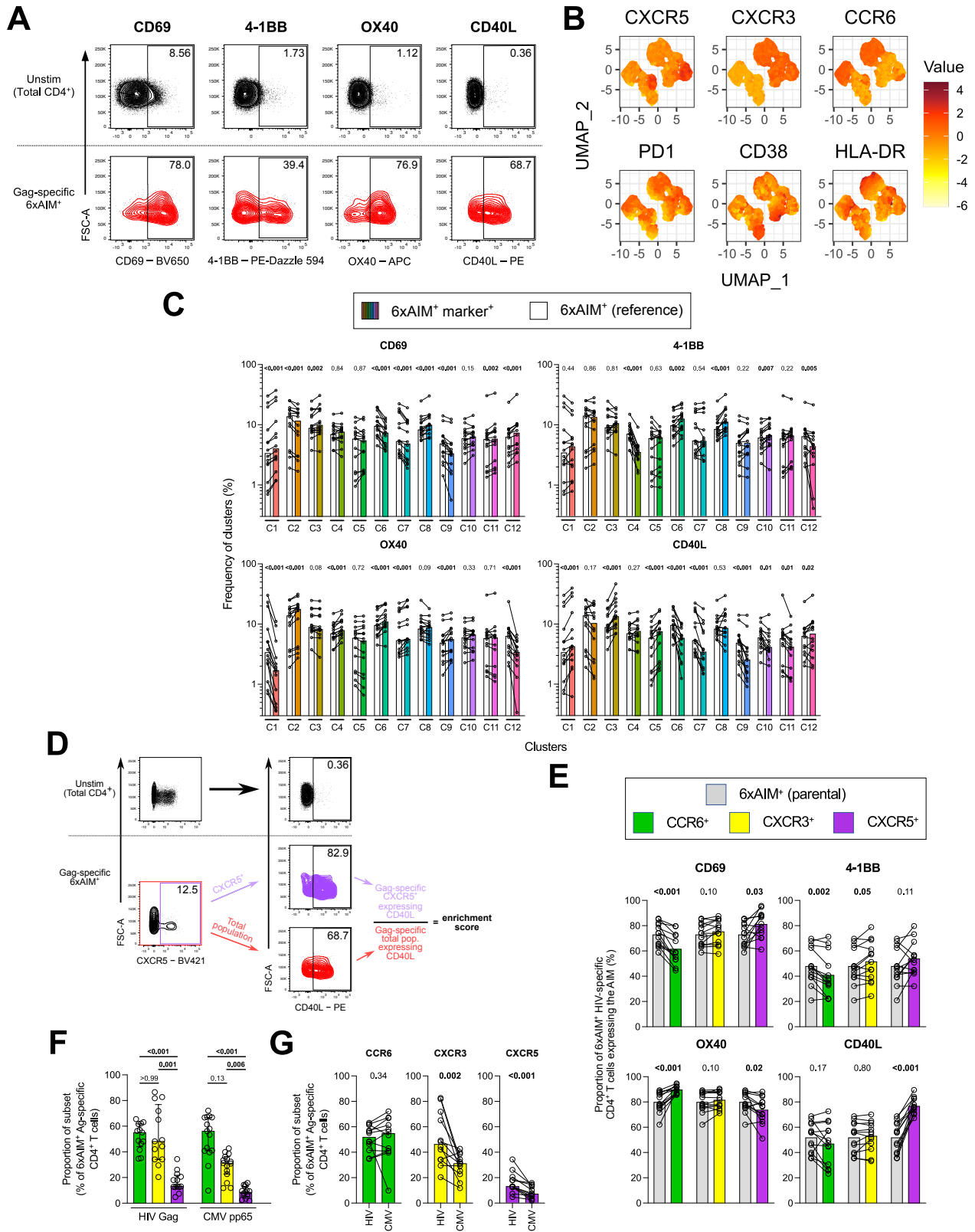
Figure 4.2.S3 – CD69, 4-1BB, OX40, and CD40L are differently expressed on HIV-specific CD4[+] T cell subsets. Related to Figure 4.2.3. (**A**) Representative FACS plots depicting gating on the CD69[+], 4-

1BB$^+$, OX40$^+$, and CD40L$^+$ populations. (**B**) Heatmap overlaid on the 6xAIM$^+$ HIV-specific CD4$^+$ T cells UMAP, showing the gradient of expression for each marker. (**C**) Differential relative frequency of each identified cluster within 6xAIM$^+$ HIV-specific CD4$^+$ marker$^+$ T cell subpopulations (colored bars) compared to within the reference parental 6xAIM$^+$ cells (white bars). (**D**) Representative strategy to calculate the enrichment scores. The red population represents the parental Gag-specific CD4$^+$ T cells, irrespectively of the polarization, expressing CD40L, while the purple one represents its subpopulation of Gag-specific CD4$^+$ CXCR5$^+$ T cells expressing CD40L. The unstimulated condition was used to gate on the AIMs (see also **Figure 4.2.S3A**). For simplicity, the example focuses on CXCR5 and CD40L. (**E**) Proportion of 6xAIM$^+$ Gag-specific CD4$^+$ T cells, irrespectively of the polarization, expressing the AIM (gray) in comparison to Gag-specific CD4$^+$ CCR6$^+$ (green), CXCR3$^+$ (yellow), and CXCR5$^+$ (purple) T cells expressing the AIM. n = 13 ART participants. (**F**) Proportion of CCR6$^+$, CXCR3$^+$, and CXCR5$^+$ T cells in 6xAIM$^+$ Gag- and CMV-specific CD4$^+$ T cells, and its (**G**) intrahost paired comparisons. (**F**) Pairwise multiple comparisons were performed using the Friedman test with the Dunn's post-hoc test. Medians ± interquartile ranges are shown. n = 13 ART participants and n = 15 CMV$^+$ ART participants. (**G**) Matched donors who had the HIV Gag stimulation and the CMV stimulation reaching >2x over unstimulated condition were kept. n = 12 CMV$^+$ ART matched participants. (**C, E, and G**) Medians and Wilcoxon tests are shown. * = p <0.05; ** = p <0.01; *** = p <0.001. To avoid contaminating phenotype profiling with excessive background, only donors who had at least one HIV stimulation (**B and C**), the HIV Gag stimulation (**E−G**), or the CMV stimulation (**F and G**) reaching >2x over unstimulated condition were kept. (**B and C**) n = 14 ART participants.
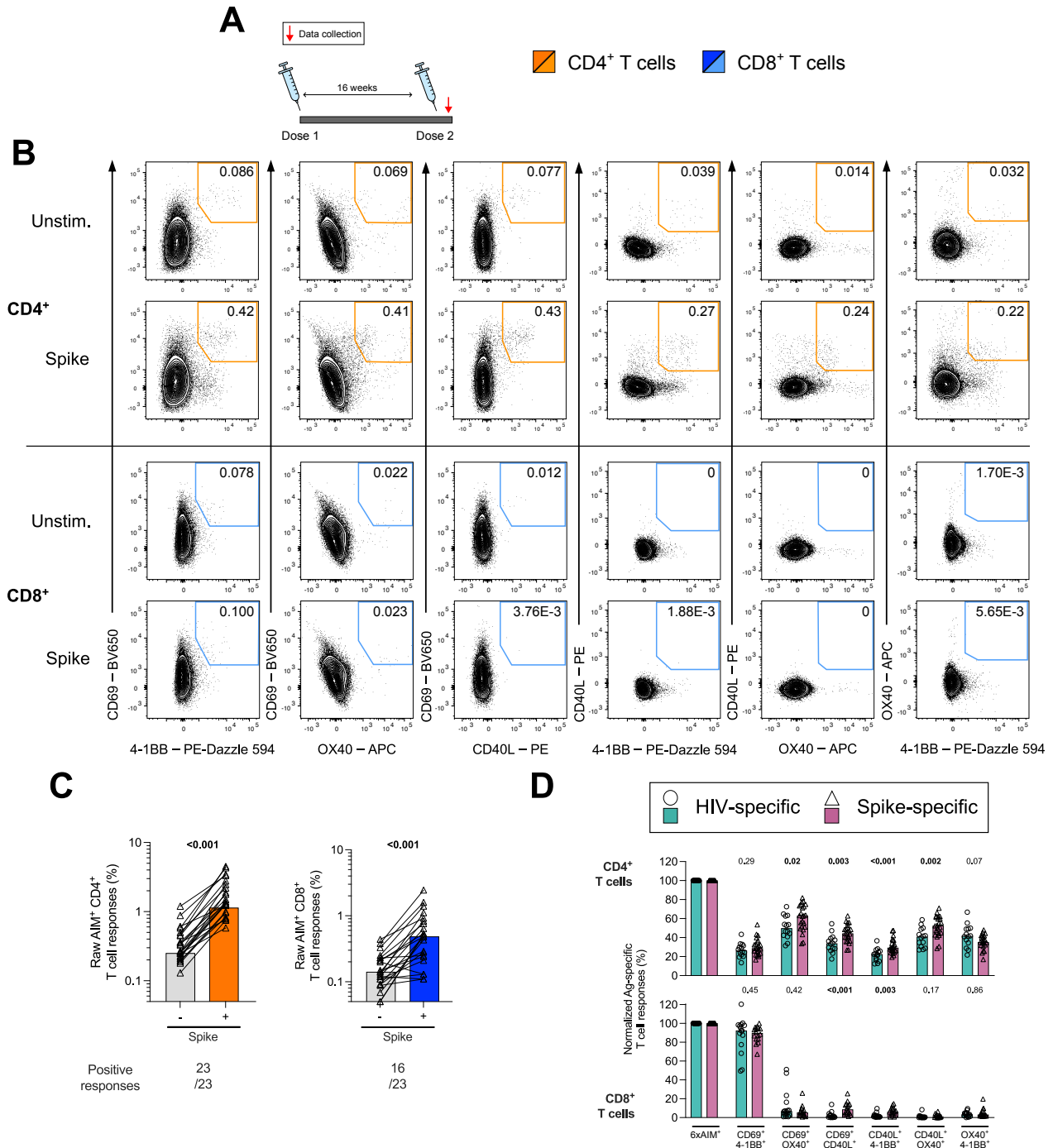
Figure 4.2.S4 – Infection-specific (HIV) and vaccine-induced (SARS-CoV-2) responses are detected by the same AIM pairs. Related to Figure 4.2.4. (**A**) Schematic representation of the study design. Blood samples were collected approximatively three weeks after dose 2 of SARS-CoV-2 mRNA vaccines. Dose administrations are indicated by a syringe. (**B**) Representative FACS plots depicting the multiplexed (6xAIM) strategy to identify SARS-CoV-2 Spike-specific CD4$^+$ (orange gate) and CD8$^+$ (blue gate) T cells. (**C**) Raw frequencies of AIM$^+$ CD4$^+$ and CD8$^+$ T cell responses following an *ex vivo* 15h stimulation of

PBMCs with a pool of SARS-CoV-2 Spike Glycoprotein peptides. PBMCs were left unstimulated as a control (gray bars). Number of responders reaching >2x over unstimulated condition are written below the histograms for each stimulation. The 6xAIM values and Wilcoxon tests are shown. n = 23 SARS-CoV-2 vaccinated participants. (**D**) Comparisons between the normalized amplitudes of net HIV- (teal) and Spike-specific (lavender) CD4$^+$ and CD8$^+$ T cell responses. Within each cohort, responses were normalized to their 6xAIM pairs values. Mann-Whitney tests are shown. Donors who had at least one HIV stimulation reaching >2x over the unstimulated condition were analyzed, as well as those whose Spike stimulation reached the same threshold. For CD4$^+$ T cells, n = 14 ART participants and n = 23 SARS-CoV-2 vaccinated participants; for CD8$^+$ T cells, n = 15 ART participants and n = 16 SARS-CoV-2 vaccinated participants. (**C and D**) Circles (HIV) and triangles (Spike) represent biologically independent samples. The bars represent median values.

4.2.8 Supplemental Tables

Table 4.2.S1 – Clinical characteristics of the study participants.[a] Related to the STAR Methods section and Figures 4.2.1 to 4.2.4.

	ART-treated n = 16	SARS-CoV-2 vaccinated n = 23
Age		
Age (years old)	55.5 (51.8 – 58)	51 (41.5 – 55)
Sex		
Male	16 (100%)	9 (39%)
Female	0 (0%)	14 (61%)
HIV infection		
Duration of infection (years)	19.9 (15.6 – 24.3)	N/A
Time before ART (years)	4.1 (2.0 – 11.2)	N/A
Time on ART (years)	10.4 (6.8 – 19.5)	N/A
CD4 counts	588 (455 – 726.5)	N/A
CD8 counts	991.5 (570.3 – 1283)	N/A
CD4/CD8 ratio	0.49 (0.35 – 1.13)	N/A
CMV infection		
Positive	15 (94%)	N/A
Negative	1 (6%)	N/A
SARS-CoV-2 vaccination		
Days between doses 1 and 2 of mRNA vaccines	N/A	111 (109 – 112)
Days after first dose	N/A	132 (130 – 139)
Days after second dose	N/A	21 (20 – 27)

[a] Values displayed are medians. Interquartile ranges (continuous variables) or percentages (categorical variables) are shown in parentheses.

Table 4.2.S2 – Flow cytometry antibody staining panel for the activation-induced marker (AIM) assays. Related to the STAR Methods section and Figures 4.2.1 to 4.2.4.

Marker – Fluorophore	Clone	Source	Catalog #
CD183 (CXCR3) – BV605	G025H7	Biolegend	353728
CD185 (CXCR5) – BV421	J25D4	Biolegend	356920
CD186 (CXCR6) – BUV805	13B 1^{E}5	BD Bosciences	748448
CD196 (CCR6) – BUV737	11A9	BD Bosciences	564377
CD3 – BUV496	UCHT1	BD Bosciences	612941
CD4 – BB630	SK3	BD Bosciences	624294 CUSTOM
CD8 – BV570	RPA-T8	Biolegend	301037
CD14 – BV480	M5E2	BD Bosciences	746304
CD19 – BV480	HIB19	BD Bosciences	746457
CD38 – BB790	HIT2	BD Bosciences	624296 CUSTOM
CD45RA – PerCP Cy5.5	HI100	BD Bosciences	563429
CD69 – BV650	FN50	Biolegend	310934
CD134 (OX40) – APC	ACT35	BD Bosciences	563473
CD137 (4-1BB) – PE-Dazzle 594	4B4-1	Biolegend	309826
CD154 (CD40L) – PE	TRAP1	BD Bosciences	555700
CD279 (PD1) – BV711	EH122H	Biolegend	329928
HLA-DR – FITC	LN3	Biolegend	327005
LIVE/DEAD Fixable dead cell	N/A	Thermo Fisher Scientific	L34960

This project aims to identify the transcriptional and epigenetic mechanisms leading the dysfunction of HIV-specific CD4$^+$ T cells. As previously described in **sections 2** and **4.1**, our first objective was to optimize the strategy of HIV-specific CD4$^+$ and CD8$^+$ T cell identification. We did so by developing the multiplexed AIM assay, presented in **section 4.2** and published in *Cell Reports Methods*. Building on this new approach, we were then able to elaborate our sorting strategy. We achieved our second objective by sorting HIV-specific CD4$^+$ and CD8$^+$ T cells for upcoming analyses of their transcriptomes and epigenomes. The experimental design describing the assays performed and the cohorts studied, the sorting strategy, and the results of the sorted Ag-specific CD4$^+$ and CD8$^+$ T cells will be detailed in **sections 4.3** (Materials and Methods) and **4.4** (Results).

4.3 Materials and Methods

4.3.1 Experimental design

Our group has shown that the transcriptomic profiles of HIV-specific CD4$^+$ T cells from elite controllers (ECs), chronic progressors (CP), and ART-suppressed participants differed, with a skewing toward Tfh for viremic individuals, at the expense of a Th1, Th17, and Th22-like profile compared to ECs (146–148). This was only partially corrected with ART. Similarly, we and others have shown that some epigenetic signatures were established early at the beginning of the infection, both in humans and in murine models (154,203), and that these signatures were not totally reversed with ART (151,152). Therefore, we focused on the Ag-specific T cell responses in two different cohorts of 12 people with HIV (PWH) each (**Figure 4.3.1A**). Participants in the first cohort of ECs were untreated and spontaneously controlled the viremia, with a viral load (VL) of <50 plasma viral RNA (vRNA) copies/mL. The second cohort was longitudinal and described former CP participants with time points before ("pre-ART"; viremic with a VL >2,000 vRNA/mL) and after ("post-ART"; aviremic with a VL of <40 copies/mL) suppressive ART. For the post-ART time points, we selected those who had been on ART for >1 year and that were virally suppressed for >6 months. To observe the direct effects of ART initiation on the transcriptomic, methylation, and chromatin accessibility profiles, we also chose the closest pre- and post-ART time points to each other. **Table 4.3.S1** (in **Appendix 7**) summarizes the characteristics of the participants.

We will investigate the genome-wide transcriptome and epigenome of HIV-specific CD4$^+$ and CD8$^+$ T cells in depth, using DNA methylation, chromatin accessibility, and transcriptomics assays. DNA methylation assays will first be performed using the Reduced representation bisulfite sequencing (RRBS) technology, allowing identification of regions enriched in CpG sites, indicative of a methylated locus. Methylation is associated with repressed transcription. To investigate epigenetic mechanisms that are or are not corrected with ART, differentially methylated loci between ECs and pre-ART participants will be identified and compared to those of post-ART participants. Tagmentation assays will then be performed with the Assay for transposase-accessible chromatin using sequencing (ATAC-Seq) to assess chromatin accessibility. "Open" chromatin (euchromatin) is more accessible and permissive to transcriptional factors than "closed" chromatin (heterochromatin). This type of analysis will allow the mapping of the transcriptional activity of promotors and/or regulatory regions associated with epigenetic modifications. Similarly, differentially accessible regions will be compared between the groups. Finally, as accessible and/or unmethylated genes are not necessarily expressed, DNA methylation and accessible chromatin profiles will be paired with analyses of the transcriptome. Transcriptomics will be performed using RNA Sequencing (RNA-Seq), allowing the comparison of differentially expressed transcripts between the three groups. All these techniques have been optimized and established by our collaborators, the laboratory of Dr. Jeremy M. Boss (Emory University, Atlanta, USA) (151,153,203–205), who will be performing the sequencing and the analyses.

As our previous results showed that the skewing in lineage differentiation could not be observed in SEB- and CMV pp65-stimulated cells (147), we also evaluated CMV-specific T cell responses to test whether the dysfunction is specifically associated with HIV. CMV is another chronic virus highly prevalent in PWH (206,207), that has been associated with a more terminally differentiated and senescent phenotype (208–210). For each participant, we targeted to sort, in this order, 2,000 HIV- and CMV-specific CD4$^+$ and CD8$^+$ T cells for the RNA-Seq, 2,000 cells for the RRBS, and 10,000 cells for the ATAC-Seq (**Figure 4.3.1B**).
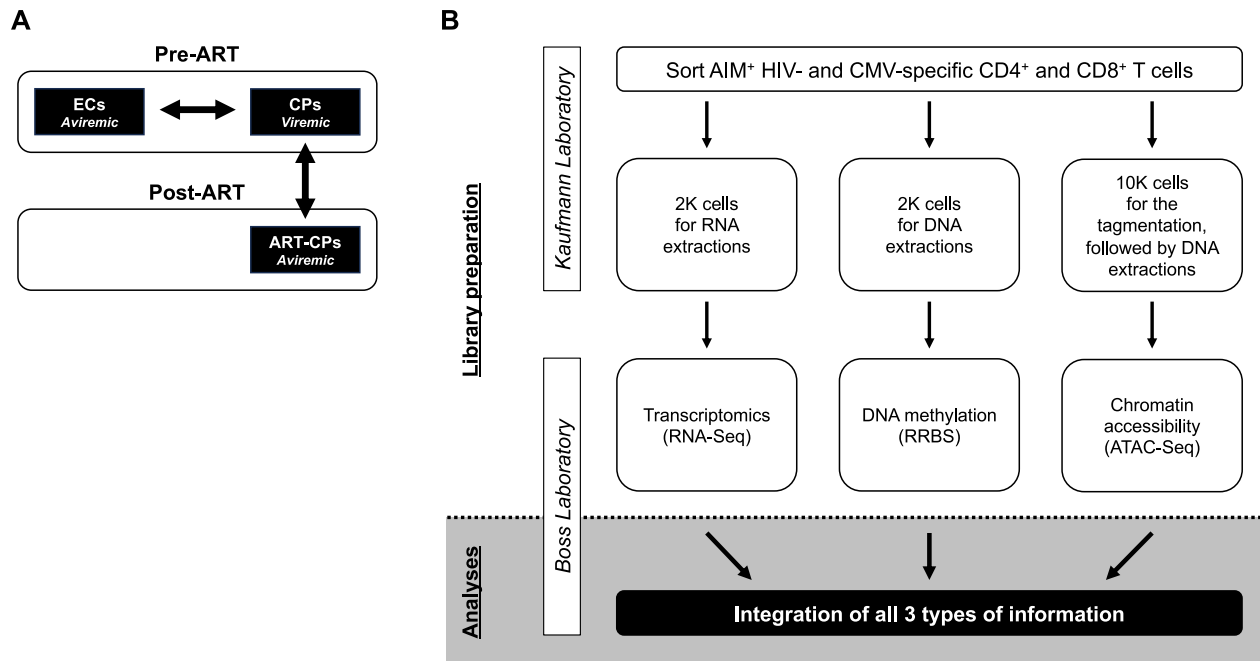
Figure 4.3.1 – Project overview. (A) Description of the cohorts that will be investigated. Arrows represent the main comparisons that will be made. **(B)** Experimental design. The sorting of AIM$^+$ cells and the RNA or DNA extractions will be performed in the Kaufmann laboratory, while the sequencing and analyses will be performed in the Boss laboratory, at Emory University. EC: Elite controllers; CP: Chronic progressors; ART-CP: ART-treated chronic progressors; RNA-Seq: RNA sequencing; RRBS: Reduced representation bisulfite sequencing; ATAC-Seq: Assay for transposase-accessible chromatin using sequencing.

4.3.2 Materials and Methods

Experimental model and subject details

Ethics statement

All work was conducted in accordance with the Declaration of Helsinki in terms of informed consent and approval by an appropriate institutional board. Blood samples were obtained from donors who consented to participate in this research project at the CHUM (13.019 and 19.381). Plasma and PBMCs were isolated by centrifugation and Ficoll gradient, and samples were stored at -80°C and in liquid nitrogen, respectively, until use.

Participants and samples

We obtained the leukaphereses from study participants at the McGill University Health Centre (Montréal, QC, Canada), and at the Centre Hospitalier de l'Université de Montréal (CHUM; Montréal, QC, Canada). The study was approved by the respective IRBs, and written informed consent was obtained from all participants before enrolment. Our cohorts of 12 untreated elite controllers (VL <50 copies/mL) and 12 longitudinal pre- (VL >2,000 copies/mL) and post-ART (VL <40 copies/mL) participants were used to measure HIV- and CMV-specific responses. The characteristics of all participants are summarized in **Table 4.3.S1**. PBMCs were isolated by the Ficoll density gradient method and stored in liquid nitrogen until use.

Method details

Multiplexed AIM assay

The multiplexed AIM assay (178) was adapted and used to quantify HIV- and CMV-specific CD4$^+$ and CD8$^+$ T cell responses. As previously described (42,169–171,178), PBMCs were thawed and rested for 3 h in 96-well flat-bottom plates in RPMI 1640 supplemented with HEPES, penicillin/streptomycin, and 10% Human Serum, Type AB (HAB). 2×10^6 PBMCs were stimulated with pools of 150 overlapping peptides of HIV-1 Gag, Pol, Env, or Nef proteins (0.5 µg/ml per peptide) (JPT), or 138 overlapping peptides of HCMVA 65 kDa phosphoprotein (pp65) (0.5 µg/ml per peptide) (JPT) for 15 h (37°C, 5% CO2). A DMSO-treated condition and a *Staphylococcus enterotoxin B* (SEB)-treated (0.5 µg/ml) condition served, respectively, as negative and positive controls. Cells were first stained for viability dye (Aquavivid, Thermofisher, 20 min, 4°C), surface markers (30 min, 4°C) (see **Table 4.3.S2** for the list of antibodies), and then fixed using 1% paraformaldehyde (15 min, 4°C) before acquisition on the BD LSRFortessa^{TM} (BD Biosciences). Analyses were performed using FlowJo (v10.10.0).

Sorting of live cells

For the live sorting of HIV- and CMV-specific CD4$^+$ and CD8$^+$ T cells, 7×10^6 PBMCs were rested and stimulated with HIV-1 Gag and HCMVA pp65 peptides (JPT) in 48-well flat-bottom plates. Wells were initially pooled, then split into individual FACS tubes and stained separately. Cells were stained as described above, except that they were not fixed with paraformaldehyde. After

staining, cells were immediately sorted (4°C, 1.5 mL microtubes) on the flow cytometer FACSAria™ Fusion (BD Biosciences). For downstream transcriptomics analyses, 2,000 AIM$^+$ CD4$^+$ and CD8$^+$ T cells were directly sorted in RLT lysis buffer (Qiagen) + 1% β-Mercaptoethanol (Thermofisher), flash freezed, and stored at -80°C until harvesting the RNA. For the DNA methylation (n = 2,000 cells) and chromatin accessibility (n = 10,000 cells) assays, cells were sorted in RPMI 1640 supplemented with HEPES, penicillin/streptomycin, and 10% HAB. The 2,000 cells were centrifuged in swinging buckets at 300 g (10 min, 4°C), the supernatant was removed, leaving approximatively 50 μL of medium, and stored at -80°C until harvesting the DNA. After sorting the 10,000 cells, we immediately proceeded with the tagmentation reaction.

Tagmentation reaction

As previously described (205), we first centrifuged the 10,000 cells in swinging buckets at 500 g for 10 mins (4°C), then removed the supernatant. We resuspended the dry pellet in 25 μL of Tagmentation Reaction mix (2X Tagmentation DNA buffer, Tn5 Tagmentation DNA enzyme, 0.02% Digitonin, 0.1% Tween-20, and molecular grade water), and incubated it at 37°C for 60 mins. We stopped the tagmentation reaction with a mix of 25 μL Tagmentation Clean-up buffer (23 μL of 2X Clean-up buffer [326 mM NaCl, 109 mM EDTA, and 0.63% SDS] + 2 μL of Proteinase-K [10 mg/mL]), which we incubated at 40 °C for 30 mins. The tagmentation reaction was stored at -20°C until harvesting the DNA.

RNA and DNA extractions

RNA was extracted on column using the *Quick*-RNA Microprep kit (Zymo Research), eluted in 15 μL of DNAse/RNAse-free water, and stored at -80°C until sequencing. For methylation and chromatin accessibility analyses, DNA was extracted on column using the *Quick*-DNA Microprep kit (Zymo Research), eluted in 55 μL (DNA methylation) or 50 μL (chromatin accessibility) of DNA elution buffer, and stored at -20°C until sequencing.

Tagmentation internal controls

At the beginning of each sort, 10,000 cells were sorted twice from the "Lymphocytes" gate in two separate microtubes. These were kept on ice until proceeding with the tagmentation reaction. After spinning the cells at 500 g for 10 mins, we removed the supernatant of the tagmentation control #1, resuspended the cells in medium, then counted the cells. We proceeded with the tagmentation reaction, as previously described, for the tagmentation control #2. After extracting the DNA, we

ran a PCR using primers specific for the human CD3 gene (HCD3OUT5′: ACTGACATGGAACAGGGGAAG; HCD3OUT3′: CCTCTCTTCAGCCATTTAAGTA) (211), the AccuStart II PCR ToughMix enzyme (Quantabio), and the following PCR cycle conditions: a 3 mins denaturation step at 95°C, 40 cycles of amplification (95°C for 30 secs, 56.5°C for 30 secs, and 70°C for 1 min), and a final elongation of 7 mins at 70°C.

Statistical analysis

Antigen-specific T cell responses are expressed in percentages of CD4$^+$ and CD8$^+$ T cells. Symbols (circles and triangles for CD4$^+$ T cells detected with the Boolean OR gating or the L gating strategies, and squares for CD8$^+$ T cells detected with the CD69$^+$4-1BB$^+$ AIM pair) represent biologically independent samples. Lines connect data from the same participant. Proportions of background were calculated by dividing the background signal (unstimulated condition, used as a negative control) by the signal of the stimulated condition. The coefficients of variation (CV) were calculated by dividing the standard deviation by the mean. Differences in responses for the same participant were performed using Wilcoxon matched pair tests. Pairwise multiple comparisons within cohorts were performed using the Friedman test with the Dunn's post hoc test, while it was the Kruskal-Wallis test (with the Dunn's post hoc test) when comparing between elite controllers, pre-ART, and post-ART participants. The Wilcoxon, Friedman, and Kruskal-Wallis tests were generated using GraphPad Prism (v10.1.1) (GraphPad, San Diego, CA). *P* values are indicated for each comparison assessed, and *p* <0.05 were considered significant. For graphical representation on a log scale (but not for statistical tests), null values were arbitrarily set at the minimal values for each assay. Graphics were generated using GraphPad PRISM version 10.1.1.

4.4 Results

Stronger Gag-specific CD4$^+$ and CD8$^+$ T cell responses were found compared to the other HIV antigens

To conduct the genome-wide transcriptomics, DNA methylation, and chromatin accessibility assays described in **section 4.3.1**, we adapted the multiplexed AIM assay (presented in **section 4.2**) for the live sorting of cells. We first discarded CD40L, as our data showed that it was the most redundant with the other markers to detect Ag-specific CD4$^+$ T cells. Thereby, its removal only had a limited impact on the detection of Tfh cells, despite the enrichment of CD40L in this subset (178). It also contributed little to capturing Ag-specific CD8$^+$ T cells. Furthermore, removing CD40L eliminated the need of a CD40 blocking antibody, which may alter cellular interactions, and optimized the stimulation time considering that CD40L is expressed early, while OX40 is upregulated later (155,165). We directly detected Ag-specific CD8$^+$ T cells with the CD69$^+$4-1BB$^+$ pair, as we showed that multiplexing the AIM assay is beneficial for CD4$^+$ T cells only (178) **(Figures 4.4.S1A and 4.4.S1B, in Appendix 8)**.

We first screened all participants to measure AIM$^+$ T cell responses following a 15 h stimulation with pools of overlapping HIV-1 (Gag, Pol, Env, and Nef) and CMV (pp65) peptides **(Figures 4.4.S1C and 4.4.S1D)**, and to estimate the number of cells required for sorting. We subtracted the background from the stimulated responses to obtain net responses **(Figures 4.4.1A and 4.4.1B)**. Consistently with data published by our group (42,178), strong AIM$^+$ CD4$^+$ and CD8$^+$ T cell responses were found in the Gag-stimulated conditions across all three groups of participants. Other antigens, notably Env, yielded significantly lower responses. Additionally, as previously reported (145,212,213), we observed a significant upregulation of PD-1, an immune checkpoint inhibitor and marker of immune activation and exhaustion, in Gag-specific CD4$^+$ T cells compared to the total CD4$^+$ T cell population **(Figures 4.4.1C and 4.4.1D)**. Gag-specific CD4$^+$ T cells of pre- and post-ART donors expressed significantly more PD-1 than elite controllers'. To assess HIV-specific CD4$^+$ and CD8$^+$ T cell responses, we thus decided to focus on T cells specific for the structural protein Gag for downstream live sorting.
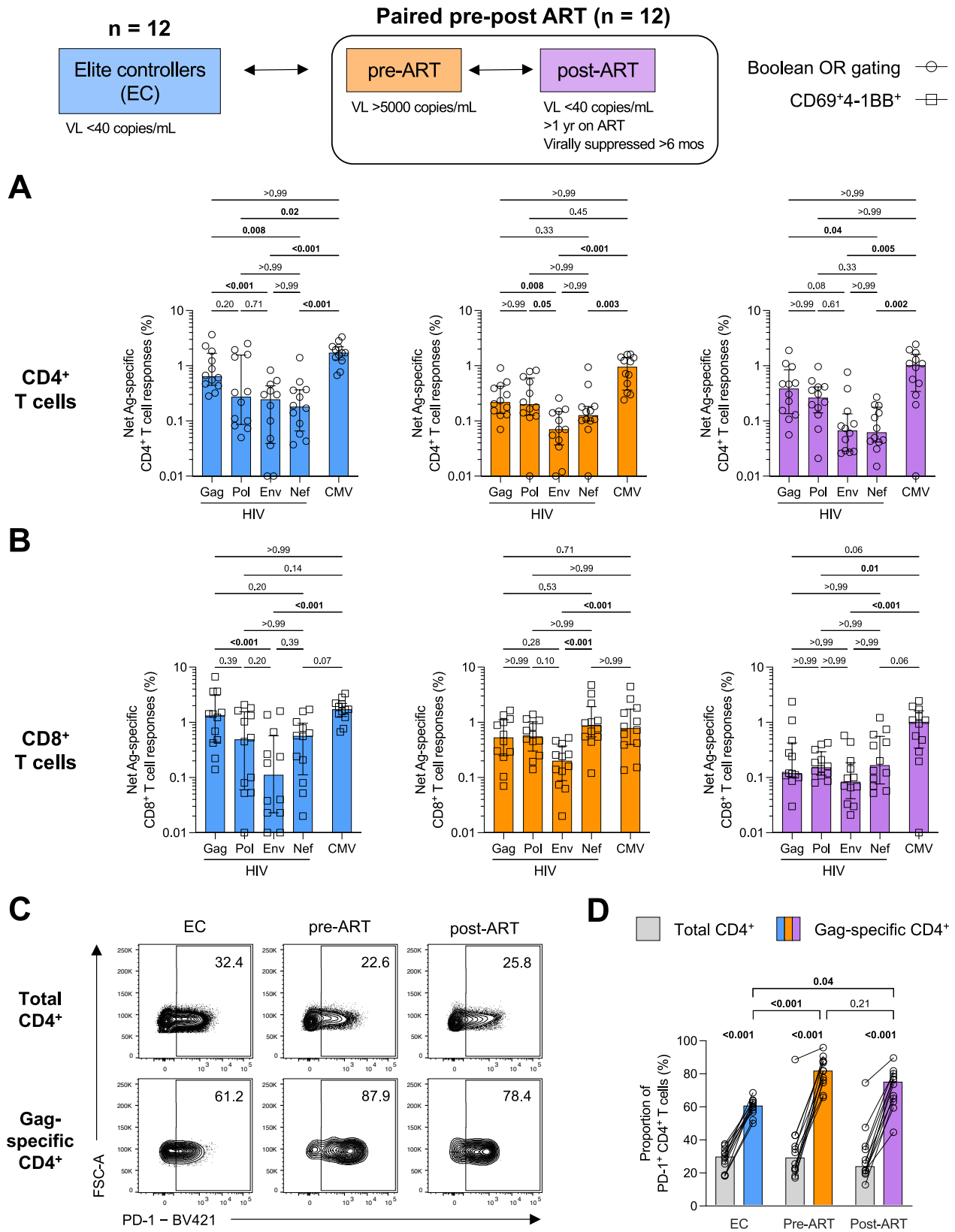
Figure 4.4.1 – Characterization of Ag-specific CD4$^+$ and CD8$^+$ T cells in ECs and pre- and post-ART participants. (**A and B**) Net Ag-specific (**A**) CD4$^+$ and (**B**) CD8$^+$ T cell responses in the elite controllers (EC; blue) and pre- (orange) and post-ART (purple) cohorts. (**C**) Example of univariate PD-1 gating on total and HIV Gag-specific CD4$^+$ T cell populations. PD-1 was gated based on the FMO. (**D**) Proportion of PD-1$^+$ cells in total (grey bars) and Gag-specific CD4$^+$ T cells. Wilcoxon tests between the total and Gag-specific populations are shown. Pairwise multiple comparisons of PD-1 expression between cohorts were performed using the Kruskal-Wallis test with Dunn's post hoc test. (**A and B**), pairwise multiple comparisons were performed using the Friedman test with Dunn's post hoc test. (**A, B, and D**), circles represent CD4$^+$ responses detected with the Boolean OR gating strategy, while squares represent CD8$^+$ T cells detected with the CD69$^+$4-1BB$^+$ pair. n = 12 ECs and n = 12 pre- and post-ART participants.

The multiplexed AIM assay was adapted to an "L gating" strategy for the live sorting of AIM$^+$ cells

Our sorter was not able to directly sort Ag-specific CD4$^+$ T cells with the Boolean AND/OR gating without resulting to software and hardware malfunctions. We thus adapted the multiplexed AIM approach to one that would be compatible for sorting, depicted in **Figures 4.4.2A and 4.4.S2A** as an "L gating" strategy. Briefly, after gating on CD4$^+$ T cells, we first selected cells based on the expression of CD69, then subsequently sorted the cells that were OX40$^+$ (CD69$^+$OX40$^+$), 4-1BB$^+$ (CD69$^+$4-1BB$^+$), or OX40$^+$4-1BB$^+$ (CD69$^+$OX40$^+$4-1BB$^+$). Simultaneously, Ag-specific CD8$^+$ T cells were directly sorted with CD69$^+$4-1BB$^+$ **(Figure 4.4.2A)**. AIM$^+$ CD4$^+$ T cells were sorted with the L gating strategy, acquired, and re-analyzed with standard AIM pairs to estimate the proportion of additional AIM$^+$ cells gained with the multiplexing. Approximatively 22.2% and 30.6% more cells were detected with the L gating compared to CD69$^+$4-1BB$^+$ and CD69$^+$OX40$^+$, respectively (**Figure 4.4.S2B**). We also showed that both strategies, the Boolean OR Gating and the L gating, were equivalent (**Figure 4.4.S2C**), and that none of the *p*-values reached statistical significance.
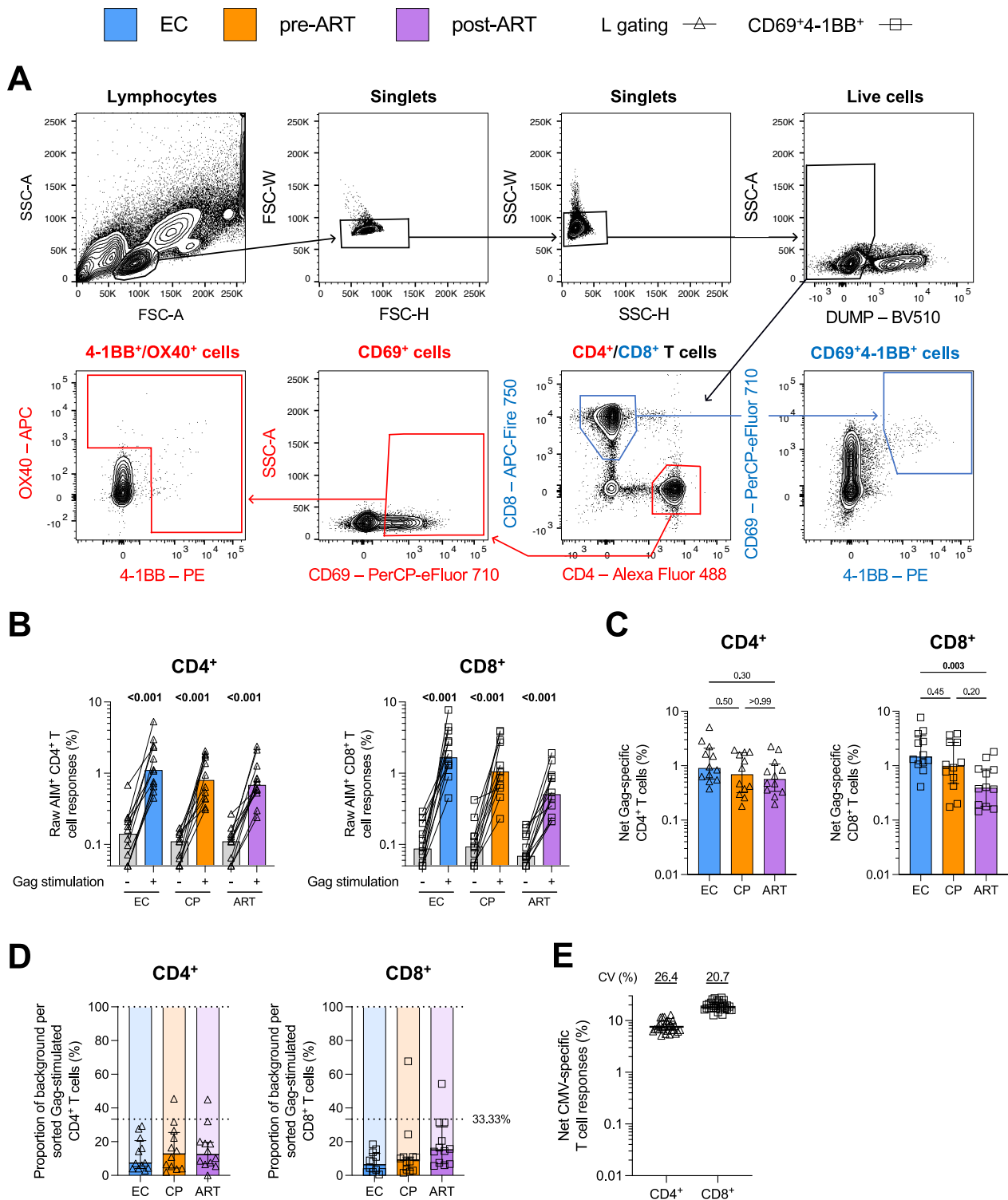
Figure 4.4.2 – Live sorting of Ag-specific CD4+ and CD8+ T cells in ECs and pre- and post-ART participants. (**A**) Live sorting strategy. The multiplexed AIM assay was adapted for the sorts of CD4+ T cells using an "L gating" strategy (red gates), while CD8+ T cells were simultaneously sorted using the CD69+4-1BB+ AIM pair (blue gates). Black gates depict the upstream gating strategy. (**B**) Raw and (**C**) net

frequencies of sorted AIM$^+$ CD4$^+$ and CD8$^+$ T cell responses in elite controllers (blue), pre-ART (orange), and post-ART (purple) participants, following an *ex vivo* 15 h stimulation of PBMCs with a pool of HIV Gag peptides. PBMCs were left unstimulated as a control (grey bars). (**B**) Wilcoxon tests and (**C**) Kruskal-Wallis tests with Dunn's correction are shown. (**D**) Proportion of background in the HIV-stimulated condition. The Ag-stimulated conditions are normalized to 100% to better represent the autologous unstimulated signal. Light colors represent the total of stimulated cells; dark colors represent the background. (**E**) Net CMV-specific CD4$^+$ and CD8$^+$ T cell responses of a single participant repeated at every sort. Coefficients of variation (CV) are shown above the graph. n = 30 sorts. (**B** − **E**), triangles represent CD4$^+$ responses detected with the L gating strategy, while squares represent CD8$^+$ T cells detected with the CD69$^+$4-1BB$^+$ pair. n = 12 ECs and n = 12 pre- and post-ART participants.

Gag-specific, now on referred to as "HIV-specific," CD4$^+$ and CD8$^+$ T cells were sorted from elite controllers and longitudinal pre- and post-ART participants (**Figure 4.4.2B**). ECs had slightly larger magnitudes of sorted HIV-specific CD4$^+$ T cells than pre- and post-ART participants, with medians of 0.90%, 0.69%, and 0.57%, respectively (**Figure 4.4.2C**). A hierarchy was more noticeable in CD8$^+$ T cells, as large and significant differences with the unpaired Kruskal-Wallis test with Dunn's correction were found between ECs (median: 1.45%) and ART-treated (median: 0.39%) participants. In contrast, longitudinal participants had higher HIV-specific CD8$^+$ T cell responses before ART (median: 0.98%) than after, although these responses were more similar in CD4$^+$ T cells, and neither differences in magnitude of CD4$^+$ and CD8$^+$ T cell responses reached statistical significance.

To make sure that the positive signal was specific, we calculated the ratio of the signal in the unstimulated vs. Ag-stimulated conditions. The median proportion of background was 12.8% for CD4$^+$ T cells and 15.1% for CD8$^+$ T cells (**Figure 4.4.2D**). Only cells whose proportion of background was below the 33.33% threshold were kept for RNA and DNA extractions. In parallel with Gag-stimulated cells, we sorted AIM$^+$ CD4$^+$ and CD8$^+$ T cells specific to CMV (**Figures 4.4.S2D, 4.4.S2E, and 4.4.S2F**). In most participants, 14,000 Gag and CMV-specific CD4$^+$ and CD8$^+$ T cells were sorted (**Figures 4.4.3 and 4.4.S3**). We also sorted 2,000 unstimulated AIM$^+$ CD4$^+$ and CD8$^+$ T cells from one individual to access the transcriptome of "background" cells. **Table 4.4.S1** presents the number of cells sorted for each participant.

To correct for potential batch effects, samples from an additional participant with high CMV responses were repeated at every sort. These data showed that the AIM assay is highly reproducible, with low variability for the detection of both Ag-specific CD4$^+$ (26.4%) and CD8$^+$ T cells (20.7%) **(Figures 4.4.2E and 4.4.S2G)**.
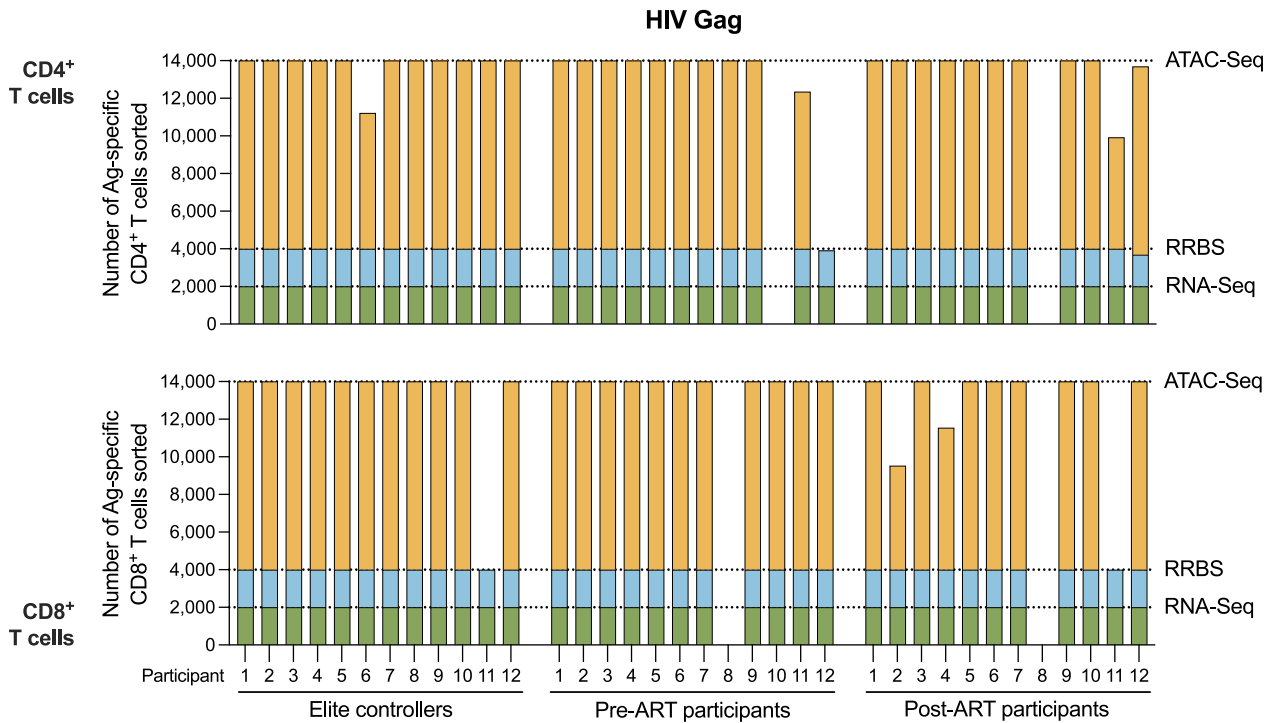
Figure 4.4.3 – Summary of HIV Gag-specific CD4$^+$ and CD8$^+$ T cells sorted in ECs and pre- and post-ART participants. Each colour represents the cells sorted for the different assays: green for the RNA-Seq, blue for the RRBS, and yellow for the ATAC-Seq. See also **Figure 4.4.S3** for sorted CMV pp65-specific CD4$^+$ and CD8$^+$ T cells. n = 12 ECs and n = 12 pre- and post-ART participants. RNA-Seq: RNA sequencing; RRBS: Reduced representation bisulfite sequencing; ATAC-Seq: Assay for transposase-accessible chromatin using sequencing.

Sorting controls yielded high purity values (96.4%) for both CD4$^+$ and CD8$^+$ T cells (**Figure 4.4.4A**). In the optimization phase of our experiments, we had generated results of poor quality for the ATAC-Seq assay. This might have been explained by accidentally aspirating or disrupting the pallet when removing the supernatant, or on the contrary, leaving too much supernatant, reducing the efficiency of the tagmentation reaction. We thus set up two internal controls to ensure that cells

were not lost during the removal of the supernatant (by counting the cells afterwards, tagmentation control #1) and that we were able to recuperate DNA (by running a PCR targeting the human CD3 gene, tagmentation control #2) (**Figure 4.4.4B**). We found a recovery of approximatively 100% in all sorts (**Figure 4.4.4C**), as well as bands corresponding to the expected size of the amplicon (356 bp) (**Figure 4.4.4D**).

As part of an initial pilot study, the HIV- and CMV-specific CD4$^+$ and CD8$^+$ T cells sorted from three ECs, three pairs of pre- and post-ART participants, and the repeated participant (n = 156 samples, 52 samples for each assay) were sent to the laboratory of Dr. Boss. Before sequencing, they first performed quality control experiments. The quality of RNA-Seq data was measured with the mapping statistics and library size patterning using a Bio/Fragment Analyzer, which shows both the size distribution of the reads and the concentration of the final sequencing library. All 52 RNA-Seq samples had a coverage of more than the 15 million reads threshold (**Figure 4.4.S4A**), broad distributions of library sizes ranging from the expected 250 – 1,000 bp (**Figure 4.4.4D**), and passed the quality control. Similarly, the distributions of library sizes for the RRBS had the distinctive "hedge-hogging" trace (**Figure 4.4.4D**). Unfortunately, we have not yet received the mapping statistics of this assay. The ATAC-Seq is a more challenging assay that requires additional metrics to evaluate the quality of the data: along with the mapping statistics and the Bio/Fragment Analyzer, the number of MACS peaks (number of enriched accessible loci identified in each sample), the Fractions of Reads in Peaks (FRIP, proportion of total reads in the enriched regions), and a manual assessment to verify that known regions of housekeeping genes are indeed enriched, were considered. Out of the 52 ATAC-Seq samples, 77% were of optimal quality, having >15 million reads (**Figure 4.4.S4B**), >10,000 MACS peaks (**Figure 4.4.S4C**), a FRIP score >5% (**Figure 4.4.S4D**), enriched regions of the Beta-actin gene (**Figure 4.4.S4E**), and a nucleosomal peak at 300 – 400 bp (**Figure 4.4.4D**). Nonetheless, the remaining 10 samples were of the CMV-stimulated conditions, including five of the repeated participant, and have been deemed by our collaborators as still usable, although they may provide limited data.

This is the initial phase of a much larger project, considering the quantity of information that will be extracted from those cells. Between July 2023 and April 2024, we sorted over 2.5 million cells, divided in 574 samples collected in 37 participants. In the next few months, our collaborators will proceed with the sequencing and the analyses of the remaining 418 samples.

While we have not yet received the results from our collaborators for the pilot study, our internal tagmentation controls and the bioinformatic quality controls show that, altogether, the data produced are of optimal quality and ready for sequencing.

A

AIM⁺ CD4⁺ T cells (L gating)

AIM$^+$ CD4$^+$ T cells (L gating)

Lymphocytes — 73.8
Singlets — 99.1
Singlets — 99.5
DUMP negative — 97.2
CD4$^+$ T cells — 0 / 94.2
CD69$^+$ — 99.2
OX40$^+$/4-1BB$^+$ — 96.4

AIM$^+$ CD8$^+$ T cells (CD69$^+$4-1BB$^+$)

Lymphocytes — 77.7
Singlets — 99.3
Singlets — 99.7
DUMP negative — 93.6
CD8$^+$ T cells — 99.9 / 0
CD69$^+$4-1BB$^+$ — 96.4

B

Sort 10K lymphocytes × 2

Control #1 (10K cells)

Count cells after removing the supernatant

Control #2 (10K cells)

Run a CD3 PCR after proceeding with the tagmentation reaction

C

Internal tagmentation control #1

Recovery (%)

Count

D

Internal tagmentation control #2

L: DNA ladder
C1: water control
C2: no enzyme control
50K: DNA previously extracted from 50K cells
10K: DNA previously extracted from 10K cells

500 bp
350 bp
300 bp
200 bp

L C1 C2
50K 10K Sorts #1-3

E

Expected library size patterning   **Pilot samples**

RNA-Seq
Optimal RNA-Seq trace showing 250 – 1,000 bp broad size distribution.

RRBS
Optimal RRBS trace showing 250 – 1,000 bp broad size distribution with "hedge-hogging."

ATAC-Seq
Optimal ATAC-Seq trace showing 300 – 400 bp nucleosomal peak.

127

Figure 4.4.4 – Quality control of the sorted cells. (A) Purity controls for the sorting of Ag-specific CD4$^+$ and CD8$^+$ T cells. (**B**) Diagram outlining our two internal tagmentation controls. (**C**) Percentage of recovery after removing the supernatant. Cell counts were estimated with a hemocytometer. n = 30 sorts. (**D**) Example of an agarose electrophoresis gel following a PCR targeting the human CD3 gene. L: DNA ladder; C1: water condition (negative control); C2: no enzyme condition (negative control); 50K and 10K: DNA previously extracted from 50K and 10K cells (positive controls); sorts #1-3: PCR performed on these samples following the tagmentation reaction. (**E**) Library size patterning using a Bio/Fragment Analyzer. Graphs on the left represent the "expected patterns" for each of the assays, while graphs on the right represent examples of three samples of the pilot study. These data were provided by the laboratory of Dr. Boss (Emory University, Atlanta, USA). RNA-Seq: RNA sequencing; RRBS: Reduced representation bisulfite sequencing; ATAC-Seq: Assay for transposase-accessible chromatin using sequencing.

Chapitre 5 – Discussion générale et perspectives

Ce mémoire comporte deux méthodes développées pour l'étude virologique et immunologique de l'infection au VIH. D'un côté, dans le laboratoire Chomont, nous avons développé un algorithme de prédiction de l'intégrité des génomes du VIH, automatisant une séquence d'analyse déjà utilisée par notre groupe. L'algorithme, nommé « IntegrityAlgorithm », permet à la fois d'analyser l'intégrité des provirus à l'aide de plusieurs outils communément utilisés dans la littérature, mais également leur clonalité. L'algorithme s'est montré cohérent avec les résultats publiés et a révélé la granularité des résultats, jusqu'alors dissimulée par la hiérarchisation.

De l'autre côté, dans le laboratoire Kaufmann, nous avons développé une version multiplexée de l'essai AIM, qui présente plusieurs avantages comparativement aux versions standards n'utilisant qu'une seule paire de AIMs. Nous avons trouvé que l'essai multiplexé augmentait la fréquence de détection des réponses T CD4$^+$ spécifiques à un antigène et qu'il réduisait les biais dans le phénotypage et dans la comparaison entre études. En revanche, il n'y avait pas d'avantage pour les lymphocytes T CD8$^+$, presque exclusivement détectés avec la paire CD69$^+$4-1BB$^+$. Nous avons appliqué cet essai multiplexé pour trier plus de 2,5 millions de cellules T CD4$^+$ et CD8$^+$ spécifiques au VIH et au CMV, dans le cadre d'analyses imminentes de leur transcriptome et de leur épigénome.

5.1 La virologie de l'infection au VIH : l'IntegrityAlgorithm

Discussion des résultats

L'IntegrityAlgorithm a produit des résultats consistants avec ceux précédemment publiés. Dans presque tous les jeux de données, le même nombre de génomes intacts, inversés, hypermutés et largement délétés a été trouvé. Les différences se trouvaient surtout au niveau des séquences classifiées avec un codon stop et des défauts dans le Ψ. Cela est dû à l'algorithme utilisé pour prédire les codons stop, soit ProSeq-IT dans les articles publiés (42,82,90,101) ou Gene Cutter dans l'IntegrityAlgorithm. La différence clé entre ces deux algorithmes semble être au niveau des alignements produits: alors que ProSeq-IT n'aligne qu'une séquence à la fois sur la séquence de

référence, Gene Cutter utilise un modèle plus complexe, entraîné sur des séquences de VIH-1, de VIH-2 ou de SIV, permettant d'analyser et de comparer plusieurs séquences de manière simultanée. Gene Cutter détecte à la fois les insertions et les délétions, mais peut aussi identifier les mutations compensatoires et ajuster l'alignement en conséquence (https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter-help.html). De manière générale, ProSeq-IT semble avoir détecté plus de codons stop que Gene Cutter, en particulier pour le jeu de données de Sannier et al., *Cell Rep.* (2021), probablement dû à son incapacité d'identifier les mutations compensatoires. Puisque les défauts sont hiérarchisés, la proportion de défauts dans le Ψ, plus bas dans la hiérarchie, a donc été modifiée. Considérant que 100% des séquences ayant des codons stop avaient également des défauts dans le Ψ, la conclusion sur l'intégrité n'a toutefois pas changé.

Au niveau des défauts plus hauts dans la hiérarchie, l'algorithme a identifié six séquences comprenant de grandes délétions de plus que les résultats publiés dans Dufour et al., *Nat Commun.* (2023) (90). Ces séquences avaient toutes un nucléotide de moins que la valeur seuil de 8 800 pb. Cela pourrait être dû aux différences subtiles entre l'alignement que j'ai créé et celui qui avait était fait à l'époque. Néanmoins, ce résultat pourrait être manuellement modifié dans les fichiers Excel produits par l'algorithme. L'algorithme a également trouvé trois séquences intactes plutôt que seulement deux, tel que publié dans Sannier et al., *Cell Rep.* (2021) (82). Selon les résultats publiés et basés sur ProSeq-IT, la séquence intacte additionnelle possédait des codons stop dans les exons 1 et 2 du gène Tat. En comparaison, en se basant sur Gene Cutter, l'algorithme n'a trouvé qu'un codon stop dans l'exon 2, ce qui n'est pas considéré critique pour l'intégrité, et expliquerait la différence obtenue.

L'IntegrityAlgorithm permet d'identifier tous les défauts dans les séquences provirales et de les hiérarchiser. La fréquence de certains défauts plus bas dans la hiérarchie, notamment les codons stop et les défauts dans le Ψ, était souvent sous-estimée. En particulier pour les séquences de Dufour et al., *Nat Commun.* (2023), presque tous les génomes provenant de cellules p24+ présentaient des défauts dans le locus du Ψ, avec une fréquence approximativement 38% plus grande qu'initialement rapportée (90). Identifier et caractériser les séquences ayant plus d'un type de défaut pourrait permettre d'avoir un portrait plus global des différents types de génomes proviraux existants, ainsi que d'informer sur les types de défauts se chevauchant fréquemment (91).

Dans les quatre jeux de données, une observation commune était que parmi les séquences ayant été classifiées comme défectives pour le Ψ, c'est-à-dire qu'elles n'avaient ni d'inversion, ni d'hypermutation, ni de grand délétion interne et ni de codon stop, peu d'entre elles avaient des ORFs défectueux. Sans ces défauts dans le Ψ, la majorité de ces génomes seraient alors « intacts ». En 2013, Ho et al. ont montré que trois virus possédant une délétion de 8 et 16 nucléotides dans la boucle SL2, impliquée dans l'encapsidation de l'ARN viral, ou une mutation au MSD, impliqué dans l'épissage des transcrits, ne pouvaient se répliquer, malgré des ORFs intacts (69). La raison biologique pour un tel enrichissement des génomes avec des défauts dans le Ψ reste à explorer, surtout qu'il a été observé en grandes proportions et à maintes reprises dans les cellules compétentes pour la traduction (70,90,185,188,214). Théoriquement, sans le Ψ ou le MSD, le virus ne devrait pas pouvoir s'encapsider et/ou produire des transcrits viraux épissés. Or, Pollack et al. ont pu détecter de l'ARN épissé de VIH-1 dans un virus défectueux pour le MSD, suggérant la présence d'un MSD alternatif (215). White et al. ont aussi montré que dans les génomes avec un Ψ défectueux, la dimérisation et l'encapsidation était réduite, mais pas absente, et qu'ils donnaient lieu à des virions non-infectieux possédant l'ARN viral et une réduction de la production de protéines d'enveloppe (188). Selon ces données, les génomes ayant uniquement des défauts dans le Ψ auraient donc été classées comme « défectueux » à juste titre, dans le sens de non compétent pour la réplication.

Tel que rapporté précédemment, les différences entre les catégories et l'ordre des critères fait varier la caractérisation des génomes intacts et défectueux (91). En comparaison aux autres outils publiquement disponibles, l'IntegrityAlgorithm a détecté le même nombre de génomes intacts que ProSeq-IT et HIVIntact. HIVSeqinR, au contraire, semble les surestimer. L'algorithme développé utilise une liste exhaustive de critères pour inférer l'intégrité, incluant notamment les inversions et l'analyse des protéines régulatrices et accessoires, ce que les autres outils ne font pas. Pourtant, certaines protéines telles que Rev sont essentielles à l'infectivité du provirus : sans elle, les transcrits non ou peu épissés, qui coderont éventuellement pour les polyprotéines virales, ne pourront pas sortir du noyau pour leur traduction (27). En plus des différences de critères, nous avons observé des différences au niveau même de la méthode de détection des défauts, spécifiquement pour les codons stop et les hypermutations, ce qui complique davantage la comparaison entre études.

L'IntegrityAlgorithm et HIVSeqinR sont les deux seuls algorithmes donnant un verdict ou une raison expliquant la défectivité du provirus. Dans les données testées, les verdicts étaient relativement similaires entre les deux outils, sauf pour les défauts dans la région Ψ. Les valeurs seuil de délétions du Ψ différentes pourraient potentiellement expliquer ces différences. Quant à eux, ProSeq-IT et HIVIntact produisent des données brutes qui nécessitent un effort supplémentaire d'analyse pour résumer les défauts dans chacune des séquences.

En ce qui concerne la clonalité, un avantage particulier de l'IntegrityAlgorithm est qu'il est la seule bibliothèque permettant d'analyser autant l'intégrité que la clonalité des provirus. En comparant les résultats publiés et ceux produits par l'IntegrityAlgorithm, les mêmes proportions de séquences unique et clonales ont été trouvées dans tous les jeux de données, sauf pour celui de Dubé et al., *Cell Host Microbe* (2023) (42). Pourtant, l'IntegrityAlgorithm et l'outil ElimDupes ont tous les deux trouvé un clone chez le participant PWH5 qui n'avait pas été rapporté précédemment, illustrant ainsi que l'algorithme limite les risques d'erreurs provenant de décisions « humaines ».

L'algorithme détecte les séquences clonales, mais également les séquences « potentiellement clonales », c'est-à-dire celles présentant un (ou plus, dépendant de la valeur seuil choisie) nucléotide de différence. En pratique, pour les utilisateurs, cet ajout permet d'identifier les différences qui relèveraient d'erreurs « humaines » et non de mutations ponctuelles, comme ce fut le cas pour ce grand clone dont la première base de l'amplicon manquait (la première base a probablement été accidentellement effacée lorsque les séquences des amorces ont été retirées). L'ajout de cette information a également permis d'identifier où se trouvaient fréquemment les mutations ponctuelles d'un seul nucléotide. Dans le jeu de données de Dufour et al., *Nat Commun.* (2023), elles étaient plutôt concentrées dans le gène Env, déjà connu pour son taux rapide de mutations (216–218).

Perspectives

À ma connaissance, aucun algorithme de prédiction de l'intégrité du génome viral n'existe pour le SIV ou le SHIV (*Simian-Human immunodeficiency virus*). Notre groupe caractérise présentement les provirus intacts et défectueux d'un clone pathogénique (AD8) de SHIV, un virus chimérique SIV possédant les protéines Tat, Rev et Env du VIH (219). Parmi QC Tool, Gene Cutter et ProSeq-IT, seul Gene Cutter pourrait être appliqué aux séquences SIV et SHIV, les autres étant spécifiques

au VIH-1. La seule manière de mesurer leur intégrité est d'utiliser Gene Cutter à deux reprises, une fois en utilisant la référence HXB2 et une fois en utilisant la référence SIV pour identifier les codons start et stop. Gene Cutter peut ensuite être combiné à Hypermut pour les hypermutations, ainsi qu'une vérification manuelle des alignements avec Geneious pour les grandes délétions internes, les défauts dans le Ψ et l'intégrité des ORFs. Il faudrait, entre autres, trouver une manière d'incorporer et d'automatiser l'outil d'Hypermut à une séquence d'analyse bioinformatique, ce qui pourrait éventuellement être avantageux pour l'IntegrityAlgorithm. J'avais tenté de reproduire les résultats de ProSeq-IT avec les séquences VIH pour ensuite les appliquer aux séquences SIV et SHIV, en identifiant avec succès la taille des séquences, les délétions et les insertions, mais sans succès pour les décalages dans le cadre de lecture.

Limites

L'algorithme développé est une alternative rapide et automatisée, réduisant ainsi les erreurs manuelles, de la séquence d'analyse publiée. Une de ses limites est qu'il est dépendant des résultats de quatre autres outils (QC Tool, Gene Cutter, ProSeq-IT et Geneious) qui sont basés en ligne et/ou sur un logiciel installé localement, compliquant ainsi son utilisation. En comparaison, ProSeq-IT seul, HIVIntact et HIVSeqinR peuvent inférer l'intégrité des génomes dans un seul et même algorithme. ProSeq-IT ne nécessite aucune expertise en bioinformatique et est un outil disponible en ligne. De plus, l'IntegrityAlgorithm ne s'applique qu'au sous-type B du VIH-1. L'algorithme est lui-même limité par la compatibilité des autres outils avec les sous-types du VIH. QC Tool n'est applicable que sur les séquences du VIH-1, puisqu'il utilise la référence HXB2. Il peut être appliqué aux autres sous-types, mais les résultats seront moins fiables. Gene Cutter peut fonctionner avec les séquences du VIH-1 (sous-type B seulement) ou du VIH-2/SIV (*Simian immunodeficiency virus*), tandis que ProSeq-IT fonctionne avec les sous-types B et C du VIH-1. Considérant que le sous-type C est la forme de VIH-1 la plus prévalente, bien que la recherche soit concentrée sur le sous-type B (220), il serait pertinent d'étendre l'utilisation de cet algorithme à d'autres sous-types du VIH-1, comme c'est le cas pour ProSeq-IT, HIVIntact et HIVSeqinR. Finalement, comme tout algorithme de prédiction, l'IntegrityAlgorithm ne peut prévoir l'impact de mutations ponctuelles qui changeraient un acide aminé pour un autre sur l'intégrité fonctionnelle du génome.

5.2 L'immunologie de l'infection au VIH : l'essai AIM multiplexé

Discussion des résultats

Dans notre article publié dans *Cell Reports Methods*, nous avons montré que le marqueur CD40L, bien qu'utilisé par plusieurs équipes (146–148,155,156,173,175,221), était le moins essentiel (178). Son absence avait un effet limité sur la détection des cellules T CD4$^+$ spécifiques au VIH, incluant sur ses sous-populations, et presque aucune influence sur les lymphocytes T CD8$^+$. CD40L a une cinétique différente des autres marqueurs utilisés (CD69, 4-1BB et OX40), lui qui est exprimé rapidement après l'activation des cellules (6-9 h) en comparaison à 16-24 h pour les trois autres (155,165). CD40L pouvait aussi biaiser le phénotype des cellules détectées, puisque nous avons trouvé que les cellules CD40L$^+$ étaient enrichies en cellules Tfh. En comparaison, OX40 s'est avéré être un marqueur plus robuste pour la détection des T CD4$^+$ spécifiques au VIH. Enlever OX40 a fortement nui à la détection des lymphocytes Th17, une sous-population pourtant importante dans l'étude de l'infection au VIH, vu la forte expression des gènes y étant associés chez les ECs en comparaison aux CPs et aux participants sous ART (146–148). L'absence d'OX40 n'a pas non plus eu un impact détrimentaire sur la détection des cellules Tfh, l'autre sous-population importante pour notre étude, vu le biais dans la différentiation des lignées vers les cellules Tfh observé chez les CPs. Nos résultats montrent que l'effet d'activation *bystander* observé dans l'essai AIM multiplexé, bien que limité, serait en partie dû au marqueur CD40L. Pour toutes ces raisons, nous avons choisi de trier les réponses T CD4$^+$ spécifiques au VIH avec l'essai AIM multiplexé utilisant CD69, 4-1BB et OX40 uniquement.

Pour reconnaître que les essais fonctionnels comme l'essai AIM engendrent inévitablement des faux positifs, nous présentons les valeurs nettes (valeurs brutes de la condition non-stimulée soustraites aux valeurs brutes des conditions stimulées), mais aussi les proportions par rapport au bruit de fond (valeurs brutes de la condition non-stimulée divisées par les valeurs brutes des conditions stimulées). Chez certains participants, les proportions de bruit de fond étaient plus élevées que ce qui avaient été rapportées précédemment avec l'essai AIM multiplexé, soit une médiane de 9% (178). Ceci est dû à l'évaluation d'un seul antigène du VIH (Gag), plutôt qu'à la somme des réponses à Gag, Pol, Env et Nef.

Nous avons élaboré notre stratégie de tri en se concentrant d'abord sur les cellules positives pour CD69, un marqueur robuste utilisé depuis des années pour mesurer l'activation des cellules T (222–226) et qui est fréquemment utilisé en combinaison avec CD40L (146–148,155,173,175,221) et OX40 (173,176,227) pour les cellules T CD4$^+$. Le tri subséquent des cellules OX40$^+$ et/ou 4-1BB$^+$ évite de compter les cellules AIM$^+$ plus d'une fois et ne nécessite pas l'utilisation d'opérateurs booléens. Nous avons montré que la stratégie du *gating* en L a détecté le plus grand nombre de cellules T spécifiques à un antigène, puisque 20 à 30% des cellules triées auraient été manquées en utilisant des paires d'AIMs standards. Toutefois, l'augmentation de la fréquence des cellules détectées ne s'est pas faite au détriment de la spécificité. Notre valeur seuil de proportion de bruit de fond (33,3%) nous assure que chaque échantillon comprend >66,6% de cellules triées spécifiques au VIH ou au CMV.

Nous avons trié plus de 2,5 millions de cellules trouvées à des petites fréquences (généralement moins de 1% des cellules totales). Les cellules T CD4$^+$ et CD8$^+$ spécifiques à un antigène étaient triées de manière simultanée. Pour les analyses de séquençage ARN, nous avons d'abord trié les 2 000 cellules de tous les échantillons directement dans un tampon de lyse et les avons gelées instantanément pour assurer la stabilité de l'ARN et du transcriptome, l'ARN étant une molécule fragile qui se dégrade généralement rapidement. Les 2 000 et 10 000 cellules triées pour les extractions d'ADN étaient, quant à elles, triées dans du sérum humain pour les garder en vie jusqu'aux extractions ou aux réactions de tagmentation. Nos résultats montrent qu'autant l'ARN que l'ADN extrait était de bonne qualité.

Autant dans la procédure de sélection (*screening*) que dans les tris eux-mêmes, de fortes réponses T CD4$^+$ et CD8$^+$ spécifiques à l'antigène Gag du VIH ont été observées chez les contrôleurs élites en comparaison aux participants pré- et post-ART. Ces résultats sont consistants avec ceux montrant des réponses T CD4$^+$ plus faibles chez les participants sous ART prolongé (228–230), ainsi qu'à d'autres montrant des réponses de magnitudes supérieures chez les contrôleurs que chez les participants virémiques (146,147). La plus grande expression de PD-1 chez les participants pré-ART comparé aux contrôleurs avait aussi été rapportée (146). Pour les participants suivis de manière longitudinale, les réponses T CD4$^+$ spécifiques au VIH étaient similaires aux deux visites, mais elles étaient supérieures avant ART pour les T CD8$^+$. Une étude, utilisant aussi un essai AIM, a montré que les réponses T CD4$^+$ entre les individus pré- et post-

traitement étaient similaires, mais un peu plus élevées après le traitement pour les T CD8[+] (231). Aucune de ces différences n'était cependant significative. Dans cette étude, l'utilisation d'un mégapool de peptides du VIH (incluant les protéines Env, Gag, Nef, Pol, Rev, Tat, Vif, Vpr et Vpu), contrairement à Gag uniquement, pourrait expliquer la différence avec nos résultats.

Des différences subtiles pouvaient être trouvées dans la magnitude de réponses T CD4[+] et CD8[+] spécifiques à Gag et CMV entre la procédure de sélection et les tris. Plusieurs facteurs pourraient expliquer ces résultats, comme la différence dans les stratégies de *gating* (*gating* booléen vs. *gating* en L), dans les cytomètres utilisés (FACSAria Fusion vs. LSRFortessa), les logiciels d'analyse (FlowJo vs. BD FACSDiva) et le format des plaques pour la stimulation (96 puits vs. 48 puits). Comme tout essai fonctionnel, il y a également une faible variabilité intra-participant et inter-expérience (178). Avec l'échantillon répété à 30 reprises, nous avons montré la reproductibilité et la robustesse de l'essai.

Perspectives

L'épuisement des lymphocytes T CD8[+] spécifiques au VIH a davantage été décrit que la dysfonction des lymphocytes T CD4[+] spécifiques. Avec les cellules que nous avons triées, nous pourrons comparer les mécanismes de la dysfonction des cellules T CD4[+] spécifiques au VIH à celle des cellules T CD8[+].

Les résultats obtenus pourront être validés avec des méthodes alternatives au séquençage. Les loci différentiellement méthylés pourront être analysés avec une combinaison de l'essai COBRA et de qPCR, tandis que les transcrits différentiellement exprimés pourront être validés avec des méthodes de cytométrie en flux, telles que le RNAflow-Fish. L'essai d'hypersensibilité à la DNAse I, utilisé auparavant par nos collaborateurs (232), pourrait être employé pour les régions différentiellement accessibles. Cette étape de validation permettrait également de tester si la déplétion d'un gène dont l'expression à la hausse ne serait pas rétablie avec le traitement ART, par exemple, peut contrer la dysfonction. Ce projet permettra donc d'identifier les mécanismes épigénétiques et transcriptionnels à l'origine de la dysfonction des lymphocytes T CD4[+] spécifiques au VIH dans le contexte de l'infection chronique au VIH. L'identification de ces mécanismes pourrait conduire à la découverte de cibles médicamenteuses modifiant l'épigénome,

et donc, éventuellement d'inverser la dysfonction des cellules T CD4$^+$ et CD8$^+$ spécifiques au VIH pour permettre une meilleure élimination des réservoirs.

Limites

Une limite de notre étude en cours est l'utilisation d'un seul antigène du VIH, soit Gag, pour définir les cellules T spécifiques au VIH. Plusieurs études ont montré que le phénotype et la fonctionnalité des cellules T spécifiques au VIH variaient selon l'antigène (typiquement Gag, Pol, Env et Nef). Bien qu'elles aient seulement été réalisées dans les cellules T CD8$^+$ spécifiques à un antigène, certaines ont montré que ces cellules produisaient des quantités différentes de cytokines dépendant de l'antigène du VIH et de la cohorte (233–236) et qu'il y avait de la variabilité intra-participant (237). Ainsi, les résultats que nous obtiendrons pourraient potentiellement être spécifiques à Gag uniquement.

Il s'agit également d'un essai fonctionnel. Puisque nous essayons d'étudier la dysfonction des cellules T CD4$^+$ et CD8$^+$ spécifiques au VIH, il se peut qu'une cellule épuisée ou dysfonctionnelle, nos cellules d'intérêts, qui ne produit pas de cytokines ou qui n'augmentera pas l'expression de ces marqueurs d'activation demeurera indétectée. Pour éviter les limites posées par un essai fonctionnel, nous pourrions réaliser un essai de tétramères, mais qui serait limité à un faible nombre de complexes peptide-CMH, avec des réponses beaucoup plus faibles et qui est difficile à utiliser chez les cellules T CD4$^+$. Bien que >66,6% de ce que nous avons trié est considéré comme « spécifique », nous ne connaissons pas la proportion de la dysfonction observée qui est spécifique au VIH, puisque nous n'avons pas de contrôle aspécifique. Finalement, nous n'avons multiplexé que l'essai AIM pour les lymphocytes T CD4$^+$. Même si nos données montrent que le multiplexage n'est pas bénéfique pour la détection des lymphocytes T CD8$^+$ spécifiques à un antigène, une petite proportion de cellules T CD8$^+$ spécifiques pourrait avoir été manquée lors des tris.

5.3 L'importance des méthodes pour caractériser les réservoirs et les réponses T spécifiques au VIH

Plusieurs méthodes, basées ou non sur le séquençage, permettent de qualifier l'intégrité des réservoirs à différents niveaux. Une mauvaise classification des provirus peut surestimer la proportion de génomes intacts contribuant au rebond viral. De ce fait, des provirus défectueux classifiés inexactement comme intacts ne seraient pas éliminés et cela pourrait sous-estimer l'efficacité des stratégies d'éradication (238). Une mauvaise classification des génomes dits « défectueux » pourrait aussi sous-estimer des provirus fonctionnellement compétents pour la transcription (42), la traduction (90) et possiblement pour la réplication.

Les génomes intacts ont principalement été étudiés, dû à leur potentiel d'être compétents pour la réplication. Cependant, une étude récemment publiée par Dubé et al. montre l'existence de production spontanée de courts transcrits d'ARN viral et, dans des rares cas, de protéines virales chez des participants sous traitement ART suppressif, ce qui maintiendrait l'activation chronique des réponses T spécifiques au VIH (42). Tous les génomes étudiés étaient pourtant défectueux. Une proportion de génomes viraux défectueux peut également présenter des antigènes du VIH à la surface des cellules T, malgré le fait qu'ils soient incapables de produire des virions infectieux, et être reconnus par les cellules T CD8$^+$ spécifiques au VIH pour engendrer une réponse cytotoxique (215). Ces résultats sont en accord avec plusieurs études soulignant désormais l'importance biologique des provirus défectueux et malgré tout, incompétents pour la réplication (239–241).

Similairement à l'étude de Dubé et al. (42), des résultats récents suggèrent que les cellules transcriptionnellement actives maintiennent la réponse active des cellules T CD8$^+$ spécifiques au VIH (41). Ces deux études suggèrent que les réservoirs engagent les cellules T spécifiques à « un jeu de la taupe », qui maintiendrait l'activation chronique des réponses immunitaires (242). À son tour, l'activation chronique persistante de la réponse T entraînerait la dysfonction des cellules T CD4$^+$ et T CD8$^+$ spécifiques au VIH.

Considérant l'étroite relation qu'entretient le virus avec le système immunitaire de l'hôte, il est crucial de continuer d'améliorer nos méthodes pour la quantification et la qualification des provirus intacts et défectueux, ainsi que des lymphocytes T CD4$^+$ et CD8$^+$ spécifiques au VIH. Une

meilleure étude des réponses T immunitaires spécifiques au VIH permettrait d'identifier les mécanismes impliqués dans leur dysfonction et, éventuellement, de les inverser et d'améliorer leur fonctionnalité et leur efficacité. À son tour, cette amélioration pourrait être combinée avec une meilleure classification des provirus. Ensemble, ils pourraient aider à cibler les antigènes exprimés par ces provirus, facilitant leur élimination par le système immunitaire et, de ce fait, l'élimination de la virémie résiduelle et une diminution de l'inflammation chronique.

Références bibliographiques

1.   Bekker LG, Beyrer C, Mgodi N, Lewin SR, Delany-Moretlwe S, Taiwo B, et al. HIV infection. Nat Rev Dis Primer [Internet]. 2023 Aug 17;9(1):42. Available from: https://doi.org/10.1038/s41572-023-00452-3

2.   German Advisory Committee Blood (Arbeitskreis Blut) S 'Assessment of PT by B. Human Immunodeficiency Virus (HIV). Transfus Med Hemotherapy [Internet]. 2016 May 9 [cited 2024 Mar 14];43(3):203–22. Available from: https://doi.org/10.1159/000445852

3.   Shaw GM, Hunter E. HIV Transmission. Cold Spring Harb Perspect Med [Internet]. 2012 Nov 1;2(11). Available from: http://perspectivesinmedicine.cshlp.org/content/2/11/a006965.abstract

4.   Grossman Z, Meier-Schellersheim M, Paul WE, Picker LJ. Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. Nat Med [Internet]. 2006 Mar 1;12(3):289–95. Available from: https://doi.org/10.1038/nm1380

5.   McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF. The immune response during acute HIV-1 infection: clues for vaccine development. Nat Rev Immunol [Internet]. 2010 Jan 1;10(1):11–23. Available from: https://doi.org/10.1038/nri2674

6.   Saba E, Grivel JC, Vanpouille C, Brichacek B, Fitzgerald W, Margolis L, et al. HIV-1 sexual transmission: early events of HIV-1 infection of human cervico-vaginal tissue in an optimized ex vivo model. Mucosal Immunol [Internet]. 2010 May 1;3(3):280–90. Available from: https://www.sciencedirect.com/science/article/pii/S1933021922014842

7.   Perez-Zsolt D, Cantero-Pérez J, Erkizia I, Benet S, Pino M, Serra-Peinado C, et al. Dendritic Cells From the Cervical Mucosa Capture and Transfer HIV-1 via Siglec-1. Front Immunol [Internet]. 2019;10. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2019.00825

8.   Cohen MS, Gay CL, Busch MP, Hecht FM. The Detection of Acute HIV Infection. J Infect Dis [Internet]. 2010 Oct 1 [cited 2024 Mar 21];202(Supplement_2):S270–7. Available from: https://doi.org/10.1086/655651

9.   Cohen MS, Shaw GM, McMichael AJ, Haynes BF. Acute HIV-1 Infection. N Engl J Med [Internet]. 2011 May 19 [cited 2024 Mar 21];364(20):1943–54. Available from: https://doi.org/10.1056/NEJMra1011874

10.   Weber J. The pathogenesis of HIV-1 infection. Br Med Bull [Internet]. 2001 Sep 1 [cited 2024 Mar 23];58(1):61–72. Available from: https://doi.org/10.1093/bmb/58.1.61

11.   Barnett D, Walker B, Landay A, Denny TN. CD4 immunophenotyping in HIV infection. Nat Rev Microbiol [Internet]. 2008 Nov 1;6(11):S7–15. Available from: https://doi.org/10.1038/nrmicro1998

12.   Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. Nat Rev Dis Primer [Internet]. 2015 Oct 1;1(1):15035. Available from: https://doi.org/10.1038/nrdp.2015.35

13.   Lopez-Galindez C, Pernas M, Casado C, Olivares I, Lorenzo-Redondo R. Elite controllers and lessons learned for HIV-1 cure. Eng Viral Resist [Internet]. 2019 Oct 1;38:31–6. Available from: https://www.sciencedirect.com/science/article/pii/S1879625719300264

14. Blankson JN. Effector mechanisms in HIV-1 infected elite controllers: Highly active immune responses? Twenty-Five Years Antiretrovir Drug Dev Prog Prospects [Internet]. 2010 Jan 1;85(1):295–302. Available from: https://www.sciencedirect.com/science/article/pii/S0166354209004525

15. UNAIDS. UNAIDS. 2023. Global HIV & AIDS statistics — Fact sheet. Available from: https://www.unaids.org/en/resources/fact-sheet

16. Marcus JL, Chao CR, Leyden WA, Xu L, Quesenberry CPJ, Klein DB, et al. Narrowing the Gap in Life Expectancy Between HIV-Infected and HIV-Uninfected Individuals With Access to Care. JAIDS J Acquir Immune Defic Syndr [Internet]. 2016;73(1). Available from: https://journals.lww.com/jaids/fulltext/2016/09010/narrowing_the_gap_in_life_expectancy_between.6.aspx

17. Fromentin R, Chomont N. HIV persistence in subsets of CD4+ T cells: 50 shades of reservoirs. Immunol HIV SIV Infect [Internet]. 2021 Jan 1;51:101438. Available from: https://www.sciencedirect.com/science/article/pii/S1044532320300543

18. Monteiro P, Gosselin A, Wacleche VS, El-Far M, Said EA, Kared H, et al. Memory CCR6+CD4+ T Cells Are Preferential Targets for Productive HIV Type 1 Infection Regardless of Their Expression of Integrin β7. J Immunol [Internet]. 2011 Apr 15 [cited 2024 Jun 4];186(8):4618–30. Available from: https://doi.org/10.4049/jimmunol.1004151

19. Planas D, Zhang Y, Monteiro P, Goulet JP, Gosselin A, Grandvaux N, et al. HIV-1 selectively targets gut-homing CCR6+CD4+ T cells via mTOR-dependent mechanisms. JCI Insight [Internet]. 2017 Aug 3;2(15). Available from: https://doi.org/10.1172/jci.insight.93230

20. Anderson JL, Khoury G, Fromentin R, Solomon A, Chomont N, Sinclair E, et al. Human Immunodeficiency Virus (HIV)–Infected CCR6+ Rectal CD4+ T Cells and HIV Persistence On Antiretroviral Therapy. J Infect Dis [Internet]. 2020 Feb 18 [cited 2024 Jun 4];221(5):744–55. Available from: https://doi.org/10.1093/infdis/jiz509

21. Burdick RC, Li C, Munshi M, Rawson JMO, Nagashima K, Hu WS, et al. HIV-1 uncoats in the nucleus near sites of integration. Proc Natl Acad Sci [Internet]. 2020 Mar 10 [cited 2024 Apr 25];117(10):5486–93. Available from: https://doi.org/10.1073/pnas.1920631117

22. Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol [Internet]. 2012 Apr 1;10(4):279–90. Available from: https://doi.org/10.1038/nrmicro2747

23. NCBI. Human immunodeficiency virus type 1 (HXB2), complete genome; HIV1/HTLV-III/LAV reference genome.

24. Frankel AD, Young JAT. HIV-1: Fifteen Proteins and an RNA. Annu Rev Biochem. 1998;67:1–25.

25. Debaisieux S, Rayne F, Yezid H, Beaumelle B. The Ins and Outs of HIV-1 Tat. Traffic [Internet]. 2012 Mar 1 [cited 2024 Mar 26];13(3):355–63. Available from: https://doi.org/10.1111/j.1600-0854.2011.01286.x

26. Frankel A. Activation of HIV transcription by Tat. Curr Opin Genet Dev [Internet]. 1992 Apr;2(2):293–8. Available from: http://europepmc.org/abstract/MED/1638124

27. Strebel K. Virus–host interactions: role of HIV proteins Vif, Tat, and Rev. AIDS [Internet]. 2003;17.

Available from: https://journals.lww.com/aidsonline/fulltext/2003/17004/virus_host_interactions__role_of_hiv_proteins_vif,.3.aspx

28. Goila-Gaur R, Strebel K. HIV-1 Vif, APOBEC, and Intrinsic Immunity. Retrovirology [Internet]. 2008 Jun 24;5(1):51. Available from: https://doi.org/10.1186/1742-4690-5-51

29. Kogan M, Rappaport J. HIV-1 Accessory Protein Vpr: Relevance in the pathogenesis of HIV and potential for therapeutic intervention. Retrovirology [Internet]. 2011 Apr 13;8(1):25. Available from: https://doi.org/10.1186/1742-4690-8-25

30. Le Rouzic E, Benichou S. The Vpr protein from HIV-1: distinct roles along the viral life cycle. Retrovirology [Internet]. 2005 Feb 22;2(1):11. Available from: https://doi.org/10.1186/1742-4690-2-11

31. Dubé M, Bego MG, Paquay C, Cohen ÉA. Modulation of HIV-1-host interaction: role of the Vpu accessory protein. Retrovirology [Internet]. 2010 Dec 22;7(1):114. Available from: https://doi.org/10.1186/1742-4690-7-114

32. Perez-Caballero D, Zang T, Ebrahimi A, McNatt MW, Gregory DA, Johnson MC, et al. Tetherin Inhibits HIV-1 Release by Directly Tethering Virions to Cells. Cell [Internet]. 2009 Oct 30;139(3):499–511. Available from: https://www.sciencedirect.com/science/article/pii/S0092867409011131

33. Apps R, Del Prete GQ, Chatterjee P, Lara A, Brumme ZL, Brockman MA, et al. HIV-1 Vpu Mediates HLA-C Downregulation. Cell Host Microbe [Internet]. 2016 May 11 [cited 2024 Apr 26];19(5):686–95. Available from: https://doi.org/10.1016/j.chom.2016.04.005

34. Foster JL, Garcia JV. Role of Nef in HIV-1 Replication and Pathogenesis. In: Advances in Pharmacology [Internet]. Academic Press; 2007. p. 389–409. Available from: https://www.sciencedirect.com/science/article/pii/S1054358907550118

35. Basmaciogullari S, Pizzato M. The activity of Nef on HIV-1 infectivity. Front Microbiol [Internet]. 2014;5. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2014.00232

36. Los Alamos National Laboratory (LANL). HIV Databases. 2024. HIV-1 Gene Map. Available from: https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html

37. Zhang J, Crumpacker C. HIV UTR, LTR, and Epigenetic Immunity. Viruses. 2022;14(5).

38. Kuzembayeva M, Dilley K, Sardo L, Hu WS. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. Virology [Internet]. 2014 Apr 1;454–455:362–70. Available from: https://www.sciencedirect.com/science/article/pii/S0042682214000324

39. Maldarelli F, Palmer S, King MS, Wiegand A, Polis MA, Mican J, et al. ART Suppresses Plasma HIV-1 RNA to a Stable Set Point Predicted by Pretherapy Viremia. PLOS Pathog [Internet]. 2007 Apr 6;3(4):e46. Available from: https://doi.org/10.1371/journal.ppat.0030046

40. Palmer S, Maldarelli F, Wiegand A, Bernstein B, Hanna GJ, Brun SC, et al. Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. Proc Natl Acad Sci [Internet]. 2008 Mar 11 [cited 2024 Apr 26];105(10):3879–84. Available from: https://doi.org/10.1073/pnas.0800050105

41.   Takata H, Mitchell JL, Pacheco J, Pagliuzza A, Pinyakorn S, Buranapraditkun S, et al. An active HIV reservoir during ART is associated with maintenance of HIV-specific CD8+ T cell magnitude and short-lived differentiation status. Cell Host Microbe [Internet]. 2023 Sep 13;31(9):1494-1506.e4. Available from: https://www.sciencedirect.com/science/article/pii/S1931312823003372

42.   Dubé M, Tastet O, Dufour C, Sannier G, Brassard N, Delgado GG, et al. Spontaneous HIV expression during suppressive ART is associated with the magnitude and function of HIV-specific CD4+ and CD8+ T cells. Cell Host Microbe [Internet]. 2023 Sep 13 [cited 2023 Sep 15];31(9):1507-1522.e5. Available from: https://doi.org/10.1016/j.chom.2023.08.006

43.   J Buzón M, Massanella M, Llibre JM, Esteve A, Dahl V, Puertas MC, et al. HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. Nat Med [Internet]. 2010 Apr 1;16(4):460–5. Available from: https://doi.org/10.1038/nm.2111

44.   Cohn LB, Chomont N, Deeks SG. The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies. Cell Host Microbe [Internet]. 2020 Apr 8;27(4):519–30. Available from: https://www.sciencedirect.com/science/article/pii/S1931312820301785

45.   Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, Yassine-Diab B, et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. Nat Med [Internet]. 2009 Aug 1;15(8):893–900. Available from: https://doi.org/10.1038/nm.1972

46.   Murray AJ, Kwon KJ, Farber DL, Siliciano RF. The Latent Reservoir for HIV-1: How Immunologic Memory and Clonal Expansion Contribute to HIV-1 Persistence. J Immunol [Internet]. 2016 Jul 15 [cited 2024 Mar 25];197(2):407–17. Available from: https://doi.org/10.4049/jimmunol.1600343

47.   Hiener B, Horsburgh BA, Eden JS, Barton K, Schlub TE, Lee E, et al. Identification of Genetically Intact HIV-1 Proviruses in Specific CD4+ T Cells from Effectively Treated Participants. Cell Rep [Internet]. 2017 Oct 17 [cited 2024 Jan 27];21(3):813–22. Available from: https://doi.org/10.1016/j.celrep.2017.09.081

48.   Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, Margolick JB, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. Nat Med [Internet]. 2003 Jun 1;9(6):727–8. Available from: https://doi.org/10.1038/nm880

49.   McMyn NF, Varriale J, Fray EJ, Zitzmann C, MacLeod H, Lai J, et al. The latent reservoir of inducible, infectious HIV-1 does not decrease despite decades of antiretroviral therapy. J Clin Invest [Internet]. 2023 Sep 1;133(17). Available from: https://doi.org/10.1172/JCI171554

50.   Wagner TA, McLaughlin S, Garg K, Cheung CYK, Larsen BB, Styrchak S, et al. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. Science [Internet]. 2014 Aug 1 [cited 2024 Mar 26];345(6196):570–3. Available from: https://doi.org/10.1126/science.1256304

51.   Maldarelli F, Wu X, Su L, Simonetti FR, Shao W, Hill S, et al. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. Science [Internet]. 2014 Jul 11 [cited 2024 Mar 26];345(6193):179–83. Available from: https://doi.org/10.1126/science.1254194

52.   Gantner P, Pagliuzza A, Pardons M, Ramgopal M, Routy JP, Fromentin R, et al. Single-cell TCR sequencing reveals phenotypically diverse clonally expanded cells harboring inducible HIV proviruses during ART. Nat Commun [Internet]. 2020 Aug 14;11(1):4089. Available from: https://doi.org/10.1038/s41467-020-17898-8

53.   Anderson EM, Maldarelli F. The role of integration and clonal expansion in HIV infection: live long and prosper. Retrovirology [Internet]. 2018 Oct 23;15(1):71. Available from: https://doi.org/10.1186/s12977-018-0448-8

54.   Liu G, Zhai Q, Schaffner DJ, Wu A, Yohannes A, Robinson TM, et al. Prevention of lethal respiratory vaccinia infections in mice with interferon-α and interferon-γ. FEMS Immunol Med Microbiol [Internet]. 2004 Apr 1 [cited 2023 Apr 20];40(3):201–6. Available from: https://doi.org/10.1016/S0928-8244(03)00358-4

55.   Bui JK, Sobolewski MD, Keele BF, Spindler J, Musick A, Wiegand A, et al. Proviruses with identical sequences comprise a large fraction of the replication-competent HIV reservoir. PLOS Pathog [Internet]. 2017 Mar 22;13(3):e1006283. Available from: https://doi.org/10.1371/journal.ppat.1006283

56.   Armani-Tourret M, Bone B, Tan TS, Sun W, Bellefroid M, Struyve T, et al. Immune targeting of HIV-1 reservoir cells: a path to elimination strategies and cure. Nat Rev Microbiol [Internet]. 2024 Feb 9; Available from: https://doi.org/10.1038/s41579-024-01010-8

57.   Kulpa DA, Chomont N. HIV persistence in the setting of antiretroviral therapy: when, where and how does HIV hide? J Virus Erad [Internet]. 2015 Apr 1;1(2):59–66. Available from: https://www.sciencedirect.com/science/article/pii/S2055664020304908

58.   Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, et al. Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy. Science [Internet]. 1997 Nov 14 [cited 2024 Mar 25];278(5341):1295–300. Available from: https://doi.org/10.1126/science.278.5341.1295

59.   Wong JK, Hezareh M, Günthard HF, Havlir DV, Ignacio CC, Spina CA, et al. Recovery of Replication-Competent HIV Despite Prolonged Suppression of Plasma Viremia. Science [Internet]. 1997 Nov 14 [cited 2024 Mar 25];278(5341):1291–5. Available from: https://doi.org/10.1126/science.278.5341.1291

60.   Whitney JB, Hill AL, Sanisetty S, Penaloza-MacMaster P, Liu J, Shetty M, et al. Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. Nature [Internet]. 2014 Aug 1;512(7512):74–7. Available from: https://doi.org/10.1038/nature13594

61.   Josefsson L, von Stockenstrom S, Faria NR, Sinclair E, Bacchetti P, Killian M, et al. The HIV-1 reservoir in eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. Proc Natl Acad Sci [Internet]. 2013 Dec 17 [cited 2024 Mar 26];110(51):E4987–96. Available from: https://doi.org/10.1073/pnas.1308313110

62.   Persaud D, Gay H, Ziemniak C, Chen YH, Piatak M, Chun TW, et al. Absence of Detectable HIV-1 Viremia after Treatment Cessation in an Infant. N Engl J Med [Internet]. 2013 Nov 7 [cited 2024 Mar 26];369(19):1828–35. Available from: https://doi.org/10.1056/NEJMoa1302976

63.   Chun TW, Engel D, Berrey MM, Shea T, Corey L, Fauci AS. Early establishment of a pool of latently infected, resting CD4+ T cells during primary HIV-1 infection. Proc Natl Acad Sci [Internet]. 1998 Jul 21 [cited 2024 Mar 26];95(15):8869–73. Available from: https://doi.org/10.1073/pnas.95.15.8869

64.   Leyre L, Kroon E, Vandergeeten C, Sacdalan C, Colby DJ, Buranapraditkun S, et al. Abundant HIV-infected cells in blood and tissues are rapidly cleared upon ART initiation during acute HIV infection. Sci Transl Med [Internet]. 2020 Mar 4 [cited 2024 Apr 26];12(533):eaav3491. Available from: https://doi.org/10.1126/scitranslmed.aav3491

65.   Douek DC, Brenchley JM, Betts MR, Ambrozak DR, Hill BJ, Okamoto Y, et al. HIV preferentially infects HIV-specific CD4+ T cells. Nature [Internet]. 2002 May 1;417(6884):95–8. Available from: https://doi.org/10.1038/417095a

66.   Geldmacher C, Ngwenyama N, Schuetz A, Petrovas C, Reither K, Heeregrave EJ, et al. Preferential infection and depletion of Mycobacterium tuberculosis–specific CD4 T cells after HIV-1 infection. J Exp Med [Internet]. 2010 Nov 29 [cited 2024 Jul 4];207(13):2869–81. Available from: https://doi.org/10.1084/jem.20100090

67.   Kim Y, Anderson JL, Lewin SR. Getting the "Kill" into "Shock and Kill": Strategies to Eliminate Latent HIV. Cell Host Microbe [Internet]. 2018 Jan 10 [cited 2024 Mar 26];23(1):14–26. Available from: https://doi.org/10.1016/j.chom.2017.12.004

68.   Bruner KM, Wang Z, Simonetti FR, Bender AM, Kwon KJ, Sengupta S, et al. A quantitative approach for measuring the reservoir of latent HIV-1 proviruses. Nature [Internet]. 2019 Feb 1;566(7742):120–5. Available from: https://doi.org/10.1038/s41586-019-0898-8

69.   Ho YC, Shan L, Hosmane NN, Wang J, Laskey SB, Rosenbloom DIS, et al. Replication-Competent Noninduced Proviruses in the Latent Reservoir Increase Barrier to HIV-1 Cure. Cell [Internet]. 2013 Oct 24;155(3):540–51. Available from: https://www.sciencedirect.com/science/article/pii/S0092867413011574

70.   Cole B, Lambrechts L, Gantner P, Noppe Y, Bonine N, Witkowski W, et al. In-depth single-cell analysis of translation-competent HIV-1 reservoirs identifies cellular sources of plasma viremia. Nat Commun [Internet]. 2021 Jun 17;12(1):3727. Available from: https://doi.org/10.1038/s41467-021-24080-1

71.   Roux H, Chomont N. Measuring Human Immunodeficiency Virus Reservoirs: Do We Need to Choose Between Quantity and Quality? J Infect Dis [Internet]. 2024 Mar 15 [cited 2024 Mar 27];229(3):635–43. Available from: https://doi.org/10.1093/infdis/jiad381

72.   Christopherson Cindy, Kidane Yorda, Conway Brian, Krowka John, Sheppard Haynes, Kwok Shirley. PCR-Based Assay To Quantify Human Immunodeficiency Virus Type 1 DNA in Peripheral Blood Mononuclear Cells. J Clin Microbiol [Internet]. 2000 Feb 1 [cited 2024 Mar 27];38(2):630–4. Available from: https://doi.org/10.1128/jcm.38.2.630-634.2000

73.   Gibellini D, Vitone F, Schiavone P, Ponti C, La Placa M, Re MC. Quantitative detection of human immunodeficiency virus type 1 (HIV-1) proviral DNA in peripheral blood mononuclear cells by SYBR green real-time PCR technique. J Clin Virol [Internet]. 2004 Apr 1;29(4):282–9. Available from: https://www.sciencedirect.com/science/article/pii/S1386653203001690

74.   Strain MC, Lada SM, Luong T, Rought SE, Gianella S, Terry VH, et al. Highly Precise Measurement of HIV DNA by Droplet Digital PCR. PLOS ONE [Internet]. 2013 Apr 3;8(4):e55943. Available from: https://doi.org/10.1371/journal.pone.0055943

75.   O'Doherty Una, Swiggard William J., Jeyakumar Deepa, McGain David, Malim Michael H. A Sensitive, Quantitative Assay for Human Immunodeficiency Virus Type 1 Integration. J Virol [Internet]. 2002 Nov 1 [cited 2024 Mar 27];76(21):10942–50. Available from: https://doi.org/10.1128/jvi.76.21.10942-10950.2002

76.   Zhang X, Chen J. HIV Reservoir: How to Measure It? Curr HIV/AIDS Rep [Internet]. 2023 Apr 1;20(2):29–41. Available from: https://doi.org/10.1007/s11904-023-00653-1

77.   Massanella M, Richman DD. Measuring the latent reservoir in vivo. J Clin Invest [Internet]. 2016

Feb 1;126(2):464–72. Available from: https://doi.org/10.1172/JCI80567

78.   Pasternak AO, Jurriaans S, Bakker M, Prins JM, Berkhout B, Lukashov VV. Cellular Levels of HIV Unspliced RNA from Patients on Combination Antiretroviral Therapy with Undetectable Plasma Viremia Predict the Therapy Outcome. PLOS ONE [Internet]. 2010 Dec 31;4(12):e8490. Available from: https://doi.org/10.1371/journal.pone.0008490

79.   Kiselinova M, Pasternak AO, De Spiegelaere W, Vogelaers D, Berkhout B, Vandekerckhove L. Comparison of Droplet Digital PCR and Seminested Real-Time PCR for Quantification of Cell-Associated HIV-1 RNA. PLOS ONE [Internet]. 2014 Jan 21;9(1):e85999. Available from: https://doi.org/10.1371/journal.pone.0085999

80.   Baxter AE, Niessl J, Fromentin R, Richard J, Porichis F, Charlebois R, et al. Single-Cell Characterization of Viral Translation-Competent Reservoirs in HIV-Infected Individuals. Cell Host Microbe [Internet]. 2016 Sep 14 [cited 2024 Mar 27];20(3):368–80. Available from: https://doi.org/10.1016/j.chom.2016.07.015

81.   Baxter AE, Niessl J, Fromentin R, Richard J, Porichis F, Massanella M, et al. Multiparametric characterization of rare HIV-infected cells using an RNA-flow FISH technique. Nat Protoc [Internet]. 2017 Oct 1;12(10):2029–49. Available from: https://doi.org/10.1038/nprot.2017.079

82.   Sannier G, Dubé M, Dufour C, Richard C, Brassard N, Delgado GG, et al. Combined single-cell transcriptional, translational, and genomic profiling reveals HIV-1 reservoir diversity. Cell Rep [Internet]. 2021 Aug 31;36(9):109643. Available from: https://www.sciencedirect.com/science/article/pii/S221112472101086X

83.   Grau-Expósito Judith, Serra-Peinado Carla, Miguel Lucia, Navarro Jordi, Curran Adrià, Burgos Joaquin, et al. A Novel Single-Cell FISH-Flow Assay Identifies Effector Memory CD4+ T cells as a Major Niche for HIV-1 Transcription in HIV-Infected Patients. mBio [Internet]. 2017 Jul 11 [cited 2024 Mar 27];8(4):10.1128/mbio.00876-17. Available from: https://doi.org/10.1128/mbio.00876-17

84.   Pardons M, Baxter AE, Massanella M, Pagliuzza A, Fromentin R, Dufour C, et al. Single-cell characterization and quantification of translation-competent viral reservoirs in treated and untreated HIV infection. PLOS Pathog [Internet]. 2019 Feb 27;15(2):e1007619. Available from: https://doi.org/10.1371/journal.ppat.1007619

85.   Procopio FA, Fromentin R, Kulpa DA, Brehm JH, Bebin AG, Strain MC, et al. A Novel Assay to Measure the Magnitude of the Inducible Viral Reservoir in HIV-infected Individuals. eBioMedicine [Internet]. 2015 Aug 1 [cited 2024 Mar 27];2(8):874–83. Available from: https://doi.org/10.1016/j.ebiom.2015.06.019

86.   Taswell C. Limiting dilution assays for the determination of immunocompetent cell frequencies. I. Data analysis. J Immunol [Internet]. 1981 Apr 1 [cited 2024 Feb 5];126(4):1614–9. Available from: https://doi.org/10.4049/jimmunol.126.4.1614

87.   Deeks SG, Lewin SR, Ross AL, Ananworanich J, Benkirane M, Cannon P, et al. International AIDS Society global scientific strategy: towards an HIV cure 2016. Nat Med [Internet]. 2016 Aug 1;22(8):839–50. Available from: https://doi.org/10.1038/nm.4108

88.   Lee GQ, Orlova-Fink N, Einkauf K, Chowdhury FZ, Sun X, Harrington S, et al. Clonal expansion of genome-intact HIV-1 in functionally polarized Th1 CD4+ T cells. J Clin Invest [Internet]. 2017 Jun

30;127(7):2689–96. Available from: https://doi.org/10.1172/JCI93289

89.   Lambrechts L, Cole B, Rutsaert S, Trypsteen W, Vandekerckhove L. Emerging PCR-Based Techniques to Study HIV-1 Reservoir Persistence. Viruses. 2020;12(2).

90.   Dufour C, Richard C, Pardons M, Massanella M, Ackaoui A, Murrell B, et al. Phenotypic characterization of single CD4+ T cells harboring genetically intact and inducible HIV genomes. Nat Commun [Internet]. 2023 Feb 27;14(1):1115. Available from: https://doi.org/10.1038/s41467-023-36772-x

91.   Ferreira FA, He Q, Banning S, Roberts-Sano O, Wilkins O, Kuritzkes DR, et al. HIV-1 proviral landscape characterization varies by pipeline analysis. J Int AIDS Soc [Internet]. 2021 Jul 1 [cited 2024 Mar 26];24(7):e25725. Available from: https://doi.org/10.1002/jia2.25725

92.   Lee GQ. Chemistry and Bioinformatics Considerations in Using Next-Generation Sequencing Technologies to Inferring HIV Proviral DNA Genome-Intactness. Viruses. 2021;13(9).

93.   Buchholtz NVEJ, Nühn MM, de Jong TCM, Stienstra TAT, Reddy K, Ndung'u T, et al. Development of a highly sensitive and specific intact proviral DNA assay for HIV-1 subtype B and C. Virol J [Internet]. 2024 Jan 31;21(1):36. Available from: https://doi.org/10.1186/s12985-024-02300-6

94.   Kinloch NN, Ren Y, Conce Alberto WD, Dong W, Khadka P, Huang SH, et al. HIV-1 diversity considerations in the application of the Intact Proviral DNA Assay (IPDA). Nat Commun [Internet]. 2021 Jan 8;12(1):165. Available from: https://doi.org/10.1038/s41467-020-20442-3

95.   Gaebler Christian, Falcinelli Shane D., Stoffel Elina, Read Jenna, Murtagh Ross, Oliveira Thiago Y., et al. Sequence Evaluation and Comparative Analysis of Novel Assays for Intact Proviral HIV-1 DNA. J Virol [Internet]. 2021 Feb 24 [cited 2024 Apr 3];95(6):10.1128/jvi.01986-20. Available from: https://doi.org/10.1128/jvi.01986-20

96.   Gaebler C, Lorenzi JCC, Oliveira TY, Nogueira L, Ramos V, Lu CL, et al. Combination of quadruplex qPCR and next-generation sequencing for qualitative and quantitative analysis of the HIV-1 latent reservoir. J Exp Med [Internet]. 2019 Jul 26 [cited 2024 Mar 5];216(10):2253–64. Available from: https://doi.org/10.1084/jem.20190896

97.   Shao W, Shan J, Hu WS, Halvas EK, Mellors JW, Coffin JM, et al. HIV Proviral Sequence Database: A New Public Database for Near Full-Length HIV Proviral Sequences and Their Meta-Analyses. AIDS Res Hum Retroviruses [Internet]. 2020 Jan 1 [cited 2024 Jan 27];36(1):1–3. Available from: https://doi.org/10.1089/aid.2019.0214

98.   Wright IA, Bale MJ, Shao W, Hu WS, Coffin JM, Van Zyl GU, et al. HIVIntact: a python-based tool for HIV-1 genome intactness inference. Retrovirology [Internet]. 2021 Jun 27;18(1):16. Available from: https://doi.org/10.1186/s12977-021-00561-5

99.   Lee GQ, Reddy K, Einkauf KB, Gounder K, Chevalier JM, Dong KL, et al. HIV-1 DNA sequence diversity and evolution during acute subtype C infection. Nat Commun [Internet]. 2019 Jun 21;10(1):2737. Available from: https://doi.org/10.1038/s41467-019-10659-2

100.   Lee GQ, Lichterfeld M. Near-Full-Length Single-Genome HIV-1 DNA Sequencing. In: Poli G, Vicenzi E, Romerio F, editors. HIV Reservoirs: Methods and Protocols [Internet]. New York, NY: Springer US; 2022. p. 357–64. Available from: https://doi.org/10.1007/978-1-0716-1871-4_23

101.   Dufour C, Ruiz MJ, Pagliuzza A, Richard C, Shahid A, Fromentin R, et al. Near full-length HIV

sequencing in multiple tissues collected postmortem reveals shared clonal expansions across distinct reservoirs during ART. Cell Rep [Internet]. 2023 Sep 26 [cited 2024 Jan 27];42(9). Available from: https://doi.org/10.1016/j.celrep.2023.113053

102.   Lambrechts L, Bonine N, Verstraeten R, Pardons M, Noppe Y, Rutsaert S, et al. HIV-PULSE: a long-read sequencing assay for high-throughput near full-length HIV-1 proviral genome characterization. Nucleic Acids Res [Internet]. 2023 Nov 10 [cited 2024 Apr 26];51(20):e102–e102. Available from: https://doi.org/10.1093/nar/gkad790

103.   Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol [Internet]. 2020 Feb 7;21(1):30. Available from: https://doi.org/10.1186/s13059-020-1935-5

104.   Kuraku S, Zmasek CM, Nishimura O, Katoh K. aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res [Internet]. 2013 Jul 1 [cited 2024 Jan 27];41(W1):W22–8. Available from: https://doi.org/10.1093/nar/gkt389

105.   Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform [Internet]. 2019 Jul 19 [cited 2024 Jan 27];20(4):1160–6. Available from: https://doi.org/10.1093/bib/bbx108

106.   Los Alamos National Laboratory (LANL). Quality Control [Internet]. 2024. Available from: https://www.hiv.lanl.gov/content/sequence/QC/index.html

107.   Los Alamos National Laboratory (LANL). Gene Cutter [Internet]. 2023. Available from: https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html

108.   Los Alamos National Laboratory (LANL). ElimDupes [Internet]. 2018. Available from: https://www.hiv.lanl.gov/content/sequence/elimdupesv2/elimdupes.html

109.   Altfeld M, Gale Jr M. Innate immunity against HIV-1 infection. Nat Immunol [Internet]. 2015 Jun 1;16(6):554–62. Available from: https://doi.org/10.1038/ni.3157

110.   Carrington M, Alter G. Innate Immune Control of HIV. Cold Spring Harb Perspect Med [Internet]. 2012 Jul 1;2(7). Available from: http://perspectivesinmedicine.cshlp.org/content/2/7/a007070.abstract

111.   Iwasaki A. Innate Immune Recognition of HIV-1. Immunity [Internet]. 2012 Sep 21 [cited 2024 Apr 2];37(3):389–98. Available from: https://doi.org/10.1016/j.immuni.2012.08.011

112.   Mogensen TH, Melchjorsen J, Larsen CS, Paludan SR. Innate immune recognition and activation during HIV infection. Retrovirology [Internet]. 2010 Jun 22;7(1):54. Available from: https://doi.org/10.1186/1742-4690-7-54

113.   Soper A, Kimura I, Nagaoka S, Konno Y, Yamamoto K, Koyanagi Y, et al. Type I Interferon Responses by HIV-1 Infection: Association with Disease Progression and Control. Front Immunol [Internet]. 2018;8. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2017.01823

114.   Demers KR, Makedonas G, Buggert M, Eller MA, Ratcliffe SJ, Goonetilleke N, et al. Temporal Dynamics of CD8+ T Cell Effector Responses during Primary HIV Infection. PLOS Pathog [Internet]. 2016 Aug 3;12(8):e1005805. Available from: https://doi.org/10.1371/journal.ppat.1005805

115.   Mouquet H. Antibody B cell responses in HIV-1 infection. Trends Immunol [Internet]. 2014 Nov 1;35(11):549–61. Available from: https://www.sciencedirect.com/science/article/pii/S1471490614001495

116.   Simek Melissa D., Rida Wasima, Priddy Frances H., Pung Pham, Carrow Emily, Laufer Dagna S., et al. Human Immunodeficiency Virus Type 1 Elite Neutralizers: Individuals with Broad and Potent Neutralizing Activity Identified by Using a High-Throughput Neutralization Assay together with an Analytical Selection Algorithm. J Virol [Internet]. 2009 Jul 15 [cited 2024 Apr 26];83(14):7337–48. Available from: https://doi.org/10.1128/jvi.00110-09

117.   Bernard A, Lamy L, Alberti I. THE TWO-SIGNAL MODEL OF T-CELL ACTIVATION AFTER 30 YEARS. Transplantation [Internet]. 2002;73(1). Available from: https://journals.lww.com/transplantjournal/fulltext/2002/01151/the_two_signal_model_of_t_cell_activatio n_after_30.11.aspx

118.   Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. Nat Rev Immunol [Internet]. 2013 Apr 1;13(4):227–42. Available from: https://doi.org/10.1038/nri3405

119.   Curtsinger JM, Mescher MF. Inflammatory cytokines as a third signal for T cell activation. Lymph Act Eff Funct ● Vaccines [Internet]. 2010 Jun 1;22(3):333–40. Available from: https://www.sciencedirect.com/science/article/pii/S0952791510000476

120.   Laidlaw BJ, Craft JE, Kaech SM. The multifaceted role of CD4+ T cells in CD8+ T cell memory. Nat Rev Immunol [Internet]. 2016 Feb 1;16(2):102–11. Available from: https://doi.org/10.1038/nri.2015.10

121.   Kaech SM, Wherry EJ, Ahmed R. Effector and memory T-cell differentiation: implications for vaccine development. Nat Rev Immunol [Internet]. 2002 Apr 1;2(4):251–62. Available from: https://doi.org/10.1038/nri778

122.   Chatzileontiadou DSM, Sloane H, Nguyen AT, Gras S, Grant EJ. The Many Faces of CD4+ T Cells: Immunological and Structural Characteristics. Int J Mol Sci. 2021;22(1).

123.   Zhu X, Zhu J. CD4 T Helper Cell Subsets and Related Human Immunological Disorders. Int J Mol Sci. 2020;21(21).

124.   Schmitt N, Ueno H. Regulation of human helper T cell subset differentiation by cytokines. Antigen Process Cytokines [Internet]. 2015 Jun 1;34:130–6. Available from: https://www.sciencedirect.com/science/article/pii/S0952791515000552

125.   Zlotnik A, Yoshie O, Nomiyama H. The chemokine and chemokine receptor superfamilies and their molecular evolution. Genome Biol [Internet]. 2006 Dec 29;7(12):243. Available from: https://doi.org/10.1186/gb-2006-7-12-243

126.   Swain SL, McKinstry KK, Strutt TM. Expanding roles for CD4+ T cells in immunity to viruses. Nat Rev Immunol [Internet]. 2012 Feb 1;12(2):136–48. Available from: https://doi.org/10.1038/nri3152

127.   Cox MA, Kahan SM, Zajac AJ. Anti-viral CD8 T cells and the cytokines that they love. 2013 Rev Issue [Internet]. 2013 Jan 5;435(1):157–69. Available from: https://www.sciencedirect.com/science/article/pii/S0042682212004539

128.   McMichael AJ, Rowland-Jones SL. Cellular immune responses to HIV. Nature [Internet]. 2001 Apr 1;410(6831):980–7. Available from: https://doi.org/10.1038/35073658

129.   Appay V, Nixon DF, Donahoe SM, Gillespie GMA, Dong T, King A, et al. HIV-Specific Cd8+ T Cells Produce Antiviral Cytokines but Are Impaired in Cytolytic Function. J Exp Med [Internet]. 2000 Jul 3 [cited 2024 Aug 5];192(1):63–76. Available from: https://doi.org/10.1084/jem.192.1.63

130.   Hazenberg MD, Otto SA, van Benthem BH, Roos MT, Coutinho RA, Lange JM, et al. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. AIDS [Internet]. 2003;17(13). Available from: https://journals.lww.com/aidsonline/fulltext/2003/09050/persistent_immune_activation_in_hiv_1_infection_is.6.aspx

131.   Hunt PW, Martin JN, Sinclair E, Bredt B, Hagos E, Lampiris H, et al. T Cell Activation Is Associated with Lower CD4+ T Cell Gains in Human Immunodeficiency Virus-Infected Patients with Sustained Viral Suppression during Antiretroviral Therapy. J Infect Dis [Internet]. 2003 May 15 [cited 2024 Aug 5];187(10):1534–43. Available from: https://doi.org/10.1086/374786

132.   Lichtfuss GF, Hoy J, Rajasuriar R, Kramski M, Crowe SM, Lewin SR. Biomarkers of immune dysfunction following combination antiretroviral therapy for HIV infection. Biomark Med [Internet]. 2011 Apr 1 [cited 2024 Apr 8];5(2):171–86. Available from: https://doi.org/10.2217/bmm.11.15

133.   Marchetti Giulia, Tincati Camilla, Silvestri Guido. Microbial Translocation in the Pathogenesis of HIV Infection and AIDS. Clin Microbiol Rev [Internet]. 2013 Jan 1 [cited 2024 Apr 8];26(1):2–18. Available from: https://doi.org/10.1128/cmr.00050-12

134.   Berg RD. Bacterial translocation from the gastrointestinal tract. Trends Microbiol [Internet]. 1995 Apr 1 [cited 2024 Apr 8];3(4):149–54. Available from: https://doi.org/10.1016/S0966-842X(00)88906-4

135.   Klatt NR, Funderburg NT, Brenchley JM. Microbial translocation, immune activation, and HIV disease. Trends Microbiol [Internet]. 2013 Jan 1 [cited 2024 Apr 8];21(1):6–13. Available from: https://doi.org/10.1016/j.tim.2012.09.001

136.   Sandler NG, Douek DC. Microbial translocation in HIV infection: causes, consequences and treatment opportunities. Nat Rev Microbiol [Internet]. 2012 Sep 1;10(9):655–66. Available from: https://doi.org/10.1038/nrmicro2848

137.   Paiardini M, Müller-Trutwin M. HIV-associated chronic immune activation. Immunol Rev [Internet]. 2013 Jul 1 [cited 2024 Apr 8];254(1):78–101. Available from: https://doi.org/10.1111/imr.12079

138.   Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med [Internet]. 2006 Dec 1;12(12):1365–71. Available from: https://doi.org/10.1038/nm1511

139.   Fenwick C, Joo V, Jacquier P, Noto A, Banga R, Perreau M, et al. T-cell exhaustion in HIV infection. Immunol Rev [Internet]. 2019 Nov 1 [cited 2024 Apr 7];292(1):149–63. Available from: https://doi.org/10.1111/imr.12823

140.   Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol [Internet]. 2015 Aug 1;15(8):486–99. Available from: https://doi.org/10.1038/nri3862

141.   El-Far M, Halwani R, Said E, Trautmann L, Doroudchi M, Janbazian L, et al. T-cell exhaustion in HIV infection. Curr HIV/AIDS Rep [Internet]. 2008 Feb 1;5(1):13–9. Available from: https://doi.org/10.1007/s11904-008-0003-7

142.   Khaitan A, Unutmaz D. Revisiting Immune Exhaustion During HIV Infection. Curr HIV/AIDS Rep [Internet]. 2011 Mar 1;8(1):4–11. Available from: https://doi.org/10.1007/s11904-010-0066-0

143.   D'Souza M, Fontenot AP, Mack DG, Lozupone C, Dillon S, Meditz A, et al. Programmed Death 1 Expression on HIV-Specific CD4+ T Cells Is Driven by Viral Replication and Associated with T Cell Dysfunction1. J Immunol [Internet]. 2007 Aug 1 [cited 2024 Aug 5];179(3):1979–87. Available from: https://doi.org/10.4049/jimmunol.179.3.1979

144.   Kaufmann DE, Kavanagh DG, Pereyra F, Zaunders JJ, Mackey EW, Miura T, et al. Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. Nat Immunol [Internet]. 2007 Nov 1;8(11):1246–54. Available from: https://doi.org/10.1038/ni1515

145.   Day CL, Kaufmann DE, Kiepiela P, Brown JA, Moodley ES, Reddy S, et al. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. Nature [Internet]. 2006 Sep 1;443(7109):350–4. Available from: https://doi.org/10.1038/nature05115

146.   Brunet-Ratnasingham E, Morou A, Dubé M, Niessl J, Baxter AE, Tastet O, et al. Immune checkpoint expression on HIV-specific CD4+ T cells and response to their blockade are dependent on lineage and function. eBioMedicine [Internet]. 2022 Oct 1;84:104254. Available from: https://www.sciencedirect.com/science/article/pii/S2352396422004364

147.   Morou A, Brunet-Ratnasingham E, Dubé M, Charlebois R, Mercier E, Darko S, et al. Altered differentiation is central to HIV-specific CD4+ T cell dysfunction in progressive disease. Nat Immunol [Internet]. 2019 Aug 1;20(8):1059–70. Available from: https://doi.org/10.1038/s41590-019-0418-x

148.   Niessl J, Baxter AE, Morou A, Brunet-Ratnasingham E, Sannier G, Gendron-Lepage G, et al. Persistent expansion and Th1-like skewing of HIV-specific circulating T follicular helper cells during antiretroviral therapy. EBioMedicine [Internet]. 2020 Apr 1;54:102727. Available from: https://www.sciencedirect.com/science/article/pii/S235239642030102X

149.   Morou A, Palmer BE, Kaufmann DE. Distinctive features of CD4+ T cell dysfunction in chronic viral infections. Curr Opin HIV AIDS [Internet]. 2014;9(5). Available from: https://journals.lww.com/co-hivandaids/fulltext/2014/09000/distinctive_features_of_cd4__t_cell_dysfunction_in.4.aspx

150.   Jansen CA, van Baarle D, Miedema F. HIV-specific CD4+ T cells and viremia: who's in control? Trends Immunol [Internet]. 2006 Mar 1 [cited 2024 Apr 8];27(3):119–24. Available from: https://doi.org/10.1016/j.it.2006.01.004

151.   Youngblood B, Noto A, Porichis F, Akondy RS, Ndhlovu ZM, Austin JW, et al. Cutting Edge: Prolonged Exposure to HIV Reinforces a Poised Epigenetic Program for PD-1 Expression in Virus-Specific CD8 T Cells. J Immunol [Internet]. 2013 Jul 15 [cited 2024 Jul 5];191(2):540–4. Available from: https://doi.org/10.4049/jimmunol.1203161

152.   Martin GE, Sen DR, Pace M, Robinson N, Meyerowitz J, Adland E, et al. Epigenetic Features of HIV-Induced T-Cell Exhaustion Persist Despite Early Antiretroviral Therapy. Front Immunol [Internet]. 2021;12. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2021.647688

153.   Scharer CD, Barwick BG, Youngblood BA, Ahmed R, Boss JM. Global DNA Methylation Remodeling Accompanies CD8 T Cell Effector Function. J Immunol [Internet]. 2013 Sep 15 [cited 2024

Jul 5];191(6):3419–29. Available from: https://doi.org/10.4049/jimmunol.1301395

154.   Sen DR, Kaminski J, Barnitz RA, Kurachi M, Gerdemann U, Yates KB, et al. The epigenetic landscape of T cell exhaustion. Science [Internet]. 2016 Dec 2 [cited 2024 Apr 21];354(6316):1165–9. Available from: https://doi.org/10.1126/science.aae0491

155.   Reiss S, Baxter AE, Cirelli KM, Dan JM, Morou A, Daigneault A, et al. Comparative analysis of activation induced marker (AIM) assays for sensitive identification of antigen-specific CD4 T cells. PLOS ONE   [Internet].   2017   Oct   24;12(10):e0186998.   Available   from: https://doi.org/10.1371/journal.pone.0186998

156.   Chattopadhyay PK, Yu J, Roederer M. A live-cell assay to detect antigen-specific CD4+ T cells with diverse cytokine profiles. Nat Med [Internet]. 2005 Oct 1;11(10):1113–7. Available from: https://doi.org/10.1038/nm1293

157.   Yu D, Rao S, Tsai LM, Lee SK, He Y, Sutcliffe EL, et al. The Transcriptional Repressor Bcl-6 Directs T Follicular Helper Cell Lineage Commitment. Immunity [Internet]. 2009 Sep 18;31(3):457–68. Available from: https://www.sciencedirect.com/science/article/pii/S1074761309003148

158.   Kroenke MA, Eto D, Locci M, Cho M, Davidson T, Haddad EK, et al. Bcl6 and Maf Cooperate To Instruct Human Follicular Helper CD4 T Cell Differentiation. J Immunol [Internet]. 2012 Apr 15 [cited 2023 Feb 23];188(8):3734–44. Available from: https://doi.org/10.4049/jimmunol.1103246

159.   Freer G, Rindi L. Intracellular cytokine detection by fluorescence-activated flow cytometry: Basic principles and recent advances. Cytokine Methods [Internet]. 2013 May 15;61(1):30–8. Available from: https://www.sciencedirect.com/science/article/pii/S1046202313001047

160.   Lamoreaux L, Roederer M, Koup R. Intracellular cytokine optimization and standard operating procedure.   Nat   Protoc   [Internet].   2006   Aug   1;1(3):1507–16.   Available   from: https://doi.org/10.1038/nprot.2006.268

161.   Horton H, Thomas EP, Stucky JA, Frank I, Moodie Z, Huang Y, et al. Optimization and validation of an 8-color intracellular cytokine staining (ICS) assay to quantify antigen-specific T cells induced by vaccination.   J   Immunol   Methods   [Internet].   2007   May   31;323(1):39–54.   Available   from: https://www.sciencedirect.com/science/article/pii/S0022175907000890

162.   Albert-Vega C, Tawfik DM, Trouillet-Assant S, Vachot L, Mallet F, Textoris J. Immune Functional Assays, From Custom to Standardized Tests for Precision Medicine. Front Immunol [Internet]. 2018;9. Available from: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2018.02367

163.   Robinson HL, Amara RR. T cell vaccines for microbial infections. Nat Med [Internet]. 2005 Apr 1;11(4):S25–32. Available from: https://doi.org/10.1038/nm1212

164.   Sims S, Willberg C, Klenerman P. MHC–peptide tetramers for the analysis of antigen-specific T cells.   Expert   Rev   Vaccines   [Internet].   2010   Jul   1;9(7):765–74.   Available   from: https://doi.org/10.1586/erv.10.66

165.   Poloni C, Schonhofer C, Ivison S, Levings MK, Steiner TS, Cook L. T-cell activation–induced marker assays in health and disease. Immunol Cell Biol [Internet]. 2023 Jul 1 [cited 2024 Mar 4];101(6):491–503. Available from: https://doi.org/10.1111/imcb.12636

166.   Sallusto F, Cassotta A, Hoces D, Foglierini M, Lanzavecchia A. Do Memory CD4 T Cells Keep

Their Cell-Type Programming: Plasticity versus Fate Commitment?: T-Cell Heterogeneity, Plasticity, and Selection in Humans. Cold Spring Harb Perspect Biol [Internet]. 2018 Mar 1;10(3):a029421. Available from: http://cshperspectives.cshlp.org/content/10/3/a029421.abstract

167.   Crotty S. Do Memory CD4 T Cells Keep Their Cell-Type Programming: Plasticity versus Fate Commitment?: Complexities of Interpretation due to the Heterogeneity of Memory CD4 T Cells, Including T Follicular Helper Cells. Cold Spring Harb Perspect Biol [Internet]. 2018 Mar 1;10(3):a032102. Available from: http://cshperspectives.cshlp.org/content/10/3/a032102.abstract

168.   Niessl J, Baxter AE, Mendoza P, Jankovic M, Cohen YZ, Butler AL, et al. Combination anti-HIV-1 antibody therapy is associated with increased virus-specific T cell immunity. Nat Med [Internet]. 2020 Feb 1;26(2):222–7. Available from: https://doi.org/10.1038/s41591-019-0747-1

169.   Nayrac M, Dubé M, Sannier G, Nicolas A, Marchitto L, Tastet O, et al. Temporal associations of B and T cell immunity with robust vaccine responsiveness in a 16-week interval BNT162b2 regimen. Cell Rep [Internet]. 2022 Jun 28;39(13):111013. Available from: https://www.sciencedirect.com/science/article/pii/S2211124722008026

170.   Sannier G, Nicolas A, Dubé M, Marchitto L, Nayrac M, Tastet O, et al. A third SARS-CoV-2 mRNA vaccine dose in people receiving hemodialysis overcomes B cell defects but elicits a skewed CD4+ T cell profile. Cell Rep Med [Internet]. 2023 Mar 21;4(3):100955. Available from: https://www.sciencedirect.com/science/article/pii/S2666379123000472

171.   Nicolas A, Sannier G, Dubé M, Nayrac M, Tauzin A, Painter MM, et al. An extended SARS-CoV-2 mRNA vaccine prime-boost interval enhances B cell immunity with limited impact on T cells. iScience [Internet]. 2023 Jan 20 [cited 2023 Jan 16];26(1):105904. Available from: https://doi.org/10.1016/j.isci.2022.105904

172.   Bacher P, Scheffold A. Flow-cytometric analysis of rare antigen-specific T cells. Cytometry A [Internet]. 2013 Aug 1 [cited 2023 Feb 10];83A(8):692–701. Available from: https://doi.org/10.1002/cyto.a.22317

173.   Ferragut F, Cruz KM, Gallardo JP, Fernández M, Hernández Vasquez Y, Gómez KA. Activation-induced marker assays for identification of Trypanosoma cruzi-specific CD4 or CD8 T cells in chronic Chagas disease patients. Immunology [Internet]. 2022 Dec 25 [cited 2023 Mar 6]; Available from: https://doi.org/10.1111/imm.13622

174.   Painter MM, Mathew D, Goel RR, Apostolidis SA, Pattekar A, Kuthuru O, et al. Rapid induction of antigen-specific CD4+ T cells is associated with coordinated humoral and cellular immunity to SARS-CoV-2 mRNA vaccination. Immunity [Internet]. 2021 Sep 14;54(9):2133-2142.e3. Available from: https://www.sciencedirect.com/science/article/pii/S1074761321003083

175.   Barham MS, Whatney WE, Khayumbi J, Ongalo J, Sasser LE, Campbell A, et al. Activation-Induced Marker Expression Identifies Mycobacterium tuberculosis–Specific CD4 T Cells in a Cytokine-Independent Manner in HIV-Infected Individuals with Latent Tuberculosis. ImmunoHorizons [Internet]. 2020 Oct 5 [cited 2023 Oct 1];4(10):573–84. Available from: https://doi.org/10.4049/immunohorizons.2000051

176.   Pallikkuth S, Williams E, Pahwa R, Hoffer M, Pahwa S. Association of Flu specific and SARS-CoV-2 specific CD4 T cell responses in SARS-CoV-2 infected asymptomatic heath care workers. Vaccine [Internet]. 2021 Oct 1;39(41):6019–24. Available from:

https://www.sciencedirect.com/science/article/pii/S0264410X2101152X

177.   Tarke A, Coelho CH, Zhang Z, Dan JM, Yu ED, Methot N, et al. SARS-CoV-2 vaccination induces immunological T cell memory able to cross-recognize variants from Alpha to Omicron. Cell [Internet]. 2022 Mar 3;185(5):847-859.e11. Available from: https://www.sciencedirect.com/science/article/pii/S0092867422000733

178.   Lemieux A, Sannier G, Nicolas A, Nayrac M, Delgado GG, Cloutier R, et al. Enhanced detection of antigen-specific T cells by a multiplexed AIM assay. Cell Rep Methods [Internet]. 2024 Jan 22 [cited 2024 Feb 9];4(1). Available from: https://doi.org/10.1016/j.crmeth.2023.100690

179.   Morcilla Vincent, Bacchus-Souffan Charline, Fisher Katie, Horsburgh Bethany A., Hiener Bonnie, Wang Xiao Qian, et al. HIV-1 Genomes Are Enriched in Memory CD4+ T-Cells with Short Half-Lives. mBio [Internet]. 2021 Sep 21 [cited 2024 Mar 20];12(5):10.1128/mbio.02447-21. Available from: https://doi.org/10.1128/mbio.02447-21

180.   Clark E, Nava B, Caputi M. Tat is a multifunctional viral protein that modulates cellular gene expression and functions. Oncotarget Vol 8 No 16 [Internet]. 2017 [cited 2017 Jan 1]; Available from: https://www.oncotarget.com/article/15174/text/

181.   Rose PP, Korber BT. Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. Bioinformatics [Internet]. 2000 Apr 1 [cited 2024 Mar 28];16(4):400–1. Available from: https://doi.org/10.1093/bioinformatics/16.4.400

182.   Edlefsen P, Labuschagne P. hypermutR [Internet]. 2020. Available from: https://github.com/philliplab/hypermutR

183.   Heibl C, Cusimano N, Krah FS. ips: Interfaces to Phylogenetic Software in R [Internet]. 2019. Available from: https://cran.r-project.org/web/packages/ips/index.html

184.   Wickham H, Hester J, Chang W, Bryan J. devtools: Tools to Make Developing R Packages Easier [Internet]. 2022. Available from: https://CRAN.R-project.org/package=devtools

185.   Pardons M, Cole B, Lambrechts L, van Snippenberg W, Rutsaert S, Noppe Y, et al. Potent latency reversal by Tat RNA-containing nanoparticle enables multi-omic analysis of the HIV-1 reservoir. Nat Commun [Internet]. 2023 Dec 18;14(1):8397. Available from: https://doi.org/10.1038/s41467-023-44020-5

186.   Jörimann L, Tschumi J, Zeeb M, Leemann C, Schenkel CD, Neumann K, et al. Absence of Proviral Human Immunodeficiency Virus (HIV) Type 1 Evolution in Early-Treated Individuals With HIV Switching to Dolutegravir Monotherapy During 48 Weeks. J Infect Dis [Internet]. 2023 Oct 1 [cited 2024 Mar 28];228(7):907–18. Available from: https://doi.org/10.1093/infdis/jiad292

187.   Simonetti FR, Zhang H, Soroosh GP, Duan J, Rhodehouse K, Hill AL, et al. Antigen-driven clonal selection shapes the persistence of HIV-1–infected CD4+ T cells in vivo. J Clin Invest [Internet]. 2021 Feb 1;131(3). Available from: https://doi.org/10.1172/JCI145254

188.   White JA, Wu F, Yasin S, Moskovljevic M, Varriale J, Dragoni F, et al. Clonally expanded HIV-1 proviruses with 5′-leader defects can give rise to nonsuppressible residual viremia. J Clin Invest [Internet]. 2023 Mar 15;133(6). Available from: https://doi.org/10.1172/JCI165245

189.   Einkauf KB, Osborn MR, Gao C, Sun W, Sun X, Lian X, et al. Parallel analysis of transcription,

integration, and sequence of single HIV-1 proviruses. Cell [Internet]. 2022 Jan 20;185(2):266-282.e15. Available from: https://www.sciencedirect.com/science/article/pii/S0092867421014495

190.   Mellors JW, Guo S, Naqvi A, Brandt LD, Su L, Sun Z, et al. Insertional activation of STAT3 and LCK by HIV-1 proviruses in T cell lymphomas. Sci Adv [Internet]. [cited 2024 Mar 28];7(42):eabi8795. Available from: https://doi.org/10.1126/sciadv.abi8795

191.   Joseph Kevin W., Halvas Elias K., Brandt Leah D., Patro Sean C., Rausch Jason W., Chopra Abha, et al. Deep Sequencing Analysis of Individual HIV-1 Proviruses Reveals Frequent Asymmetric Long Terminal Repeats. J Virol [Internet]. 2022 Jun 8 [cited 2024 Mar 28];96(13):e00122-22. Available from: https://doi.org/10.1128/jvi.00122-22

192.   Cassidy NAJ, Fish CS, Levy CN, Roychoudhury P, Reeves DB, Hughes SM, et al. HIV reservoir quantification using cross-subtype multiplex ddPCR. iScience [Internet]. 2022 Jan 21;25(1):103615. Available from: https://www.sciencedirect.com/science/article/pii/S2589004221015856

193.   Wright IA, Delaney KE, Katusiime MGK, Botha JC, Engelbrecht S, Kearney MF, et al. NanoHIV: A Bioinformatics Pipeline for Producing Accurate, Near Full-Length HIV Proviral Genomes Sequenced Using the Oxford Nanopore Technology. Cells. 2021;10(10).

194.   Bbosa N, Ssemwanga D, Nsubuga RN, Salazar-Gonzalez JF, Salazar MG, Nanyonjo M, et al. Phylogeography of HIV-1 suggests that Ugandan fishing communities are a sink for, not a source of, virus from general populations. Sci Rep [Internet]. 2019 Jan 31;9(1):1051. Available from: https://doi.org/10.1038/s41598-018-37458-x

195.   Méndez-Ortega MC, Restrepo S, Rodríguez-R LM, Pérez I, Mendoza JC, Martínez AP, et al. An RNAi in silico approach to find an optimal shRNA cocktail against HIV-1. Virol J [Internet]. 2010 Dec 20;7(1):369. Available from: https://doi.org/10.1186/1743-422X-7-369

196.   De Scheerder MA, Vrancken B, Dellicour S, Schlub T, Lee E, Shao W, et al. HIV Rebound Is Predominantly Fueled by Genetically Identical Viral Expansions from Diverse Reservoirs. Cell Host Microbe [Internet]. 2019 Sep 11;26(3):347-358.e7. Available from: https://www.sciencedirect.com/science/article/pii/S1931312819303683

197.   Aloi F, Riolo M, Sanzani SM, Mincuzzi A, Ippolito A, Siciliano I, et al. Characterization of Alternaria Species Associated with Heart Rot of Pomegranate Fruit. J Fungi. 2021;7(3).

198.   Prigigallo MI, Abdelfattah A, Cacciola SO, Faedda R, Sanzani SM, Cooke DEL, et al. Metabarcoding Analysis of Phytophthora Diversity Using Genus-Specific Primers and 454 Pyrosequencing. Phytopathology® [Internet]. 2016 Mar 1 [cited 2024 Mar 28];106(3):305–13. Available from: https://doi.org/10.1094/PHYTO-07-15-0167-R

199.   Trento Alfonsina, Ábrego Leyda, Rodriguez-Fernandez Rosa, González-Sánchez Maria Isabel, González-Martínez Felipe, Delfraro Adriana, et al. Conservation of G-Protein Epitopes in Respiratory Syncytial Virus (Group A) Despite Broad Genetic Diversity: Is Antibody Selection Involved in Virus Evolution? J Virol [Internet]. 2015 Jul 8 [cited 2024 Mar 28];89(15):7776–85. Available from: https://doi.org/10.1128/jvi.00467-15

200.   Datta S, Das B, Budhauliya R. Distinct begomovirus complexes infecting adjacently growing plants in Northeast India-A report. Indian Phytopathol [Internet]. 2019 Sep 1;72(3):523–30. Available from: https://doi.org/10.1007/s42360-019-00170-5

201.   Bengyella L, Hetsa BA, Fonmboh DJ, Jose RC. Assessment of damage caused by evolved fall armyworm on native and transgenic maize in South Africa. Phytoparasitica [Internet]. 2021 Feb 1;49(1):1–12. Available from: https://doi.org/10.1007/s12600-020-00862-z

202.   Adamowicz SJ, Marinone MC, Menu-Marque S, Martin JW, Allen DC, Pyle MN, et al. The Hyalella (Crustacea: Amphipoda) species cloud of the ancient Lake Titicaca originated from multiple colonizations. Mol Phylogenet Evol [Internet]. 2018 Aug 1;125:232–42. Available from: https://www.sciencedirect.com/science/article/pii/S1055790317301549

203.   Scharer CD, Bally APR, Gandham B, Boss JM. Cutting Edge: Chromatin Accessibility Programs CD8 T Cell Memory. J Immunol [Internet]. 2017 Mar 15 [cited 2024 Jul 5];198(6):2238–43. Available from: https://doi.org/10.4049/jimmunol.1602086

204.   Barwick BG, Scharer CD, Bally APR, Boss JM. Plasma cell differentiation is coupled to division-dependent DNA hypomethylation and gene regulation. Nat Immunol [Internet]. 2016 Oct 1;17(10):1216–25. Available from: https://doi.org/10.1038/ni.3519

205.   Scharer CD, Blalock EL, Barwick BG, Haines RR, Wei C, Sanz I, et al. ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. Sci Rep [Internet]. 2016 Jun 1;6(1):27030. Available from: https://doi.org/10.1038/srep27030

206.   Lichtner M, Cicconi P, Vita S, Cozzi-Lepri A, Galli M, Lo Caputo S, et al. Cytomegalovirus Coinfection Is Associated With an Increased Risk of Severe Non–AIDS-Defining Events in a Large Cohort of HIV-Infected Patients. J Infect Dis [Internet]. 2015 Jan 15 [cited 2023 Nov 4];211(2):178–86. Available from: https://doi.org/10.1093/infdis/jiu417

207.   Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, Devleesschauwer B, et al. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. Rev Med Virol [Internet]. 2019 May 1 [cited 2024 Mar 21];29(3):e2034. Available from: https://doi.org/10.1002/rmv.2034

208.   Booiman T, Wit FW, Girigorie AF, Maurer I, De Francesco D, Sabin CA, et al. Terminal differentiation of T cells is strongly associated with CMV infection and increased in HIV-positive individuals on ART and lifestyle matched controls. PLOS ONE [Internet]. 2017 Aug 14;12(8):e0183357. Available from: https://doi.org/10.1371/journal.pone.0183357

209.   van den Berg SPH, Pardieck IN, Lanfermeijer J, Sauce D, Klenerman P, van Baarle D, et al. The hallmarks of CMV-specific CD8 T-cell differentiation. Med Microbiol Immunol (Berl) [Internet]. 2019 Aug 1;208(3):365–73. Available from: https://doi.org/10.1007/s00430-019-00608-7

210.   Ballegaard V, Brændstrup P, Pedersen KK, Kirkby N, Stryhn A, Ryder LP, et al. Cytomegalovirus-specific T-cells are associated with immune senescence, but not with systemic inflammation, in people living with HIV. Sci Rep [Internet]. 2018 Feb 28;8(1):3778. Available from: https://doi.org/10.1038/s41598-018-21347-4

211.   Vandergeeten Claire, Fromentin Rémi, Merlini Esther, Lawani Mariam B., DaFonseca Sandrina, Bakeman Wendy, et al. Cross-Clade Ultrasensitive PCR-Based Assays To Measure HIV Persistence in Large-Cohort Studies. J Virol [Internet]. 2014 Nov 1 [cited 2024 Apr 19];88(21):12385–96. Available from: https://doi.org/10.1128/jvi.00609-14

212.   Hofmeyer KA, Jeon H, Zang X. The PD-1/PD-L1 (B7-H1) Pathway in Chronic Infection-Induced Cytotoxic T Lymphocyte Exhaustion. Curtsinger J, editor. J Biomed Biotechnol [Internet]. 2011 Sep

25;2011:451694. Available from: https://doi.org/10.1155/2011/451694

213.   Porichis F, Kaufmann DE. Role of PD-1 in HIV Pathogenesis and as Target for Therapy. Curr HIV/AIDS Rep [Internet]. 2012 Mar 1;9(1):81–90. Available from: https://doi.org/10.1007/s11904-011-0106-4

214.   Mohammadi A, Etemad B, Zhang X, Li Y, Bedwell GJ, Sharaf R, et al. Viral and host mediators of non-suppressible HIV-1 viremia. Nat Med [Internet]. 2023 Dec 1;29(12):3212–23. Available from: https://doi.org/10.1038/s41591-023-02611-1

215.   Pollack RA, Jones RB, Pertea M, Bruner KM, Martin AR, Thomas AS, et al. Defective HIV-1 Proviruses Are Expressed and Can Be Recognized by Cytotoxic T Lymphocytes, which Shape the Proviral Landscape. Cell Host Microbe [Internet]. 2017 Apr 12 [cited 2024 Apr 6];21(4):494-506.e4. Available from: https://doi.org/10.1016/j.chom.2017.03.008

216.   Sun N, Yang J, Yau SST. Identification of HIV Rapid Mutations Using Differences in Nucleotide Distribution over Time. Genes. 2022;13(2).

217.   Liu Y, Rao U, McClure J, Konopa P, Manocheewa S, Kim M, et al. Impact of Mutations in Highly Conserved Amino Acids of the HIV-1 Gag-p24 and Env-gp120 Proteins on Viral Replication in Different Genetic Backgrounds. PLOS ONE [Internet]. 2014 Apr 8;9(4):e94240. Available from: https://doi.org/10.1371/journal.pone.0094240

218.   Zanini F, Brodin J, Thebo L, Lanz C, Bratt G, Albert J, et al. Population genomics of intrapatient HIV-1 evolution. Chakraborty AK, editor. eLife [Internet]. 2015 Dec 10;4:e11282. Available from: https://doi.org/10.7554/eLife.11282

219.   Shingai M, Donau OK, Schmidt SD, Gautam R, Plishka RJ, Buckler-White A, et al. Most rhesus macaques infected with the CCR5-tropic SHIVAD8 generate cross-reactive antibodies that neutralize multiple HIV-1 strains. Proc Natl Acad Sci [Internet]. 2012 Nov 27 [cited 2024 Apr 6];109(48):19769–74. Available from: https://doi.org/10.1073/pnas.1217443109

220.   Gartner MJ, Roche M, Churchill MJ, Gorry PR, Flynn JK. Understanding the mechanisms driving the spread of subtype C HIV-1. eBioMedicine [Internet]. 2020 Mar 1 [cited 2024 Apr 5];53. Available from: https://doi.org/10.1016/j.ebiom.2020.102682

221.   Busà R, Sorrentino MC, Russelli G, Amico G, Miceli V, Miele M, et al. Specific Anti-SARS-CoV-2 Humoral and Cellular Immune Responses After Booster Dose of BNT162b2 Pfizer-BioNTech mRNA-Based Vaccine: Integrated Study of Adaptive Immune System Components. Front Immunol [Internet]. 2022;13. Available from: https://www.frontiersin.org/articles/10.3389/fimmu.2022.856657

222.   Lindsey WB, Lowdell MW, Marti GE, Abbasi F, Zenger V, King KM, et al. CD69 expression as an index of T-cell function: assay standardization, validation and use in monitoring immune recovery. Cytotherapy [Internet]. 2007 Jan 1 [cited 2024 Apr 7];9(2):123–32. Available from: https://doi.org/10.1080/14653240601182838

223.   Maino VC, Suni MA, Ruitenberg JJ. Rapid flow cytometric method for measuring lymphocyte subset activation. Cytometry [Internet]. 1995 Jun 1 [cited 2024 Apr 7];20(2):127–33. Available from: https://doi.org/10.1002/cyto.990200205

224.   Mardiney III M, Brown MR, Fleisher TA. Measurement of T-cell CD69 expression: A rapid and efficient means to assess mitogen- or antigen-induced proliferative capacity in normals. Cytometry

[Internet]. 1996 Dec 15 [cited 2024 Apr 7];26(4):305–10. Available from: https://doi.org/10.1002/(SICI)1097-0320(19961215)26:4<305::AID-CYTO11>3.0.CO;2-V

225. Kilmartin DJ, Fletcher ZJ, Almeida JA, Liversidge J, Forrester JV, Dick AD. CD69 Expression on Peripheral CD4+ T Cells Parallels Disease Activity and Is Reduced by Mycophenolate Mofetil Therapy in Uveitis. Invest Ophthalmol Vis Sci. 2001 May 1;42(6):1285–92.

226. Testi R, D'Ambrosio D, De Maria R, Santoni A. The CD69 receptor: a multipurpose cell-surface trigger for hematopoietic cells. Immunol Today [Internet]. 1994 Oct 1;15(10):479–83. Available from: https://www.sciencedirect.com/science/article/pii/0167569994901937

227. Naaber P, Tserel L, Kangro K, Sepp E, Jürjenson V, Adamson A, et al. Dynamics of antibody response to BNT162b2 vaccine after six months: a longitudinal prospective study. Lancet Reg Health - Eur [Internet]. 2021 Nov 1;10:100208. Available from: https://www.sciencedirect.com/science/article/pii/S266677622100185X

228. Ferre April L., Hunt Peter W., McConnell Delandy H., Morris Megan M., Garcia Juan C., Pollard Richard B., et al. HIV Controllers with HLA-DRB1*13 and HLA-DQB1*06 Alleles Have Strong, Polyfunctional Mucosal CD4+ T-Cell Responses. J Virol [Internet]. 2010 Nov 1 [cited 2024 Apr 8];84(21):11020–9. Available from: https://doi.org/10.1128/jvi.00980-10

229. Pitcher CJ, Quittner C, Peterson DM, Connors M, Koup RA, Maino VC, et al. HIV-1-specific CD4+ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged viral suppression. Nat Med [Internet]. 1999 May 1;5(5):518–25. Available from: https://doi.org/10.1038/8400

230. Seth N, Kaufmann D, Lahey T, Rosenberg ES, Wucherpfennig KW. Expansion and Contraction of HIV-Specific CD4 T Cells with Short Bursts of Viremia, but Physical Loss of the Majority of These Cells with Sustained Viral Replication1. J Immunol [Internet]. 2005 Nov 15 [cited 2024 Aug 5];175(10):6948–58. Available from: https://doi.org/10.4049/jimmunol.175.10.6948

231. Al-kolla R, Grifoni A, Crotty S, Sette A, Gianella S, Dan J. Design and validation of HIV peptide pools for detection of HIV-specific CD4+ and CD8+ T cells. PLOS ONE [Internet]. 2022 Aug 16;17(8):e0268370. Available from: https://doi.org/10.1371/journal.pone.0268370

232. Austin JW, Lu P, Majumder P, Ahmed R, Boss JM. STAT3, STAT4, NFATc1, and CTCF Regulate PD-1 through Multiple Novel Regulatory Regions in Murine T Cells. J Immunol [Internet]. 2014 May 15 [cited 2024 Jul 5];192(10):4876–86. Available from: https://doi.org/10.4049/jimmunol.1302750

233. Hersperger AR, Pereyra F, Nason M, Demers K, Sheth P, Shin LY, et al. Perforin Expression Directly Ex Vivo by HIV-Specific CD8+ T-Cells Is a Correlate of HIV Elite Control. PLOS Pathog [Internet]. 2010 May 27;6(5):e1000917. Available from: https://doi.org/10.1371/journal.ppat.1000917

234. Turk Gabriela, Ghiglione Yanina, Falivene Juliana, Socias María Eugenia, Laufer Natalia, Coloccini Romina Soledad, et al. Early Gag Immunodominance of the HIV-Specific T-Cell Response during Acute/Early Infection Is Associated with Higher CD8+ T-Cell Antiviral Activity and Correlates with Preservation of the CD4+ T-Cell Compartment. J Virol [Internet]. 2013 Jul 1 [cited 2024 Apr 7];87(13):7445–62. Available from: https://doi.org/10.1128/jvi.00865-13

235. Betts MR, Nason MC, West SM, De Rosa SC, Migueles SA, Abraham J, et al. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. Blood [Internet]. 2006 Jun 15 [cited 2024 Jul 5];107(12):4781–9. Available from: https://doi.org/10.1182/blood-2005-12-4818

236.   Huang S, Dunkley-Thompson J, Tang Y, Macklin EA, Steel-Duncan J, Singh-Minott I, et al. Deficiency of HIV-Gag-Specific T Cells in Early Childhood Correlates with Poor Viral Containment1. J Immunol [Internet]. 2008 Dec 1 [cited 2024 Jul 5];181(11):8103–11. Available from: https://doi.org/10.4049/jimmunol.181.11.8103

237.   Betts Michael R., Ambrozak David R., Douek Daniel C., Bonhoeffer Sebastian, Brenchley Jason M., Casazza Joseph P., et al. Analysis of Total Human Immunodeficiency Virus (HIV)-Specific CD4+ and CD8+ T-Cell Responses: Relationship to Viral Load in Untreated HIV Infection. J Virol [Internet]. 2001 Dec 15 [cited 2024 Apr 7];75(24):11983–91. Available from: https://doi.org/10.1128/jvi.75.24.11983-11991.2001

238.   Reeves DB, Gaebler C, Oliveira TY, Peluso MJ, Schiffer JT, Cohn LB, et al. Impact of misclassified defective proviruses on HIV reservoir measurements. Nat Commun [Internet]. 2023 Jul 13;14(1):4186. Available from: https://doi.org/10.1038/s41467-023-39837-z

239.   Imamichi H, Dewar RL, Adelsberger JW, Rehm CA, O'Doherty U, Paxinos EE, et al. Defective HIV-1 proviruses produce novel protein-coding RNA species in HIV-infected patients on combination antiretroviral therapy. Proc Natl Acad Sci [Internet]. 2016 Aug 2 [cited 2024 Apr 6];113(31):8783–8. Available from: https://doi.org/10.1073/pnas.1609057113

240.   Imamichi H, Smith M, Adelsberger JW, Izumi T, Scrimieri F, Sherman BT, et al. Defective HIV-1 proviruses produce viral proteins. Proc Natl Acad Sci [Internet]. 2020 Feb 18 [cited 2024 Apr 6];117(7):3704–10. Available from: https://doi.org/10.1073/pnas.1917876117

241.   Zhao L, Lythgoe KA. The social role of defective viral genomes in chronic viral infections: a commentary on Leeks et al. 2023. J Evol Biol [Internet]. 2023 Nov 1 [cited 2024 Jun 4];36(11):1577–81. Available from: https://doi.org/10.1111/jeb.14244

242.   Herrera A, Jones RB. Whack-a-virus: HIV-specific T cells play an exhausting game. Cell Host Microbe [Internet]. 2023 Sep 13;31(9):1427–30. Available from: https://www.sciencedirect.com/science/article/pii/S1931312823003384

Annexes

Annexe 1. Example d'un fichier « Analyzed_QCTool.csv » produit à partir du jeu de données de Dubé et al. (2023)

SeqName	StopCodons	stop_comments	IncompleteCodons	Hypermutation
PWH3_L1p1_A6_C0_433	0		1	NotDetected
PWH3_L1p1_B4_C0_392	1	Pol	0	NotDetected
PWH3_L1p1_B5_C0_422	0		0	NotDetected
PWH3_L1p1_C6_C0_390	0		0	NotDetected
PWH3_L1p1_D8_C0_366	0		0	NotDetected
PWH3_L1p1_D9_C0_436	0		0	NotDetected
PWH3_L1p1_E10_C0_430	1	Pol	2	NotDetected
PWH3_L1p1_E12_C0_413	3	Tat, Tat2, Nef	4	NotDetected
PWH3_L1p1_F3_C0_418	2	Tat, Tat2	5	NotDetected
PWH3_L1p1_F7_C0_350	1	Pol	0	NotDetected
PWH3_L1p1_G1_C0_430	0		0	NotDetected
PWH3_L1p1_G3_C0_365	1	Pol	3	NotDetected
PWH3_L1p1_H7_C0_263	0		0	NotDetected
PWH3_L1p2_A10_C0_42	0		0	NotDetected
PWH3_L1p2_B10_C0_371	0		0	NotDetected
PWH3_L1p2_C11_C0_329	1	Pol	2	NotDetected
PWH3_L1p2_C2_C0_366	0		0	NotDetected
PWH3_L1p2_C8_C0_432	0		0	NotDetected
PWH3_L1p2_D12_C0_394	0		0	NotDetected
PWH3_L1p2_D3_C0_399	0		0	NotDetected
PWH3_L1p2_F2_C0_412	12	Gag, Pol, Vif, Vpr, Tat, Tat1, Tat2, Rev, Rev2, Vpu, Env, Nef	5	Possible
PWH3_L1p2_G5_C0_392	0		0	NotDetected
PWH3_L1p2_H12_C0_390	0		0	NotDetected
PWH3_L1p2_H4_C0_378	0		0	NotDetected
PWH3_L1p3_A3_C0_418	0		0	NotDetected
PWH3_L1p3_A5_C0_417	0		0	NotDetected
PWH3_L1p3_B11_C0_402	0		0	NotDetected
PWH5_L2p1_A10_C0_341	0		0	NotDetected
PWH5_L2p1_B4_C0_391	0		2	NotDetected
PWH5_L2p1_B7_C0_325	0		0	NotDetected
PWH5_L2p1_C10_C0_473	3	Pol, Tat, Tat2	4	NotDetected
PWH5_L2p1_C12_PCR2_C12_C0_440	2	Tat, Tat2	5	NotDetected
PWH5_L2p1_C6_PCR2_C6_C0_474	1	Gag	1	NotDetected
PWH5_L2p2_D10_C0_432	2	Tat, Tat2	4	NotDetected
PWH5_L2p2_E10_C0_386	0		0	NotDetected
PWH5_L2p2_E3_C0_463	2	Tat, Tat2	4	NotDetected
PWH5_L2p2_E9_PCR2_E9_C0_454	2	Tat, Tat2	4	NotDetected
PWH5_L2p2_F6_C0_290	0		0	NotDetected
PWH5_L2p3_A9_PCR2_G9_C0_455	0		0	NotDetected
PWH9_L3p1_A2_C0_434	0		1	NotDetected

Annexe 2. Example d'un fichier « Analyzed_GeneCutter.csv » produit à partir du jeu de données de Dubé et al. (2023)

Name	start_codon	stop_codon
PWH3_L1p1_A6_C0_433	Gag	
PWH3_L1p1_B4_C0_392	Gag	Pol, Nef
PWH3_L1p1_B5_C0_422	Gag	Nef
PWH3_L1p1_C6_C0_390	Gag	Nef
PWH3_L1p1_D8_C0_366	Gag	Nef
PWH3_L1p1_D9_C0_436	Gag	Nef
PWH3_L1p1_E10_C0_430	Gag	Pol, Nef
PWH3_L1p1_E12_C0_413	Gag, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Tat, Tat2, Nef
PWH3_L1p1_F3_C0_418	Gag, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Tat, Tat2
PWH3_L1p1_F7_C0_350	Gag	Pol, Nef
PWH3_L1p1_G1_C0_430	Gag	
PWH3_L1p1_G3_C0_365	Gag, Vif, Vpr, Tat, Rev, Vpu, Env	Pol, Nef
PWH3_L1p1_H7_C0_263	Gag	Nef
PWH3_L1p2_A10_C0_42	Gag	Nef
PWH3_L1p2_B10_C0_371	Gag	Nef
PWH3_L1p2_C11_C0_329	Gag	Pol, Nef
PWH3_L1p2_C2_C0_366	Gag, Vif	
PWH3_L1p2_C8_C0_432	Gag	Nef
PWH3_L1p2_D12_C0_394	Gag	Nef
PWH3_L1p2_D3_C0_399	Gag, Nef	Pol, Nef
PWH3_L1p2_F2_C0_412	Rev, Vpu	Gag, Pol, Vif, Vpr, Tat, Tat1, Tat2, Rev, Rev2, Vpu, Env, Nef
PWH3_L1p2_G5_C0_392	Gag	Nef
PWH3_L1p2_H12_C0_390	Gag	
PWH3_L1p2_H4_C0_378	Gag	Nef
PWH3_L1p3_A3_C0_418	Gag	Nef
PWH3_L1p3_A5_C0_417	Gag	Nef
PWH3_L1p3_B11_C0_402	Gag	Nef
PWH5_L2p1_A10_C0_341	Gag	
PWH5_L2p1_B4_C0_391	Gag, Vif, Vpr, Tat, Rev, Vpu, Env	
PWH5_L2p1_B7_C0_325	Gag	
PWH5_L2p1_C10_C0_473	Gag, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Pol, Tat, Tat2
PWH5_L2p1_C12_PCR2_C12_C0_440	Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Tat, Tat2
PWH5_L2p1_C6_PCR2_C6_C0_474	Gag	Gag
PWH5_L2p2_D10_C0_432	Gag, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Tat, Tat2
PWH5_L2p2_E10_C0_386	Gag	
PWH5_L2p2_E3_C0_463	Gag, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Tat, Tat2
PWH5_L2p2_E9_PCR2_E9_C0_454	Gag, Vif, Vpr, Tat, Rev, Vpu, Env, Nef	Tat, Tat2
PWH5_L2p2_F6_C0_290	Gag	
PWH5_L2p3_A9_PCR2_G9_C0_455	Gag, Nef	
PWH9_L3p1_A2_C0_434	Gag, Nef	Nef

Annexe 3. Example d'un fichier « Analyzed_ProseqIT.csv » produit à partir du jeu de données de Dubé et al. (2023)

Name	seq_length	Large_inter_delet	Psi_defects	Psi_defects_comments	Gag_defects
PWH3_L1p1_A6_C0_433	2053	1	0		0
PWH3_L1p1_B4_C0_392	4688	1	0		0
PWH3_L1p1_B5_C0_422	1804	1	0		1
PWH3_L1p1_C6_C0_390	2821	1	0		0
PWH3_L1p1_D8_C0_366	4634	1	0		0
PWH3_L1p1_D9_C0_436	2355	1	0		0
PWH3_L1p1_E10_C0_430	4939	1	0		1
PWH3_L1p1_E12_C0_413	8927	0	1	msd_status, package_deletion	0
PWH3_L1p1_F3_C0_418	6539	1	0		0
PWH3_L1p1_F7_C0_350	4688	1	0		0
PWH3_L1p1_G1_C0_430	2797	1	0		0
PWH3_L1p1_G3_C0_365	7006	1	0		0
PWH3_L1p1_H7_C0_263	2337	1	0		0
PWH3_L1p2_A10_C0_42	4694	1	0		0
PWH3_L1p2_B10_C0_371	5030	1	0		0
PWH3_L1p2_C11_C0_329	4939	1	0		1
PWH3_L1p2_C2_C0_366	4736	1	0		0
PWH3_L1p2_C8_C0_432	4693	1	0		0
PWH3_L1p2_D12_C0_394	4693	1	0		0
PWH3_L1p2_D3_C0_399	5050	1	0		0
PWH3_L1p2_F2_C0_412	8960	0	0		1
PWH3_L1p2_G5_C0_392	3140	1	0		0
PWH3_L1p2_H12_C0_390	1336	1	0		1
PWH3_L1p2_H4_C0_378	4693	1	0		0
PWH3_L1p3_A3_C0_418	4693	1	0		0
PWH3_L1p3_A5_C0_417	4694	1	0		0
PWH3_L1p3_B11_C0_402	4965	1	0		0
PWH5_L2p1_A10_C0_341	4160	1	0		0
PWH5_L2p1_B4_C0_391	7400	1	0		0
PWH5_L2p1_B7_C0_325	2772	1	0		0
PWH5_L2p1_C10_C0_473	8989	0	1	msd_status, package_deletion	0
PWH5_L2p1_C12_PCR2_C12_C0_440	8915	0	1	msd_status, package_deletion	1
PWH5_L2p1_C6_PCR2_C6_C0_474	1408	1	0		1
PWH5_L2p2_D10_C0_432	8909	0	1	msd_status, package_deletion	0
PWH5_L2p2_E10_C0_386	2899	1	0		0
PWH5_L2p2_E3_C0_463	8997	0	1	msd_status, package_deletion	0
PWH5_L2p2_E9_PCR2_E9_C0_454	8909	0	1	msd_status, package_deletion	0
PWH5_L2p2_F6_C0_290	1399	1	0		1
PWH5_L2p3_A9_PCR2_G9_C0_455	1953	1	0		1
PWH9_L3p1_A2_C0_434	2568	1	0		1

Gag_defects_comments	Pol_defects	Pol_defects_comments	Vif_defects	Vif_defects_comments
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion, pol_frameshift, pol_stop_codon	1	StartCodon, vif_deletion
gag_deletion	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
gag_frameshift	1	pol_deletion, pol_frameshift, pol_stop_codon	1	StartCodon, vif_deletion
	0		0	
	1	pol_deletion, pol_frameshift	0	
	1	pol_deletion, pol_frameshift, pol_stop_codon	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_frameshift, pol_stop_codon	0	
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
gag_frameshift	1	pol_deletion, pol_frameshift, pol_stop_codon	1	StartCodon, vif_deletion
	0		1	vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion, pol_stop_codon	1	StartCodon, vif_deletion
StartCodon, gag_stop_codon	1	gag_start_codon, pol_stop_codon	1	StartCodon, vif_frameshift, vif_stop_codon
	1	pol_deletion	1	StartCodon, vif_deletion
gag_deletion	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_deletion	1	StartCodon, vif_deletion
	0		0	
	1	pol_deletion	1	StartCodon, vif_deletion
	1	pol_stop_codon	0	
StartCodon, U5_gag_pair_R2_deletion, 2base_before_gag_status, gag_frameshift	1	gag_start_codon	0	
gag_deletion, gag_frameshift, gag_stop_codon	1	pol_deletion	1	StartCodon, vif_deletion
	0		0	
	1	pol_deletion	1	StartCodon, vif_deletion
	0		0	
	0		0	
gag_deletion	1	pol_deletion	1	StartCodon, vif_deletion
gag_deletion	1	pol_deletion	1	StartCodon, vif_deletion
gag_deletion	1	pol_deletion	1	StartCodon, vif_deletion

163

Vpr_defects	Vpr_defects_comments	Tat_defects	Tat_defects_comments	Rev_defects	Rev_defects_comments
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion, rev_frameshift
0		1	tat2_stop_codon	0	
0		1	tat2_stop_codon	0	
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
0		1	tat_deletion	1	rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion, rev_frameshift
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_stop_codon	1	StartCodon, tat1_stop_codon, tat2_stop_codon	1	rev_stop_codon
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
0		1	tat_deletion	1	rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
0		1	tat2_stop_codon	0	
0		1	tat2_stop_codon	0	
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
0		1	tat2_stop_codon	0	
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
0		1	tat2_stop_codon	0	
0		1	tat2_stop_codon	0	
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion
1	StartCodon, vpr_deletion	1	StartCodon, tat_deletion	1	StartCodon, rev_deletion, rev_frameshift

Vpu_defects	Vpu_defects_comments	Nef_defects	Nef_defects_comments
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_stop_codon
0		1	nef_stop_codon
0		1	nef_deletion
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
0		1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	nef_stop_codon
1	vpu_stop_codon	1	StartCodon, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion, nef_frameshift, nef_stop_codon
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
0		1	StartCodon, nef_deletion
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
0		0	
0		0	
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
0		0	
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
0		0	
0		0	
1	StartCodon, vpu_deletion	1	StartCodon, nef_deletion
1	StartCodon, vpu_deletion	0	
1	StartCodon, vpu_deletion	1	nef_stop_codon

Env_defects	Env_defects_comments	rre_status
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion, env_frameshift	1
0		0
0		0
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion, env_frameshift	1
1	env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion, env_frameshift	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion, env_frameshift	1
1	StartCodon, env_stop_codon	0
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	env_deletion	1
1	StartCodon, env_deletion	1
0		0
0		0
1	StartCodon, env_deletion	1
0		0
1	StartCodon, env_deletion	1
0		0
0		0
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion	1
1	StartCodon, env_deletion, env_frameshift	1

Annexe 4. Example d'un fichier « intactness_detailedsummary.csv » produit à partir du jeu de données de Dubé et al. (2023)

Name	intactness	n_main_defects	main_defect	defects_comments
PWH3_L1p1_A6_C0_433	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p1_B4_C0_392	defective	3	large_intern_delet	large_intern_delet, stop_codon, small_intern_delet
PWH3_L1p1_B5_C0_422	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p1_C6_C0_390	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p1_D8_C0_366	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p1_D9_C0_436	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p1_E10_C0_430	defective	3	large_intern_delet	large_intern_delet, stop_codon, small_intern_delet
PWH3_L1p1_E12_C0_413	defective	1	psi_defects	psi_defects
PWH3_L1p1_F3_C0_418	defective	3	inversions	inversions, large_intern_delet, small_intern_delet
PWH3_L1p1_F7_C0_350	defective	3	large_intern_delet	large_intern_delet, stop_codon, small_intern_delet
PWH3_L1p1_G1_C0_430	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p1_G3_C0_365	defective	4	inversions	inversions, large_intern_delet, stop_codon, small_intern_delet
PWH3_L1p1_H7_C0_263	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_A10_C0_42	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_B10_C0_371	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_C11_C0_329	defective	3	large_intern_delet	large_intern_delet, stop_codon, small_intern_delet
PWH3_L1p2_C2_C0_366	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_C8_C0_432	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_D12_C0_394	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_D3_C0_399	defective	3	large_intern_delet	large_intern_delet, stop_codon, small_intern_delet
PWH3_L1p2_F2_C0_412	defective	3	hypermutations	hypermutations, stop_codon, small_intern_delet
PWH3_L1p2_G5_C0_392	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_H12_C0_390	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p2_H4_C0_378	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p3_A3_C0_418	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p3_A5_C0_417	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH3_L1p3_B11_C0_402	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH5_L2p1_A10_C0_341	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH5_L2p1_B4_C0_391	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH5_L2p1_B7_C0_325	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH5_L2p1_C10_C0_473	defective	3	stop_codon	stop_codon, psi_defects, small_intern_delet
PWH5_L2p1_C12_PCR2_C12_C0_440	defective	2	psi_defects	psi_defects, small_intern_delet
PWH5_L2p1_C6_PCR2_C6_C0_474	defective	3	large_intern_delet	large_intern_delet, stop_codon, small_intern_delet
PWH5_L2p2_D10_C0_432	defective	1	psi_defects	psi_defects
PWH5_L2p2_E10_C0_386	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH5_L2p2_E3_C0_463	defective	1	psi_defects	psi_defects
PWH5_L2p2_E9_PCR2_E9_C0_454	defective	1	psi_defects	psi_defects
PWH5_L2p2_F6_C0_290	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH5_L2p3_A9_PCR2_G9_C0_455	defective	2	large_intern_delet	large_intern_delet, small_intern_delet
PWH9_L3p1_A2_C0_434	defective	2	large_intern_delet	large_intern_delet, small_intern_delet

seq_length	inversions	hypermutations	large_intern_delet	stop_codon	psi_defects	small_intern_delet
2053	0	0	1	0	0	1
4688	0	0	1	1	0	1
1804	0	0	1	0	0	1
2821	0	0	1	0	0	1
4634	0	0	1	0	0	1
2355	0	0	1	0	0	1
4939	0	0	1	1	0	1
8927	0	0	0	0	1	0
6539	1	0	1	0	0	1
4688	0	0	1	1	0	1
2797	0	0	1	0	0	1
7006	1	0	1	1	0	1
2337	0	0	1	0	0	1
4694	0	0	1	0	0	1
5030	0	0	1	0	0	1
4939	0	0	1	1	0	1
4736	0	0	1	0	0	1
4693	0	0	1	0	0	1
4693	0	0	1	0	0	1
5050	0	0	1	1	0	1
8960	0	1	0	1	0	1
3140	0	0	1	0	0	1
1336	0	0	1	0	0	1
4693	0	0	1	0	0	1
4693	0	0	1	0	0	1
4694	0	0	1	0	0	1
4965	0	0	1	0	0	1
4160	0	0	1	0	0	1
7400	0	0	1	0	0	1
2772	0	0	1	0	0	1
8989	0	0	0	1	1	1
8915	0	0	0	0	1	1
1408	0	0	1	1	0	1
8909	0	0	0	0	1	0
2899	0	0	1	0	0	1
8997	0	0	0	0	1	0
8909	0	0	0	0	1	0
1399	0	0	1	0	0	1
1953	0	0	1	0	0	1
2568	0	0	1	0	0	1

n_small_inter_delet	Gag_defects	Pol_defects	Vif_defects	Vpr_defects	Tat_defects	Rev_defects	Vpu_defects
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
10	1	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
10	1	1	1	1	1	1	1
2	0	0	0	0	1	0	0
3	0	1	0	0	1	0	0
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
6	0	1	0	0	1	1	0
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
10	1	1	1	1	1	1	1
8	0	0	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	1	1	1	1	1	1	1
9	0	1	1	1	1	1	1
10	1	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
9	0	1	1	1	1	1	1
5	0	0	0	0	1	1	0
9	0	1	1	1	1	1	1
2	0	1	0	0	1	0	0
3	1	1	0	0	1	0	0
10	1	1	1	1	1	1	1
1	0	0	0	0	1	0	0
9	0	1	1	1	1	1	1
1	0	0	0	0	1	0	0
1	0	0	0	0	1	0	0
10	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1

Nef_defects	Env_defects	rre_status
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	0	0
1	0	0
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	0
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
0	0	0
0	0	0
1	1	1
0	0	0
1	1	1
0	0	0
0	0	0
1	1	1
0	1	1
1	1	1

Annexe 5. Example d'un fichier « intactness_summary.csv » produit à partir du jeu de données de Dubé et al. (2023)

Name	intactness	seq_length	n_main_defects	inversions	hypermutations
PWH3_L1p1_A6_C0_433	defective	2053	2	0	0
PWH3_L1p1_B4_C0_392	defective	4688	3	0	0
PWH3_L1p1_B5_C0_422	defective	1804	2	0	0
PWH3_L1p1_C6_C0_390	defective	2821	2	0	0
PWH3_L1p1_D8_C0_366	defective	4634	2	0	0
PWH3_L1p1_D9_C0_436	defective	2355	2	0	0
PWH3_L1p1_E10_C0_430	defective	4939	3	0	0
PWH3_L1p1_E12_C0_413	defective	8927	1	0	0
PWH3_L1p1_F3_C0_418	defective	6539	3	1	0
PWH3_L1p1_F7_C0_350	defective	4688	3	0	0
PWH3_L1p1_G1_C0_430	defective	2797	2	0	0
PWH3_L1p1_G3_C0_365	defective	7006	4	1	0
PWH3_L1p1_H7_C0_263	defective	2337	2	0	0
PWH3_L1p2_A10_C0_42	defective	4694	2	0	0
PWH3_L1p2_B10_C0_371	defective	5030	2	0	0
PWH3_L1p2_C11_C0_329	defective	4939	3	0	0
PWH3_L1p2_C2_C0_366	defective	4736	2	0	0
PWH3_L1p2_C8_C0_432	defective	4693	2	0	0
PWH3_L1p2_D12_C0_394	defective	4693	2	0	0
PWH3_L1p2_D3_C0_399	defective	5050	3	0	0
PWH3_L1p2_F2_C0_412	defective	8960	3	0	1
PWH3_L1p2_G5_C0_392	defective	3140	2	0	0
PWH3_L1p2_H12_C0_390	defective	1336	2	0	0
PWH3_L1p2_H4_C0_378	defective	4693	2	0	0
PWH3_L1p3_A3_C0_418	defective	4693	2	0	0
PWH3_L1p3_A5_C0_417	defective	4694	2	0	0
PWH3_L1p3_B11_C0_402	defective	4965	2	0	0
PWH5_L2p1_A10_C0_341	defective	4160	2	0	0
PWH5_L2p1_B4_C0_391	defective	7400	2	0	0
PWH5_L2p1_B7_C0_325	defective	2772	2	0	0
PWH5_L2p1_C10_C0_473	defective	8989	3	0	0
PWH5_L2p1_C12_PCR2_C12_C0_440	defective	8915	2	0	0
PWH5_L2p1_C6_PCR2_C6_C0_474	defective	1408	3	0	0
PWH5_L2p2_D10_C0_432	defective	8909	1	0	0
PWH5_L2p2_E10_C0_386	defective	2899	2	0	0
PWH5_L2p2_E3_C0_463	defective	8997	1	0	0
PWH5_L2p2_E9_PCR2_E9_C0_454	defective	8909	1	0	0
PWH5_L2p2_F6_C0_290	defective	1399	2	0	0
PWH5_L2p3_A9_PCR2_G9_C0_455	defective	1953	2	0	0
PWH9_L3p1_A2_C0_434	defective	2568	2	0	0

large_intern_delet	stop_codon	psi_defects	small_intern_delet	Gag_defects	Pol_defects	Vif_defects
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	1	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	1	1	1
0	0	1	0	0	0	0
0	0	0	0	0	1	0
1	0	0	0	0	1	1
1	0	0	0	0	1	1
0	0	0	0	0	1	0
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	1	1	1
1	0	0	0	0	0	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
0	0	0	0	1	1	1
1	0	0	0	0	1	1
1	0	0	0	1	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	1	1
1	0	0	0	0	0	0
1	0	0	0	0	1	1
0	1	0	0	0	1	0
0	0	1	0	1	1	0
1	0	0	0	1	1	1
0	0	1	0	0	0	0
1	0	0	0	0	1	1
0	0	1	0	0	0	0
0	0	1	0	0	0	0
1	0	0	0	1	1	1
1	0	0	0	1	1	1
1	0	0	0	1	1	1

Vpr_defects	Tat_defects	Rev_defects	Vpu_defects	Nef_defects	Env_defects	rre_status
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
0	1	0	0	1	0	0
0	1	0	0	1	0	0
1	1	1	1	1	1	1
1	1	1	1	1	1	1
0	1	1	0	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	0
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
1	1	1	1	1	1	1
0	1	1	0	1	1	1
1	1	1	1	1	1	1
0	1	0	0	0	0	0
0	1	0	0	0	0	0
1	1	1	1	1	1	1
0	1	0	0	0	0	0
1	1	1	1	1	1	1
0	1	0	0	0	0	0
0	1	0	0	0	0	0
1	1	1	1	1	1	1
1	1	1	1	0	1	1
1	1	1	1	1	1	1

Annexe 6. Example d'un fichier « PWH3_ClonalityAnalysis.csv » produit à partir du jeu de données de Dubé et al. (2023)

unique_sequence	nclones	clones	npotential_clones	potential_clones
PWH3_L1p1_A6_C0_433	0		0	
PWH3_L1p1_B4_C0_392	1	PWH3_L1p1_F7_C0_350	0	
PWH3_L1p1_B5_C0_422	0		0	
PWH3_L1p1_C6_C0_390	0		0	
PWH3_L1p1_D8_C0_366	0		0	
PWH3_L1p1_D9_C0_436	0		0	
PWH3_L1p1_E10_C0_430	1	PWH3_L1p2_C11_C0_329	0	
PWH3_L1p1_E12_C0_413	0		0	
PWH3_L1p1_F3_C0_418	0		0	
PWH3_L1p1_G1_C0_430	0		0	
PWH3_L1p1_G3_C0_365	0		0	
PWH3_L1p1_H7_C0_263	0		0	
PWH3_L1p2_A10_C0_42	0		0	
PWH3_L1p2_B10_C0_371	0		0	
PWH3_L1p2_C2_C0_366	0		0	
PWH3_L1p2_C8_C0_432	0		0	
PWH3_L1p2_D12_C0_394	2	PWH3_L1p2_H4_C0_378, PWH3_L1p3_A3_C0_418	0	
PWH3_L1p2_D3_C0_399	0		0	
PWH3_L1p2_F2_C0_412	0		0	
PWH3_L1p2_G5_C0_392	0		0	
PWH3_L1p2_H12_C0_390	0		0	
PWH3_L1p3_A5_C0_417	0		0	
PWH3_L1p3_B11_C0_402	0		0	

Annexe 7. Tableaux supplémentaires de la section 4.3 « *Methods* »

Table 4.3.S1 – Clinical characteristics of the study participants.[a]

	Elite controllers n = 12	Pre-ART participants n = 12	Post-ART participants n = 12
Age			
Age (years old)	49 (44 – 57.5)	38 (32 – 40.5)	40.5 (35 – 42.25)
Sex			
Male	8 (67%)	11 (92%)	11 (92%)
Female	4 (33%)	1 (8%)	1 (8%)
HIV infection			
Duration of infection (yrs)	16.71 (6.15 – 19.1)	0.4 (0.19 – 6.01)	3.28 (2.45 – 7.84)
Time before ART (yrs)	N/A	N/A	0.90 (0.50 – 6.69)
Time on ART (yrs)	N/A	N/A	1.25 (1.15 – 2.13)
CD4 counts	646 (597.75 – 690)	446 (315 – 688.25)	655 (561.5 – 731.25)
CD8 counts	589.5 (466.25 – 848.75)	967 (686.5 – 1463.5)	715 (498.75 – 957.75)
CD4/CD8 ratio	1.13 (0.86 – 1.26)	0.52 (0.29 – 0.82)	1.16 (0.58 – 1.39)
Log VL (copies/mL)	<1.70	4.56 (3.84 – 5.00)	<1.60
CMV infection			
Positive	7 (58%)	12 (100%)	12 (100%)
Negative	N/A	0 (0%)	0 (0%)
Missing information	5 (42%)	N/A	N/A

[a] Values displayed are medians. Interquartile ranges (continuous variables) or percentages (categorical variables) are shown in parentheses.

Table 4.3.S2 – Flow cytometry antibody staining panel for the activation-induced marker (AIM) assays.

| Marker – Fluorophore	Clone	Source	Identifier
CD4 – Alexa Fluor 488	RPA-T4	BD Biosciences	Cat#557695; Lot: 3326999; RRID: AB_396804
CD8 – APC-Fire 750	SK1	Biolegend	Cat#344746; Lot: B359145; RRID: AB_2572095
CD14 – BV510	MφP9	BD Biosciences	Cat#563079; Lot: 3240065; RRID: AB_2737993
CD16 – BV510	3G8	BD Biosciences	Cat#563830; Lot: 3124250; RRID: AB_2938676
CD19 – BV510	H1B19	Biolegend	Cat#302242; Lot: B399472; RRID: AB_2561668
CD69 – PerCP-eFluor 710	FN50	eBioscience	Cat#46-0699-42; Lot: 2778230, 2823740; RRID: AB_2573694
CD134 (OX40) – APC	ACT35	BD Biosciences	Cat#563473; Lot: 3131767; RRID: AB_2738230
CD137 (4-1BB) – PE	4B4-1	Biolegend	Cat#309804; Lot: B367290; RRID: AB_314783
CD279 (PD1) – BV421	EH12.2H7	Biolegend	Cat#329920; Lot: B398160; RRID: AB_10960742
LIVE/DEAD Fixable dead cell	N/A	Thermo Fisher Scientific	Cat#L34966

# Annexe 8. Figures et tableaux supplémentaires de la section 4.4 « *Results* »
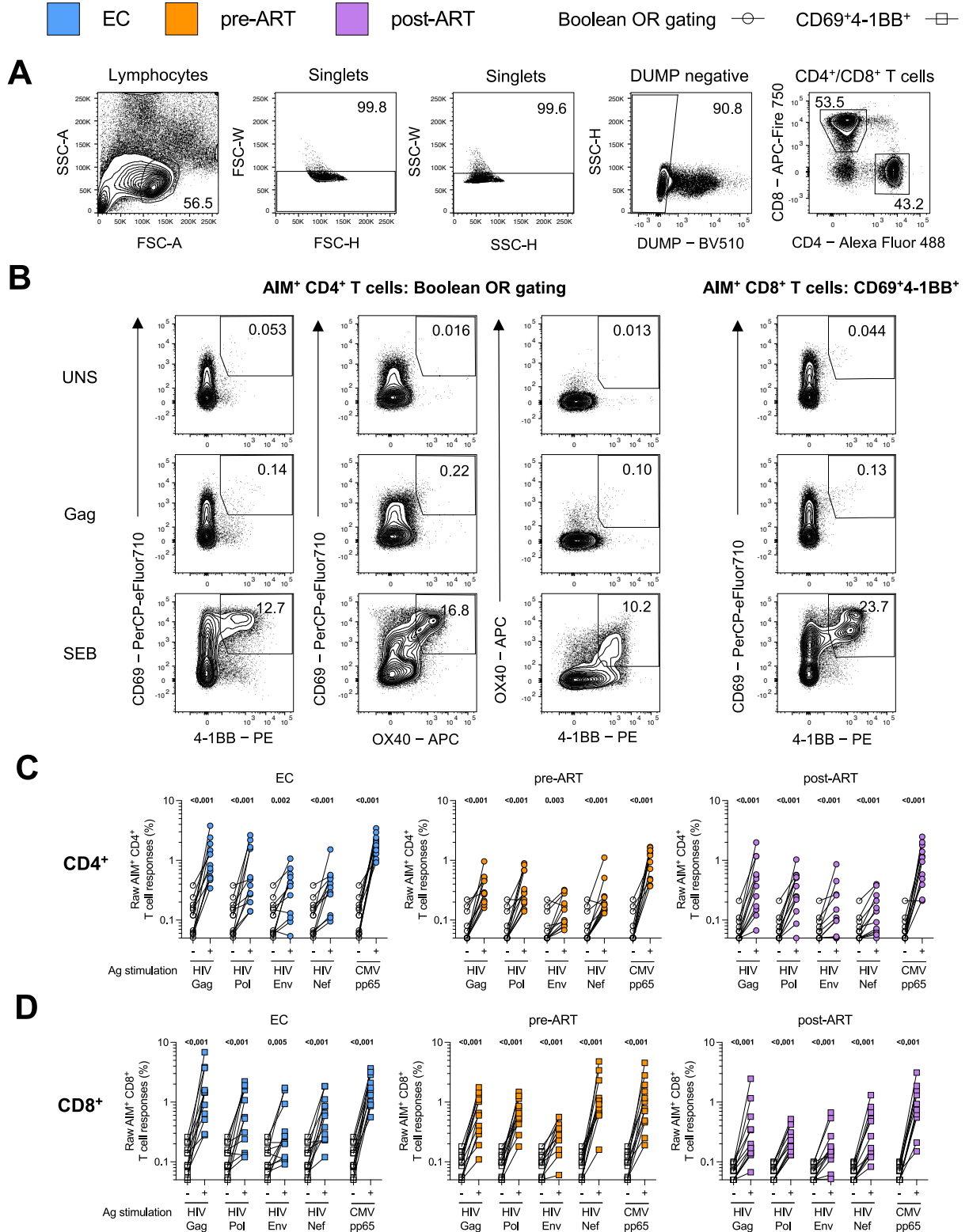
Figure 4.4.S1 – Characterization of Ag-specific CD4$^+$ and CD8$^+$ T cells in ECs and pre- and post-ART participants. Related to Figure 4.4.1. (**A**) Representative upstream generic gating. (**B**) FACS plots depicting the strategies to identify Ag-specific CD4$^+$ and CD8$^+$ T cells. A DMSO-treated (unstimulated) condition and a *Staphylococcus enterotoxin B* (SEB)-treated condition served, respectively, as negative and positive controls. For simplicity, the example focuses on the HIV Gag stimulation. (**C and D**) Raw frequencies of AIM$^+$ (**C**) CD4$^+$ and (**D**) CD8$^+$ T cell responses following an *ex vivo* 15 h stimulation of PBMCs with a pool of HIV Gag, Pol, Envelope (Env), Nef, or HCMVA pp65 peptides. PBMCs were left unstimulated as a control (clear dots). The AIM assay was performed on a cohort of elite controllers (EC; blue) and a cohort of pre- (orange) and post- (purple) ART participants. Wilcoxon tests are shown. Circles represent CD4$^+$ responses detected with the Boolean OR gating strategy, while squares represent CD8$^+$ T cells detected with the CD69$^+$4-1BB$^+$ pair. n = 12 ECs and n = 12 pre- and post-ART participants.
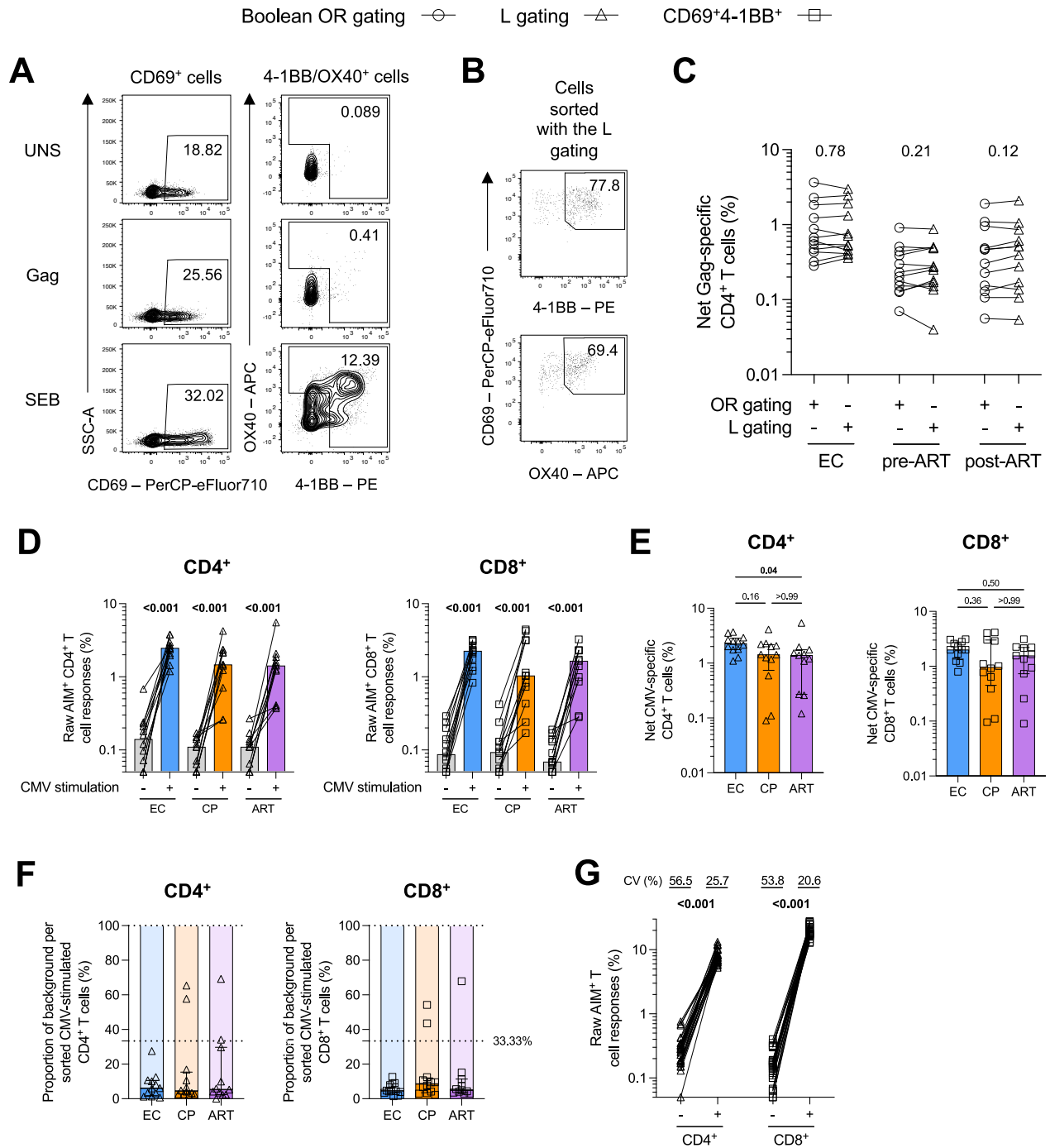
Figure 4.4.S2 – Live sorting of Ag-specific CD4⁺ and CD8⁺ T cells in ECs and pre- and post-ART participants. Related to Figure 4.4.2. (**A**) FACS plots depicting the L gating strategy to sort Ag-specific CD4⁺ T cells. A DMSO-treated (unstimulated) condition and a *Staphylococcus enterotoxin B* (SEB)-treated condition served, respectively, as negative and positive controls. For simplicity, the example focuses on the HIV Gag stimulation. For both CD69⁺ and 4-1BB⁺/OX40⁺ cells, values shown are expressed in percentages of CD4⁺ T cells. (**B**) Proportion of L gating-sorted CD4⁺ T cells detected by standard AIM pairs. (**C**)

Comparison between the net Ag-specific CD4$^+$ T cell responses detected by the Boolean OR gating vs. the L gating. Wilcoxon tests are shown. For simplicity, the example focuses on Gag responses. (**D**) Raw and (**E**) net frequencies of sorted AIM$^+$ CD4$^+$ and CD8$^+$ T cell responses in elite controllers (blue), pre-ART (orange), and post-ART (purple) participants, following an *ex vivo* 15 h stimulation of PBMCs with a pool of HCVMA pp65 peptides. PBMCs were left unstimulated as a control (grey bars). (**D**) Wilcoxon tests and (**E**) Kruskal-Wallis tests with Dunn's correction are shown. (**F**) Proportion of background in the CMV-stimulated condition. The Ag-stimulated conditions are normalized to 100% to better represent the autologous unstimulated signal. Light colors represent the total of stimulated cells; dark colors represent the background. (**G**) Raw AIM$^+$ CMV-stimulated CD4$^+$ and CD8$^+$ T cell responses of a single participant repeated at every sort. Coefficients of variation (CV) are shown above the graph. n = 30 sorts. (**C** – **G**), circles represent CD4$^+$ responses detected with the Boolean OR gating strategy, triangles represent CD4$^+$ responses detected with the L gating strategy, and squares represent CD8$^+$ T cells detected with the CD69$^+$4-1BB$^+$ pair. n = 12 ECs and n = 12 pre- and post-ART participants.
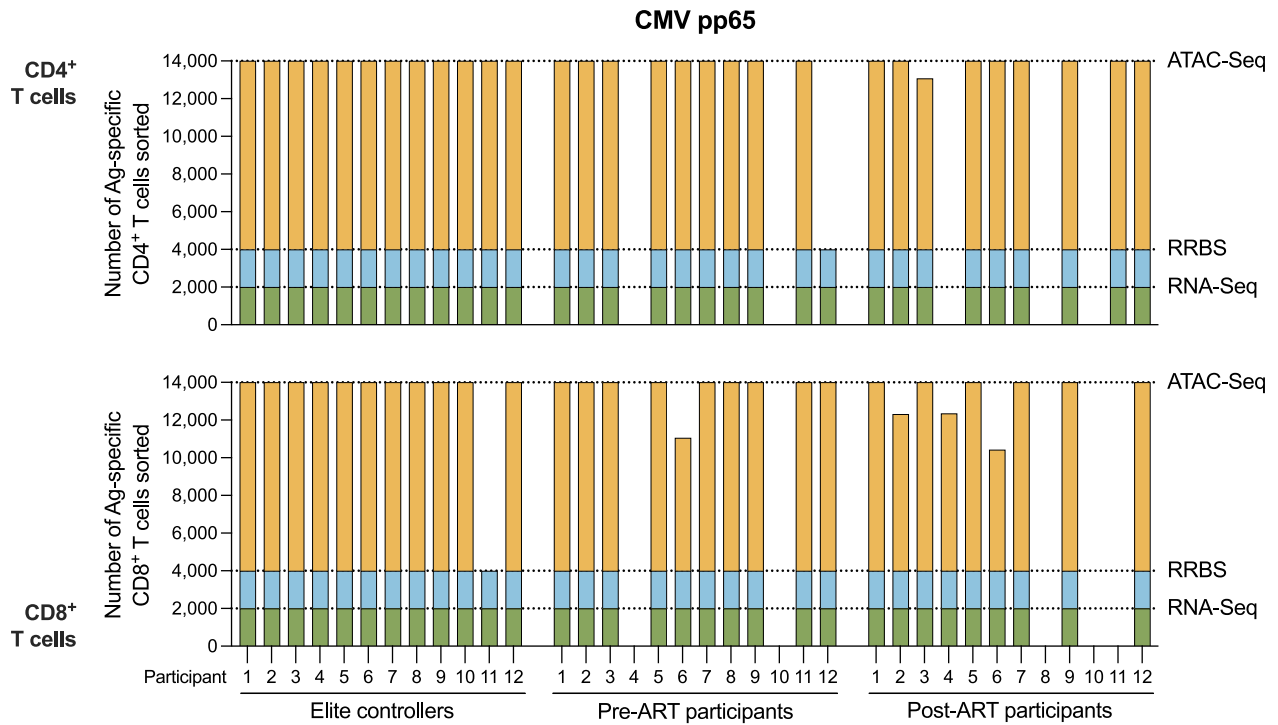
Figure 4.4.S3 –Summary of CMV pp65-specific CD4$^+$ and CD8$^+$ T cells sorted in ECs and pre- and post-ART participants. Related to Figure 4.4.3. Each colour represents the cells sorted for the different assays: green for the RNA-Seq, blue for the RRBS, and yellow for the ATAC-Seq. n = 12 ECs and n = 12 pre- and post-ART participants. RNA-Seq: RNA sequencing; RRBS: Reduced representation bisulfite sequencing; ATAC-Seq: Assay for transposase-accessible chromatin using sequencing.
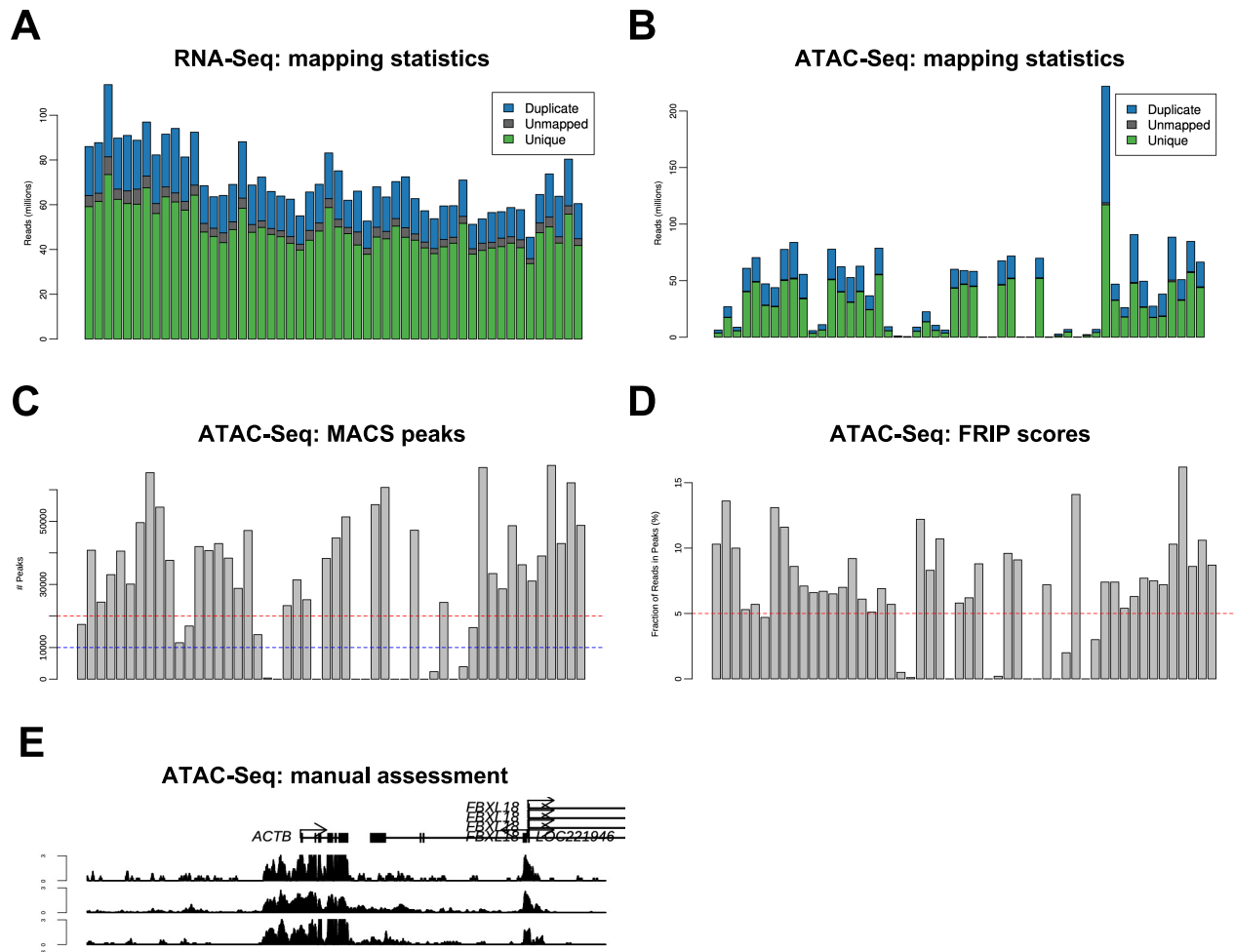
Figure 4.4.S4 – Quality control of the sorted cells. Related to Figure 4.4.4. Mapping statistics of the (**A**) RNA-Seq and (**B**) ATAC-Seq data. Number of (**C**) MACS peaks and (**D**) FRIP scores for the quality control of the ATAC-Seq data. (**E**) Examples of the manual assessment done on the Beta-actin (ATCB) gene. Peaks represent regions with enriched accessible chromatin. (**A and B**), blue, black, and green bars represent, respectively, the duplicated, unmapped, and unique reads. (**A – D**), n = 52 samples for each assay. All the graphs have been provided by the laboratory of Dr. Boss (Emory University, Atlanta, USA). RNA-Seq: RNA sequencing; RRBS: Reduced representation bisulfite sequencing; ATAC-Seq: Assay for transposase-accessible chromatin using sequencing.

Table 4.4.S1 – Number of cells sorted per participant. Samples from a single participant with high CMV responses (identified as "repeated participant") were repeated at every sort. n = 12 ECs, n = 12 pre- and post-ART participants, and n = 30 samples from the repeated participant.

Participant	Specificity	RNA CD4$^+$	RNA CD8$^+$	DNA CD4$^+$	DNA CD8$^+$	Tagmentation CD4$^+$	Tagmentation CD8$^+$
EC 1	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 2	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 3	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 4	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 5	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 6	HIV Gag	2,000	2,000	2,000	2,000	7,222	10,000
EC 7	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 8	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 9	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 10	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
EC 11	HIV Gag	2,000	2,000	2,000	2,000	10,000	N/A[d]
EC 12	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 1	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 2	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 3	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 4	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 5	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 6	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 7	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 8	HIV Gag	2,000	N/A[a]	2,000	N/A[a]	10,000	N/A[a]
Pre-ART 9	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 10	HIV Gag	N/A[a]	2,000	N/A[a]	2,000	N/A[a]	10,000
Pre-ART 11	HIV Gag	2,000	2,000	2,000	2,000	8,349	10,000
Pre-ART 12	HIV Gag	2,000	2,000	2,000	2,000	N/A[d]	10,000
Post-ART 1	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 2	HIV Gag	2,000	2,000	2,000	2,000	10,000	5,524
Post-ART 3	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 4	HIV Gag	2,000	2,000	2,000	2,000	10,000	7,544
Post-ART 5	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 6	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 7	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 8	HIV Gag	N/A[a]	N/A[a]	N/A[a]	N/A[a]	N/A[a]	N/A[a]
Post-ART 9	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 10	HIV Gag	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 11	HIV Gag	2,000	N/A[a]	2,000	N/A[a]	10,000	N/A[a]
Post-ART 12	HIV Gag	2,000	2,000	1,686	2,000	10,000	10,000
EC 1	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 2	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 3	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 4	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 5	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 6	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000

EC 7	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 8	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 9	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 10	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
EC 11	CMV pp65	2,000	2,000	2,000	2,000	10,000	N/A[d]
EC 12	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 1	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 2	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 3	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 4	CMV pp65	2,000	2,000	2,000	2,000	N/A[a]	N/A[a]
Pre-ART 5	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 6	CMV pp65	2,000	2,000	2,000	2,000	10,000	7,055
Pre-ART 7	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 8	CMV pp65	2,000	2,000	2,000	2,000	N/A[a]	N/A[a]
Pre-ART 9	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 10	CMV pp65	2,000	2,000	2,000	2,000	N/A[b]	N/A[a]
Pre-ART 11	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Pre-ART 12	CMV pp65	2,000	2,000	2,000	2,000	N/A[b]	10,000
Post-ART 1	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 2	CMV pp65	2,000	2,000	2,000	2,000	10,000	8,318
Post-ART 3	CMV pp65	2,000	2,000	2,000	2,000	9,077	10,000
Post-ART 4	CMV pp65	2,000	2,000	2,000	2,000	NA[a]	8,352
Post-ART 5	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 6	CMV pp65	2,000	2,000	2,000	2,000	10,000	6,420
Post-ART 7	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 8	CMV pp65	2,000	2,000	2,000	2,000	N/A[a]	N/A[a]
Post-ART 9	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 10	CMV pp65	2,000	2,000	2,000	2,000	N/A[a]	N/A[a]
Post-ART 11	CMV pp65	2,000	2,000	2,000	2,000	10,000	N/A[b]
Post-ART 12	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Post-ART 10	Unstimulated	2,000	2,000	2,000	2,000	N/A	N/A
Repeated participant 1	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 2	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 3	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 4	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 5	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 6	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 7	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 8	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 9	CMV pp65	2,000	2,000	2,000	2,000	8,344	10,000
Repeated participant 10	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 11	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 12	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 13	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 14	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 15	CMV pp65	2,000	N/A[c]	2,000	N/A[c]	10,000	N/A[c]
Repeated participant 16	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 17	CMV pp65	N/A[c]	N/A[c]	N/A[c]	N/A[c]	10,000	10,000
Repeated participant 18	CMV pp65	2,000	2,000	2,000	2,000	7,131	10,000
Repeated participant 19	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000

Repeated participant 20	CMV pp65	2,000	2,000	2,000	2,000	N/A[c]	N/A[c]
Repeated participant 21	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 22	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 23	CMV pp65	2,000	2,000	2,000	2,000	N/A[c]	10,000
Repeated participant 24	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000
Repeated participant 25	CMV pp65	N/A[c]	N/A[c]	N/A[c]	N/A[c]	10,000	10,000
Repeated participant 26	CMV pp65	N/A[c]	N/A[c]	N/A[c]	N/A[c]	N/A[c]	10,000
Repeated participant 27	CMV pp65	2,000	2,000	2,000	2,000	10,000	10,000

[a] Cells were not sorted due to a high proportion of background (>33.3%), either during the sort or as seen in the screening.

[b] CMV-specific cells were omitted as HIV-specific cells, our primary focus, were not sorted.

[c] CMV-specific cells of the repeated participant were omitted, as HIV- and/or CMV-specific cells of the participant of interest were not sorted.

[d] Will not be sorted. Not enough vials left.

Annexe 9. Contribution de la candidate à des manuscrits supplémentaires

Dans le cadre de cette maîtrise, j'ai effectué les analyses non-supervisées des manuscrits suivants :

- Shaaban Kabakibo, T., Arnold, E., Padhan, K., **Lemieux, A.**, Delgado, G.G., Routy, J.-P., Shoukry, N., Dubé, M., & Kaufmann, D.E. *An artificial antigen-presenting cell system reveals CD28 co-stimulation's role in modulating T cell effector functions during HIV infection*. En révision.

- Nayrac, M., **Lemieux, A.**, Brassard, N., Trifone, C., Fromentin, R., Sannier, G., Nekorchuk, M., Busman-Sahay, K., Terry, M., Paiardini, M., Estes, J.D., Chomont, N., Dubé, M., & Kaufmann, D.E. *Transcriptional and translational SIV profiles of circulating and lymphoid CD4$^+$ T cells of viremic and ART-suppressed Rhesus macaques*. En préparation.

- Sannier, G., **Lemieux, A.**, Delgado, G.G., Cloutier, R., Pagliuzza, A., Duchesne, M., Brassard, N., Tremblay, C., Bernard, N., Routy, J.-P., Chomont, N., Dubé, M., & Kaufmann, D.E. *Single-cell relationships between HIV-1-specific immune responses in untreated individuals and viral transcription and translation*. En préparation.

L'IntegrityAlgorithm développé sera notamment utilisé dans les manuscrits suivants:

- Dufour, C., **Lemieux, A.**, Pagliuzza, A., Fromentin, R., & Chomont, N. *Development and validation of highly effective method to sequence and analyze near-full length HIV genomes*. En préparation.

- Espinosa Ortiz, A., **Lemieux, A.**, Fromentin, R., Massé, C., Routy, J.-P., Falcone, E.L., & Chomont, N. *Role of bacterial antigens in HIV persistence*. En préparation.

- Roux, H., Pagliuzza, A., **Lemieux, A.**, Fromentin, R., Routy, J.-P., & Chomont, N. *Phenotypic, genotypic and functional analyses of HIV-1 reservoir cells and proviruses*. En préparation.

Le jeu de données du projet de la dysfonction des lymphocytes T CD4$^+$ spécifiques au VIH sera entre autres utilisé dans le manuscrit suivant :

- **Lemieux, A.**, Duchesne, M., Rajah, M.M., Hicks, S., Delgado, G.G., Sharer, C., Dubé, M., Boss, J.M., & Kaufmann, D.E. *The molecular basis of persistent HIV-specific CD4 T cell dysfunction*. En préparation.