



HAL
open science

Définition du translatome du VIH-1 dans le but d'identifier de nouveaux antigènes conservés reconnus par les lymphocytes T

Lisa Bertrand

► **To cite this version:**

Lisa Bertrand. Définition du translatome du VIH-1 dans le but d'identifier de nouveaux antigènes conservés reconnus par les lymphocytes T. Maladies infectieuses. Université Paris-Saclay, 2023. Français. NNT : 2023UPASQ026 . tel-04690001

HAL Id: tel-04690001

<https://theses.hal.science/tel-04690001v1>

Submitted on 6 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Définition du translatome du VIH-1 dans le but d'identifier de nouveaux antigènes conservés reconnus par les lymphocytes T

Defining the translatome of HIV-1 to identify novel conserved T-cell antigens

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°569 : innovation thérapeutique : du fondamental à l'appliqué (ITFA)
Spécialité de doctorat : Immunologie
Graduate School : Santé et médicaments. Référent : Faculté de pharmacie

Thèse préparée dans l'unité de recherche **Institute for Integrative Biology of the Cell (I2BC)** (Université Paris-Saclay, CEA, CNRS),
sous la direction de **Arnaud Moris**, DR

Thèse soutenue à Paris-Saclay, le 07 juillet 2023, par

Lisa BERTRAND

Composition du Jury

Membres du jury avec voix délibérative

Bernard MAILLÈRE DR, Université Paris-Saclay	Président
Loredana SAVEANU DR, Université Paris-Cité	Rapportrice et examinatrice
Victor APPAY DR, Université de Bordeaux	Rapporteur et examinateur
Sebastien APCHER CR, Université Paris-Saclay	Examineur
Behazine COMBADIÈRE DR, Sanofi R&D Vaccins, Sorbonne Université	Examinatrice

« Essayer d'aimer *les questions elles-mêmes* [...] *Vivez* maintenant les questions. Peut-être, alors, cette vie, peu à peu, un jour lointain, sans que vous le remarquiez, vous fera-t-elle entrer dans la réponse. »

Rainer Maria Rilke, 1903

REMERCIEMENTS

Je tiens tout d'abord à remercier chacun des membres du jury, **Loredana Saveanu, Behazine Combadière, Victor Appay, Sebastien Apcher, et Bernard Maillère** d'avoir accepté d'évaluer mes travaux de thèse.

Évidemment, je remercie **Arnaud Moris**, mon directeur de thèse, de m'avoir offert l'opportunité et sa confiance pour prendre en main ce projet de thèse. Merci de m'avoir si bien accueillie dès le début de mon contrat d'ingénieure. Ta disponibilité, ta convivialité, ton savoir scientifique sont des éléments qui ont participé à mon envie de m'engager dans cette thèse. Je me souviens encore le jour où je suis rentrée dans ton bureau, en te disant que j'aimerais bien continuer le projet en thèse, c'était déjà mon petit bébé à l'époque, tu avais l'air ravi, et c'était parti pour 3 ans d'aventures ! Le moins qu'on puisse dire, étant donné tout ce qui nous est arrivé professionnellement et personnellement dans nos vies respectives. Je suis contente qu'on ait pu tisser des affinités, discuter de nos week-ends, et de nos points de vue personnels sur les choses de la vie. Merci pour ton encadrement, tes enseignements, et tes « ah bon ? mais pourquoi tu stresses ! ». Merci de m'avoir toujours impliquée dans les projets du laboratoire, de m'avoir toujours poussée à soumettre des abstracts pour des congrès, même si timidité oblige, je n'étais pas toujours initialement enchantée de l'idée. Inconsciemment ou non, tu m'as laissé de la place pour devenir autonome. Je me sens une jeune chercheuse accomplie et prête pour la suite, alors merci.

Je voudrais également remercier, **Sidaction, l'ANRS** et le **CNRS** pour avoir soutenu ma candidature et financé ce projet. J'ai de la gratitude envers **Géraldine Schlecht-Louf, Hervé Bernard, et Marianne Mangeney** pour m'avoir ouvert les portes de leurs formations, de leurs laboratoires de recherche, et de m'avoir accompagnée dans la découverte du métier. Une pensée particulière pour **Vincent Feuillet**, j'ai relu plusieurs fois l'ouvrage de Rainer Maria Rilke depuis.

Je ne peux écrire ces remerciements sans évoquer mon indispensable voisine de bureau, collègue et amie, **Sylvie Grégoire-Ferret**. Sylvie, malgré une légère différence d'âge, j'ai de la gratitude d'être devenue ton amie (ou ta fille pour certains.es !). Merci pour ton humour et ta légèreté. Ton écoute et ton soutien sans faille ont été d'une aide précieuse à certains moments délicats de ma thèse (et de ma vie !). C'est l'occasion pour moi de te présenter mes excuses pour les fois où j'ai radoté. On sait toutes les deux que tu gagnerais plus d'argent à devenir psy. Je suis contente d'avoir partagé de beaux moments professionnels et personnels avec toi, l'imprimante du bâtiment 23A suivie de l'andouillette gratinée s'en souviendra, et moi aussi ! J'espère qu'il y en aura pleins de nouveaux. Surtout ne change rien, et continue de croire en la belle et forte personne que tu es.

Arnaud, Sylvie, le fait que vous ayez chacun le même jour de naissance que mes parents ne peut pas juste être une coïncidence, c'est un signe ! Père et mère scientifiques !

Merci à l'équipe **HERPES** d'avoir un nom pareil. Merci à **Eva Hernandez-Djebali, Marion Lussignol, Guillaume Beauclair, Audrey Esclatine, et Damion Glon** de nous avoir accueillis dans des conditions optimales suite au premier déménagement à Châtenay. Merci pour votre bonne humeur et votre aide. Merci **Cecilia Ramirez** pour m'avoir transmis ton savoir en virologie et biologie moléculaire. Merci **Guillaume** d'avoir toujours volé à mon secours quand je m'arrachais les cheveux avec R ou quand j'avais des baisses de régimes. Tu as toujours su me poser des questions scientifiques qui me posaient une colle, ce qui, bizarrement, m'a toujours plu.

Je remercie **Olivier Namy**, pour ce projet, sa disponibilité, et sa gentillesse. Merci d'avoir pris du temps pour de me donner des conseils pour la suite de ma carrière.

Je remercie **Isabelle Hatin**, pour sa bonne humeur, sa disponibilité et sa pédagogie. Merci de m'avoir formée sur la technique du ribosome profiling.

Je tiens à exprimer ma gratitude à **Emiliano Ricci** pour avoir accepté de faire partie de mon comité de mi-thèse. Ta contribution, disponibilité, et bienveillance tout au long de cette thèse ont été importantes pour moi. J'ai beaucoup apprécié nos échanges scientifiques, que j'ai toujours trouvés stimulants. Je remercie également **Emmanuel Labaronne**.

Je remercie **Anne Lopes**, pour sa pédagogie et son amour de la traduction et de la génétique. Merci pour nos discussions qui m'ont toujours motivées et donné envie d'approfondir davantage ce domaine de ma thèse.

Je remercie **Emmanuelle Quemin** pour sa disponibilité, et ses retours précieux sur mes démarches professionnelles.

Merci à **Antoine Gross**, **Frédéric Subra**, et **Clémence Richetta**, pour leur gentillesse et disponibilité.

Je dois évidemment remercier tous mes petits anges partis trop tôt (désolée Rémy de te piquer cette punchline, mais je l'adore. Tu sais bien que tu es mon mentor *blague*). Il est difficile de faire de mentions spéciales, mais je vais quand même essayer.

Loïse Valentin et **Chriss Mouanga**, toutes les trois, nous reformons le trio des Totally Spies. Nous savons toutes qu'il n'y a pas de débat sur l'attribution des personnages, bien que je n'assume pas toujours le mien. Je me souviens du premier Intermède avec toi, Chriss, tu avais les cheveux rouges à l'époque. Cela fait donc déjà plus de 5 ans que nous nous connaissons, que nous nous tapons des barres de rires, que l'on se confie nos histoires de cœur, nos doutes, et que l'on traverse des moments plus difficiles ensemble. À nos repas de Noël (personne ne fait d'aussi belles cartes de vœux que nous), d'anniversaires, nos chorégraphies endiablées, nos parties de crapette & Co. Merci pour ce fabuleux voyage en Corse, qui, nous serons sûrement toutes d'accord, a scellé notre amitié (je sais que la photo du sceau ressortira un jour ou l'autre, RIP). Nous nous complétons parfaitement, chacune avec nos aspérités. Vous êtes des amies en or. Je vous aime, merci d'être là et de répondre présentes dès qu'il le faut. We are the sun !

Mathias Pereira, que dire, je ne sais même pas par où commencer ! Je suis arrivée en 2018, dans une équipe pour le moins minuscule, avec deux mecs qui avaient déjà tissé une super relation d'amitié. Et pourtant, dès le début, toi et **Raphaël Jeger-Madiot** vous avez été bienveillants, inclusifs, et à l'écoute. Notre premier afterwork, 3 mois après notre rencontre, mémorable ! Ça a marqué le début d'une belle amitié. **Mathias**, merci pour nos discussions et nos philosophies scientifiques parfois alambiquées, nos pauses hebdomadaires mais indispensables, ta patience et ta gentillesse, envers moi et toutes les personnes que tu as encadrées. Tu as un cœur encore plus gros que ce manuscrit de thèse. Je ne sais pas si j'ai le droit de dire que tu étais le meilleur des représentants doc/post-doc... Merci d'être devenu un ami, toujours à l'écoute et toujours prêt à dégainer une blague. Je me souviendrais longtemps de tes roulements de tête, tes NANYYY. A chaque fois que j'entends « Can't hold us » je pense à toi et à ton fameux pas de danse, que tout le monde sait légendaire. Même si la localisation ne nous permet plu de nous voir aussi souvent, tu es un véritable ami. **Raphaël**, merci de m'avoir tant fait rire durant toutes ces heures passées au L2. J'ai découvert grâce à toi plus de genres de musique qu'il en existe dans le monde.

Vient le tour du fabuleux, du légendaire, de l'héroïque **Rémy Villette**. Merci d'être toi tout simplement. Mon syndicat préféré. Le contribuable sera ravi de savoir que la boule disco et les enceintes nous ont soulagés des heures astronomiques passées au labo. Merci de nous avoir faits boire et danser maintes fois dans nos locaux, dans les bars etc. ! À tes sondages qui font toujours débats. Merci d'avoir été à la fois mon conseiller R et dans la vie. Merci de m'avoir botté les fesses plus d'une fois et remise sur le droit

chemin quand je voulais tout plaquer pour devenir serveuse de bar de plage, et vivre d'amour et d'eau fraîche. Merci Rémi de m'avoir dit que la capote ça marchait quand même bien mieux que le ribosome profiling. Tu as toujours le mot juste, c'est notamment pour cela que l'on a tous besoin de quelqu'un comme toi à nos côtés.

Eoghann White, merci pour ta bonne humeur à toute épreuve, les nombreux fous rires qu'on a partagés, les soirées alcoolisées et les longues discussions à discuter de tout et de rien, et comme tu le dis si bien, surtout de rien. Raccoon team for ever.

Soledad Marsile-Medun, merci d'avoir été une si extraordinaire stagiaire de master 2. J'ai l'impression que tu m'as appris plus de chose que l'inverse. Ça a été une chance inouïe que de se retrouver à partager une semaine ensemble lors de l'édition des jeunes chercheurs de Sidaction 2022. Quelle semaine ! Ta tolérance, ta gentillesse, et ton cœur sur la main te caractérisent le mieux.

Merci à **Clémence Riffard, Jordane Divoux, Marylou Panouillot, Marie Laviron, Morgane Hilaire, Elina Rosoy et Camille Baudesson** pour avoir contribué à cette folle ambiance ! **Jordane**, une pensée pour nos trajets RER B Paris-Orsay. **Clémence**, a quand notre randonnée dans les montagnes ?!

Merci donc du fond du cœur aux membres du CIMI. Je ne suis pas sûre qu'il y soit d'autres milieux que la recherche qui offre la possibilité de créer des amitiés aussi riches entre collègues.

Merci à mes amies de longue date, **Angelina LASTRA** et **Sonia ARNAUD**. Vous m'avez toujours soutenues et guidées avec délicatesse vers ce qui est le meilleur pour moi, que ce soit dans cette aventure ou dans d'autres ! On ne peut rêver mieux comme amies. Vous êtes mon refuge dans la tempête. Je déteste l'Australie de t'avoir emmenée si loin, Angelina ! Mais je remercie la vie (et les supers parents), de m'avoir désignée marraine de baby Tiago. En parallèle d'essayer d'être marraine la bonne fée, je lui ouvrirai un compte en Suisse, et lui offrirai son premier shot !

Merci à **Marine YVON**, nos chemins se sont récemment rapprochés, et j'en suis extrêmement reconnaissante. Merci pour ta bonne humeur, ta sensibilité, ta spiritualité, ton sens de la fête, et tes podcasts interminables. Notre amour pour les débats philosophiques de la vie, et nos aspérités communes, ont forgé cette chouette relation.

Merci **Frank Marcourt**, de m'avoir offert la chance de vivre et d'écrire ma thèse dans le calme et la sérénité. Promis, je ne deviendrai pas un de ces fonctionnaires qui fou rien.

Une pensée pour mon BISON, **Alexandra Zak**, avec qui j'ai partagé un bout de ma scolarité et de chemin de vie. Être ton amie est une chance incroyable.

Merci aux I2BC copains, **Gabriela, Émilie, Benjamin, Benoit, Sophie, Hugo, Maxime, Kenza, Paul, Marie et pleins d'autres**, qui ont donné un second souffle à ma fin de thèse. Merci **Hugo** d'avoir toléré mes allées et venues pour venir toquer à ta porte (qui est toujours ouverte) pour toutes mes questions sur l'analyse bio-informatique.

Des remerciements plus généraux à ma **team BLL**, ma team d'amour, vous savez que vous êtes ma deuxième famille, mon socle, mon repère, ma sous-pape de décompression. Depuis 12 ans vous m'offrez ce lot de très jolis souvenirs, innombrables soirées et vacances inoubliables. Je vous aime.

Globalement, merci à tous de me supporter toutes les fois où je suis chiante, parce que je sais très bien l'être.

Merci à Bovis. Après 2 déménagements professionnels et 3 personnels, grâce à vous, je sais maintenant dans quoi je peux me recycler si la recherche ne fonctionne pas pour moi.

À tous les auteurs de revues, je vous remercie d'avoir rendu mes idées plus claires !

Merci à tous ceux qui m'ont inspiré d'une manière ou d'une autre, par leurs ambitions, leurs spiritualités, leurs authenticités, et leurs amours.

Je finis bien sûr par remercier **ma famille**. Merci à ma "**mommy**", sans qui je ne serais clairement pas là où j'en suis aujourd'hui. Tu nous as toujours soutenues dans ce qu'on aimait faire, quoi que ce fût. Merci de m'avoir donné les clés pour être indépendante et débrouillarde, d'avoir toujours cru en moi et remonter le moral. Merci de m'aimer si fort. Merci à **Papa** et toi de m'avoir donné le goût du travail, et celui de la conscience professionnelle. Si j'aime autant travailler et y trouve un épanouissement c'est grâce à vous. A ma **grande sœur**, Lolo, qui fait partie des personnes les plus fortes que je connaisse et qui a toujours fait en sorte de me protéger. Je vous aime.

TABLES DES MATIÈRES

LISTE DES ABRÉVIATIONS	1
TABLE DES ILLUSTRATIONS	5
TABLE DES ANNEXES	7
INTRODUCTION	9
CHAPITRE 1 : LA TRADUCTION EUCARYOTE	11
1.1 Qu'est-ce qu'un gène ?	11
1.2 Le code génétique et les ARN de transferts	12
1.3 Le ribosome : la « nanomachine » de décodage	14
1.4 L'initiation de la traduction	15
1.4.1 Les complexes d'initiation	16
1.4.2 Le scanning et la reconnaissance du codon start AUG	17
1.5 L'élongation	20
1.6 La terminaison et le recyclage	22
1.7 Le mal aimé génome non codant	23
1.8 Nomenclature des régions génomiques non canoniques	24
1.9 Les mécanismes non canoniques de traduction	25
1.9.1 Le leaky scanning	26
1.9.2 Le codon read-through	27
1.9.3 Le ribosome shunting	27
1.9.4 La ré-initiation	27
1.9.5 Les IRES	29
1.9.6 Le décalage de phase de lecture ou ribosomal frame shifting (FS)	30
1.9.7 Les codons start alternatifs ou codons non-AUG	31
1.10. Le ribosome profiling	33
1.10.1 Historique et technique	33
1.10.2 Analyse des données et périodicité	35
1.10.3 Identification des sites d'initiations	37
1.10.3.1 Harringtonine	37
1.10.3.2 Lactimidomycine	38
1.10.4 Les applications du ribosome profiling	38
1.10.4.1 Translatome et annotations des génomes	38
1.10.4.2 Identification du protéome alternatif	42
CHAPITRE 2 : LA PRESENTATION ANTIGENIQUE AUX LYMPHOCYTES T	47
2.1 Les molécules du CMH de classe I	47

2.1.1 Génétique _____	47
2.1.2 Structure et localisation _____	49
2.1.3 Nature des peptides présentés _____	49
2.2 Les molécules du CMH de classe II _____	51
2.2.1 Génétique _____	51
2.2.2 Structure et localisation _____	51
2.2.3 Nature des peptides présentés _____	52
2.3 La restriction HLA aux lymphocytes T CD8 ⁺ et CD4 ⁺ _____	53
2.4 Apprêtement et présentation des antigènes _____	54
2.4.1 La voie de présentation des antigènes par les molécules du CMH-I _____	54
2.4.1.1 Présentation des antigènes endogènes : la voie canonique _____	54
2.4.1.2 La présentation croisée : la voie alternative _____	57
2.4.2 La voie de présentation des antigènes par les molécules du CMH-II _____	60
2.4.2.1 Présentation des antigènes exogènes : la voie canonique _____	60
2.4.2.2 Les voies alternatives de présentation antigénique classe II _____	63
2.5. Les sources d'antigènes présentés par les molécules du CMH _____	66
2.5.1 La traduction canonique comme source antigénique _____	66
2.5.1.1 Les protéines retraitées VS les DRiPs _____	66
2.5.1.2 Le peptide splicing _____	67
2.5.1.3 Les néoantigènes _____	68
2.5.2 La traduction non canonique comme source antigénique _____	69
2.5.2.1 La traduction cytosolique _____	69
2.5.2.2 La traduction nucléaire _____	70
2.6. La spectrométrie de masse en tandem appliquée à l'immunopeptidome _____	72
2.6.1 Définition de l'immunopeptidomique _____	72
2.6.2 Principe de la spectrométrie de masse _____	74
2.6.3 Identification des peptides _____	75
2.6.4 Applications de l'immunopeptidomique à l'identification de peptides non canoniques _____	77
2.6.4.1 Application dans le cancer _____	77
2.6.4.2 Application dans les maladies infectieuses _____	77
2.6.7 Du ribosome profiling à l'immunopeptidomique : la combinaison d'outils puissants _____	79
CHAPITRE 3 : Le VIH-1 _____	83
3.1 Contexte épidémiologique de l'infection à VIH _____	83
3.2 Origine et diversité génétique du VIH _____	84
3.3 Structure du virion et du génome viral _____	85

3.4 Pathogénèse de l'infection à VIH	87
3.4.1 La primo-infection	87
3.4.1.1 La phase d'éclipse	87
3.4.1.2 Propagation et prolifération du VIH au GALT	88
3.4.1.3 Le viral set-point	88
3.4.2 La phase chronique	89
3.4.3 La phase SIDA	89
3.5 Le cycle viral du VIH-1	90
3.5.1 Attachement et fusion	90
3.5.2 La transcription inverse	91
3.5.3 L'intégration	92
3.5.4 La traduction	93
3.5.4.1 Les différents transcrits	93
3.5.4.2 La traduction dépendante de la coiffe	94
3.5.4.3 Le frameshifting	95
3.5.4.4 La traduction dépendante d'IRES	96
3.5.4.5 Impact de l'infection virale sur les protéines cellulaires	97
3.5.5 Assemblage, bourgeonnement et maturation	98
3.6 Les réservoir viraux	99
3.7 La réponse lymphocytaire anti-VIH	100
3.7.1 La réponse humorale	100
3.7.2 La réponse lymphocytaire T	101
3.7.2.1 La réponse lymphocytaire T CD4+ anti-VIH	101
3.7.2.2 La réponse lymphocytaire T CD8+ anti-VIH	103
3.8 Les différents statuts cliniques	103
3.8.1 Les Elite Controllers	104
3.8.1.1 Caractéristiques virologiques	104
3.8.1.2 Caractéristiques génétiques	104
3.8.1.3 Caractéristiques immunologiques	104
3.9 Les stratégies thérapeutiques	107
3.9.1 Le traitement antirétroviral	107
3.9.1.1 Les antirétroviraux	107
3.9.1.2 L'échec thérapeutique	107
3.9.2 Le traitement prophylactique	108
3.9.3 La vaccination VIH	108

3.9.4 Les stratégies curatives _____	109
3.10 Les ARFs du VIH _____	112
3.10.1 Les ARFs identifiés au laboratoire _____	112
3.10.2 Les ARFs publiés par les autres laboratoires _____	113
3.10.2.1 Identification par approche computationnelle _____	113
3.10.2.2 Identification par HLA footprint _____	114
3.10.2.3 Identification par immunopeptidomique _____	115
MISE EN CONTEXTE, OBJECTIFS DE THÈSE, & RÉSUMÉ DES RÉSULTATS _____	117
RÉSULTATS _____	121
DISCUSSION _____	181
1. Les cadres alternatifs de lecture du VIH _____	189
2. La conservation des ARFs _____	193
3. Détection des polypeptides issus des ARFs à l'aide du système immunitaire _____	196
4. Détection par immunopeptidomique des peptides dérivés d'ARFs présentés par les molécules du CMH à la surface des cellules infectées _____	198
5. Les lymphocytes T spécifiques des peptides dérivés des ARFs _____	200
6. Perspectives sur les ARFs du VIH : un modèle exploratoire _____	202
6.1 Réservoir antigénique : cause ou conséquence de la traduction des régions non codantes ? _____	202
6.2 Réponses lymphocytaires T spécifiques des peptides dérivés d'ARFs et présentation antigénique _____	205
6.3 Les applications thérapeutiques _____	206
CONCLUSION GÉNÉRALE _____	208
ANNEXES _____	212
Mise en contexte et contribution aux travaux publiés en collaboration : Labaronne et al. (Annexe 4) _____	218
BIBLIOGRAPHIE _____	272

LISTE DES ABRÉVIATIONS

AA : Acide aminé

ADCC : Antibody-dependent cellular cytotoxicity

ADN : Acide désoxyribonucléique

APC : Cellule présentatrice d'antigène

APOBEC3G : Apolipoprotein B mRNA editing enzyme, catalytic subunit 3G

ARF : Cadre alternatif de lecture

ARN : Acide ribonucléique

ARNm : Acide ribonucléique messenger

ARNr : ARN ribosomique

ARNt : ARN de transfert

ARNt-Méth : ARNt méthionine

ART : Antiretroviral therapy

ARV : Antirétroviraux

ASP : Antisense Protein

ATP : Adenosine triphosphate

β 2m : β 2-microglobuline

BCR : Récepteur des cellules B

bNAbs : Anticorps neutralisants à large spectre

CA : Capside

CDR : Complementarity diversity receptor

CDS : Coding DNA sequence

CHX : Cycloheximide

CMH : Complexe majeur d'histocompatibilité

CRF : Forme recombinante circulante

CTL : Lymphocyte T cytotoxique

DC : Cellule dendritique

DDX3 : DEAD-box helicase 3

dORF : Downstream ORF

DRiPs : Defective ribosomal products

EBV : Virus d'Epstein-Barr

EC: Elite controller
eIF: Facteur d'initiation eucaryote
eEF: Facteur d'élongation eucaryote
EN : Elite neutralizer
Env : Enveloppe
ERAP1/2 : Endoplasmic reticulum aminopeptidase
eRF : Facteur de terminaison eucaryote
FcγR : Récepteur à la partie Fc des immunoglobulines de type G
FDR : False Discovery Rate
FS : Frame shifting
GALT : Gut associated lymphoid tissues
GTP : Guanosine triphosphate
HAR : Harringtonine
HBV : Virus de l'hépatite B
HCMV : Cytomégalovirus humain
HCV : Virus de l'hépatite C
hERV : Rétrovirus endogène humain
HPV : Papillomavirus humain
HTLV : Virus T-lymphotrope humain
HSH : Hommes ayant des rapports sexuels avec des hommes
IAV : Virus de la grippe A
IBV : Virus de la grippe B
IN : Intégrase
INNTI : Inhibiteur non nucléosidique de la transcriptase inverse
INTI : Inhibiteur nucléosidique de la transcriptase inverse
IP : Inhibiteur de protéase
ITAF : IRES-transacting factor
IRES : Internal ribosome entry site
ISR : Réponse intégrée au stress
KSVH : Virus de l'herpès associé au sarcome de Kaposi
LB : Lymphocyte B

LC : Cellules de Langerhans
lncRNA : Long ARN non-codant
LT : Lymphocyte T
LTM : Lactimidomycine
LTR : Long terminal repeat sequence
MA : Matrice
MIIC : Compartiment de chargement des molécules de classe II
miRNA : micro ARN
MLV : Virus de la leucémie murine
MVB : Corps multivésiculaires
NC : Nucléocapside
NGS : Séquençage nouvelle génération
OMS : Organisation mondiale de la santé
ORF : Cadre ouvert de lecture
pCMH : Complexe CMH-peptide
PIC : Complexe de pré-initiation 43S
PICv : Complexe de pré-intégration viral
PLC : Complexe de chargement des peptides
PR : Protéase
PrEP : Prophylaxie pré-exposition
PTC : Centre de peptidyl-transférase
PVVIH : Personne vivant avec le VIH
RE : Réticulum endoplasmique
RiboSeq : Ribosome profiling
RNAseq : Séquençage de l'ARN
RPF : Ribosome protected fragment
RRE : Rev-responsive element
RT : Transcriptase inverse
SARS-CoV-2 : Coronavirus 2 du syndrome respiratoire aigu sévère
SIDA : Syndrome de l'immunodéficience acquise
SIV : Virus de l'immunodéficience simienne

sORF : small ORF
uAUG : Codon AUG situé en 5'
uORF : Upstream ORF
TAP : Transporter associated with antigen processing
TARV : Traitement antirétroviral
TCR : Récepteur des lymphocytes T
TC : Complexe de pré-terminaison
TE : Efficacité de traduction
TF : Transmitted founder virus
Tfh : Lymphocyte T folliculaire
TIL : Lymphocyte T infiltrant la tumeur
TLR : Toll-like receptor
TURBS : Terminal upstream ribosome binding site
UPS : Système ubiquitine-protéasome
UTR : Untranslated region
VIH : Virus de l'immunodéficience humaine

TABLE DES ILLUSTRATIONS

Figure 1. Code génétique et table des codons _____	12
Figure 2. Acides aminés : classification et propriétés générales _____	13
Figure 3. Exemple d'un ARN de transfert histine _____	14
Figure 4. Caractéristiques générales de la structure du ribosome et positions des ARNt _____	15
Figure 5. Illustration des 3 phases de lecture d'un ARNm _____	16
Figure 6. Modèle d'une voie canonique d'initiation de la traduction eucaryote _____	18
Figure 7. Modèle de l'élongation de la traduction eucaryote _____	21
Figure 8. Modèle de terminaison de la traduction eucaryote _____	22
Figure 9. Modèle de recyclage des acteurs de la traduction _____	23
Figure 10. Nomenclature des sORFs _____	25
Figure 11. Exemples de mécanismes de traduction non canoniques _____	26
Figure 12. Classification des IRES viraux _____	29
Figure 13. Exemple de sites de décalage programmé du ribosome (FS) du VIH-1 _____	30
Figure 14. La technique du ribosome profiling _____	33
Figure 15. Différences entre le RiboSeq et RNAseq _____	34
Figure 16. Profil d'empreintes ribosomales ; le translátome _____	35
Figure 17. Etapes clés de l'analyse bio-informatique des données de ribosome profiling _____	36
Figure 18. Périodicité _____	37
Figure 19. Schéma représentatif des différents modes d'action de la CHX et LTM _____	38
Figure 20. Fresque non exhaustive des applications du ribosome profiling de son développement à ce jour _____	39
Figure 21. Données qualitatives et quantitatives obtenues par ribosome profiling _____	40
Figure 22. Carte génétique du locus HLA sur le chromosome 6 _____	48
Figure 23. Structure moléculaire d'une molécule du CMH-I _____	49
Figure 24. Motif peptidique de fixation à la molécule HLA-A0201 _____	50
Figure 25. Structure moléculaire d'une molécule du CMH-II _____	52
Figure 26. Motifs de fixation à la molécule HLA-DR β 1*0101 _____	53
Figure 27. Représentation schématique de la structure du TCR et de l'engagement des différents récepteurs _____	53
Figure 28. Voie de présentation des antigènes de classe I _____	54
Figure 29. Représentation schématique du protéasome 20S et 26S _____	55

Figure 30. Représentation schématique de l'immunoprotéasome _____	56
Figure 31. La voie vacuolaire et cytosolique de présentation croisée vacuolaire par les molécules du CMH-I _____	58
Figure 32. Les différentes sources de molécules de classe I dans la présentation croisée	
Figure 33. Voie de présentation des antigènes de classe II _____	61
Figure 34. Les différentes voies et acteurs moléculaires utilisés par le VIH-1 à l'intérieur de la DC ____	62
Figure 35. Les différentes voies non canoniques de présentation des antigènes intracellulaires par les molécules du CMH-II _____	64
Figure 36. Les diverses sources non canoniques d'antigènes _____	71
Figure 37. Etapes clés de l'immunopeptidomique _____	73
Figure 38. Représentation schématique d'un spectromètre de masse _____	74
Figure 39. Représentation schématique d'un spectromètre de masse en tandem MS-MS _____	75
Figure 40. Les différentes bases de données utilisées pour l'identification de peptides non canoniques par LC-MS/MS _____	76
Figure 41. Les différentes procédures de validation de l'immunopeptidomique _____	76
Figure 42. Distribution géographique des sous-types du VIH-1 _____	85
Figure 43. Structure et organisation génomique du VIH-1 _____	86
Figure 44. Evolution naturelle de l'infection à VIH _____	87
Figure 45. Cycle répliatif du VIH _____	90
Figure 46. Etapes clés de la transcription inverse du VIH _____	92
Figure 47. Les différents transcrits du VIH _____	93
Figure 48. Le -1 ribosomal frameshifting du VIH _____	95
Figure 49. Représentation schématique de la structure d'ARN du -1 FS _____	96
Figure 50. Les IRES du VIH _____	96
Figure 51. Les différents facteurs viraux et cellulaires impliqués dans le contrôle de la traduction du VIH _____	98
Figure 52. Assemblage, bourgeonnement et maturation des virions _____	99
Figure 53. Les ARFs du VIH : résultats du projet et modèles proposés	
Figure 52. Assemblage, bourgeonnement et maturation des virions _____	Erreur ! Signet non défini.
Figure 53. Les ARFs du VIH : résultats du projet et modèles proposés _____	203

TABLE DES ANNEXES

Annexe n°1 : Table des peptides non canoniques chargés sur les molécules du CMH-I mis en évidence par immunopeptidomique_____	214
Annexe n°2 : Tableau récapitulatif de la littérature sur les ARFs du VIH-1_____	215
Annexe n°3 : ARFs et réponses T spécifiques détectés dans notre étude : comparaison avec la littérature_____	216
Annexe n°4 : Les travaux en collaboration, Labaronne et al 2022_____	220

INTRODUCTION

CHAPITRE 1 : LA TRADUCTION EUCARYOTE

1.1 Qu'est-ce qu'un gène ?

Dans les cellules eucaryotes, l'information génétique est portée par l'acide désoxyribonucléique (ADN). Cet acide nucléique héberge le génome humain défini comme l'ensemble des gènes nécessaires au développement, fonctionnement, et à la régulation de l'ensemble des cellules. Bien qu'il soit fastidieux de donner une définition exacte de ce qu'est un gène, il a été initialement décrit par le D^r Mendel en 1865 comme un ensemble d'éléments cellulaires portant l'information génétique qui se transmet à travers les générations (Mendel, 1965; Dunn, 2003). La dénomination « gène » apparaît une quarantaine d'années plus tard suite aux découvertes du D^r Johannsen suivies des théories du D^r Muller (Johannsen, 1909; Muller, 1922). Les travaux postérieurs du D^r Muller précisent que les gènes sont les unités essentielles de l'évolution et de la vie (Muller, 1927; Carlson, 1966; Muller, 1966). La recombinaison génétique vient ensuite démontrer que les gènes ont une notion de taille et qu'ils peuvent être segmentés. Un gène est alors caractérisé d'unité fondamentale sur la base de quatre critères : (1) la transmission héréditaire, (2) la recombinaison génétique, (3) la mutation et (4) la fonction (Portin and Wilkins, 2017).

Grâce à la découverte du code génétique, les scientifiques démontrent que c'est la séquence en nucléotides de l'ADN (Adénine, Thymine, Cytosine, et Guanine) qui détermine la séquence des protéines. La définition du gène se précise et devient alors une séquence spécifique d'ADN transcrite en acide ribonucléique (ARN) puis traduite en protéine (Portin and Wilkins, 2017). A partir de cette définition naît ce que l'on appelle les cadres ouverts de lecture ou « Open Reading Frames » en anglais (ORFs). Les ORFs sont des séquences du génome codants pour des protéines fonctionnelles. Ils ont pour caractéristiques d'être localisés dans une phase de lecture spécifique, d'avoir une séquence en acide aminé conservée et une taille minimale de 100 acides aminés (Basrai, Hieter and Boeke, 1997; Couso and Patraquim, 2017; Finkel, Stern-Ginossar and Schwartz, 2018). L'ORF est apparenté au terme « Coding DNA Sequence » (CDS).

La transcription, l'épissage alternatif, la traduction et les modifications post-traductionnelles sont, entre autres, des mécanismes qui régulent et permettent l'expression des gènes. L'expression génique débute dans le noyau par la transcription de l'ADN en ARN précurseur. Sous sa forme précurseur, le transcrit, est composé d'un enchaînement de séquences nucléiques codantes et non codantes appelées exons et introns, respectivement. L'élimination des introns par épissage alternatif va permettre l'assemblage des différents exons. Après ajout de plusieurs éléments à l'ARN épissé, il devient ARN messenger (ARNm). Il est alors exporté dans le cytoplasme où il va pouvoir être traduit en protéine par la machinerie cellulaire, ce phénomène est appelé la traduction. Bien que la traduction soit l'étape finale de l'expression génique, elle n'est en revanche que la première étape de la synthèse d'une protéine structurée et fonctionnelle.

Etant au cœur de mon sujet de thèse, je fais le choix de ne développer dans ce manuscrit uniquement la traduction eucaryote. C'est un processus cellulaire essentiel puisque les protéines représentent 2/3 de la matière organique d'une cellule. La traduction est un

processus très conservé et couteux en énergie. Les acides ribonucléiques interviennent dans la traduction sous 3 natures différentes : l'ARNm, l'ARN ribosomal (ARNr) et l'ARN de transfert (ARNt). L'ARNm, cet enchaînement simple brin de nucléotides où l'uracile substitue la thymine, sert de modèle pour la formation du polymère d'acides aminés. Chaque acide aminé correspond à l'enchaînement de trois nucléotides (codon) spécifié par le code génétique. Ce décryptage ne peut se faire qu'avec l'aide des ARNt qui servent d'adaptateurs entre le codon et l'acide aminé à incorporer dans la protéine. Le ribosome est le catalyseur de la synthèse protéique par son activité enzymatique et par ses liaisons à l'ARNm et aux ARNt.

1.2 Le code génétique et les ARN de transferts

Durant la traduction, le ribosome traduit les transcrits par triplets de nucléotides définis comme codon. La lecture successive des codons va donc permettre l'incorporation des acides aminés correspondant au sein de la protéine naissante. Chaque codon est spécifique d'un acide aminé particulier. Étant donné qu'il y a 4 bases azotées, il y a 64 triplets de bases possibles, et donc, 64 codons. Mais alors, pourquoi n'avons-nous pas 64 acides aminés ? Tout simplement parce qu'un même acide aminé peut être codé par différents codons. Le code génétique est donc composé de 4 bases, 64 codons et 20 acides aminés (Figure 1). Une protéine est une chaîne de 50 acides aminés minimum. Couramment, on considère comme polypeptide et peptide un enchaînement maximal de 50 et 20 acides aminés, respectivement.

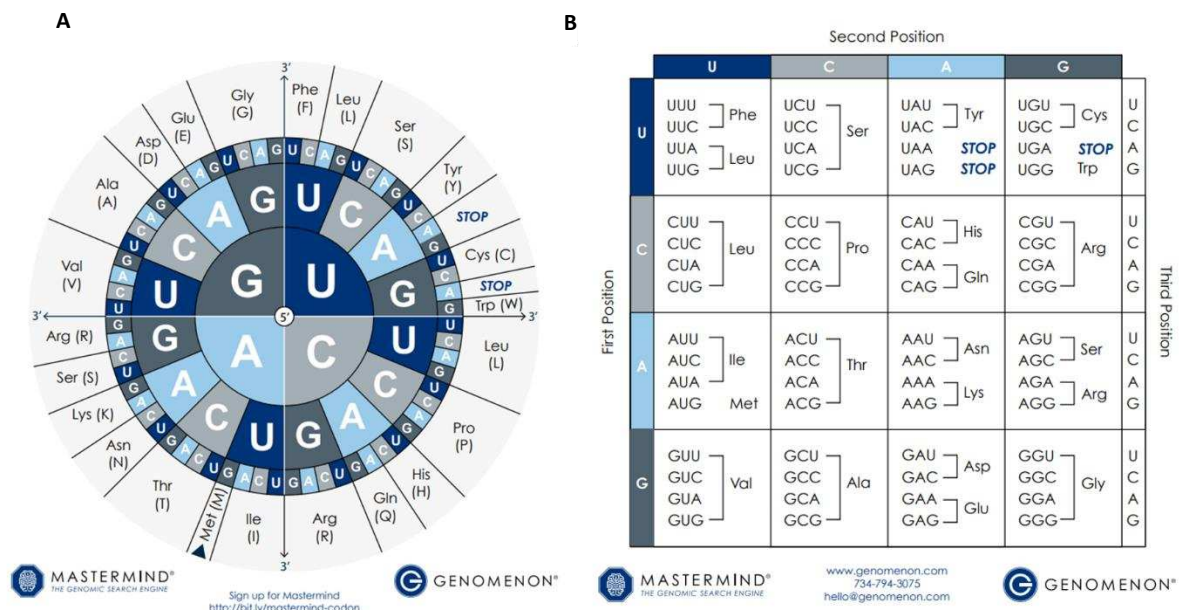


Figure 1. Code génétique et table des codons.

(A) Le code génétique de l'ARNm est composé de quatre nucléotides différents : U, C, A et G, générant au total 64 combinaisons différentes de triplet nucléotidique. (B) Le tableau des codons résume les 64 codons et précise quel codon spécifie quel acide aminé.

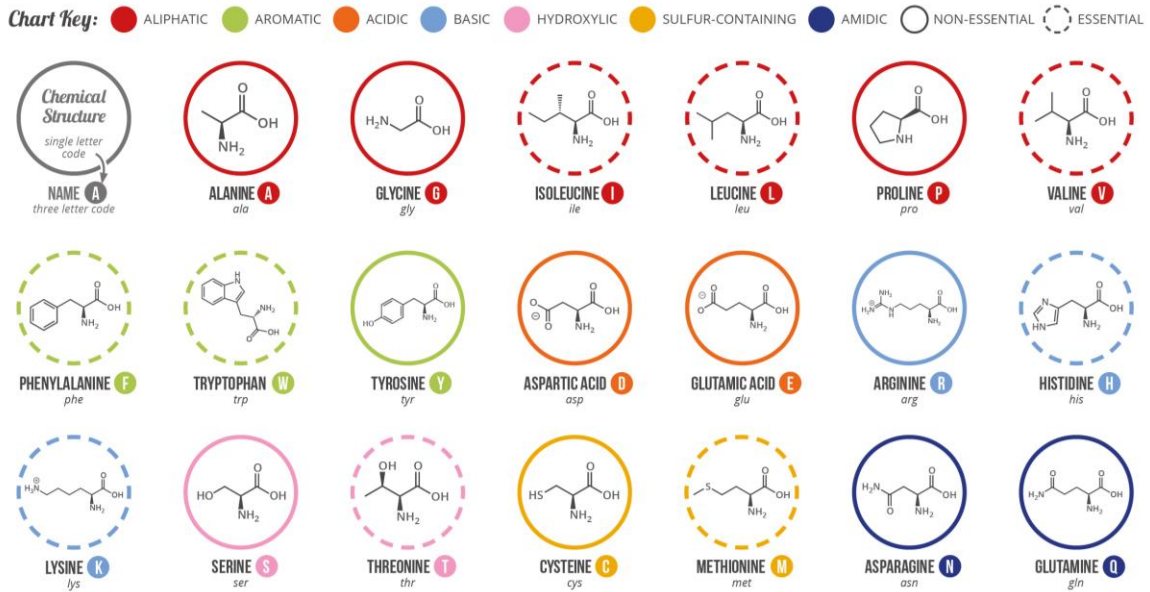


Figure 2. Acides aminés : classification et propriétés générales (figure adaptée de compoundchem.com)
 Le code génétique humain code 20 acides aminés. Les acides aminés essentiels doivent être obtenus par l'alimentation, tandis que les acides aminés non essentiels peuvent être synthétisés par l'organisme. Les acides aminés hydrophobes sont représentés en rouge, en vert et en orange. Les acides aminés hydrophiles sont représentés par les autres couleurs. Cette carte renseigne uniquement les acides aminés du code génétique. La sélénocystéine est considérée comme le vingt et unième acide aminé présent chez les eucaryotes mais elle est codée par un mécanisme différent. C'est un acide aminé rare que l'on retrouve uniquement dans quelques enzymes.

Les acides aminés (AA) sont classés selon leurs propriétés physiques et chimiques (Figure 2). On retrouve deux groupes majeurs : les acides aminés hydrophiles et les hydrophobes. Le groupe hydrophile est divisé en 3 sous- groupes : les acides aminés basiques, acides et neutres.

Les codons sont répertoriés selon leur ordre de lecture de base. Il est intéressant de noter que dans le cadre des mutations synonymes, silencieuses et neutres d'un point de vue de l'évolution, la mutation s'effectue sur le troisième nucléotide du triplet dû à la redondance du code génétique. Chez les eucaryotes, la traduction canonique est initiée par la méthionine défini par le triplet « AUG » : également appelé codon start. En revanche, trois combinaisons de nucléotides peuvent annoncer l'arrêt de la traduction, on retrouve « UAA », « UAG », et « UGA » comme codon de terminaison (Figure 1 et 2).

Le code génétique est décodé grâce aux ARNt qui s'apparient aux codons. L'ARN de transfert est une molécule d'ARNr, simple brin, composée d'environ 80 nucléotides. En trois dimensions, l'ARNt adopte une structure en « L » (Sharp *et al.*, 1985). La boucle inférieure contient l'anticodon : un enchaînement de 3 nucléotides complémentaires aux codons des transcrits. L'extrémité 3' de la tige réceptrice de l'ARNt possède une séquence conservée « CCA » qui sert d'accepteur pour l'acide aminé (Figure 3). Puisqu'un anticodon est spécifique d'un codon on s'attendrait à avoir 64 ARNt. Le « wobble », phénomène caractérisé par le Dr Crick en 1966 (Crick, 1966), permet une variation des nucléotides situés à la troisième position des codons de l'ARNm, ainsi que ceux localisés à la première position des anticodons de l'ARNt. Cette souplesse d'appariement permet de réduire le nombre d'ARNt nécessaire à la lecture du code génétique.

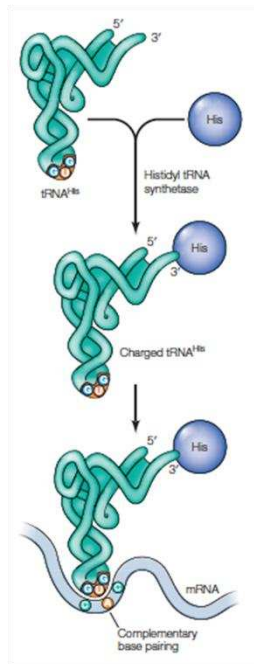


Figure 3. Exemple d'un ARN de transfert histine (figure originale Cooper 2019).

1.3 Le ribosome : la « nanomachine » de décodage

Le ribosome eucaryote est un complexe très sophistiqué et dynamique composé de protéines et d'ARNr structurés en deux sous-unités : une grande et une petite appelées chacune 60S et 40S, respectivement (Figure 4). La grand sous-unité 60S se compose de trois ARNr (28S, 5S et 5.8S) et de 47 protéines, alors que la petite est composée massivement que d'un seul ARNr (18S) et de 33 protéines (Khatter *et al.*, 2015). L'ensemble des deux sous-unités forme un complexe dénommé « 80S » qui se réfère à son coefficient de sédimentation. Sa structure asymétrie a été mise en évidence par cristallographie chez la levure en 2010 par Adam Ben-Shem et al, un an après l'attribution du prix Nobel à leur confrères Ramakrishnan, Steitz et Yonath pour la découverte de la structure du ribosome bactérien (Rodnina and Wintermeyer, 2010; Ben-Shem *et al.*, 2011). L'architecture dans l'espace des deux sous-unités l'une au-dessus de l'autre génère une cavité inter-sous-unité, assimilée à un tunnel, qui permet le passage des transcrits et des ARNt. Les deux sous-unités du ribosome fonctionnent en binôme même si leurs rôles diffèrent. Elles diffèrent d'abord par leur forme : la petite sous-unité est plus allongée avec plus de domaines structuraux alors que la grande est plus ronde et homogène. La sous-unité 60S peut être considérée comme le siège de la synthèse protéique puisqu'elle héberge le centre de peptidyl-transférase (PTC). Le PTC est responsable de la formation des liaisons peptidiques et du contrôle de la qualité des polypeptides naissants (Schlunzen *et al.*, 2000; Pelletier, Thomas and Volarević, 2018). Néanmoins, la petite sous unité n'a pas à rougir car c'est elle qui possède le centre de décodage de l'information génétique. Elle initie l'engagement de l'ARNm, décode l'information, contrôle ensuite la qualité de l'appariement de l'anticodon au codon et gouverne la translocation des ARNt aux différents sites du ribosome (Schlunzen *et al.*, 2000). Le ribosome héberge trois sites de liaisons aux ARNt : le site E, P et

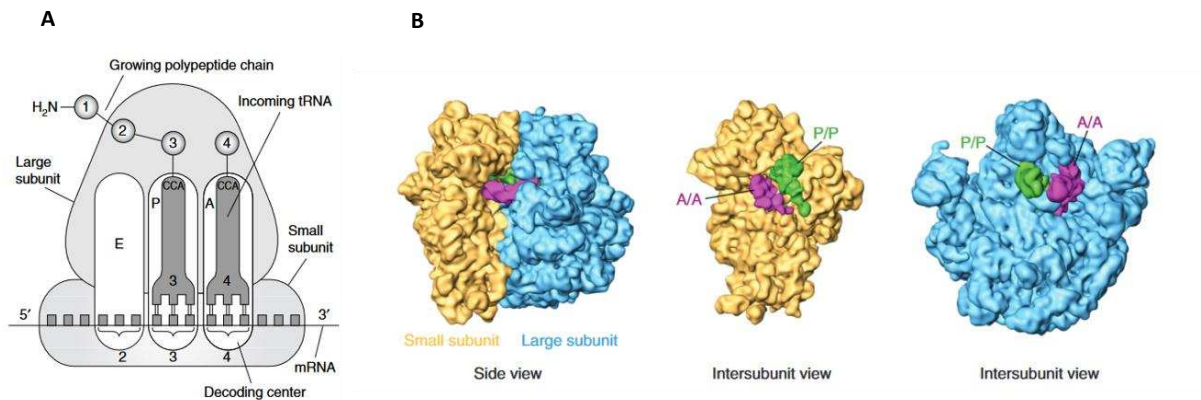


Figure 4. Caractéristiques générales de la structure du ribosome et positions des ARNt (figures adaptées de Frank 2003 et Jobe et al 2009).

(A) Structure schématique du ribosome 80S avec ses trois sites de liaisons aux ARNt : E, P, et A. (B) La grande sous-unité apparaît en bleu et la petite sous-unité en jaune. L'ARNt du site A (A/A) est en rose, et l'ARNt du site P (P/P) en vert.

A, qui sont répartis sur l'ensemble des deux sous-unités (Figure 4). Chaque site possède sa propre fonction. Le site A « aminoacyl » permet le décryptage de la séquence d'ARNm en codon et un appariement de l'anti codon correspondant. Le site P « peptidyl », lui, associe l'acide aminé au polypeptide naissant. Pour finir, le site E « exit » permet l'expulsion des ARNt désacétylés. (Frank, 2003). Chez les eucaryotes on retrouve deux populations de ribosomes : ceux qui sont libres dans le cytosol et ceux qui sont liés au réticulum endoplasmique, spécialisés dans la synthèse de protéines de la voie sécrétoire.

La synthèse des protéines se divise en 4 étapes cruciales : l'initiation, l'élongation, la terminaison, et le recyclage des ribosomes et des différents facteurs. Chacune de ces étapes est détaillée dans les sections suivantes.

1.4 L'initiation de la traduction

L'ARN mature est exporté du noyau, suite à l'épissage alternatif et à l'ajout d'une coiffe en 5' et d'une polyadénylation en 3', pour être traduit dans le cytoplasme. Avant d'aller plus loin dans l'initiation de la traduction, la question de la phase de lecture doit être traitée. Nous avons vu que les transcrits sont lus par triplets de nucléotides correspondant à des codons. De ce fait, il y a 3 façons de lire les codons : à partir du premier nucléotide, du deuxième, ou bien du troisième. Chaque transcrit contient alors 3 phases de lecture possibles. Puisque les transcrits possèdent une région non traduite, « 5' untranslated region » (5'UTR), en amont de leur séquence codante, c'est le codon start qui définit le début de la traduction et la phase de lecture définissant alors l'ORF (Figure 5). Pour démarrer la traduction d'un transcrit, le ribosome doit donc trouver le codon start. C'est sur cet objectif que repose l'étape d'initiation de la traduction.

L'initiation de la traduction chez les eucaryotes est un processus complexe comparé à celle du règne procaryote. En effet, bien que la petite sous-unité du ribosome soit l'acteur principal de cette étape, celle-ci ne peut s'effectuer sans l'implication des facteurs d'initiation de la traduction (eIF) et du métabolisme. L'initiation, qui est la phase la plus longue de la traduction,

comporte au moins 4 étapes : la formation des différents complexes d'initiation, le scanning, la reconnaissance du codon initiateur, et le recrutement de la large sous-unité.



Figure 5. Illustration des 3 phases de lecture d'un ARNm. Les flèches bleues représentent les trois positions possibles de démarrage de lecture des codons : à partir du premier nucléotide, deux deuxième, ou du troisième. Ce morcellement permet d'obtenir trois phases de lecture avec un enchaînement nucléotidique différent. La traduction de chaque phase démarre avec la reconnaissance du codon start AUG et chaque phase code pour un polypeptide différent.

1.4.1 Les complexes d'initiation

Il existe deux complexes d'initiation de la traduction. D'une part, le complexe de liaison à la coiffe, eIF4F, qui permet la reconnaissance et la bonne conformation du transcrit ; de l'autre le complexe de pré-initiation 43S (PIC) qui permet le chargement de la petite sous-unité avec un ARNt portant la méthionine (ARNt-Méth). La formation du complexe PIC s'effectue en deux temps. D'un côté, l'ARNt-Méth va venir s'associer avec le facteur eIF2 sous sa forme GTP. eIF2 est une protéine G qui a une affinité 10 fois supérieure pour l'ARNt-Méth quand elle est sous sa forme GTP plutôt que sa forme GDP (Erickson and Hannig, 1996; Kapp and Lorsch, 2004). De leur côté, les facteurs eIF1, -1A, eIF3, et eIF5 lient la petite sous-unité du ribosome. Des études chez les mammifères et la levure montrent que la liaison de l'ARNt-Méth/eIF2-GTP à la petite sous unité est rendue possible par l'action de eIF1, -1A, et eIF3, qui auraient des rôles cumulatifs (Merrick and Pavitt, 2018). Ces dernières interactions stabilisent la bonne conformation de la sous-unité 40S dans le but de fixer le transcrit et de rendre le site P accessible à la fixation de l'ARNt-Méth/eIF2-GTP (Fraser, 2015; Merrick and Pavitt, 2018). L'assemblage de la sous-unité 40S complexée à un ARNt-Méth/eIF2-GTP avec les facteurs eIF1, -1A, eIF3, et eIF5 forme le complexe de pré-initiation PIC 43S (Figure 6A).

Le deuxième complexe d'initiation se nomme le complexe de liaison à la coiffe pour sa dépendance à la reconnaissance de la guanosine méthylée en 5'. C'est d'ailleurs pour cela que l'on considère chez les eucaryotes la traduction comme majoritairement dépendante de la coiffe. Le complexe de liaison à la coiffe, désigné eIF4F, regroupe différents facteurs d'initiation qui sont impliqués dans la reconnaissance des deux extrémités du transcrit, le recrutement du PIC, et le déroulement de l'ARNm. Il se construit par l'engagement des facteurs eIF4E, -4G, -4A, -4B et de la « poly(A)-binding protein » (PABP). Premièrement, le facteur d'initiation eIF4E interagit avec la coiffe du transcrit en 5' et la PABP lie la queue polyadénylée en 3'. Simultanément, le facteur eIF4G est recruté par eIF4E. Grâce à ses différents domaines de liaisons, eIF4G lie aussi PABP et permet le rapprochement des deux extrémités du transcrit, lui donnant ainsi une forme circulaire (Figure 6B) (Hinnebusch, 2014). Le facteur eIF4G est considéré comme le squelette du complexe de pré-initiation dû à ses nombreux domaines de liaison. Deuxièmement, les facteurs eIF4A et eIF4B sont recrutés par eIF4G. Le facteur eIF4A est un partenaire prépondérant dans le complexe de pré-initiation. Son activité hélicase, assurée

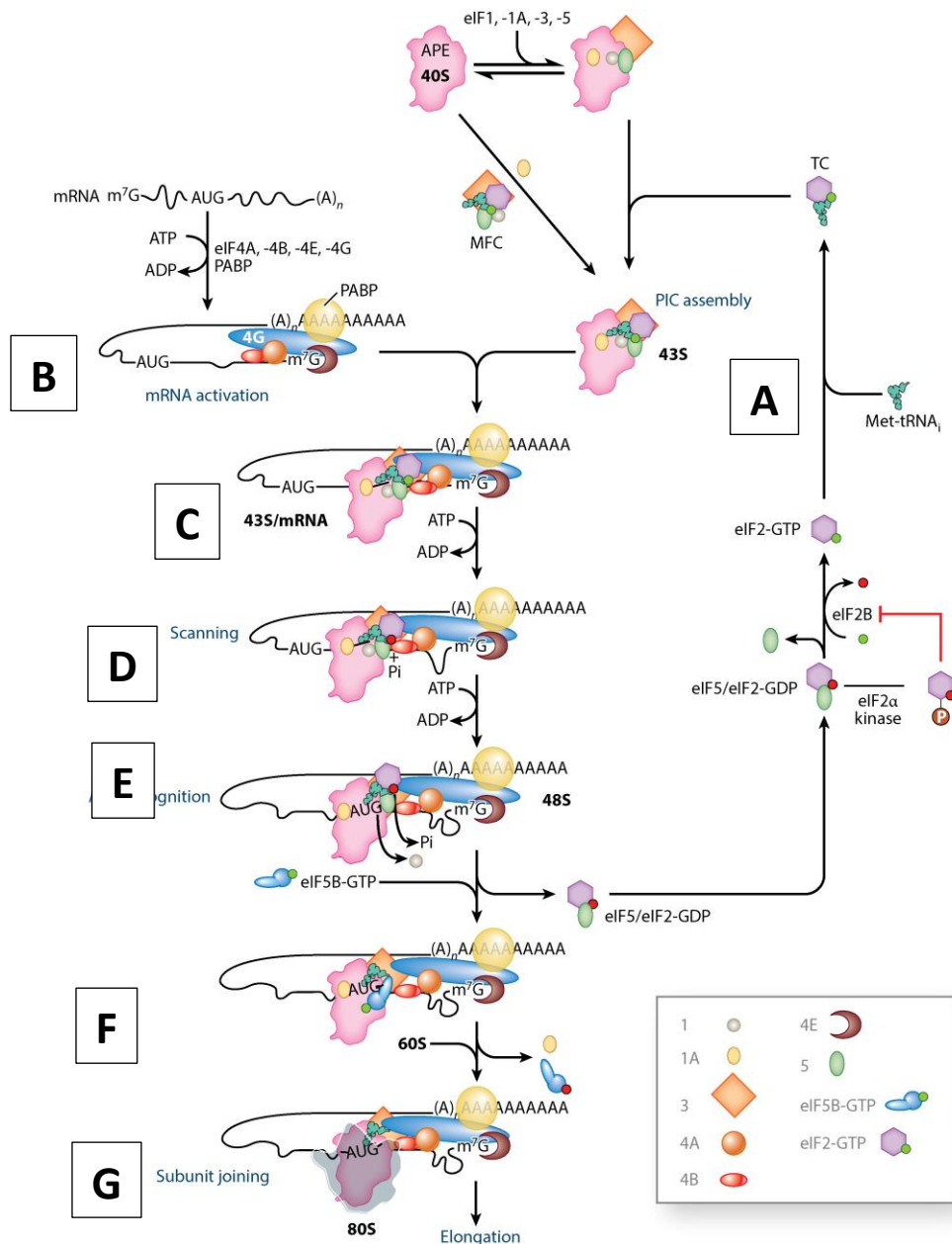
par l'hydrolyse de l'ATP, lui permet de dénouer les structures secondaires de l'ARNm et de faire progresser le complexe de pré-initiation le long du transcrit. L'association du complexe PIC au complexe eIF4F/ARNm est déclenchée par les facteurs eIF3 et eIF4B. Le complexe de pré-initiation est alors mature et débute le phénomène de scanning (Figure 6C).

1.4.2 Le scanning et la reconnaissance du codon start AUG

Le scanning est un mécanisme par lequel le complexe de pré-initiation mature balaye l'ARNm jusqu'à rencontrer un codon initiateur AUG. C'est le D^r Kozak qui en 1978 a été le premier à proposer ce modèle général d'initiation de la traduction connu sous le nom de scanning (Kozak, 1978, 1989). Le modèle a été par la suite confirmé par un grand nombre de recherche dans un contexte d'initiation canonique (Jean-Jean, Cassan and Rousset, 1993). Le scanning est un mécanisme rapide puisqu'il aurait une vitesse de lecture de 8 bases par seconde (Hinnebusch, 2014). Le scanning repose sur plusieurs paramètres : 1) la présence du complexe de pré-initiation mature, 2) une bonne conformation de la sous-unité 40S pour pouvoir se déplacer en avant, 3) la présence d'un contexte fort de traduction, et 4) un déroulement des structures de l'ARNm.

Les facteurs eIF1, -1A, et eIF5 jouent un rôle important dans la conformation du complexe PIC et dans la fidélité de la reconnaissance du codon initiateur au site P (Fraser, 2015). eIF1 et eIF1A favorisent une conformation ouverte, propice au scanning, de la sous-unité 40S du complexe PIC. Cette conformation facilite l'inspection des codons au site P pour la complémentarité avec l'anticodon de l'ARNt-Méth (Thakur, Marler and Hinnebusch, 2019). Pendant le scanning, le GTP lié à eIF2 peut- être hydrolysé par l'activité GTPase de eIF5 mais le phosphate inorganique (Pi) est libéré uniquement à la rencontre de l'AUG. C'est le facteur eIF1 qui bloque la libération du Pi aux codons non-AUG (Figure 6D, E, et F) (Thakur, Marler and Hinnebusch, 2019).

Un contexte fort de traduction est défini par la présence d'une séquence consensus en amont du codon AUG. La présence d'un contexte fort de traduction autour du premier codon AUG va permettre d'initier la traduction à cette position inhibant de ce fait la traduction sur les codons AUG situés en aval. La séquence varie selon l'embranchement phylogénétique mais la présence en position -3 d'une base azotée de la famille des purines est conservée (par convention, le A de l'AUG est numéroté +1) (Jean-Jean, Cassan and Rousset, 1993; Hinnebusch, 2014). Les codons « ACC » et « GCC » sont également utilisés de façon préférentielle aux alentours du start. Cependant, l'unique présence d'une purine en position -3 ne suffit pas à initier au premier codon AUG. La nature des résidus en -1, -2, -6 et +4 joue également un rôle prépondérant. La présence d'une guanine en position +4 est souvent associée à un contexte fort de traduction. L'étude de ces contextes chez le virus de l'hépatite B (HBV) montre en revanche que ce n'est pas une règle absolue (Jean-Jean *et al.*, 1989; Fouillot *et al.*, 1993). Dans ce cas de figure, la présence d'une purine en position -3 et d'une guanine en +4 autour de l'AUG du gène P n'est



H Hinnebusch AG. 2014.
A Annu. Rev. Biochem. 83:779–812

Figure 6. Modèle d'une voie canonique d'initiation de la traduction eucaryote (figure adaptée de Hinnebusch 2014)

L'initiation démarre avec la formation des complexes de pré-initiation 43S (A) et du complexe de liaison à la coiffe (B). Le 43S PIC est ensuite chargé sur un transcrite complexé au facteur eIF4F regroupant eIF4-A, -B, -E, -G, et PABP (C). L'activité hélicase du facteur eIF4A permet ensuite de dérouler les structures de l'ARNm permettant le scanning du transcrite jusqu'à la rencontre d'un codon AUG (D). La reconnaissance du codon AUG déclenche une cascade d'étapes, notamment la libération du Pi et la dissociation d'eIF1 (E). Le complexe PIC adopte alors une conformation fermée. eIF5B, sous sa forme liée au GTP, favorise la jonction de la sous-unité 60S au complexe PIC (F). Le complexe mature d'initiation 80S est prêt pour la phase d'élongation de la synthèse protéique (G).

pas caractérisée comme un contexte fort de traduction. Sur ce transcrite bi-cistronique, 5% des ribosomes ne s'arrêtent pas à l'AUG, ce qui permet la traduction d'un cadre de lecture situé en aval. Ce contexte de traduction est alors défini comme faible. Un contexte faible de traduction au premier AUG laisse passer les complexes d'initiation et favorise ainsi la traduction des codon AUG présents en aval dans la même phase de lecture. C'est le « leaky scanning ». Ce mécanisme

permet d'engager l'initiation à différents codons dans la même phase de lecture et donc de synthétiser plusieurs isoformes d'une même protéine. Le scanning du complexe de pré-initiation mature le long du transcrite nécessite de l'ATP uniquement si l'ARNm est pourvu de structures secondaires. eIF4A exerce sa fonction d'hélicase grâce au support des facteurs eIF4G et eIF4B (Hinnebusch, 2014; Merrick and Pavitt, 2018). Le complexe eIF4F au complet permet donc la fixation du PIC mais améliore également le scanning si les structures rencontrées ne sont pas trop complexes (de type tige-boucle).

- Focus sur les hélicases du complexe d'initiation

eIF4A n'est pas la seule hélicase ou « DEAD-box RNA hélicases » requise pour le déroulement de l'ARNm. Il a été montré chez les eucaryotes que les hélicases, Ded1, Dhx29 et DDX3 participent également à ce mécanisme notamment lors de la rencontre de structures complexes étendues dans la région 5'UTR de l'ARNm (Hinnebusch, 2014; Fraser, 2015). C'est par leur association avec le complexe eIF4F et plus particulièrement avec le facteur eIF4G que DDX3 et Ded1 augmentent le taux d'initiation des ARNm qui possèdent des structures secondaires étendues (Hilliker *et al.*, 2011; Soto-Rifo *et al.*, 2012). Chez la levure, la déplétion de Ded1 entraîne une forte inhibition de la traduction alors que le rôle de DDX3, son homologue humain, n'est pas si clair (Berthelot *et al.*, 2004; Marsden *et al.*, 2006). Dans les cellules humaines, la déplétion de DDX3 n'impact pas la traduction de façon globale mais impact uniquement certains ARNm structurés (Lai, Lee and Tarn, 2008). Une conséquence directe de la persistance d'une structure complexe proche de la coiffe dans les transcrits cellulaires et viraux est la diminution de la fixation du complexe PIC. Un très bel exemple est le virus de l'immunodéficience humaine 1 (VIH-1) qui possède une tige-boucle très stable à proximité de sa coiffe. Soto et al. ont démontré que les transcrits du VIH-1 sont bien traduits de façon coiffe dépendante et que DDX3 participe à ce phénomène de manière spécifique et temporelle (Soto-Rifo *et al.*, 2012). DDX3 intervient dans la traduction des transcrits viraux uniquement lorsque ceux-ci présentent une tige-boucle proximale. Plusieurs études suggèrent un rôle essentiel mais ponctuel de DDX3 dans le scanning contrairement à eIF4A qui effectue plusieurs cycles d'hélicase. DDX3 déstabiliserait localement l'ARNm afin de générer une région monocaténaire assez longue pour le recrutement du complexe PIC (Rajagopal *et al.*, 2012; Soto-Rifo *et al.*, 2012).

DDX3 est aussi impliquée dans la régulation de la traduction des cadres de lectures situés en 5' UTR appelés « upstream reading frames » (uORFs) (Chen *et al.*, 2018; Guenther *et al.*, 2018; Linsalata *et al.*, 2019). Les uORFs sont connus pour réprimer la traduction de la séquence canonique située en aval (voir section 1.10.4.2). Comme exposé dans l'Annexe n°4, les travaux de nos collaborateurs Labaronne et al. montre l'intervention de DDX3 dans la levée de l'inhibition de la traduction des protéines canoniques exercée par les uORFs du VIH-1 (Labaronne *et al.*, 2022). Indépendamment de sa fonction hélicase, DDX3 semble aussi être essentielle au recrutement de la large sous-unité du ribosome au complexe d'initiation (Geissler, Golbik and Behrens, 2012). Malgré le fait qu'il persiste encore quelques points d'ombre sur leur analogie, Ded1 et DDX3 présentent une forte conservation et une même localisation cellulaire (Senissar *et al.*, 2014).

Ces hélicases à ARN ont de multiples fonctions dans la régulation des gènes, le métabolisme de l'ARN, la tumorigenèse, et les infections virales. Dans l'oncogenèse, il a été rapporté des propriétés anti et pro-tumorale de DDX3 en fonction du type cellulaire. Par exemple, DDX3 favorise la transformation, la migration cellulaire ainsi que les métastases en modulant l'expression de la cadhérine E, du facteur de transcription ATF4 et la voie de signalisation Rac1 (Ariumi, 2014; Chen *et al.*, 2018). Dans les infections virales, DDX3 est impliquée dans la réplication du VIH-1, du virus de l'hépatite C (HCV) alors qu'elle restreint celle du virus de l'HBV (Schröder, 2010; Ariumi, 2014). DDX3 contribue également à l'induction de la voie de signalisation TBK1/IKK ϵ suite à l'engagement des Toll-like receptors et des hélicases de type RIG.

L'initiation de la traduction se termine avec le recrutement de la large sous unité 60S catalysée par eIF5B. Le complexe de pré-initiation devient alors le complexe d'initiation 80S et peut démarrer l'élongation (Figure 6G).

1.5 L'élongation

Durant l'élongation, la petite et la grande sous-unité du ribosome travaillent de manière coordonnée conduisant à la synthèse de la chaîne polypeptidique. La première décode le transcrit au site A et la deuxième synthétise la chaîne polypeptidique. L'élongation est une série de cycles d'incorporation d'acides aminés au sein de la chaîne peptidique naissante jusqu'à la rencontre du codon de terminaison (Figure 7). Elle repose sur différentes étapes : 1) le décodage du transcrit au site A, 2) la formation de la liaison peptidique, et 3) la translocation du ribosome.

Dans le cas de figure où le complexe 80S d'initiation a reconnu l'AUG, l'ARNt-Méth se situe au site P. Comme le site E, le site A est alors libre et le second codon du cadre ouvert de lecture (ORF) est prêt à accueillir un nouvel ARNt. L'élongation débute avec le facteur eEF1A qui forme un complexe tertiaire par sa liaison à une molécule GTP et à un ARNt porteur d'un acide aminé. Ce complexe nouvellement formé rejoint le site A. L'appariement entre le codon et l'anticodon est effectué grâce au centre de décodage de la sous-unité 40S et à l'hydrolyse du GTP par eEF1A (Dever, Dinman and Green, 2018). A cet instant, les sites P et A sont occupés et le site E est libre (Figure 7.1).

La formation d'une liaison peptidique entre la méthionine du site P et l'acide aminé du site A est rendue possible par plusieurs acteurs. C'est d'abord le facteur d'initiation eIF5A qui par sa boucle liaison au site E et à la tige acceptrice de l'ARNt permet la stabilisation et le bon positionnement du substrat porté par l'ARNt du site P (Dever, Dinman and Green, 2018). Après cette accommodation, la peptidyl-transférase du PTC de la sous-unité 60S catalyse la formation des liaisons peptidiques entre les acides aminés permettant ainsi l'élongation de la chaîne peptidique. L'ARNt du site A, précédemment décrit comme porteur d'un acide aminé unique devient alors l'ARNt porteur de la chaîne polypeptidique naissante. L'ARNt du site P est désacétylé (Figure 7.2).

Une étape clé de l'élongation est la translocation du ribosome. Elle permet le décalage du ribosome de trois nucléotides en avant conduisant ainsi au déplacement de l'ensemble des

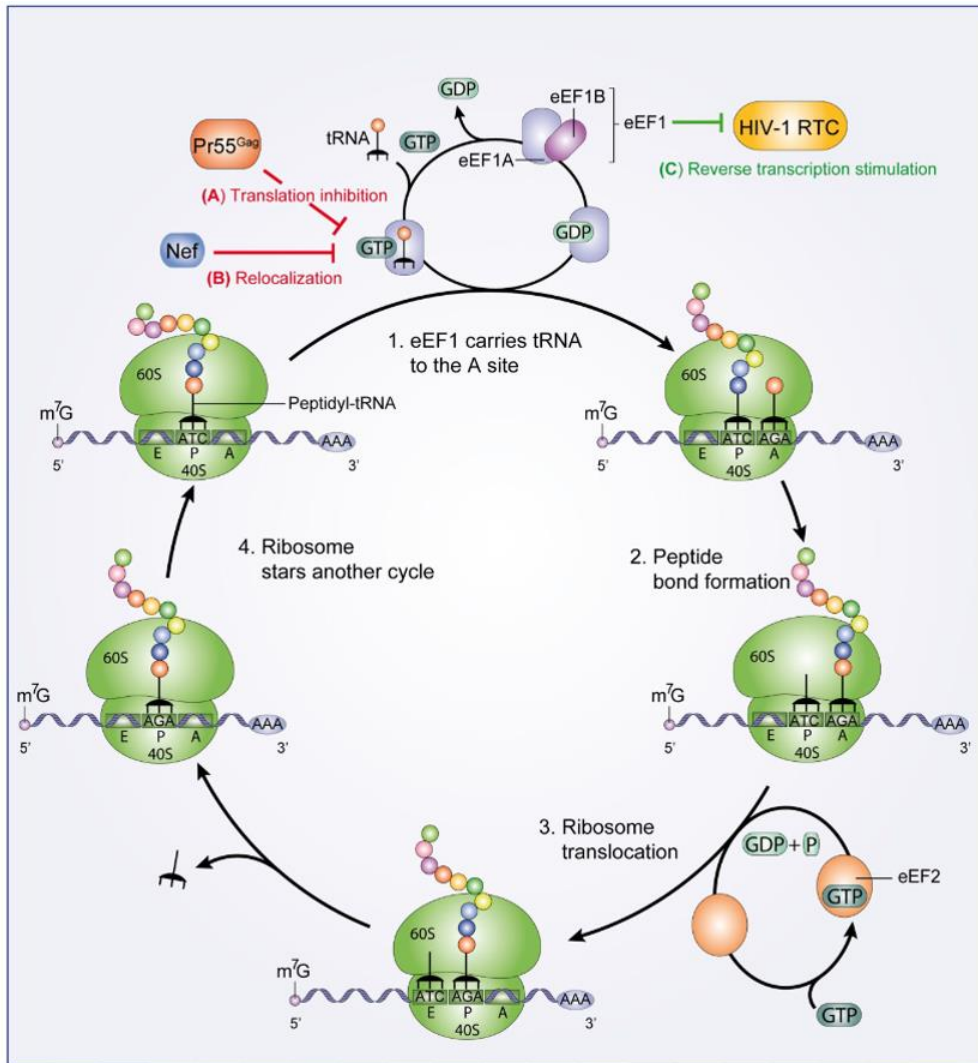


Figure 7. Modèle de l'élongation de la traduction eucaryote (figure adaptée de Guerrero et al. 2015)
1. Formation du complexe ternaire eEF1A-GTP/ARNt-acide aminé. L'accommodation du complexe ternaire au site A survient après l'hydrolyse du GTP, eEF1A-GDP qui est alors libéré est recyclé par eEF1B. **2.** Catalyse de la liaison peptidique, les ARNt du site P et A sont stabilisés par eIF5A. **3.** Suite à la formation de la liaison peptidique, le facteur eEF2 avec sa modification diptamide se lie au site A et induit la translocation des ARNt aux sites P et E. L'ARNt désacétylé du site E est ensuite libéré. **4.** Le cycle suivant d'élongation peut commencer selon le même modèle. En bleu, rouge et orange sont indiqués les protéines virales du VIH qui modulent l'étape d'élongation. RTC : reverse transcription complexe.

ARNt. Le facteur eEF2 complexée au GTP est essentiel à ce mécanisme car il assure le maintien de la phase de lecture de l'ORF pendant la translocation (Figure 7.3). Il a été mis en évidence que c'est la diptamide, modification post-traductionnelle unique à eEF2, qui lui confère ce contrôle. Liu et al. ont démontré que l'absence de diptamide diminue la traduction et augmente les phénomènes de changement de phase de lecture appelés « frame shifting » (FS) (Liu *et al.*, 2012). Plusieurs travaux ont prouvé la nécessité de mouvement du ribosome dans la translocation, majoritairement reportés comme dépendant de la petite sous-unité (Zhang 2009). Klahoz et al révèlent que la petite sous-unité subit plusieurs changements de conformation du ribosome entre l'état de pré-translocation et l'état de post-translocation. Trois types de mouvements sont référencés : la rotation par rapport à la grande sous-unité, connue pour faciliter le positionnement des ARNt, le pivotement en direction du site E et le roulement qui permet une légère ouverture libérant le site de fixation des GTPases (Frank and

Agrawal, 2000; Valle *et al.*, 2003; Zhang, Dunkle and Cate, 2009; Budkevich *et al.*, 2014; Khatter *et al.*, 2015). L'ARN désacétylé du site P et l'ARNt porteur de la chaîne polypeptidique du site A transloque ensuite aux sites E et P, respectivement. Le site A est à nouveau libre pour le démarrage d'un nouveau cycle d'élongation (Figure 7.4).

1.6 La terminaison et le recyclage

La terminaison de la traduction repose sur un principe simple : l'entrée au site A d'un codon de terminaison. Dans le code génétique, trois séquences sont connues comme codon stop : « UAG », « UGA », et « UAA ». Ces codons n'ont pas d'acide aminés correspondant mais leur reconnaissance induit la libération de la chaîne polypeptidique. La terminaison est composée de trois étapes : 1) la reconnaissance d'un codon stop, 2) la libération de la chaîne polypeptidique, et 3) le recyclage des facteurs de traduction. Contrairement aux autres étapes de la traduction elle apparaît selon moi comme l'étape la moins complexe. Elle est médiée par les deux facteurs de libération eRF1 et eRF3. eRF1 est un facteur multifonction. Par son extrémité C-terminale, il lie les facteurs de terminaison et de recyclage, eRF3 et ABCE1, respectivement. Dans un premier temps, eRF1 lie eRF3, une GTPase qui catalyse la libération de la chaîne polypeptidique. L'association de eRF1 à eRF3 inhibe la dissociation du GTP ce qui rend stable le complexe eRF1-eRF3-GTP appelé alors complexe de pré-terminaison (TC) (Hellen, 2018a) (Figure 8). C'est par l'extrémité N-terminale de eRF1 que le TC reconnaît ensuite le codon de terminaison et se fixe au site A du ribosome. Comme l'initiation, l'efficacité de la reconnaissance du codon stop est favorisée par la présence d'un contexte de terminaison : une purine en position +4, ou +5, si le résidu en +4 est une pyrimidine (Tate *et al.*, 1995). C'est ensuite le domaine central de eRF1 qui stimule l'hydrolyse du GTP par eRF3. La chaîne polypeptidique est alors libérée.

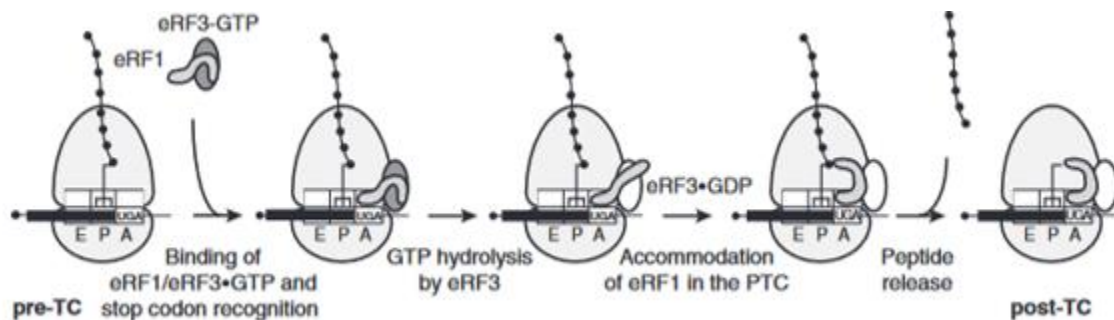


Figure 8. Modèle de terminaison de la traduction eucaryote (figure originale de Hellen 2018)

La protéine ABCE1 est très conservée et intervient dans le recyclage des éléments de traduction, plus précisément dans le détachement des deux sous-unités du ribosome. Suite à sa liaison à eRF1 et au ribosome, ABCE1 est capable de fixer l'ATP. L'hydrolyse de cet ATP engendre différents changements de conformations d'ABCE1 qui peut alors agir comme un levier pour libérer les deux sous-unités (Figure 9) (Hellen, 2018a). ABCE1 peut également agir sur les sous-unités 80S libres ou sur les ribosomes en pause lors de l'élongation.

La terminaison annonce la fin de la synthèse protéique. Le recyclage permet aux ribosomes et aux divers acteurs de participer à de multiples cycles de traduction.

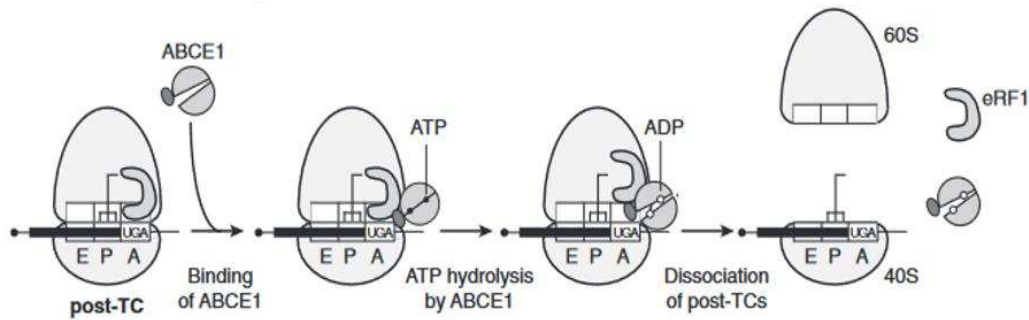


Figure 9. Modèle de recyclage des acteurs de la traduction (figure originale Hellen 2018).

1.7 Le mal aimé génome non codant

La traduction d'acides aminés en macromolécules protéiques est décrite comme la dernière étape de l'émergence d'un gène (Papadopoulos *et al.*, 2021). Nous avons vu en début de chapitre que la définition initiale du gène consiste en une séquence d'ADN transcrite en ARN puis traduite en protéine. Néanmoins, sa définition a évolué depuis la découverte du séquençage puisque le Projet Génome Humain (HGP en anglais) révèle, en 2001, l'existence de nombreux gènes transcrits en ARN non codants. Les ARN non codants regroupent les « long ARN non-codant » (lncRNA), les microARN (miARNs), les pseudogènes, les ARN antisens, les courts ARN etc. Globalement, leurs fonctions biologiques ne sont pas bien caractérisées. La littérature rapporte un rôle régulateur des ARN non codants dans l'expression génique. Certains lncRNA régulent l'architecture des chromosomes, la transcription, la stabilité de l'ARN, et la traduction grâce à leurs domaines de liaison à l'ARN (Ji *et al.*, 2015; Yao, Wang and Chen, 2019). Les pseudogènes, homologues non fonctionnels d'un gène, peuvent réguler l'expression des gènes en agissant comme des concurrents des miARNs (Pink *et al.*, 2011).

L'estimation du nombre de gène humain est aussi délicate que sa propre définition. Si l'on considère la définition initiale, qui est une séquence spécifique d'ADN transcrite en ARNm puis traduite en protéine, le nombre de gène est estimé aux alentours de 20 000. Dans le cas où l'on considère comme gène la totalité des transcrits traduits ou non en protéine, le nombre s'élève alors à environ 200 000 gènes (Salzberg, 2018). Vingt-deux ans après la première publication du génome humain, nos connaissances sur la capacité de codage du génome humain restent encore incomplètes.

Durant de nombreuses années, le génome non codant a été ignoré, considéré comme non fonctionnel. Le dogme était : « Un gène = un ARN = une protéine » ; et le terme provocateur d'ADN poubelle pour les séquences non codantes fait même son apparition dans les années 60. En 1977, les groupes de recherche des D^{rs} Sharp et Roberts sont les premiers à remettre en cause le dogme avec la découverte de l'épissage alternatif (Berget, Moore and Sharp, 1977; Chow *et al.*, 1977). L'élimination des introns lors de la maturation des transcrits donne lieu à plusieurs recombinaisons possibles entre les exons, permettant ainsi le codage de différentes protéines à partir d'un gène unique. C'est d'ailleurs un des mécanismes qui augmente la capacité de codage du génome humain. Malgré sa grande taille (environ 3 milliards de paires

de bases), il a été estimé que seulement 1.5% de notre génome coderait pour des protéines fonctionnelles (ENCODE Project Consortium, 2012; Nurk *et al.*, 2022; Rhie *et al.*, 2022).

En plus du dogme, le génome humain a longtemps été annoté selon les caractéristiques des ORFs notamment celle d'une taille de minimum 100 acides aminés. Ce critère repose sur l'hypothèse qu'une courte séquence peptidique ne peut pas se replier et adopter une structure stable ; et donc, ne peut pas avoir de fonction. En conséquence, une multitude de régions génomiques de moins de 100 acides aminés, appelées « small ORFs » (sORFs), n'ont pas été annotées. Pour la majorité d'entre elles, l'absence d'une démonstration fonctionnelle appuya d'autant plus leur caractère non codant (Basrai, Hieter and Boeke, 1997; Ruiz-Orera and Albà, 2019; Chen *et al.*, 2020; Papadopoulos, 2022). Au courant des années 2010 plusieurs équipes pionnières dans le domaine révèlent, d'abord chez la levure puis chez le poisson zèbre, que ces courtes régions génomiques ont une signature traductionnelle (Ingolia *et al.*, 2009a; Bazzini, Lee and Giraldez, 2012; Smith *et al.*, 2014). Cela s'étend chez les mammifères, avec l'identification d'une activité traductionnelle en dehors des gènes codant pour des protéines annotées (Ingolia, Lareau and Weissman, 2011; Ingolia *et al.*, 2014; Ruiz-Orera *et al.*, 2014; Ji *et al.*, 2015; Raj *et al.*, 2016; Ruiz-Orera and Albà, 2019; Martinez *et al.*, 2020). On ne peut plus ignorer l'existence et l'intérêt biologique de ces séquences. Si elles sont traduites c'est qu'elles ont peut-être une fonction, ou du moins, confèrent un avantage biologique.

1.8 Nomenclature des régions génomiques non canoniques

Suite à ces découvertes, plusieurs groupes de recherches se sont mis à étudier ces régions non canoniques, et par manque d'une dénomination universelle, la confusion entre les diverses appellations est vite arrivée. Le terme sORFs regroupe donc toutes les séquences non canoniques, courtes, initialement décrites comme inférieures à 100 acides aminés, bien que certaines études rapportent une taille supérieure de certains sORFs (Andrews and Rothnagel, 2014). Les sORFs sont répartis en plusieurs sous-groupes. Parmi eux, on retrouve les sORFs codés dans les lncRNAs que je ne développerais pas dans la suite de ce chapitre. Je souhaite focaliser mon travail bibliographique sur toutes les sORFs qui ont un lien direct avec une CDS/ORF. Pour les classer, la première subtilité est alors de bien faire la dichotomie entre les séquences qui sont en phase avec la phase de lecture de l'ORF et celles qui ne le sont pas (Figure 10). On retrouve alors :

Les sORFs traduits en phase de lecture de l'ORF, appelés aussi « in-frame ORFs » :

- Ce sont des séquences étendues ou déplétées en région N- ou C-terminale de l'ORF, c'est-à-dire des isoformes de protéines canoniques.

Les sORFs traduits dans une phase de lecture différente de celle de l'ORF, appelés aussi « out-of-frame ORFs » :

- Ce sont des séquences qui résultent d'un changement de phase de lecture par rapport à l'ORF. Leur phase de lecture chevauche alors celle de l'ORF. Elles sont communément appelées « ARFs » pour « alternative reading frames ». Leur phase de lecture étant décalée de 1 ou 2 nucléotides par rapport à l'ORF, les produits de traduction issus des ARFs ont une séquence en acides aminés totalement différente de celle de l'ORF. Les

ARFs se divisent en 3 sous-groupes : ceux localisés à l'intérieur d'un ORF, ceux qui s'étendent au-delà de l'ORF en 3' sur une séquence non annotée, et ceux qui chevauchent un épissage alternatif.

- Autre subtilité, il existe des sORFs localisées au sein des 5'UTR et 3'UTR, que l'on appelle « upstream reading frames » (uORFs) et « downstream reading frames » (dORFs), respectivement. Ces séquences ne sont donc pas dépendantes de la CDS, et sont communément classées dans les « out-of-frame ». En revanche, les uORFs et dORFs, peuvent être aussi considérés comme « in-frame » s'ils s'avèrent transcrits dans la même phase de lecture que leur ORF respectif. Ces sORFs se divisent en deux groupes : ceux spécifiquement localisés dans de la région UTR et ceux qui chevauchent l'ORF, créant une séquence chimère uORF-ORF ou ORF-dORF.

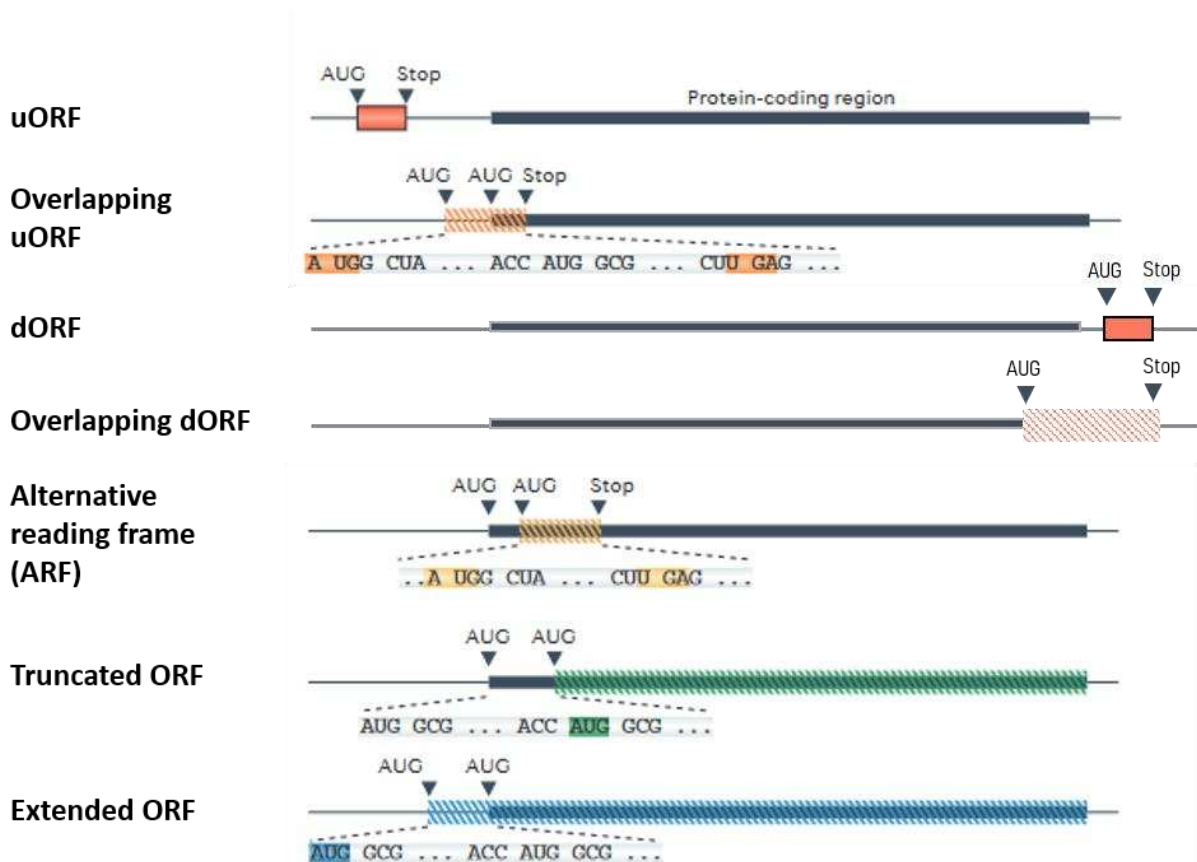


Figure 10. Nomenclature des sORFs (figure adaptée de Ingolia et al. 2014). Des sites alternatifs d'initiation de la traduction peuvent être trouvés dans le 5'UTR et peuvent initier la traduction d'un court cadre de lecture ouvert en amont (rectangle orange). Ces uORF peuvent chevaucher le codon initiateur du gène codant pour la protéine canonique, ce qui entraîne une traduction hors cadre qui s'étend dans la séquence codante. La traduction dans un cadre de lecture alternatif peut également commencer au sein du gène codant pour la protéine (rectangle jaune). Un codon AUG en phase de lecture d'un gène peut conduire à une protéine tronquée (rectangle vert), tandis qu'un AUG en phase lecture situé en amont d'un gène sans codon stop intermédiaire peut produire un produit protéique étendu (rectangle bleu). Les régions ombragées représentent les séquences alternatives d'ARNm qui sont traduites.

1.9 Les mécanismes non canoniques de traduction

Une des raisons pour laquelle les régions non annotées ont longtemps été considérées comme impertinentes c'est qu'elles n'initient pas nécessairement leur traduction avec un codon AUG.

En effet, même si la traduction eucaryote est caractérisée comme coiffe dépendante, d'autres mécanismes non canoniques permettent l'expression et la régulation des transcrits. La section suivante détaille l'expression des transcrits selon les mécanismes non canoniques de traduction (Figure 11). Beaucoup de ces découvertes reposent sur l'étude des génomes viraux et de leur traduction.

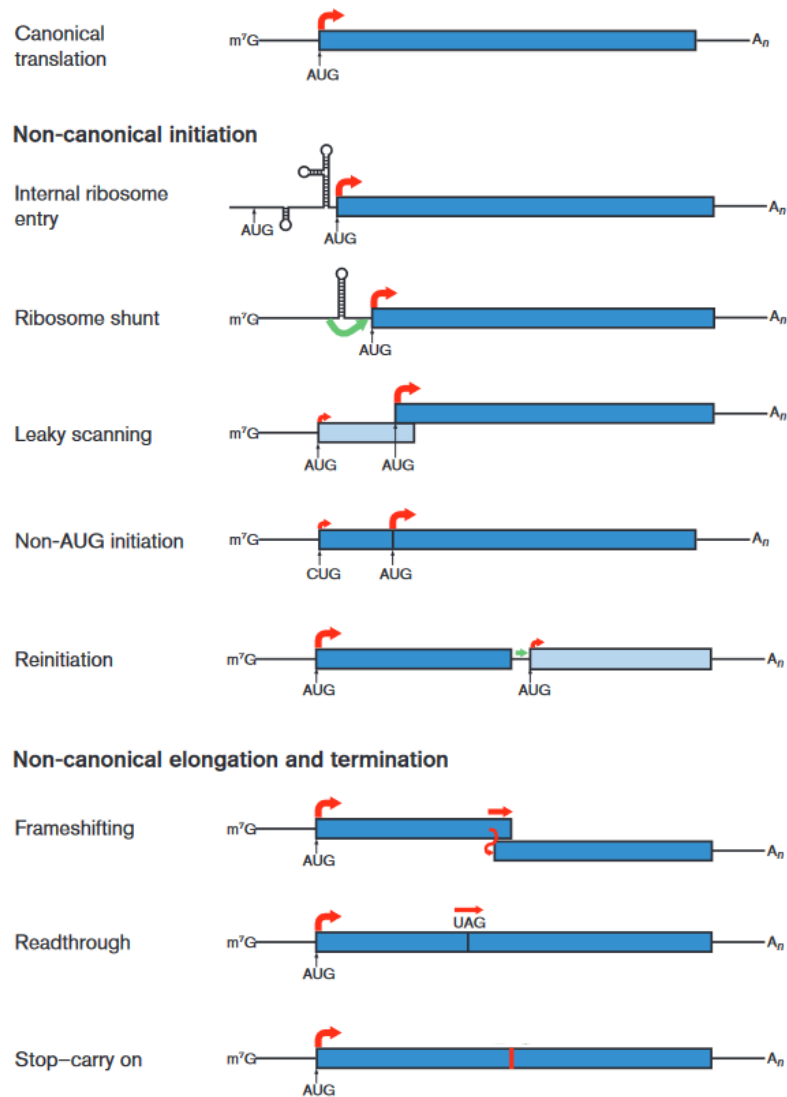


Figure 11. Exemples de mécanismes de traduction non canoniques (figure originale de Firth and Brierley 2012). Les flèches rouges indiquent l'initiation de la synthèse protéique. L'épaisseur des flèches est proportionnelle à la voie empreintée par les ribosomes. Les flèches vertes indiquent le déplacement probable des sous-unités 40S de manière non canonique. Lorsque deux polypeptides distincts sont synthétisés, les ORFs sont représentés dans des nuances de bleu différentes ; lorsqu'un événement de recodage au cours de l'élongation entraîne l'extension C-terminale d'un polypeptide, les deux ORFs sont représentés dans la même couleur. Il est important de noter ici, que l'ORF traduit après un frame shifting peut également encoder un polypeptide distinct (qui pourrait être représenté avec une nuance de bleu différente).

1.9.1 Le leaky scanning

Le leaky scanning correspond à une absence de reconnaissance du codon AUG par le ribosome. Dans le cas d'un contexte défavorable de traduction, le complexe mature d'initiation 80S continue son scanning au-delà du premier codon AUG de la CDS ou de l'uORF pour initier au

codon start suivant (Benitez-Cantos *et al.*, 2020). Le leaky scanning est aussi favorisé par la présence de codons start alternatifs, ou « near-cognate AUG », connus pour initier la traduction de manière suboptimale. Le mécanisme de leaky scanning est d'avantage retrouvé dans l'expression de protéines isoformes tronquées en N-ter (Firth and Brierley, 2012; Andreev *et al.*, 2022). C'est un des mécanismes les plus couramment utilisé par les virus à ARN pour traduire des protéines polycistroniques. Par exemple, dans le VIH-1, les protéines Vpu et Env sont codées par le même transcrit. C'est l'ignorance par les ribosomes du codon AUG de *vpu*, présent dans un contexte faible de traduction, qui permet la traduction du codon start de *env* situé en aval (Schwartz, Felber, Fenyö, *et al.*, 1990; Anderson *et al.*, 2007).

1.9.2 Le codon read-through

Le codon read-through peut être considéré comme un leaky scanning intervenant à l'étape de terminaison de la traduction. C'est un mécanisme alternatif où la traduction s'étend au-delà du codon stop ce qui permet l'expression de transcrits non annotés en aval et en phase de l'ORF. L'absence de reconnaissance du codon stop est rendue possible par l'incorporation d'un ARNt « near-cognate » analogue avec le codon stop. Cette analogie permet de décoder le codon stop et de continuer la lecture jusqu'au prochain codon de terminaison. Chez la levure, le site A du ribosome a montré une sensibilité au mauvais appariement d'un ARNt near-cognate. Les deux résidus en position +4 et +5 influencent la sélection de ces ARNt homologues chez la levure comme chez l'humain (Hellen, 2018b). Toutefois, tous les fonctionnements du codon read-through n'ont pas été élucidés. Dans certains modèles, le codon read-through se base sur le recrutement de facteurs alors que dans d'autres, c'est la présence d'un motif particulier (autre que le résidu en +4) ou d'une structure complexe qui permet la pervasivité de la terminaison (Firth and Brierley, 2012; Hellen, 2018a). Le codon read-through est donc un mécanisme qui agit en cis- et en trans- pour permettre l'expression d'ORF alternatives prolongées en C-terminale.

1.9.3 Le ribosome shunting

Le ribosome shunting est un mécanisme dans lequel la traduction démarre à partir d'un court uORF situé en amont d'une longue région 5'UTR structurée en tige-boucle. Lors de la terminaison de la traduction de l'uORF, le ribosome est alors capable de contourner la structure en tige-boucle permettant la traduction d'un AUG en aval sans à avoir à dérouler la structure de l'ARN. C'est donc un mécanisme dépendant de la coiffe mais partiellement dépendant du scanning. Il est largement utilisé par les génomes viraux tels que : le virus de la mosaïque du chou-fleur, l'adénovirus humain 5, et le papillomavirus humain (HPV). Il est également proposé comme modèle pour initier la traduction de la protéine de l'enveloppe du VIH-1 à partir de la séquence 5'UTR du gène *vpu* (Krummheuer *et al.*, 2007; Firth and Brierley, 2012; Guerrero *et al.*, 2015; Sorokin *et al.*, 2021).

1.9.4 La ré-initiation

La ré-initiation a été mise en évidence pour la première fois il y a une quarantaine d'années par le Dr Kozak (Kozak, 1984, 1987). Comme son nom l'indique, elle permet de recruter le ribosome pour une nouvelle initiation. Elle est en grande partie décrite dans les uORFs bien

qu'elle soit aussi effective pour réinitier à la fin d'une CDS. Les données de la littérature suggèrent que la ré-initiation est majoritairement utilisée par les virus pour initier des régions génomiques chevauchantes. Les différents mécanismes de ré-initiation, dont un grand nombre restent à élucider, sont présentés succinctement ci-dessous :

- La ré-initiation canonique impliquant les facteurs d'initiations. Dans ce cas de figure, la ré-initiation se produit uniquement si la sous-unité 40S reste associée au transcrit après l'action de ABCE1 dans la libération des facteurs de recyclage. Cette ré-initiation s'effectue généralement suite à la traduction d'une région génomique très courte. Le taux de ré-initiation a été montré comme négativement corrélé à la longueur de l'uORF et à la durée de l'élongation. La distance qui sépare les deux régions à traduire est également importante pour permettre la ré-initiation. Il faut que le ribosome ait le temps de réacquérir certains facteurs d'initiation comme le complexe ternaire PIC 43S. eIF3, retenu pendant l'élongation, joue aussi un rôle majeur dans cette ré-initiation en stabilisant l'association 40S/ARNm (Firth and Brierley, 2012; Hellen, 2018b).
- La ré-initiation impliquant des facteurs d'initiation non canoniques tels que DENR-MCTS1 et eIF2D. DENR-MCTS1 et eIF2D ont été précédemment décrits comme des acteurs spécialisés dans la ré-initiation, agissant sur un sous-ensemble d'ARNm liés à des fonctions spécifiques (Janich *et al.*, 2015; Castelo-Szekely *et al.*, 2019; Bohlen *et al.*, 2020). Par exemple, DENR-MCTS1 s'est avéré favoriser la ré-initiation de transcrits oncogènes tel que ATF4 (Bohlen *et al.*, 2020). L'implication de son homologue, eIF2D, dans la ré-initiation de transcrits possédant des uORFs a également été montrée (Vasudevan *et al.*, 2020). Ces facteurs jouent un rôle clé en déplaçant l'ARNt désacétylé du site P lors de la terminaison et en recrutant un nouvel ARNt initiateur (Dmitriev *et al.*, 2010). Malgré les avancées, de nombreux aspects du mécanisme de ré-initiation médié par ces facteurs non canoniques restent à élucider, notamment : les transcrits ciblés et la spécificité de l'activité de ré-initiation (est-elle limitée aux uORFs ?). De plus, il reste à déterminer si ces facteurs sont capables d'induire le scanning et quels sont leurs partenaires potentiels.
- La ré-initiation par l'implication de « Termination upstream ribosome binding site » (TURBS). Dans ce processus, le ribosome est forcé de réinitier au lieu de se détacher du transcrit. La présence de ces sites de liaisons du ribosome à proximité du codon stop capture la petite sous-unité 40S. Dans le cas du virus de la grippe B (IBV), cela permet une ré-initiation si un codon start est disponible à proximité (Sorokin *et al.*, 2021).
- La ré-initiation par l'implication de protéines virales. Elle est principalement décrite dans les virus de plantes. En bref, la protéine virale TAV des pararétrovirus de plantes interagit dès l'initiation avec le facteur eIF3 engendrant une cascade d'activation de facteurs cellulaires. Les facteurs cellulaires recrutés lient les 2 sous-unités 40S et 60S permettant leur maintien et une meilleure ré-initiation. La rétention du facteur eIF3 durant l'élongation favorise également l'attachement de la sous-unité 40S au transcrit après l'étape de terminaison. (Sorokin *et al.*, 2021).

1.9.5 Les IRES

Les « internal ribosome entry sites » (IRES) induisent l'initiation de la traduction par le recrutement des ribosomes à des positions internes sur les transcrits. Ces structures complexes de type tige-boucle ont été initialement identifiées en 1988 chez le poliovirus et le virus de l'encéphalomyocardite (Jang *et al.*, 1988; Pelletier and Sonenberg, 1988). Plus largement, les IRES sont ensuite retrouvés chez le VIH, l'HCV, et le virus de la fièvre aphteuse (Buck *et al.*, 2001; Yang and Wang, 2019). Les virus utilisent les IRES à leur profit pour 1) initier la traduction de leur protéine indépendamment de l'initiation des transcrits cellulaire, 2) poursuivre la traduction des ARNm viraux quand la traduction de l'hôte est inhibée, et 3) accéder à des ORFs qui seraient autrement inaccessibles, c'est-à-dire qui ne possèdent pas de coiffe à leur extrémités 5'UTR (Firth and Brierley, 2012).


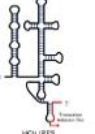
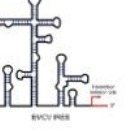

	Group I	Group II	Group III	Group IV
IFs	None	eIF2 eIF3	eIF2 eIF3 eIF4A eIF4B eIF4G	eIF2 eIF3 eIF4A eIF4G
ITAFs	None	None	PTB ITAF ₄₅	PTB PCBP1 PCBP2
40S recruitment	Direct binding to 40S ribosomal subunit, no IFs requirement	Direct binding to 40S ribosomal subunit, 40S does not scan	Binding to 40S ribosomal subunit through ITAFs, 40S does not scan	Binding to 40S ribosomal subunit through ITAFs, 40S scanning to the downstream AUG
Examples	CrPV, PSIV, TSV	HCV, CSFV, PTV-1	EMCV, FMDV, TMEV	PV, rhinovirus
Secondary Structure				

Figure 12. Classification des IRES viraux (figure originale de Yang and Wang 2019). Quatre groupes d'IRES viraux sont répertoriés. Ils sont subdivisés selon leurs structures et l'exigence en ITAF et eIF.

L'IRES initie la traduction indépendamment de la coiffe mais requière souvent la participation de facteurs supplémentaires tels que les facteurs d'initiation et les facteurs agissant sur l'IRES appelés « IRES-transacting factors » (ITAFs). Les IRES sont classés en 4 sous-groupes selon leur degré de dépendance vis-à-vis des facteurs d'initiation et des ITAFs (Figure 12). A l'image d'une traduction canonique, les IRES du groupe II, III et IV sont dépendants de la présence d'un codon AUG, du recrutement d'un ARNt-Méth, des facteurs eIF, et des ITAFs pour recruter la sous-unité 40S et démarrer la traduction. La différence entre le groupe III et IV réside dans le besoin ou non de scanning. Les IRES de groupe III recrute et fixe la petite sous unité du ribosome directement au codon initiateur, le scanning n'étant alors pas nécessaire. Pour les IRES du groupe IV, la petite sous unité balaie la région non traduite dans une direction 5' à 3' jusqu'à la rencontre du codon AUG (Yang and Wang, 2019). A contrario, les IRES du groupe I ont une forte activité et sont capables d'initier sans codon AUG, ITAFs ou bien même sans complexe ternaire eIF2/ARNt-Méth/GTP. Les IRES du groupe II, nécessitent l'implication de quelques

facteurs d'initiation mais n'implique pas l'ARNt-Méth. Plus la structure d'un IRES est complexe moins elle dépend de trans-facteurs pour initier la traduction (Yang and Wang, 2019).

En condition de stress, la traduction des protéines eucaryotes dépendantes de la coiffe est inhibée. Les gènes cellulaires susceptibles à une traduction par IRES sont donc majoritairement ceux impliqués dans la réponse à un stress cellulaire (apoptose, mitose, hypoxie, restriction nutritionnelle) ou à une infection virale. A l'inverse des IRES viraux, les IRES cellulaires partagent peu d'homologie de structure rendant leur classification plus difficile. Ils sont néanmoins catégorisés en 2 groupes en fonction de leur dépendance ou non aux ITAFs pour le recrutement de la petite sous-unité du ribosome. Une étude de 2016 a montré par l'utilisation de rapporteur *in vivo* qu'environ 10% des transcrits de humains contiennent des éléments structuraux similaires aux IRES (Weingarten-Gabbay *et al.*, 2016).

1.9.6 Le décalage de phase de lecture ou ribosomal frame shifting (FS)

Le décalage de phase de lecture est un décalage programmé du ribosome d'un nucléotide en aval (+1 FS) ou en amont (-1 FS) de la phase de l'ORF. Ce décalage génère donc un changement de phase de lecture par rapport à l'ORF. Grâce à ce changement de phase de lecture, un gène peut coder différentes protéines, l'une codée dans la phase de l'ORF, et l'autre, codée dans une des deux phases alternatives. Les deux protéines issues de ce même ORF sont généralement fonctionnelles. C'est un phénomène que l'on retrouve dans tous les génomes et qui permet d'enrichir le protéome alternatif (Namy *et al.*, 2004; Ketteler, 2012). Historiquement le FS a été mis en évidence dans les génomes viraux et les rétrotransposons. De nombreux virus à ARN requièrent le FS pour l'expression de leurs protéines (le plus souvent une réplicase) et leur infectiosité. Un exemple très connu est la traduction de la polyprotéine Gag-Pol du VIH-1 qui nécessite un -1 FS (le FS du VIH-1 est davantage développé dans le chapitre 3). C'est d'ailleurs un mécanisme fréquent chez les rétrovirus pour la synthèse du précurseur Gag-Pol à partir duquel dérive la transcriptase inverse (Brierley, 1995).

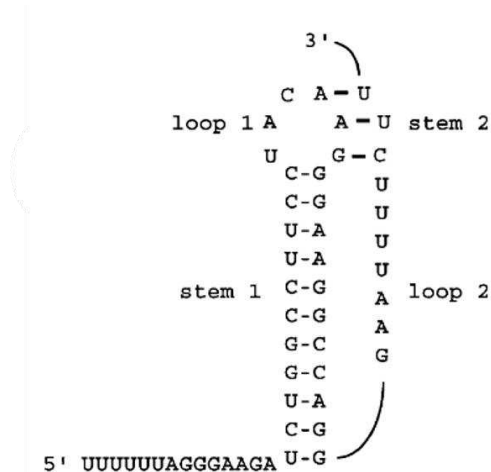


Figure 13. Exemple de sites de décalage programmé du ribosome (FS) du VIH-1 (figure originale de Brierley and Dos Ramos 2006). La structure en tige-boucle originale proposée par Jacks et al. 1988b est représentée avec la séquence glissante UUUUUU en amont.

L'utilisation du FS est également bien renseignée chez les coronavirus responsables du syndrome respiratoire sévère (SARS-CoV) (Brierley and Dos Ramos, 2006). Le FS dépend de la présence d'une séquence glissante de 1 à 15 nucléotides et d'une structure secondaire de type tige-boucle ou pseudonoeud (« pseudoknot » en anglais) (Figure 13). La séquence glissante est composée de deux triplets homopolymériques construits selon ce motif : **XXXYYYZ** où X peut-être n'importe quel nucléotide, Y un A ou un U, et Z n'est pas un G (Brierley and Dos Ramos, 2006).

La combinaison de ces deux éléments perturbe le décodage ce qui génère une pause du ribosome puis un changement de phase de lecture.

Le décalage de phase en -1 et +1 FS est également employé dans de nombreux gènes cellulaires. C'est le cas par exemple pour le gène codant l'IL10 et les antienzymes (Jacks *et al.*, 1988; Namy *et al.*, 2004; Ketteler, 2012).

1.9.7 Les codons start alternatifs ou codons non-AUG

Comme leur nom l'indique, les codons start alternatifs sont des codons initiateurs qui portent un autre acide aminé que la méthionine. Les deux acides aminés les plus répandus sont : la leucine et la valine, principalement codés par les triplets « CUG », « UUG », et « GUG », respectivement. De nombreuses études définissent le codon « CUG » comme le codon le plus efficace pour une initiation non-AUG. Contrairement aux génomes bactériens l'initiation de la traduction par codons alternatifs a longtemps été considérée comme rare chez l'humain. La communauté de la traduction des ARNm s'est longtemps accrochée au modèle de scanning et d'initiation canonique (Apcher *et al.*, 2022). Le laboratoire du D^r Shastri est le premier à avoir mis en évidence qu'après la lecture d'un codon « CUG » il y a incorporation d'une leucine, et non pas d'une méthionine comme précédemment admit dans la communauté (Schwab *et al.*, 2004; Starck *et al.*, 2012). Il dira par la suite avoir rencontré beaucoup de difficultés à convaincre les rapporteurs du papier décrivant ce mécanisme (Apcher *et al.*, 2022). Plus tard dans le modèle bactérien, l'analyse *in silico* de plus de 600 génomes révèle que 12% des gènes annotés initient avec un GUG, et 8% avec « UUG » (Villegas and Kropinski, 2008). L'étude de l'efficacité de traduction par l'utilisation de gènes rapporteurs a montré que les codons « UUG » et « GUC » ont une efficacité de traduction de 12 à 15% par rapport à l'AUG ; les codons CUG, AUU, AUC, AUA, et ACG ont des rendements protéique plus faible allant de 1 à 3% (Sussman, Simons and Simons, 1996). Dans une étude de screening, l'efficacité de traduction s'est révélée meilleure quand un U est à la deuxième position du triplet (Hecht *et al.*, 2017). Chez les eucaryotes, l'utilisation de gènes rapporteurs ultra-sensibles démontrait une efficacité de ces codons alternatifs allant de 1.7 à 19.5% dans des cellules 293T (Ivanov *et al.*, 2010). Par le biais de différentes constructions et d'immunoprécipitations, plusieurs équipes de recherches ont démontré que les protéines eucaryotes FGF-2, VEGF, et c-myc sont initiées non pas par un codon méthionine mais bien à partir d'un codon leucine « CUG » (Kearse and Wilusz, 2017; Cao and Slavoff, 2020).

L'initiation non-AUG peut se produire lorsque le contexte d'initiation AUG est faible (typiquement le leaky scanning si le codon en aval est un codon alternatif), ou s'il y a un déficit en ARNt-Méth. Le mécanisme d'initiation non-AUG se base sur la reconnaissance et l'appariement d'ARNt near-cognates aux codons alternatifs de démarrage. L'exactitude du mécanisme reste à élucider. Cependant, la présence 1) d'un contexte nucléotidique, 2) des facteurs d'initiations eIF1 et eIF5, 3) des hélicases, et 4) le taux d'élongation, impactent directement la lecture de codons non-AUG (Andreev *et al.*, 2022).

Le contexte d'initiation est différent entre un codon AUG et un non-AUG. Ici, la nature des résidus en position -4 et +4 est cruciale pour une bonne efficacité d'initiation contrairement à l'initiation AUG où se sont les positions -3 et +4 qui prédominent. De récentes études mettent en lumière que dans un contexte d'initiation optimal, l'efficacité de traduction d'un non-AUG est comparable à celle d'un AUG (Andreev *et al.*, 2022). Un contexte d'initiation optimal

correspondrait à l'enchaînement de nucléotides suivant : **(A/C), (A/G), (C/A)**, (A/C/G), CUGG. Une étude phylogénétique eucaryote soutient l'importance de la présence des résidus C, A, et G en position -4, -3, et +4, respectivement (Loughran *et al.*, 2020). Cette observation remet en cause l'opinion dominante jusqu'à présent décrite comme quoi les codons non-AUG initieraient seulement à une efficacité de 5-10% de celle de l'AUG (Andreev *et al.*, 2022).

La littérature est unanime sur l'influence de la quantité des facteurs eIF1 et eIF5 dans la discrimination des codons AUG et non-AUG. Chacun des deux facteurs s'autorégule par boucle rétroactive négative mais ce réseau de régulation est interconnecté. eIF1 favorise l'initiation AUG alors que eIF5 favorise la traduction des codons non-AUG (Arias *et al.*, 2014). La balance entre les deux facteurs est donc capitale et finement régulée. Cette balance entre eIF1 et eIF5 permet de réguler et de rendre l'initiation plus stricte en réponse au stress. Une altération dans eIF1 engendre une diminution de sa liaison à eIF2 et favorise la reconnaissance de codon non-AUG leucine et valine par le complexe PIC (Andreev *et al.*, 2008; Fijałkowska *et al.*, 2017; Thakur, Marler and Hinnebusch, 2019). L'initiation par la reconnaissance d'un codon « CUG » a été montrée comme indépendante du facteur d'initiation eIF2 mais dépendante du facteur eIF2A (Schwab *et al.*, 2004; Starck *et al.*, 2012).

Les hélicases influencent également l'initiation non-AUG. Elles permettent le déroulement des structures de l'ARNm. Lorsque ces structures sont caractérisées de « difficiles » ou qu'il y a une absence d'hélicases, cela augmente le pausing du ribosome. La présence de codons rares, l'indisponibilité d'ARNt chargés ou le ralentissement médié par la formation de la chaîne polypeptique sont aussi des événements qui induisent des pauses ribosomiques. Le pausing a été montré comme favorisant la ré-initiation aux codons non-AUG en amont (Andreev *et al.*, 2022). Lié au pausing, une vitesse d'élongation faible engendrera également une initiation en amont de la séquence traduite.

Les études récentes sont catégoriques : la capacité de codage des génomes dépasse largement nos précédentes estimations. En 2009, l'arrivée du ribosome profiling bouleverse la communauté scientifique. Par l'obtention d'empreintes ribosomales *in vivo*, Ingolia et al démontrent que 20% des régions traduites dans le 5'UTR de la levure initient avec un codon non-AUG (Ingolia *et al.*, 2009a). Ils rapportent même que les uORFs, en condition privative, ont une efficacité de traduction six fois supérieure à celle retrouvée en phase de croissance. Cette efficacité de traduction est plus forte que celle retrouvée pour les ORFs. Le phénomène est d'autant plus fort si l'initiation se fait par un codon alternatif ; principalement les codons Ile « AUA » et leucine « CUG ». En 2011, la même équipe reproduit ces résultats dans le modèle murin (Ingolia, Lareau and Weissman, 2011). En 2012, Lee et al décrivent qu'environ la moitié des transcrits humains traduits sont initiés avec un codon start alternatif avec une prédominance de 16% pour la leucine « CUG » (Lee *et al.*, 2012).

La sensibilité au nucléotide près du ribosome profiling en fait un formidable outils pour analyser, *in vivo* et à l'échelle du génome, les mécanismes de traduction et la diversité du protéome. Etant au centre de ma thèse, la technique du ribosome profiling et ses apports scientifiques sont davantage développés dans la section suivante.

1.10. Le ribosome profiling

1.10.1 Historique et technique

L'analyse de l'expression génique a longtemps reposé sur la détermination de la quantité des transcrits par la technique de Northern blot et plus récemment par séquençage de l'ARN (RNA seq). L'importance de la traduction dans l'expression des gènes a motivé le développement d'approches expérimentales pouvant répondre aux nombreuses questions mécanistiques et régulatrices de la traduction. L'avènement de la génomique et du profilage de l'expression génique a inspiré la conception de méthodes similaires appliquées à la traduction. Le Ribosome Profiling permet l'identification *in vivo* de tous les ORFs et sORFs (uORF, ARFs, etc.) traduits à l'intérieur d'une cellule. On définit le translatome comme l'ensemble de ces événements de traduction. Comme le disent Gloria A Bar et Jonathan S Weissman dans leur revue, le ribosome profiling indique quelles, quand, où et comment les protéines sont traduites (Bar and Weissman, 2015).

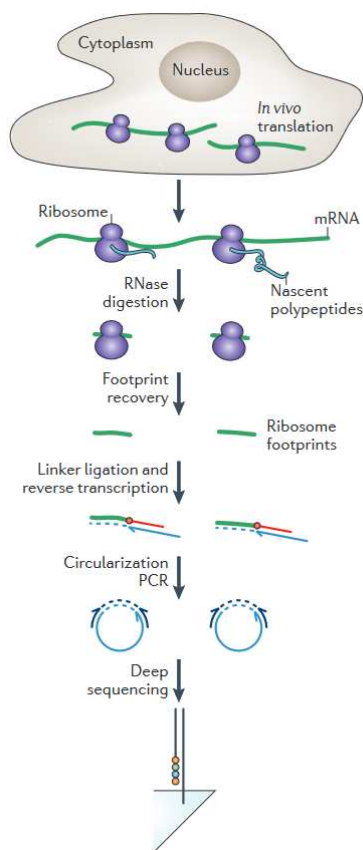


Figure 14. La technique du ribosome profiling (figure originale de Ingolia 2014).

Le ribosome profiling se base sur la méthode moléculaire classique d'isolation d'empreintes ribosomales *in vitro* développée dans le courant des années soixante-dix (Steitz, 1969; Wolin and Walter, 1988). La méthode classique consiste à traiter à la RNase des transcrits isolés *in vitro* dans le but d'éliminer les régions qui ne sont pas protégées par les ribosomes. Ces premières investigations ont dévoilé que les fragments protégés par les ribosomes, appelés « Ribosomal Protected Fragments » (RPF), mesurent en moyenne 30 nucléotides. Mais les études ultérieures ont permis de préciser que leur longueur varie entre 28 et 32 nucléotides. En 2009, l'équipe du D^r Weissman est la première à combiner cette méthode avec l'analyse de séquençage à haut-débit (Ingolia *et al.*, 2009a). Le ribosome profiling repose sur l'isolation directe et *in vivo* des transcrits en cours de traduction. Une banque d'ADN complémentaire suivit d'un séquençage de nouvelle génération (NGS) sont ensuite réalisés sur les RPF (Figure 14). *S.cerevisiae* a été le premier organisme étudié par ribosome profiling (Ingolia *et al.*, 2009a).

En 2012, D^r Nicholas T Ingolia, affilié au laboratoire du D^r Weissman, rend accessible le protocole du ribosome profiling alors publié dans Nature Protocols (Ingolia *et al.*, 2012). L'utilisation du séquençage à haut débit est commune entre le RNAseq et le ribosome profiling. Les deux techniques diffèrent dans la nature des fragments isolés. Dans le RNAseq l'ensemble des transcrits, qu'ils soient ou non en cours de lecture, sont analysés. Dans le ribosome profiling, l'isolation des fragments uniquement protégés par les ribosomes permet

d'obtenir l'ensemble des transcrits en cours de traduction *in vivo*. Les différences entre les deux techniques sont présentées dans la figure 15.

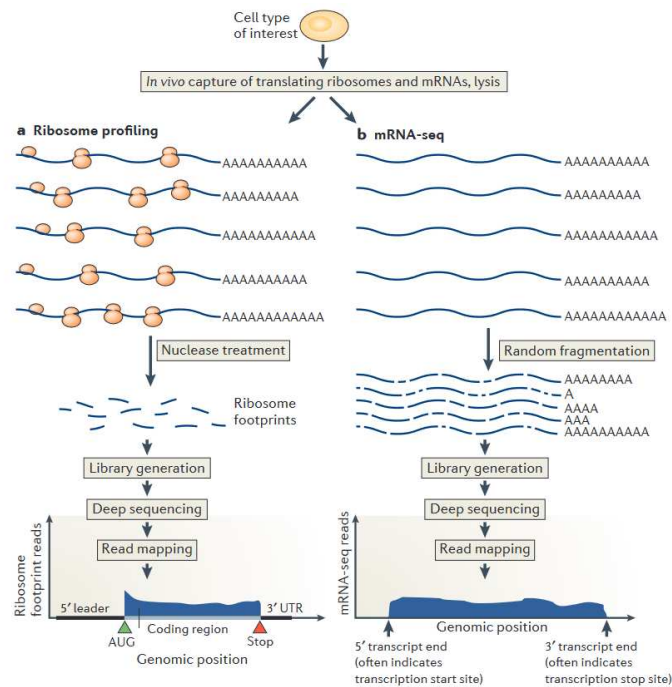


Figure 15. Différences entre le RiboSeq et RNAseq (figure originale de Brar and Weissman 2015). (A) Les ARNm liés aux ribosomes sont isolés en fonction de leur taille et traités avec une nucléase non spécifique (typiquement la RNase I), générant des fragments d'ARNm protégés appelés "empreintes". Ces empreintes ribosomales sont isolées et rétrotranscrites en une banque d'ADN complémentaire qui est ensuite séquencée à haut-débit. Les empreintes ribosomales permettent d'identifier de façon précise le codon de départ et le codon d'arrêt d'un gène, ce qui facilite l'identification des régions génomiques. En comparaison, le RNAseq capture des fragments aléatoires couvrant la totalité des transcrits. L'information positionnelle déterminée par un RNAseq standard permet de déterminer approximativement les limites du transcrit, mais elle est moins précise que celle recueillie par le profilage des ribosomes, en raison de la perte des extrémités 5' et 3' pendant la méthode de génération des fragments qui est généralement utilisée lors du RNAseq.

Dans un premier temps tous les ARNm, qu'ils soient nus ou protégés par les ribosomes, sont récupérés à partir d'un lysat cellulaire. Un ARNm qui est couvert par un seul ribosome est appelé monosome et polysome s'il est couvert de plusieurs. Le traitement à la RNase permet ensuite l'élimination des fragments d'ARNm qui ne sont pas protégés par les ribosomes. Dans un second temps, une dénaturation des ribosomes isole les RPF qui sont ensuite retro-transcrits en une banque d'ADN complémentaire puis séquencés (NGS), on parle alors de « RiboSeq ». Le ribosome profiling se termine avec le traitement puis l'alignement des empreintes ribosomales sur le génome correspondant. Cette dernière étape est réalisée par le biais d'outils bio-informatique déjà disponibles pour le RNAseq (Figure 16).

Le principal défi du ribosome profiling est d'inhiber rapidement la traduction afin d'avoir une vision d'ensemble des ribosomes à un instant et pour un état physiologique donné. La cycloheximide (CHX) est un antibiotique qui a été largement utilisé dans les premières expériences de ribosome profiling pour figer et stabiliser les ribosomes en cours d'élongation. La CHX fixe le site E de la large sous-unité, proche de la position où se lie l'ARNt désacétylé. La présence de la CHX au site E bloque la translocation et la libération de l'ARNt désacétylé figeant alors l'élongation. L'utilité et l'impact de la CHX sur les événements de traduction a soulevé un

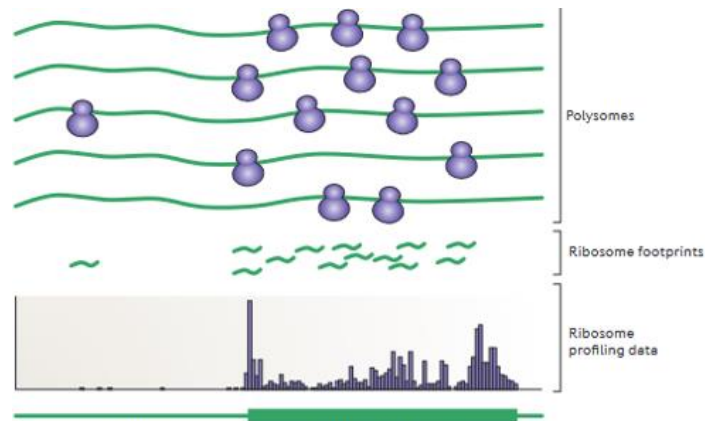


Figure 16. Profil d'empreintes ribosomales ; le translome (figure originale de Ingolia 2014). Les fragments d'ARNm protégés par les ribosomes sont obtenus par un traitement avec une ou plusieurs nucléases. Les fragments d'ARN protégés sont ensuite isolés des ribosomes par dénaturation. Les empreintes ribosomales reflètent la densité des ribosomes à chaque position sur le transcrit. L'annotation du génome est représenté par la ligne verte inférieure. La partie épaisse représente la séquence codante pour la protéine et la ligne fine représente la région annotée comme non codante.

désaccord dans la communauté. En effet, si l'inhibition de l'élongation par la CHX n'est pas complète, cela peut entraîner une accumulation artificielle des ribosomes à des positions spécifiques. Cette pause, bien qu'artificielle, actionne le mécanisme de ré-initiation non-AUG et biaise le signal en amont (Kearse *et al.*, 2019; Santos *et al.*, 2019). Désormais plus couramment employée, la congélation rapide en azote liquide du culot cellulaire permet un arrêt instantané de la traduction et une distribution globale des ribosomes (Brar and Weissman, 2015).

Le ribosome profiling est une méthode d'une grande sensibilité mais dépendante de la profondeur de séquençage : jusqu'à 10 millions d'empreintes possibles dans une cellule de mammifère, permettant ainsi la détection des événements de traduction même les plus rares.

1.10.2 Analyse des données et périodicité

L'analyse bio-informatique contient plusieurs obstacles tels que l'alignement sur les régions épissées du génome, la possibilité d'alignements génomique multiples et la contamination en ARNr. Chaque laboratoire utilise son propre code informatique pour palier à ces problèmes et traiter les données. En revanche, ils reposent tous sur des étapes communes présentées ci-dessus (Figure 17).

Lors d'une expérience de ribosome profiling, il est important de vérifier si le signal obtenu est bien spécifique à un signal de traduction et pas à un signal de transcription comme lors d'un RNAseq. Pour cela, nous cherchons à identifier où se trouvent les sites de décodage du ribosome (A ou P) sur les ARNm en cours de traduction. Ceci s'appuie sur la détermination du « P-site offset » pour chaque taille de fragment, c'est-à-dire la longueur en nucléotides qui sépare l'extrémité 5' des RPF de la position du site P du ribosome correspondant (Figure 18A).

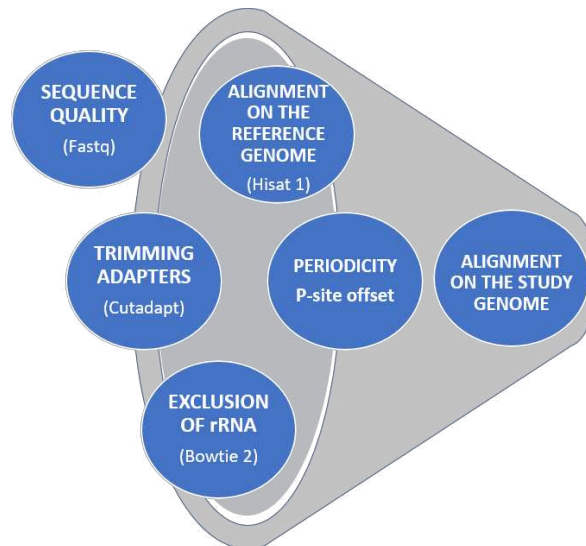


Figure 17. Etapes clés de l'analyse bio-informatique des données de ribosome profiling. Un exemple d'outil bio-informatique est indiqué entre parenthèse pour chaque étape. La dernière étape d'alignement est optionnelle si le génome de référence est différent du génome étudié. Typiquement, dans l'étude des génomes viraux par ribosome profiling le génome de référence, sur lequel est établi la périodicité, correspond le plus souvent au génome de l'hôte.

Pour une ORF donnée, tous les RPF sont alignés sur un génome de référence. Le but est de récupérer les RPF qui s'alignent sur le codon start AUG de l'ORF. L'initiation de la traduction étant un processus plus long que le décodage d'un codon sur le reste de l'ARNm, une surreprésentation des reads sur cet AUG est attendue. Ainsi, la taille des RPF pour laquelle il y a le plus fort enrichissement est retenue car elle correspond aux reads dont le site P du ribosome se trouve sur le codon start. La distance entre l'extrémité 5' de ces RPF et le début du codon start déterminera le P-site offset au sein de l'échantillon pour une taille de RPF donnée. Cet offset sera ensuite appliqué à l'ensemble des reads de même taille, quel que soit leur position sur le génome. Une fois ces P-site offsets déterminés, chaque read peut alors être représenté par une unique position correspondant à la première base de son site P théorique. Le décodage de l'ARNm par les ribosomes s'effectuant par triplets de nucléotides, la densité ribosomale devrait être enrichie tous les trois nucléotides et dans la phase de lecture des ORF. Les deux autres phases présumées non codantes seraient donc minoritairement représentées. (Figure 18B). L'expérimentateur devrait donc être capable d'observer ce signal périodique du ribosome. La périodicité est non seulement un contrôle qualité interne, mais aussi une analyse essentielle pour détecter des événements de traduction non canoniques tels que le décalage de phase de lecture ou la traduction des uORFs.

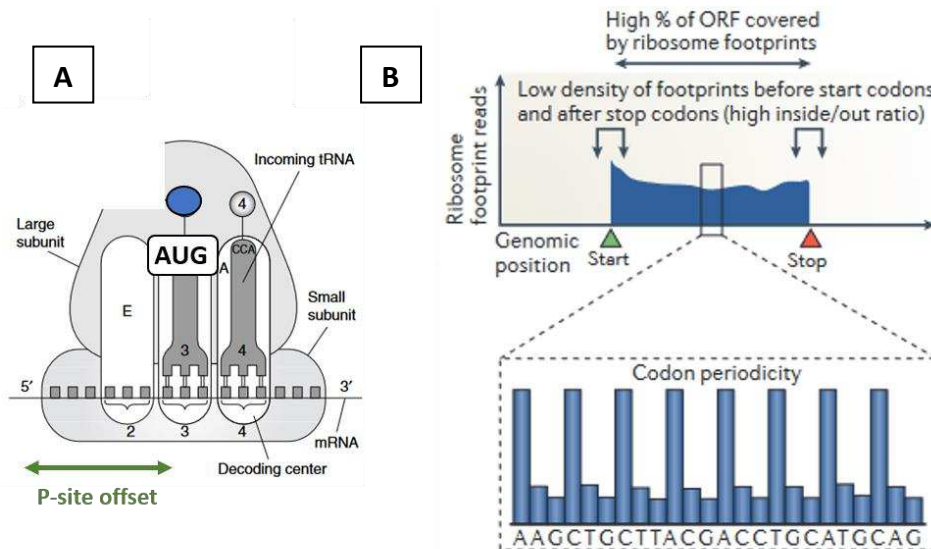


Figure 18. Périodicité (figure adaptée de Brar and Weissman 2015). (A) Détermination du P-site offset : distance qui sépare l'extrémité 5' du site P où se situe l'AUG. (B) La densité de l'empreinte ribosomale sur les ORFs commence au codon de départ et se termine au codon d'arrêt. Le signal est périodique si la couverture ribosomale est enrichie sur la phase de l'ORF, correspondant au déplacement codon par codon du ribosome sur le transcrit. Les deux autres phases alternatives sont moins représentées.

1.10.3 Identification des sites d'initiations

Le ribosome profiling permet également de déterminer et d'étudier les sites d'initiation, qu'ils soient situés dans des régions annotées ou non. Le ribosome effectue une pause dans son scanning suite à la rencontre des codons d'initiation et de terminaison. Classiquement, sur le profil d'empreinte ribosomale cela se retranscrit par une accumulation des reads aux codons start et stop, et une diminution des reads entre ces deux codons correspondant à l'élongation (Ingolia *et al.*, 2009a). L'utilisation d'antibiotiques, avant la congélation de l'échantillon, permet un enrichissement accru des RPF aux sites d'initiation. Deux antibiotiques sont couramment utilisés : l'harringtonine (HAR) et la lactimidomycine (LTM).

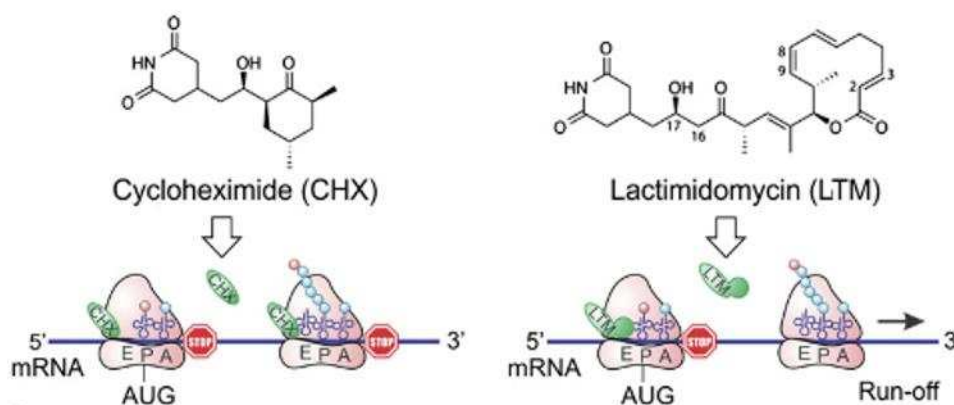
1.10.3.1 Harringtonine

L'harringtonine inhibe la traduction par sa fixation aux sous-unités 60S libres des ribosomes. Les sous-unités 60S libres fixent l'harringtonine et forment un complexe 80S qui sera bloqué aux sites d'initiation (Fresno, Jiménez and Vázquez, 1977; Ingolia, Lareau and Weissman, 2011). Ce mécanisme d'action sous-entend que les sous-unités 60S qui ont déjà formé le complexe mature d'initiation 80S sont résistantes à l'harringtonine. C'est pourquoi, l'harringtonine agit sur les ribosomes qui initient la traduction mais n'a pas d'impact sur ceux qui sont en cours d'élongation. Pour obtenir un enrichissement des ribosomes aux sites d'initiation, il est crucial de leur laisser suffisamment de temps pour terminer leur cycle d'élongation. Au moment du recyclage, l'harringtonine fixe alors la sous-unité 60S devenue libre qui ne pourra plus réinitier.

Un des artefacts majeurs de l'utilisation de l'harringtonine est le pausing du ribosome. Un ribosome qui s'arrête à une position spécifique et qui n'a pas le temps de finir son élongation pourra être interpréter comme un site initiateur.

1.10.3.2 Lactimidomycine

La lactimidomycine inhibe la traduction par sa fixation au site E du ribosome. Contrairement à la CHX, la LTM fixe le site E uniquement si celui-ci est vide. De ce fait, la LTM n'a accès au site E uniquement lors des premières étapes de l'initiation où l'ARNt initiateur entre au site P (Steitz, 2008; Schneider-Poetsch *et al.*, 2010) (Figure 19). Comme l'HAR, la LTM arrête l'initiation de la traduction tout en permettant aux ribosomes déjà engagés de finir l'élongation.



La différence majeure entre l'HAR et la LTM est que la LTM lie le ribosome sous sa forme complexée 80S contrairement à l'HAR qui ne lie uniquement les sous-unités 60S libres (Fresno, Jiménez and Vázquez, 1977; Schneider-Poetsch *et al.*, 2010). Puisque la LTM et l'HAR laissent les ribosomes terminer leur cycle d'élongation, les deux traitements antibiotiques présentent les mêmes artefacts.

1.10.4 Les applications du ribosome profiling

1.10.4.1 Translatome et annotations des génomes

De manière globale, le ribosome profiling permet une annotation de l'ensemble de génome, renseigne de manière quantitative sur la synthèse protéique et fournit des informations sur le contrôle et la régulation des transcrits. Les propriétés biologiques presque universelles des ribosomes à travers les espèces et l'absence de manipulation génétique rend le ribosome profiling applicable à une grande diversité de modèles. Le ribosome profiling a été utilisé pour étudier une large variété d'organismes tels que les bactéries, les virus, les plantes, les levures, les protozoaires, les poissons zèbres, les drosophiles, les nématodes et l'humain (Brar and Weissman, 2015). L'application du ribosome profiling aux génomes humain et viraux sera d'avantage détaillé dans cette section. Une fresque non exhaustive retraçant les diverses applications du ribosome profiling est représentée en figure 20.

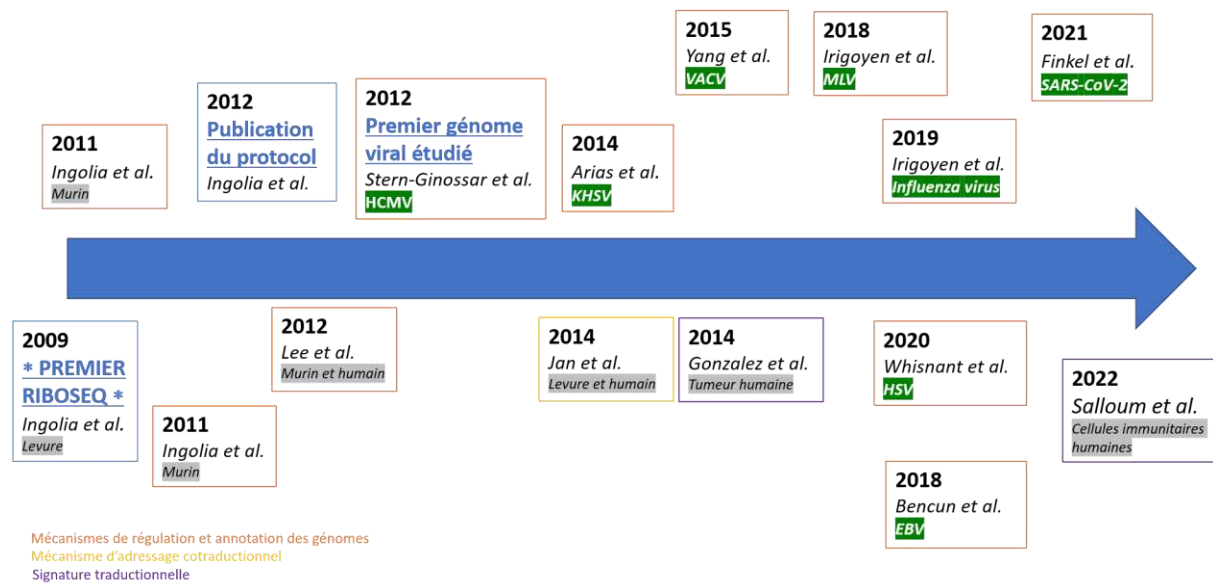


Figure 20. Fresque non exhaustive des applications du ribosome profiling de son développement à ce jour.
La couleur du cadre indique l'axe d'étude. Les modèles d'étude des génomes viraux sont mis en évidence grâce à un surlignage de couleur verte, les autres sont surlignés en gris.

Analyse qualitative et quantitative de la synthèse protéique

Une des premières applications du ribosome profiling est l'analyse qualitative et quantitative de la synthèse protéique (Figure 21). Puisque la densité ribosomale obtenue est indicative du nombre de ribosomes à une position donnée, on peut corrélérer la densité ribosomale avec le niveau de la synthèse protéique. En complément d'une analyse RNAseq, le ribosome profiling permet d'obtenir l'efficacité de traduction que l'on appelle « TE » pour « translational efficiency ». L'activité traductionnelle d'un ORF est évaluée par la division de la couverture ribosomale par la quantité d'ARNm. Un transcrite abondant peut ne pas être couvert par les ribosomes, reflétant une activité traductionnelle basse. A l'inverse, un transcrite présent en petite quantité peut être fortement traduit si les empreintes ribosomales associées à cette région sont nombreuses. La combinaison des deux techniques permet également d'étudier les gènes épissés. La couverture ribosomale apportée par le RiboSeq permet d'affirmer ou d'infirmer la traduction des régions introniques révélées par le RNAseq (Ingolia *et al.*, 2009a; Brar and Weissman, 2015) (Figure 21).

Le RiboSeq permet également de mettre en lumière les aspects mécanistiques de la traduction dans chaque organisme et dans n'importe quel état cellulaire. Dans leur papier princeps, Ingolia et al montrent que les uORFs sont majoritairement représentés dans le translatome de la levure (Ingolia *et al.*, 2009a). En plus de révéler qu'environ 20% des régions traduites dans le 5'UTR initient avec un codon non-AUG, ils mettent en évidence que l'inhibition de la traduction d'un gène en aval exercée par la traduction de certains uORFs peut-être levée en condition privative. Le gène GCN4 est régulé par la présence de 4 uORFs successifs. En phase de croissance, le premier uORF est majoritairement traduit. Les ribosomes ré-initient alors sur les uORFs en aval inhibant ainsi la traduction de l'ORF GCN4. En condition privative, la ré-initiation des uORFs est contournée pour se faire directement sur l'ORF GCN4. Dans le modèle murin, la même équipe démontrent qu'une grande partie des sites d'initiation, notamment dans la région 5'UTR, sont construits de codons non-AUG (Ingolia, Lareau and Weissman, 2011). Les

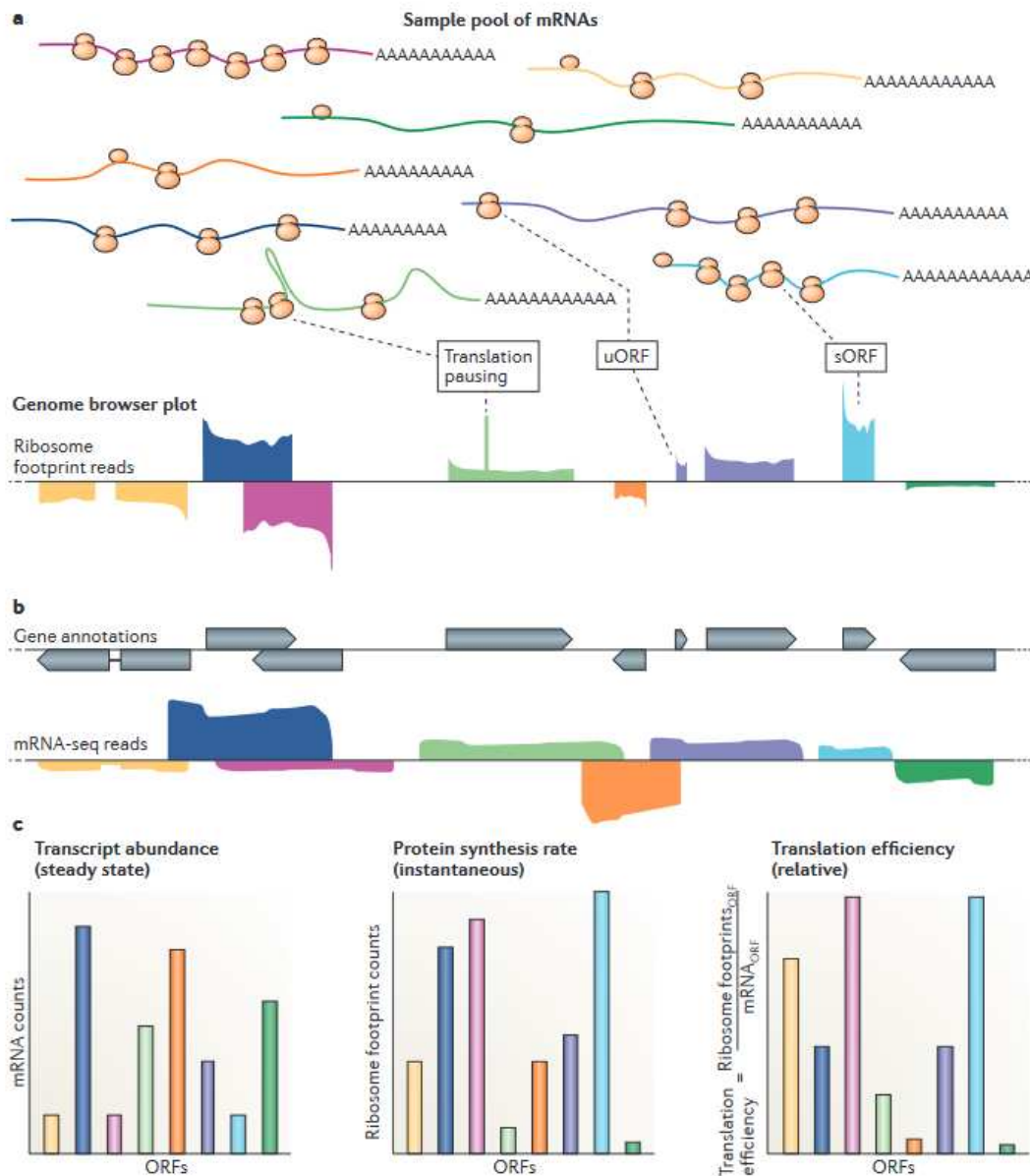


Figure 21. Données qualitatives et quantitatives obtenues par ribosome profiling (figure originale de Brar and Weissman 2015).

codons d'initiation alternatifs leucine et valine sont décrits comme les plus répandus. L'année suivante, l'équipe du D' Qian confirme ces observations sur le génome humain avec une prédominance pour la leucine « CUG » (Lee *et al.*, 2012). Courant des années 2015, c'est au tour des mécanismes de terminaison d'être étudiés par ribosome profiling. Une centaine de dORFs sont alors mis en évidence dans le génome humain (Bazzini *et al.*, 2014; Guydosh and Green, 2014; Ji *et al.*, 2015).

Avec le ribosome profiling il est également possible de mesurer *in vivo* la traduction de façon localisée dans des structures subcellulaires. La combinaison d'une purification par biotinylation au RiboSeq permet de cibler une catégorie spécifique de ribosomes. Par cette approche, Jan et al ont démontré *in vivo* chez la levure et l'humain, que la majorité des protéines sécrétoires sont transloquées dans le réticulum endoplasmique de manière co-traductionnelle (Jan, Williams and Weissman, 2014).

Annotation des génomes viraux

Le ribosome profiling a été pionnier dans l'annotation des génomes viraux et l'identification de nouveaux cadres de lecture. En 2012, Stern-Ginossar et al (affilié à l'équipe du D^r Weissman) décodent le génome du cytomégalovirus humain (HCMV) et mettent en évidence la traduction d'une centaine de cadres ouverts de lectures précédemment non annotés (Stern-Ginossar *et al.*, 2012). L'analyse de la densité ribosomale indique une traduction de sORFs en phase de lecture (produits de traduction tronqués en N-ter), de uORFs, d'ARFs et de sORFs dans des transcrits non annotés. La majorité de ces séquences non canoniques sont décrites comme très courtes, codant des polypeptides de moins de 20 acides aminés. Par l'utilisation de l'HAR et de la LTM, ils démontrent qu'une partie de ces sORFs initient avec un codon start alternatif, là encore, majoritairement par le biais d'une leucine « CUG ». Pour finir, ils mettent en lumière que ces sites d'initiations influencent l'expression des protéines virales dans le temps. Le HCMV a été le premier génome viral étudié par ribosome profiling.

En 2014, le RiboSeq apporte une meilleure connaissance du génome du virus de l'herpès associé au sarcome de Kaposi (KSHV) (Arias *et al.*, 2014). Les auteurs étudient les modifications transcriptionnelles et traductionnelles durant le cycle lytique. En combinant le RNAseq au RiboSeq, ils démontrent que les ribosomes lient rapidement les transcrits indépendamment de leur quantité. La régulation des gènes lytiques du KSHV est alors principalement transcriptionnelle. L'étude identifie aussi un nombre important de sORFs et uORFs qui augmente de 45% la capacité de codage du KSHV. En concordance avec les précédentes publications, ces séquences sont souvent traduites à partir de codon start alternatifs. Ils ne communiquent cependant pas sur la prédominance du codon utilisé.

Le premier rétrovirus à être étudié par ribosome profiling est le virus de la leucémie murine (MLV) (Irigoyen *et al.*, 2018). Contrairement au VIH-1, les ORFs codant pour *gag* et *pol* du MLV sont localisés dans la même phase de lecture. Dans ce cas de figure, la traduction de *pol* n'est donc pas dépendante d'un FS mais d'un codon-readthrough (Yoshinaka *et al.*, 1985). Irigoyen et al montrent *in vivo* qu'environ 7% des ribosomes continuent au-delà du stop du *gag* conduisant à la traduction de *pol*. Parallèlement, ils rapportent une signature traductionnelle aux régions 5'UTR des transcrits de *gag* et *env*. La traduction étendue en N-terminale de *gag*, dénommée « Glyco-Gag », peut-être initiée par la valine « GUG » en plus de la leucine « CUG » déjà décrite dans la littérature. Cette double initiation permet d'augmenter les variants de Glyco-gag. Une augmentation de la densité ribosomale est observée au site accepteur de l'épissage de *env*. Deux pics d'initiation correspondant à deux codons CUG sont observés dans cette région. Cette observation laisse sous-entendre un rôle des codons starts alternatifs dans la modulation de la traduction de *env* en plus de l'épissage canonique. L'ensemble de ces données a contribué à une meilleure connaissance du paysage traductionnel des rétrovirus et à une caractérisation du changement de cadre de lecture *gag-pol* médié par codon read-through.

En comparant *in silico* l'adaptation spécifique de diverses souches virales de la grippe à leur hôte, ici chez les mammifères, Machkovech et al montrent une sélection négative des sites non canoniques d'initiation « CUG » dans les souches virales de la grippe au cours du temps

(Machkovech, Bloom and Subramaniam, 2019). Parallèlement, le ribosome profiling révèle dans les transcrits viraux plusieurs sites d'initiation alternatifs situés en aval des codons canoniques initiateurs. De façon étonnante, tous les transcrits tronqués en partie N-terminale initient avec un codon canonique AUG. La suppression des 14 acides aminés en N-terminale provoque une perte de la queue cytosolique de la protéine N1 neuraminidase et d'une partie de son domaine transmembranaire. L'étude a montré que cette neuraminidase tronquée est active enzymatiquement et est exprimée à la surface des cellules infectées. Ils n'investissent pas cependant la question de son rôle immunologique ou viral. Concernant l'impact de l'infection virale sur la traduction de l'hôte, les auteurs ne rapportent pas de variation globale dans l'utilisation des codons starts alternatifs des transcrits humain. Avec un faible échantillonnage (n=8), ils montrent néanmoins qu'un sous-ensemble de gènes régulés à la hausse par l'infection virale sont enrichis en codon starts alternatifs.

Le génome du SARS-CoV-2 est le dernier génome viral en date étudié par ribosome profiling. Dans leur article publié dans Nature, Finkel et al, dévoilent la capacité de codage du SARS-CoV-2 (Finkel *et al.*, 2021). Premièrement, ils observent pour la plupart des ORFs, une corrélation entre leur abondance et leur couverture ribosomale, reflétant une efficacité de traduction similaires entre les ORFs à des temps précoces d'infection. L'efficacité du FS permettant la traduction de l'ORF1b est estimée à 57% et plusieurs sites potentiels de pausing sont identifiés en amont de ORF1b. Grâce au ribosome profiling, les auteurs mettent en évidence vingt-trois nouvelles régions du génome couverte par les ribosomes. Ces régions sont distribuées sur l'ensemble de génome et sont localisées « in-frame » ou « out-of-frame ». Ces 23 nouveaux ORFs comprennent : des sORFs étendus en N-ter et C-ter, des uORFs, des séquences leaders et des ARFs. Au total, 14 uORFs et 6 ARFs sont identifiés dont 2 ARFs chevauchant le gène codant la protéine spike. Avec une belle densité ribosomale, les auteurs mettent en évidence les zones d'initiations non canoniques des uORFs et des ARFs identifiés. En comparant l'abondance et la densité ribosomale, ils ont démontré que les séquences non canoniques sont exprimées à des niveaux similaires à ceux de leur ORF respectif. Certains ARFs sont très conservés parmi les génomes de coronavirus et présentent des caractéristiques structurelles qui supposent une fonction des polypeptides traduits.

Si l'on se concentre maintenant sur le modèle humain, plus particulièrement dans le cas des cancers, le ribosome profiling a permis en 2014 et 2016 d'établir les signatures traductionnelles des tumeurs cérébrales et des lymphocytes B isolés de leucémie lymphoïde chronique, respectivement (Gonzalez *et al.*, 2014; Sbarrato *et al.*, 2016). L'utilisation du ribosome profiling sur des cellules cancéreuses humaines montre qu'environ 85% des produits de traductions ne sont pas annotés dans la base du génome humain et que 50% d'entre eux correspondent à des produits alternatifs ne dérivant pas de séquences codantes (Ruiz Cuevas *et al.*, 2021; Yewdell, 2022).

1.10.4.2 Identification du protéome alternatif

Comme nous l'avons vu, l'utilisation du ribosome profiling a permis de révéler la diversité du protéome par l'identification de nouveaux cadres de lectures. La fonctionnalité et l'impact biologique de ces sORFs ne sont pas encore bien caractérisés. Pour cause, la plupart des études

qui ont identifiés les sORFs par ribosome profiling n'ont pas encore poursuivi leurs travaux par une étude fonctionnelle. Cependant certaines études, combinant le plus souvent plusieurs technologies à large échelle, ont déjà commencé à étudier leurs fonctions. Les sORFs jouent un rôle important dans la régulation trans-transcriptionnelle et cis-traductionnelle, le développement embryonnaire, la régulation des processus cytoplasmiques et mitochondrial, le métabolisme, et certains pourraient être des proto-gènes (Carvunis *et al.*, 2012; Couso and Patraquim, 2017; Cao and Slavoff, 2020; Schlesinger and Elsässer, 2022; Wright *et al.*, 2022).

Les uORFs

Les uORFs sont des éléments *cis* connus pour réprimer l'expression des gènes situés en aval. La reconnaissance d'un codon AUG en 5'UTR séquestre les différents acteurs de la traduction alors moins disponibles pour traduire la séquence canonique (Orr *et al.*, 2020). Ce mécanisme de régulation médié par les uORFs est très répandu dans les familles de virus à ADN et ARN. Par exemple, la traduction de la protéine virale UL4 du HCMV et des ORF35 et 36 du KHSV est sous le contrôle de uORFs. Pour le virus Ebola, la présence de uAUG est essentielle à sa réplication puisqu'il régule la traduction de sa polymérase L en réponse au stress (Shabman *et al.*, 2013). Dans les transcrits bicistroniques *vpu/env* du VIH-1, la présence d'un uORF en amont du gène *vpu* a un effet stimulateur sur la traduction du gène *env* (Krummheuer *et al.*, 2007). En d'autres termes, la présence de l'uORF régule négativement la traduction du gène *vpu*, mais régule positivement la traduction du gène *env*. De manière plus complexe, dans le modèle simien du virus de l'immunodéficience (SIV), 5 uORFs présents dans différents variant épissés régulent de façon différentielle l'expression des protéines *rev* et *env* (van der Velden *et al.*, 2012, 2013).

Chez l'humain, les uORFs régulent négativement 30 à 80% des transcrits en aval (Calvo, Pagliarini and Mootha, 2009). Le rôle des uORFs dans la traduction varie selon plusieurs critères : la longueur du 5'UTR, la structure de cette région, et la contenance en GC (Somers, Pöyry and Willis, 2013). De récentes études bio-informatiques indiquent la présence d'au moins un uORF dans environ 50% des transcrits humains (Chen and Tarn, 2019). De par leur rôle capital dans le contrôle de la traduction et leur grand nombre, la modification des fonctionnalités des uORFs entraîne le développement de nombreuses maladies.

Un exemple bien renseigné dans la littérature est celui de la protéine MYC. MYC est un proto-oncogène qui régule la prolifération cellulaire et la transformation. Deux isoformes de la protéine MYC sont traduites : une à partir du codon canonique AUG et l'autre à partir d'un uORF « CUG » générant une protéine avec une extrémité N-terminale distincte. Cette extrémité lui confère des capacités supplémentaires de liaisons à l'ADN ce qui lui permet de réguler les gènes qui contrôlent la croissance cellulaire. La surexpression de cette protéine non canonique inhibe la croissance cellulaire. Sa suppression ou sa mutation corrèle avec le développement de certaines tumeurs comme le lymphome de Burkitt (Hann *et al.*, 1988; Kearse and Wilusz, 2017). Dans d'autres type de cancers, la perturbation de la transcription des gènes suppresseurs de tumeurs génère une augmentation de la longueur du 5'UTR inhibiteur, des structures secondaires, et des codons AUG en 5'UTR. Cela engendre un recrutement accru des

ribosomes en 5'UTR et réduit l'expression des protéines antitumorales codées en aval (Somers, Pöyry and Willis, 2013).

L'étude de la perte de cheveux héréditaire (Hypotrichose héréditaire de Maria Unna) a également rapporté un rôle des uORFs (Wen *et al.*, 2009). Les uORFs interviennent aussi dans les troubles psychiques. Le polymorphisme dans le deuxième uORF de la sérotonine HTR3A a été significativement lié au trouble bipolaire (Niesler *et al.*, 2001).

Comme décrit précédemment, un grand nombre de uORFs sont facteurs agissant en *cis* qui inhibent la traduction de l'ORF en aval par le recrutement des ribosomes sur leur propre codon initiateur. Cependant, certains uORFs régulent la traduction en *cis* par le biais du polypeptide qu'ils synthétisent. Ces courts peptides, appelés « peptidoswitch », répondent à de petites molécules spécifiques de l'environnement cellulaire (Andrews and Rothnagel, 2014). Les petites molécules interagissent avec le peptide naissant par liaison directe ou par le biais d'un intermédiaire. La fixation de la petite molécule environnementale au peptide naissant bloque le ribosome au niveau de l'uORF, ce qui empêche sa progression et inhibe la traduction de la séquence codante en aval. Par exemple, chez l'humain, la polyamine fixe le peptidoswitch SAMDC, inhibant ainsi la traduction de l'ORF SAMDC (Law *et al.*, 2001). Le peptidoswitch est une des premières fonctions caractérisées des peptides dérivés des uORFs (Fang *et al.*, 2004; Andrews and Rothnagel, 2014; Dever, Ivanov and Sachs, 2020).

Les ARFs

L'étude fonctionnelle des sORFs s'est davantage concentrée sur les sous-groupes des lncRNA et des uORFs contrairement aux ARFs qui peinent à être caractérisés chez l'humain. Voici cependant quelques exemples rapportés au génome humain :

- L'ARF « AltPRP » est une protéine alternative de la protéine du prion. L'AltPrP est un peptide de 73 résidus codé par un ARF chevauchant le cadre de lecture de la protéine du prion. La traduction du peptide alternatif est initiée par un codon AUG entouré d'une séquence Kozak optimale. AltPrP présente une conservation et une homologie importante dans un large éventail de mammifères. Elle est localisée aux mitochondries et est régulée positivement par le stress du réticulum endoplasmique et l'inhibition du protéasome (Vanderperre *et al.*, 2011).
- L'ARF « AltFUS », résultant d'un décalage de cadre de lecture en +1 de l'ORF « FUS », inhibe l'autophagie et renforce ainsi la toxicité médiée par la protéine « FUS » (Brunet *et al.*, 2021). Les variants du gène FUS ont montré une importance clinique dans les troubles neurodégénératifs (Deng, Gao and Jankovic, 2014).
- L'ARF « Alt-RPL36 » régule négativement la voie de signalisation PI3K-AKT-mTOR impliquée dans la synthèse protéique et la croissance cellulaire (Cao and Slavoff, 2020; Cao *et al.*, 2021).
- L'ARF « AltB2R », marqueur du cancer du sein et de la prostate, régule positivement l'activité du récepteur à la bradykinine de type B2 (Gagnon *et al.*, 2021).

A ma connaissance, aucune étude n'a rapporté de fonction virologique des ARFs identifiés dans les divers génomes viraux.

Même si la plupart des polypeptides issus de sORFs (en particulier les ARFs, mais cela n'a pas été suffisamment étudié pour être certain) ne produisent probablement pas de protéines fonctionnelles, ils constituent néanmoins une partie importante du répertoire immunologique. Le laboratoire du D^r Nilabh Shastri, novateur dans le domaine de la présentation antigénique non canonique, montre en 1999 que les produits issus des ARFs peuvent être une source de substrats peptidique pour les molécules du CMH-I (Malarkannan *et al.*, 1999). En effet, les peptides dérivés d'une initiation « CUG » peuvent servir d'épitopes aux cellules T CD8⁺ (Dolstra *et al.*, 1999; Starck *et al.*, 2008, 2012). C'est notamment le cas dans le carcinome rénal où le peptide issu de l'initiation CUG du VEGF active la réponse T CD8⁺ (Weinzierl *et al.*, 2008). De façon complémentaire, il a été reporté que la réponse immunitaire augmente l'initiation « CUG » (Prasad, Starck and Shastri, 2016). Ce mécanisme peut-être une stratégie de l'hôte pour générer des épitopes suite à un stress cellulaire.

Les nombreux travaux du D^r Yewdell dans ce domaine définissent les produits défectueux de traduction (« DRiPs » pour « defective ribosomal products ») comme source majoritaire d'épitopes liés au complexe majeur d'histocompatibilité de classe I (CMH-I) (Yewdell, Antón and Bennink, 1996; Yewdell and Nicchitta, 2006; Antón and Yewdell, 2014). Les DRiPs peuvent être issus de protéines canoniques mal repliées durant l'élongation ou de polypeptides non canoniques. La caractéristique majeure des DRiPs est la rapidité : dans leur apparition et dans leur dégradation. L'incidence des DRiPs dans la présentation antigénique est d'avantage développée dans le chapitre 2 du manuscrit.

Certains peptides dérivés d'ARFs ou de lncRNA stimulent efficacement une réponse T spécifique. C'est le cas par exemple du transcrit beta 2.7 du génome HCMV précédemment désigné comme non codant. Ingolia et al montrent chez quatre donneurs infectés une forte réponse T contre les peptides issus des huit cadres de lectures nouvellement identifiés (Ingolia *et al.*, 2014). Chez le SARS-CoV-2, des réponses T ont également été mise en évidence contre des peptides non canoniques. L'un d'eux, dérivé d'un ARF chevauchant l'ORF de la nucléoprotéine stimule une réponse CD8⁺ avec une magnitude plus forte que celles retrouvées contre tous les peptides de classe I testés (Weingarten-Gabbay *et al.*, 2021). Dans le modèle simien et humain du virus de l'immunodéficience, une réponse T CD8⁺ contre des peptides alternatifs est observée en phase aiguë et chronique d'infection bien que ceux-ci n'aient pas été identifiés par ribosome profiling (Maness *et al.*, 2007, 2010; Garrison *et al.*, 2009; Bansal *et al.*, 2010; Berger *et al.*, 2010; Champiat *et al.*, 2012). Notre équipe de recherche a montré en 2004, une réponse T CD8⁺ contre six peptides dérivés d'ARFs chez des individus infectés asymptomatiques et traités (Cardinaud *et al.*, 2004).

Depuis les années 90, les chercheurs ont constaté que certains lymphocytes T infiltrant les tumeurs reconnaissent des antigènes tumoraux non canoniques et que cette reconnaissance est associée à une régression tumorale (Uenaka *et al.*, 1994; Wang *et al.*, 1996; Mayrand, Schwarz and Green, 1998; Dolstra *et al.*, 1999; Rosenberg *et al.*, 2002; Laumont and Perreault, 2018). Des études très récentes et prometteuses de thérapies personnalisées utilisent le ribosome profiling pour identifier l'ensemble des ORFs et ARFs spécifiques à la tumeur. Combiné avec de la spectrométrie de masse, cela permet l'identification de néoantigènes présentés à la surface des cellules cancéreuses. Cette approche permet l'identification et

l'isolation des lymphocytes T infiltrant la tumeur spécifiques de ces peptides d'intérêts. L'expansion *in vitro* de ces lymphocytes T montrent une bonne immunogénicité quand ils sont réinjectés chez le patient (Arnaud *et al.*, 2022; Chong, Coukos and Bassani-Sternberg, 2022).

L'ensemble des études énoncées dans ce chapitre (et d'autres non discutées) confirment l'intérêt d'étudier le génome préalablement considéré comme non codant ainsi que les mécanismes menant à l'expression de nouveaux cadres de lectures. Les nouvelles études qui combinent plusieurs analyses à haut débit telles que le ribosome profiling et la spectrométrie de masse confirment le rôle des événements non conventionnels de traduction dans la présentation antigénique. Ils sont une source sous-estimée d'antigènes immunogènes. Le chapitre suivant présente les différentes voies canoniques et non canoniques de présentation des antigènes. Je me concentrerai ensuite sur les sources alternatives d'antigènes et les techniques pour les mettre en évidence, qui forment un des axes principaux de mes travaux de thèse.

CHAPITRE 2 : LA PRESENTATION ANTIGENIQUE AUX LYMPHOCYTES T

Le locus CMH contient les gènes codant les molécules du CMH de classe I (CMH-I) et de classe II (CMH-II). Les molécules du CMH-I sont exprimées par toutes les cellules nucléées tandis que les molécules du CMH-II sont exprimées par les cellules présentatrices d'antigènes (APCs). Les APCs regroupent les cellules dendritiques (DCs), les lymphocytes B (LBs), les macrophages, les granulocytes mais aussi d'autres types cellulaires comme les fibroblastes lors d'un contexte inflammatoire (Lin and Loré, 2017; Eiz-Vesper and Schmetzer, 2020). Cette diversité classe I et classe II des molécules du CMH permet la présentation d'épitopes de natures différentes et l'activation des différents acteurs de l'immunité innée et adaptative. Les lymphocytes T (LTs) cytotoxiques CD8⁺ reconnaissent des épitopes présentés sur les molécules du CMH-I, alors que les LTs auxiliaires CD4⁺ reconnaissent des épitopes sur les molécules du CMH-II. La reconnaissance du complexe épitope/CMH par le TCR (« T cell receptor ») des lymphocytes T est une étape clé de l'activation des lymphocytes (Pishesha, Harmand and Ploegh, 2022). Chez l'Homme, le locus CMH est nommé HLA et se positionne sur le chromosome 6. Le chromosome 6 héberge en plus des gènes codant les molécules HLA, des gènes associés aux fonctions des molécules HLA, mais également des gènes non codants et des pseudogènes (Shiina *et al.*, 2009).

Dans ce chapitre, je présenterai dans un premier temps les voies canoniques et non canoniques de présentation des antigènes. Je focaliserai ensuite mon attention sur les sources alternatives d'antigènes qui constituent l'un des axes majeurs de mes recherches doctorales.

2.1 Les molécules du CMH de classe I

2.1.1 Génétique

Deux types de gènes du CMH codent pour les molécules du HLA-I : les gènes dits « classiques » et ceux considérés comme « non classiques ». Ils ont été subdivisés en deux catégories selon leurs propriétés génétique, structurales, et fonctionnelles. Les gènes classiques sont des gènes très polymorphiques qui codent chacun pour les protéines distinctes HLA-A, -B, et -C. Les gènes non classiques sont moins polymorphiques et n'ont pas toujours un rôle défini. Ils codent pour les protéines HLA-E, -F et -G. Mon propos se focalisera sur les molécules HLA-I classiques. Les gènes du CMH-I présentent le degré de polymorphisme le plus élevé du génome humain (Shiina *et al.*, 2009). Plus de 9 100 allèles ont été décrits à ce jour pour le locus HLA-B, 7 700 pour HLA-A, 7 600 pour HLA-C, la plupart des individus étant hétérozygotes (Robinson *et al.*, 2020). Les variations de séquences à l'origine de ce polymorphisme sont principalement localisées proche du site d'ancrage des peptides. Les gènes du CMH-I sont localisés sur le bras court du chromosome 6, les locus B et C en position centromérique et le locus A en position télomérique (Figure 22). Cette région classe I, du bras court du chromosome,

héberge également des locus codant pour des gènes associés aux fonctions du CMH-I, des gènes non-codants, et des pseudogènes.

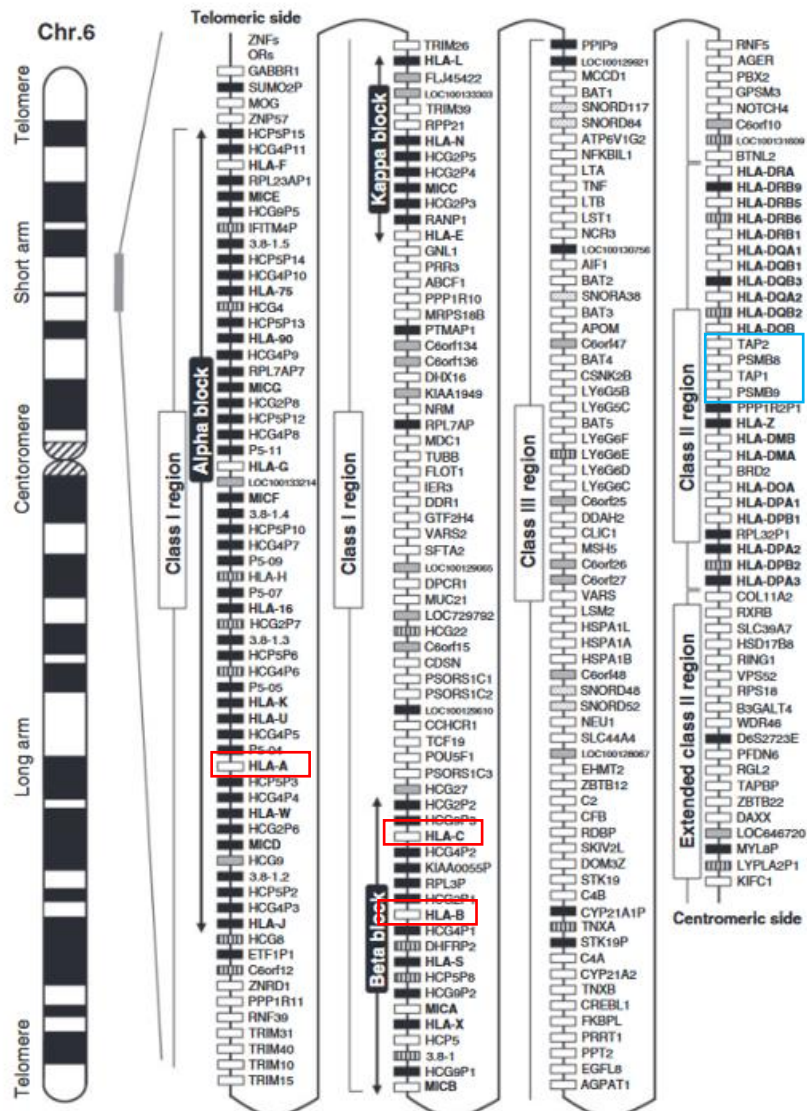


Figure 22. Carte génétique du locus HLA sur le chromosome 6 (figure originale de Shiina et al 2009). La région HLA classe I, classe II, et classe III sont séparées par des flèches et des encadrés. La région classe III ne code pour aucune molécule HLA mais contient des gènes codants pour des composants du complément et des cytokines impliquées dans les réponses immunitaires. Les rectangles blancs, gris, rayés et noirs indiquent respectivement les gènes exprimés, les gènes candidats, et les gènes non codants et pseudogènes. Les rectangles rouges indiquent les gènes codants HLA-A, -B, et -C. Les gènes codants les protéines TAP et les sous-unités inductibles du protéasome sont représentés par un rectangle bleu. Les divers gènes codants les chaînes α et β des molécules du CMH-II sont renseignés par des rectangles blancs au sein la région « Class II region ».

2.1.2 Structure et localisation

Les molécules du CMH-I sont des complexes protéiques de type immunoglobuline composés d'une chaîne lourde α liée de manière non covalente à la chaîne invariante β_2 -microglobuline (β_2m). Les gènes HLA-A, -B, et -C, codent la chaîne α tandis que la β_2m est codée par un gène se trouvant en dehors du locus CMH (Figure 22). La chaîne α est constituée d'une extrémité C-terminale constante, d'un domaine transmembranaire, et d'une extrémité N-terminale variable. L'extrémité N-terminale est composée des deux sous-unités α_1 et α_2 qui forment un sillon accueillant les peptides antigéniques. Le sillon peptidique est composé de feuilletts β et d'hélices α (Figure 23). Une fois synthétisée de manière co-translationnelle dans le réticulum endoplasmique (RE), la molécule du CMH-I chargée avec le peptide migre à la surface des cellules nucléées. Le complexe peptide-CMH-I (pCMH) est ensuite reconnu par les TCR des LTs $CD8^+$ permettant une détection rapide des cellules tumorales ou infectées. La molécule de costimulation $CD8$ exprimée par les LTs $CD8^+$ renforce l'interaction par sa liaison au sous-domaine α_3 de la molécule du CMH-I.

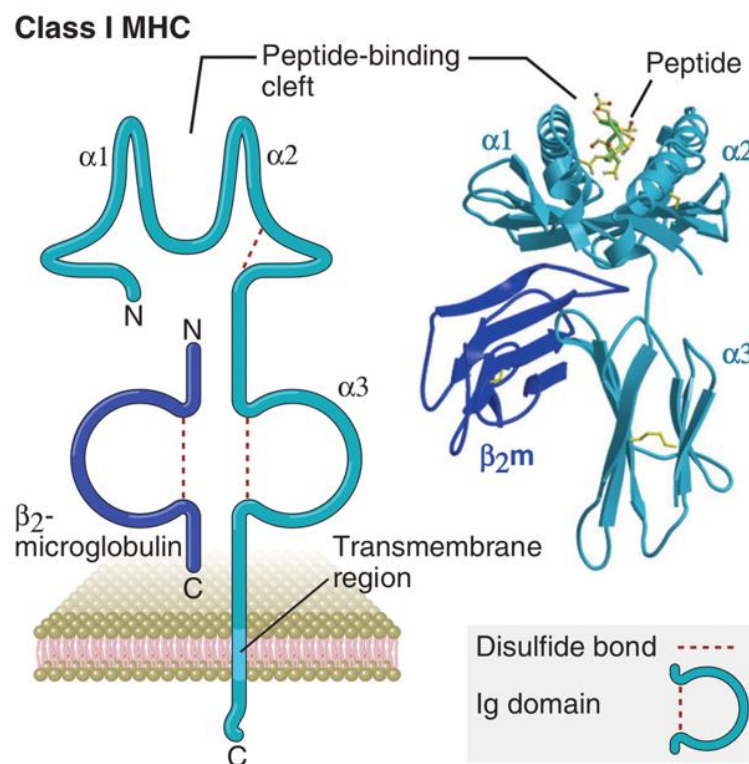


Figure 23. Structure moléculaire d'une molécule du CMH-I (figure originale de Abbas et al 2020).

Représentation schématisée et tridimensionnelle d'une molécule du CMH-I. Le sillon peptidique de fixation est formé par les deux domaines α_1 et α_2 .

2.1.3 Nature des peptides présentés

Les extrémités du sillon peptidique du CMH-I affichent une conformation fermée qui limite la taille des peptides fixés de 8 à 10 acides aminés. La plupart des molécules HLA-I ont une forte préférence pour la liaison de 9mers. L'analyse structurale du HLA-A2, premier HLA à avoir été cristallisé, révèle l'existence de six « sous poches » distinctes au sein d'un sillon peptidique HLA-A2 (Saper, Bjorkman and Wiley, 1991). Les résidus qui composent chaque région sont approximativement conservés dans toutes les structures définies du CMH de classe I.

2.2 Les molécules du CMH de classe II

2.2.1 Génétique

Les gènes classiques et non classiques du CMH-II codent les protéines HLA-DP, -DQ, et -DR ; et HLA-DM et -DO, respectivement. Comparativement aux gènes du CMH-I, ceux du CMH-II sont moins polymorphiques avec environ 4 400 allèles décrits à ce jour pour le locus HLA-DR, 3 000 pour HLA-DQ, 540 pour HLA-DP (Robinson *et al.*, 2020). Le locus du CMH-II code aussi pour des protéines impliquées dans la voie de présentation antigénique du CMH de classe I comme les protéines TAP (« transporter associated with antigen processing ») et les sous-unités inductibles du protéasome (Figure 22). Les gènes classiques du CMH-II possèdent plusieurs copies dont certaines sont sous forme de pseudogènes. Les molécules HLA de classe II sont des hétérodimères composés d'une chaîne α codée un gène A (DRA1, DPA1, et DQA1) et d'une chaîne β codée par un gène B (DRB1, DPB1, et DQB1). A la différence des autres molécules HLA où la chaîne β est codée par un gène unique, la chaîne β de la molécule HLA-DR peut être codée par 3 gènes supplémentaires (DRB3, DRB4, et DRB5) (Delbos, Malard and Congy, 2018). Dans le cas du gène codant HLA-DR, plus de 3 000 allèles de la chaîne β sont retrouvés (Robinson *et al.*, 2020). Malgré cette grande diversité dans la chaîne β , sept allèles sont surreprésentés dans la population caucasienne représentant environ 60% des allèles HLA-DR β 1 (Texier *et al.*, 2000). Les allèles HLA-DR β 1*01 et -DR β 1*03 sont les plus prépondérants dans la population française couvrant 20% de la fréquence des allèles du locus HLA-DR β 1 (Gonzalez-Galarza *et al.*, 2020). L'expression du CMH-II est finement régulée par le facteur de transcription CIITA (Chang *et al.*, 1994; Steimle *et al.*, 1994).

2.2.2 Structure et localisation

Les molécules du CMH-II forment donc un complexe protéique de type immunoglobuline composé d'une chaîne lourde α et β associées de manière non covalente. Cependant, contrairement au CMH-I, les molécules du CMH-II ne contiennent pas de β 2-microglobuline. Chacune des chaînes est constituée d'une partie intracellulaire C-terminale, un domaine transmembranaire et deux domaines extracellulaires (α 1, α 2 ou β 1, β 2) (Figure 25). Les domaines α 1 et β 1 qui contiennent des résidus polymorphes forment le sillon peptidique des molécules du CMH de classe II. Le sillon peptidique des molécules du MCH-II a une conformation ouverte par rapport à celle des molécules du CMH-I ce qui lui permet de fixer des peptides plus longs allant de 12 à 25 acides aminés (Rammensee, Friede and Stevanović, 1995). Une fois synthétisée de manière co-translationnelle dans le RE, les molécules du CMH-II chargées avec le peptide transitent par le compartiment de chargement des molécules de classe II (MIIC) avant d'être acheminées à la membrane plasmique. Le complexe pCMH est ensuite reconnu par le TCR des LTs CD4⁺ qui, à leur tour, activeront les divers acteurs de

l'immunité innée et adaptative. La molécule de costimulation CD4 exprimée par le LT CD4⁺, par sa liaison au sous-domaine β_2 , renforce l'interaction avec la molécule du CMH.

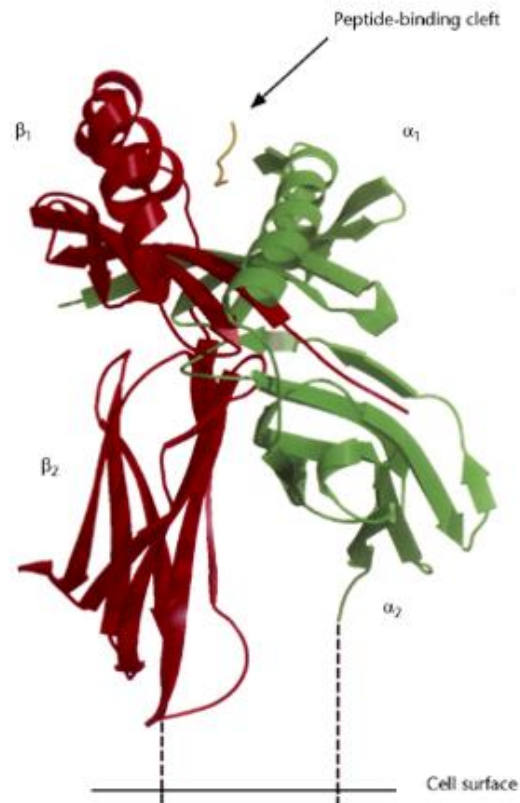


Figure 25. Structure moléculaire d'une molécule du CMH-II (figure originale de Liu and Gao 2011).

Représentation tridimensionnelle d'une molécule du CMH-II. Le sillon peptidique de fixation est formé par les deux domaines α_1 et β_1 .

2.2.3 Nature des peptides présentés

Bien que le sillon peptidique des molécules de classe II contienne davantage de résidus d'ancrage, les ligands des molécules du CMH-II semblent avoir une affinité plus faible pour le sillon comparé aux ligands fixant le sillon des molécules du CMH-I. Il a été rapporté que les molécules du CMH-II sont capables de scanner, grâce à leur conformation ouverte, un polypeptide long et désordonné jusqu'à trouver des régions avec un motif de séquence optimal (Liu and Gao, 2011). Comme pour les molécules du CMH de classe I, les résidus polymorphes sont regroupés sous forme d'une série de six poches expliquant la spécificité allélique des motifs de liaison peptidique (Texier *et al.*, 2000). Pour l'ensemble des molécules du CMH-II, trois à quatre de ces poches qui hébergent les résidus d'ancrage interviennent dans la fixation des peptides. Par exemple, les résidus d'ancrages de HLA-DR sont les positions P1, P4, P6, et P9. Bien qu'il y ait plusieurs sites d'ancrage présents au sein des sillons, une position d'ancrage prédominante est généralement observée : la position P1 pour DR1 et la position P4 pour DR17 (Rammensee, Friede and Stevanović, 1995; Liu and Gao, 2011). A l'image des molécules du CMH-I, l'existence de serveurs compilant un vaste ensemble de données de spectrométrie de masse permet de prédire les motifs de fixation des peptides aux molécules du CMH-II. Les ligands de HLA-DR β 1 possèdent préférentiellement en position 1, 4, 6 et 9, une

phénylalanine, une leucine, une alanine et une valine, respectivement (Reynisson *et al.*, 2020) (Figure 26).

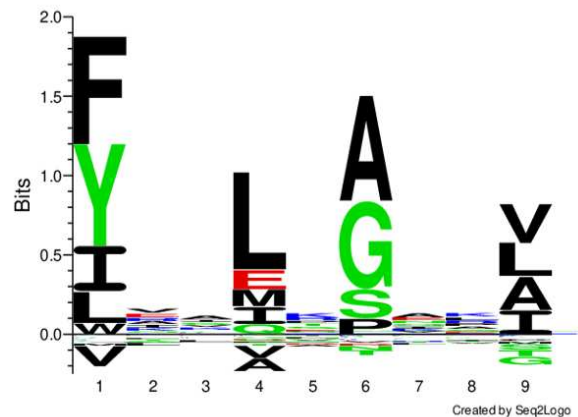


Figure 26. Motifs de fixation à la molécule HLA-DRβ1*0101 (figure originale de NetMHCIIpan-3.2). Position, nature, et fréquence des résidus qui composent les ligands de HLA-DRβ1*0101. Compilation des nombreuses données obtenues par immunopeptidomique. La taille de la lettre détermine sa prévalence à la position indiquée pour l'ensemble des séquences peptidiques de la base de données.

2.3 La restriction HLA aux lymphocytes T CD8⁺ et CD4⁺

La restriction HLA est découverte en 1974 par les D^{rs} Zinkernagel et Doherty qui obtiennent conjointement le prix Nobel de physiologie ou médecine en 1996 pour leurs travaux (Zinkernagel and Doherty, 1974, 1997). La restriction HLA fait référence à la spécificité du TCR pour une molécule HLA donnée outre la nature du peptide présenté (La Gruta *et al.*, 2018). La majorité des lymphocytes T, cytotoxiques ou auxiliaires, expriment un TCR αβ qui a pour fonction la reconnaissance du complexe peptide-CMH (pCMH). Cette reconnaissance est cruciale pour la production de molécules effectrices et de cytokines par les lymphocytes T. Le TCR est un hétérodimère de chaînes α et β. Chaque hétérodimère est constitué d'un domaine constant et d'un domaine variable, ce dernier étant formé de trois boucles hypervariables

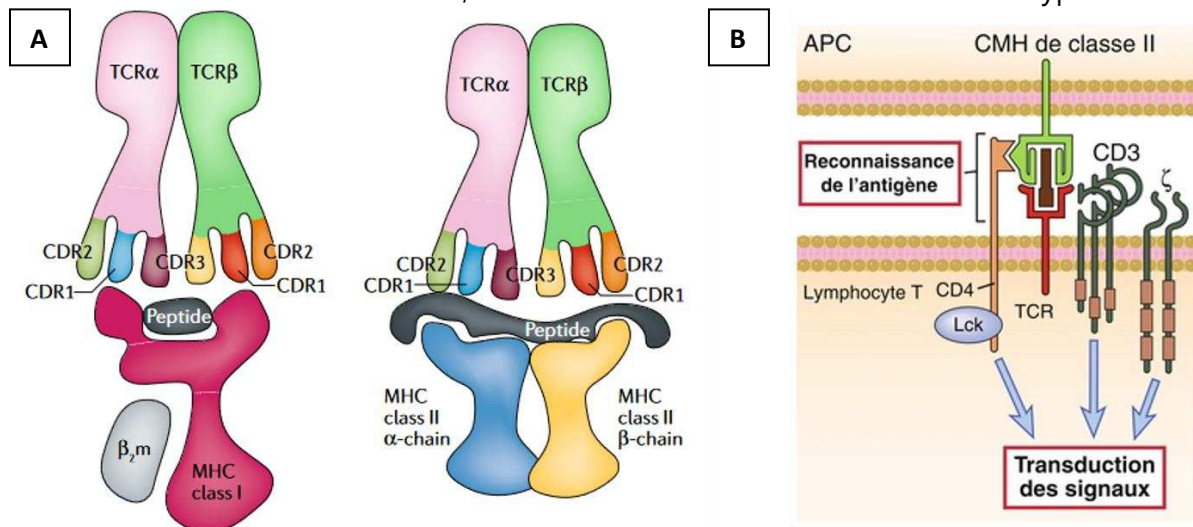


Figure 27. Représentation schématique de la structure du TCR et de l'engagement des différents récepteurs (figures originales de La Gruta et al 2018 et Abbas et al 2020). (A) Schéma décrivant les interactions typiques entre les boucles CDR du TCR et le pCMH de classe I (gauche) ou de classe II (droite). (B) Représentation simplifiée du corécepteur CD4 et des molécules de costimulation du TCR impliquées dans la transduction du signal.

distinctes, à savoir CDR1, CDR2 et CDR3 (« complementarity diversity receptor »). Ces boucles hypervariables reconnaissent et interagissent avec le pCMH (Figure 27). Après la reconnaissance du complexe pCMH et du corécepteur, la chaîne CD3 liée de manière non covalente aux chaînes α et β du TCR, engage la signalisation du TCR suite à la phosphorylation de ses motifs ITAM (« immunoreceptor-based activation motif ») (Figure 27).

2.4 Apprêtement et présentation des antigènes

2.4.1 La voie de présentation des antigènes par les molécules du CMH-I

2.4.1.1 Présentation des antigènes endogènes : la voie canonique

Les ligands des molécules du CMH de classe I dérivent de protéines de source endogène c'est-à-dire des protéines cytosoliques issues de la traduction cellulaire. Toutes protéines cellulaires dont la durée de vie est dépassée constitue une source antigénique. La voie canonique de présentation des antigènes de classe I repose sur le processus de protéolyse qui est assuré par le protéasome. Après avoir assuré leurs fonctions, les protéines à dégrader sont conjuguées à

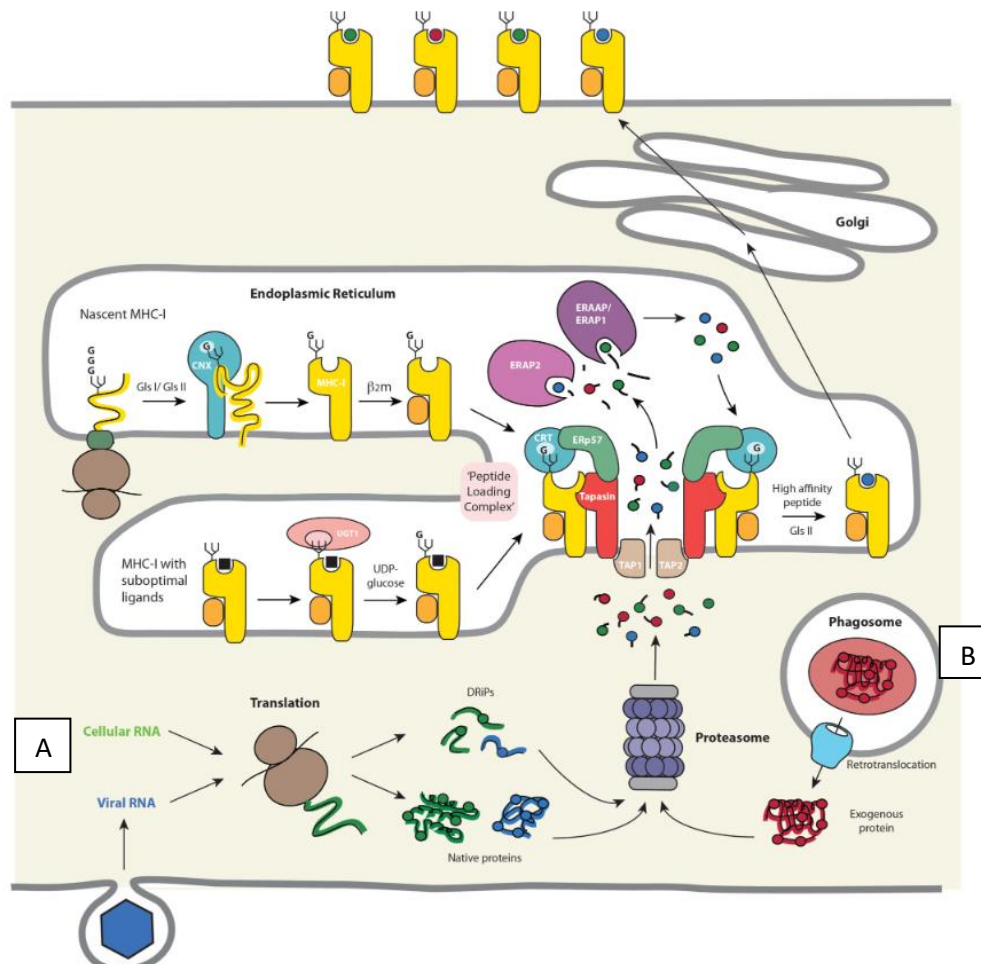


Figure 28. Voie de présentation des antigènes de classe I (figure adaptée de Blum et al 2013).
 (A) Voie de biosynthèse et d'apprêtement des antigènes sur les molécules du CMH-I : la voie canonique. Les antigènes dérivent de la traduction de transcrits cellulaires ou viraux. Ils sont dégradés par le protéasome puis les peptides générés sont acheminés au RE grâce aux pompes TAP. Les peptides sont chargés directement ou découpés par des aminopeptidases de la famille ERAP avant le chargement sur les molécules du CMH-I. Les complexes pCMH transitent par l'appareil de Golgi et sont acheminés à la surface des cellules. (B) La voie non canonique : la présentation croisée. Les antigènes dérivent de pathogènes exogènes ayant échappés à la phagocytose.

l'ubiquitine qui est ensuite reconnue par le protéasome. Les peptides résultant de la dégradation protéolytique sont ensuite transférés dans le RE, lieu de synthèse des molécules du CMH-I, grâce aux pompes TAP. Dans le RE, les molécules du CMH-I liées à la face interne des molécules TAP, fixent les peptides et sont ensuite transportées à la surface de la cellule (Figure 28).

Le protéasome

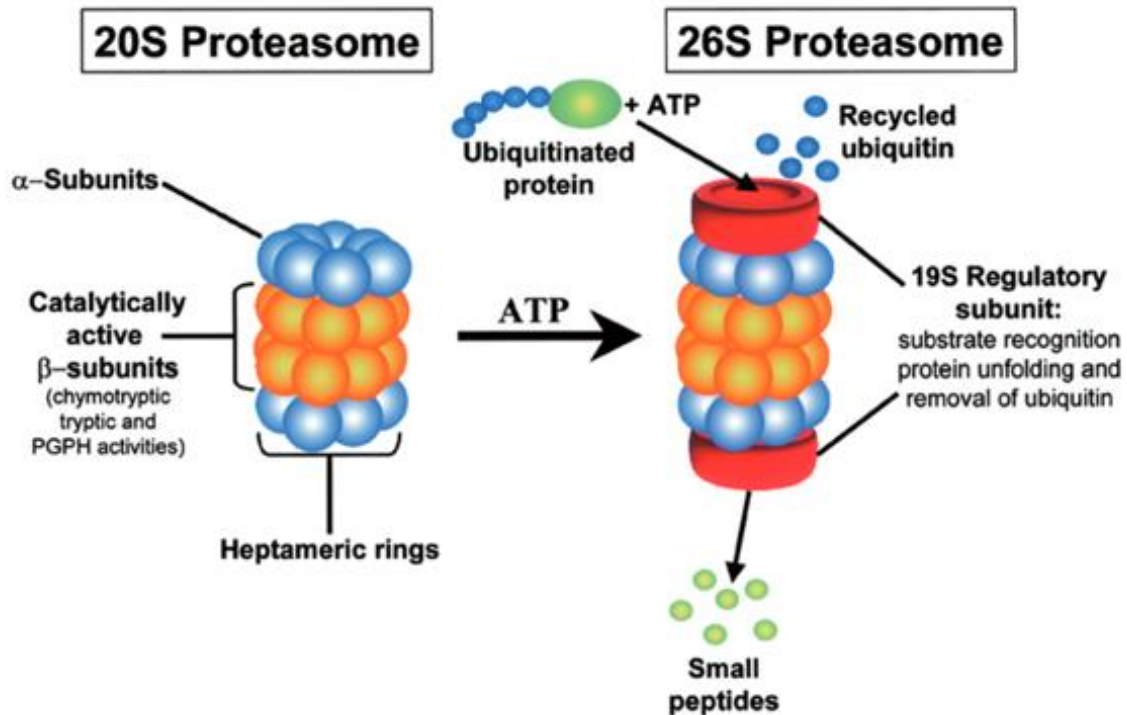


Figure 29. Représentation schématique du protéasome 20S et 26S (figure originale de Almond and Cohen 2021).

Découvert en 1980, le système ubiquitine-protéasome (UPS) permet la dégradation des protéines cytosoliques (Ciechanover *et al.*, 1980; Hershko *et al.*, 1980). Ce système implique deux étapes successives : 1) le marquage du substrat par la fixation covalente de plusieurs molécules d'ubiquitine, et 2) la dégradation de la protéine marquée par le protéasome avec libération d'ubiquitine libre et réutilisable (Glickman and Ciechanover, 2002). Il existe une distinction entre le protéasome constitutif, exprimé de manière ubiquitaire dans toutes les cellules, et l'immunoprotéasome, exprimé dans de nombreuses cellules immunitaires (LTs, DCs, LBs). L'immunoprotéasome est inductible par les cytokines pro-inflammatoires telles que l'IFN γ , l'IFN de type 1, et le TNF α (Ferrington and Gregerson, 2012; Abi Habib *et al.*, 2022). Le protéasome constitutif, aussi appelé protéasome 26S, résulte de l'assemblage du protéasome 20S avec deux sous-unités 19S (Figure 29). Le protéasome 20S est constitué de quatre anneaux heptamériques empilés que l'on appelle « ring domain ». Les deux anneaux externes sont identiques et composés de sept sous-unités α , tandis que les deux anneaux internes sont identiques mais composés de sept sous-unités β (Abi Habib *et al.*, 2022). Les sous-unités β forment la chambre catalytique dans laquelle sont dégradées les protéines. L'ajout des deux sous-unités 19S à chaque extrémité du protéasome 20S permet la reconnaissance des protéines ubiquitinées et leur translocation dans la sous-unité 20S. Le protéasome génère des peptides sous la forme de larges précurseurs qui seront ensuite raccourcis dans leur région N-

terminale par les aminopeptidases présentes dans le RE. L'extrémité C-terminale des précurseurs est définie par le protéasome.

L'immunoprotéasome diffère du protéasome 26S par la substitution de certains monomères du β -ring domain (Figure 30). En réponse aux cytokines pro-inflammatoires ou au stress cellulaire, les sous-unités LMP2, MELC-1 et LMP7 sont exprimées et préférentiellement incorporées dans les protéasomes 26S. Ces nouvelles sous-unités catalytiques confèrent à l'immunoprotéasome des propriétés de clivage distinctes et modulent le répertoire de peptides présentés par les molécules du CMH-I. LMP2, MELC-1 et LMP7 augmentent le clivage des protéines après un résidu hydrophobe en position C-terminale. Les polypeptides sont donc davantage susceptibles de se lier aux molécules du CMH-I puisque les résidus hydrophobes de l'extrémité carboxyle constituent un site d'ancrage à la position 9 du sillon des molécules du CMH-I (Falk *et al.*, 1991; Tanaka, 2009; Ferrington and Gregerson, 2012). De manière globale, la modification de ces trois sous-unités catalytiques affecte le répertoire des peptides du CMH-I et les réponses T CD8⁺ antivirales (Chen *et al.*, 2001; de Verteuil *et al.*, 2010).

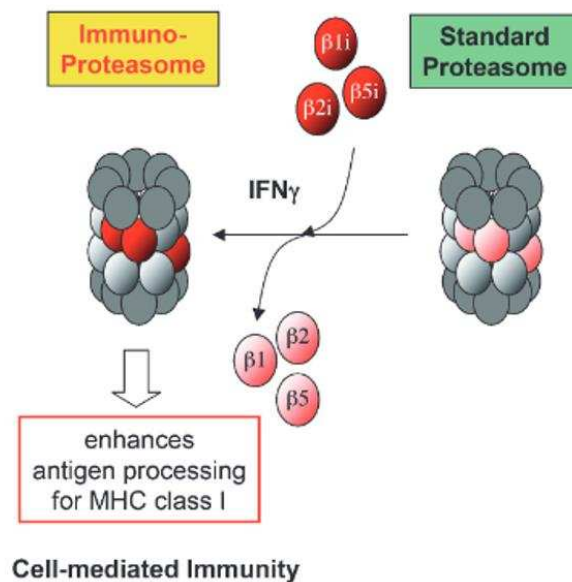


Figure 30. Représentation schématique de l'immunoprotéasome (figure adaptée de Ferrington and Gregerson 2012). LMP2, MELC-1, et LMP7 sont représentées par les sous-unités (rouges foncées) $\beta 1i$, $\beta 2i$, et $\beta 5i$, respectivement. En réponse à l'IFN γ et TNF α elles sont préférentiellement incorporées à la place des sous unités (rouges claires) $\beta 1$, $\beta 2$, et $\beta 5$, du protéasome 20S.

Transport des peptides vers le RE et assemblage des complexes peptides-molécule du CMH de classe I

Le complexe de chargement des peptides (PLC) est localisé dans le RE et joue un rôle fondamental dans le processus de chargement des peptides sur les molécules du CMH de classe I. Le PLC est, entre autres, responsable de la sélection, du traitement et du chargement des peptides de haute affinité sur les molécules du CMH-I (Margulies *et al.*, 2022). Le PLC est composé des molécules du CMH de classe I nouvellement synthétisées, des chaperonnes tapasine, calréticuline, ERp57 ainsi que de la protéine TAP (Blees *et al.*, 2017).

Une fois dégradé par le système UPS, le précurseur peptidique est transporté dans le RE par le biais de la protéine TAP. La protéine TAP est un hétérodimère composé des deux sous-unités TAP1 et TAP2 ancrées à la membrane du RE. La protéine TAP permet une translocation efficace dans le RE de peptides cytosoliques de 8 à 16 acides aminés (van Endert *et al.*, 1994; Koopmann *et al.*, 1996; Momburg and Hämmerling, 1998; Burgevin *et al.*, 2008). Quand ils pénètrent dans le RE, les précurseurs peptidiques produits par le protéasome ont principalement une taille minimale de 15 acides aminés (Rock and Goldberg, 1999; Reits *et al.*, 2004). Ils sont donc trop longs pour se fixer directement sur les molécules du CMH-I et nécessitent une ultime étape de protéolyse pour pouvoir se fixer efficacement dans le sillon des molécules du CMH-I.

Chez l'Homme, deux aminopeptidases, ERAP1 (« endoplasmic reticulum aminopeptidase ») et ERAP2 sont impliquées dans ce processus. Elles coupent le précurseur dans sa partie N-terminale (Saric *et al.*, 2002; Saveanu *et al.*, 2005). ERAP1 et ERAP2 partagent une structure similaire mais ont une spécificité propre de clivage des ligands des molécules du CMH-I. ERAP1 a plus d'affinité pour les résidus hydrophobes tandis qu'ERAP2 montre une préférence pour les résidus basiques (Saveanu *et al.*, 2005). De plus, ERAP1 favorise la production d'épitopes de huit à neuf acides aminés, taille requise pour une liaison stable aux molécules du CMH-I (York *et al.*, 2002). ERAP2 interviendrait davantage dans la prise en charge des précurseurs complexes ou lorsque l'extension N-terminale est composée de résidus basiques. Ces deux aminopeptidases ne font pas partie du PLC mais ont un rôle essentiel dans la diversité des peptides générés capables de se fixer aux molécules CMH-I. Elles sont d'ailleurs inductibles par l'IFN γ ce qui augmente le répertoire d'épitopes en réponse à un stress cellulaire. La déplétion d'ERAP chez la souris a un impact direct sur le répertoire de peptides endogènes présentés (Hammer *et al.*, 2007). En outre, les cellules déficientes en ERAP présentent à leur surface des complexes pCMH instables ayant la capacité d'activer les LTs CD8⁺ et les cellules B. Les aminopeptidases ERAP1 et ERAP2 sont donc essentielles à la diversité du répertoire antigénique et jouent un rôle dans la tolérance immunitaire définie comme l'absence d'activation des cellules immunitaires suite à la reconnaissance des peptides du soi.

Une fois générés, les épitopes sont chargés sur les molécules du CMH-I et soumis à un contrôle qualité avant d'être exportés du RE. La tapasine qui est une protéine transmembranaire du PLC assure la sélection de peptides de haute affinité, processus nommé édition des peptides ou « peptide proofreading » en anglais (Thomas and Tampé, 2017, 2021). La tapasine, en association avec la chaperonne ERp57, favorise la dissociation des peptides sous optimaux fixés sur les molécules du CMH-I et la fixation de peptides plus affins (Wearsch and Cresswell, 2007). Le complexe pCMH devient alors stable, se dissocie du PLC, et est exporté à la membrane plasmique en transitant par l'appareil de Golgi.

2.4.1.2 La présentation croisée : la voie alternative

Découvert en 1976 par le Dr Bevan, la cross-présentation est un mécanisme par lequel une source antigénique exogène est dégradée puis présentée sur les molécules du CMH-I (Bevan, 1976). Ce processus est spécifique aux APCs, et les DCs sont particulièrement efficaces pour présenter des antigènes de manière croisée en raison de leur capacité de dégradation moins importante par rapport aux autres APCs (Rodriguez *et al.*, 1999; Savina *et al.*, 2006). La

présentation croisée a une place importante dans l'immunité adaptative puisqu'elle permet aux DCs de présenter aux LTs CD8⁺ des peptides dérivés de pathogènes qui ne les infectent pas. En plus de l'activation des LTs CD8⁺, la présentation croisée est aussi impliquée dans la tolérance immunitaire. C'est la plasticité cellulaire des DCs et leur habilité à reconnaître des signaux de danger via leur PRR (« pattern recognition receptors ») qui leur permet de passer d'un statut tolérogène à immunogène (Adiko *et al.*, 2015).

Il existe deux voies principales pour la présentation des antigènes exogènes sur les molécules du CMH-I lors de la présentation croisée (Adiko *et al.*, 2015). La première est la voie vacuolaire qui est indépendante des protéines TAP et du protéasome tandis que la seconde, appelée voie cytosolique, dépend de ces dernières (Figure 31). A ce jour, nous n'avons pas une compréhension complète des implications de la voie vacuolaire et cytosolique dans la présentation croisée des antigènes *in vivo*.

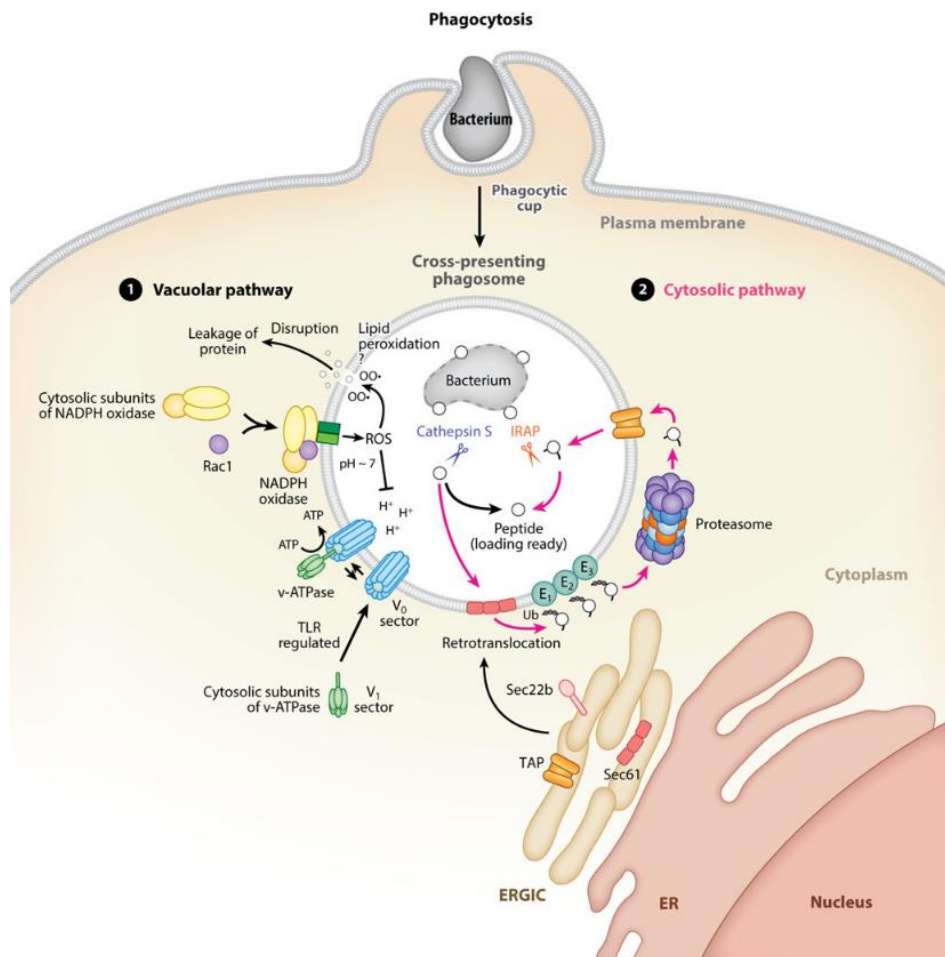


Figure 31. La voie vacuolaire et cytosolique de présentation croisée vacuolaire par les molécules du CMH-I (figure originale de Blander 2018). (1) La voie vacuolaire dégrade les antigènes phagocytés par le biais des cathepsines S. Le pH optimal est maintenu par les ROS générés par la NADPH oxydase. (2) La voie cytosolique repose sur la rétro-translocation au cytosol de l'antigène phagocyté. L'antigène alors présent dans le cytosol est dégradé par le protéasome puis acheminé dans le RE où il sera chargé sur les molécules du CMH-I nouvellement synthétisées comme lors de la présentation endogène des antigènes de classe I.

La voie vacuolaire

Après son internalisation, l'antigène est séquestré dans des vésicules d'endocytose où il est ensuite dégradé puis chargé sur les molécules du CMH-I qui seront ensuite exportées à la membrane plasmique. Cette localisation exclusivement endosomale est spécifique à la voie vacuolaire (Blander, 2018).

Les molécules du CMH-I recrutées aux vésicules endolysosomales proviennent de différentes sources subcellulaires (Figure 32). Par exemple, les molécules du CMH-I recyclées depuis la membrane plasmique dans les endosomes peuvent être directement chargées avec des peptides issus de l'endocytose ou bien servir de réserve de molécules du CMH-I en formant le compartiment de recyclage endosomal (Aderem, 2002; Lizée *et al.*, 2003; Blander, 2018). Deuxièmement lors de la phagocytose, l'activation des signaux du TLR (« Toll-like receptor ») induit le recrutement des molécules du CMH-I du compartiment de recyclage endosomal au phagosome ce qui permet de mettre en contact l'antigène avec les molécules du CMH-I (Nair-Gupta *et al.*, 2014). D'autres sources subcellulaires des molécules du CMH-I que je ne détaille pas dans cet écrit sont possibles (Blander, 2018).

L'antigène est dégradé par le biais de protéases vacuolaires dont la plus importante est la cathepsine S de par son activité à pH alcalin (Shen *et al.*, 2004). En effet, la production de ROS dans la présentation croisée par les DCs permet de maintenir le pH du phagosome alcalin.

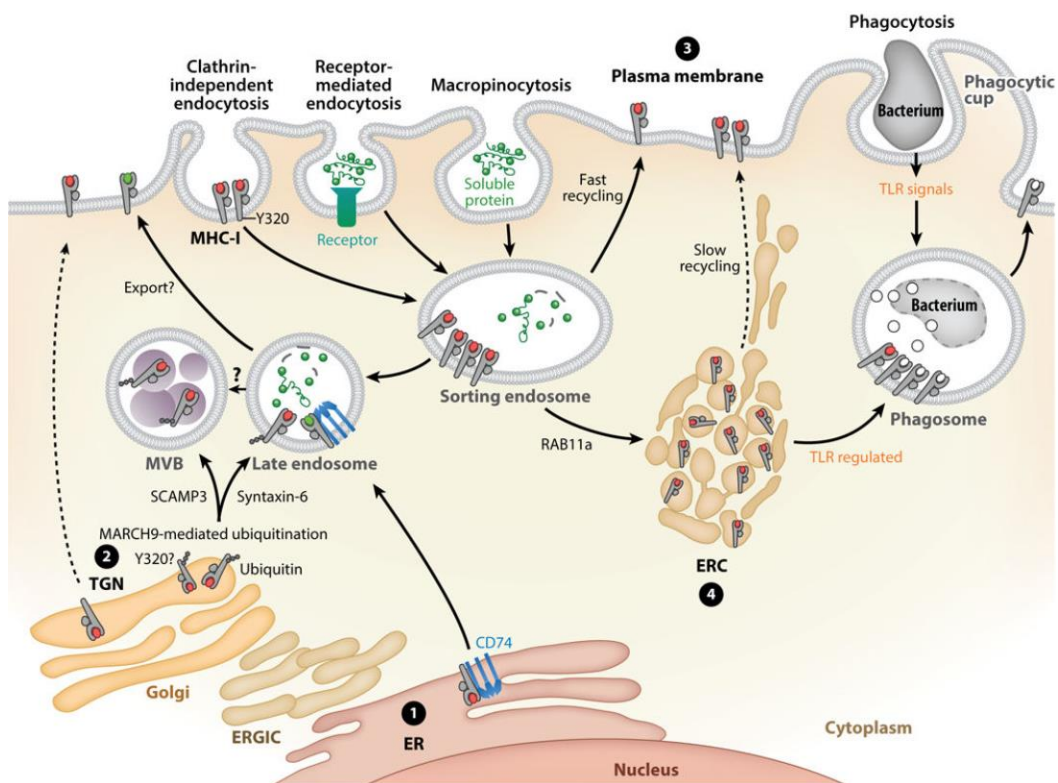


Figure 32. Les différentes sources de molécules de classe I dans la présentation croisée (figure originale de Blander 2018). Quatre sources distinctes de molécules du CMH-I peuvent être utilisées par les DCs. (1) Trafic direct des molécules du CMH-I du RE vers les compartiments endolysosomaux contenant le cargo phagocyté. (2) Le détournement de la voie sécrétoire des molécules du CMH-I vers des particules endosomales. (3) Le recrutement des molécules du CMH-I de la membrane aux compartiments endolysosomaux. (4) Le recrutement de molécules du CMH-I recyclés par la signalisation des TLR.

Le chargement des épitopes sur les molécules du CMH-I est suivie de l'exportation des pMHC vers la membrane plasmique.

La voie cytosolique

La voie cytosolique est dépendante de la protéine TAP et du protéasome. A l'inverse de la voie vacuolaire, le cargo du phagosome est libéré dans le cytoplasme par rétro-translocation (Figures 28 et 31). L'exportation des antigènes vers le cytosol peut se faire selon deux modèles : par le biais d'un transporteur ou par la perturbation des membranes du phagosome (Gros and Amigorena, 2019). Dans l'hypothèse de la molécule de transport, les antigènes doivent être dépliés avant leur translocation. Cette étape est assurée par l'enzyme GILT, une thiol réductase phagosome exprimée constitutivement par les DCs, qui réduit les ponts disulfures des antigènes. Les D^{rs} Singh and Cresswell ont montré que l'initiation des réponses LTs CD8⁺ spécifiques de l'HSV-1 ou du virus de la grippe A est fortement dépendante de GILT (Singh and Cresswell, 2010). La protéine Hsp90 est l'une des protéines de transport impliquée dans l'export de l'antigène déplié au cytosol (Imai *et al.*, 2011). Une fois dans le cytosol, l'antigène est soit replié par le biais de Hsp90 qui restaure son activité, soit dégradé par le protéasome et chargé sur les molécules du CMH-I néosynthétisées (Giodini and Cresswell, 2008).

Ces deux voies non canoniques de présentations des antigènes de classe I impliquent des facteurs moléculaires et de voies intracellulaires complexes dont certaines restent incomprises ou controversées. Bien que certaines questions doivent être résolues, la présentation croisée contribue significativement à l'immunité en exposant les LTs CD8⁺ à de nombreux antigènes viraux, bactériens et tumoraux.

2.4.2 La voie de présentation des antigènes par les molécules du CMH-II

2.4.2.1 Présentation des antigènes exogènes : la voie canonique

Historiquement, les molécules du CMH-II présentent des peptides de sources exogènes (bactéries, champignons) qui ont été internalisés depuis le milieu extracellulaire dans les APCs. La voie canonique de présentation des antigènes de classe II repose sur l'engagement de la voie vésiculaire et fait intervenir de nombreux facteurs moléculaires. Différentes voies d'endocytose peuvent être utilisées pour la capture et l'apprêtement des antigènes telles que la macropinocytose, la phagocytose, et l'endocytose modulée par les récepteurs de surface. Les cargos internalisés dans les endosomes précoces fusionnent avec les lysosomes et sont ensuite acheminés jusqu'au compartiment de chargement des molécules du CMH-II (MIIC). Les antigènes y subissent leur dernier traitement protéolytique et l'épitope le plus affiné est chargé dans le sillon de fixation des peptides des molécules du CMH-II. Le pCMH est exporté à la surface cellulaire pour interagir avec les LTs CD4⁺ (Figure 33).

L'endocytose modulée par les récepteurs de surface

Les APCs possèdent des récepteurs de surface qui favorisent l'endocytose de particules exogènes dans le MIIC. Les récepteurs de types lectine, tel que le récepteur au mannose, sont exprimés à la surface des DCs et des LB. Ils fixent les résidus carbohydrates des pathogènes et les internalisent soit par endocytose dépendante des clathrines soit par phagocytose (Roche and Furuta, 2015). Les DCs matures et les LB expriment également à leur surface des récepteurs

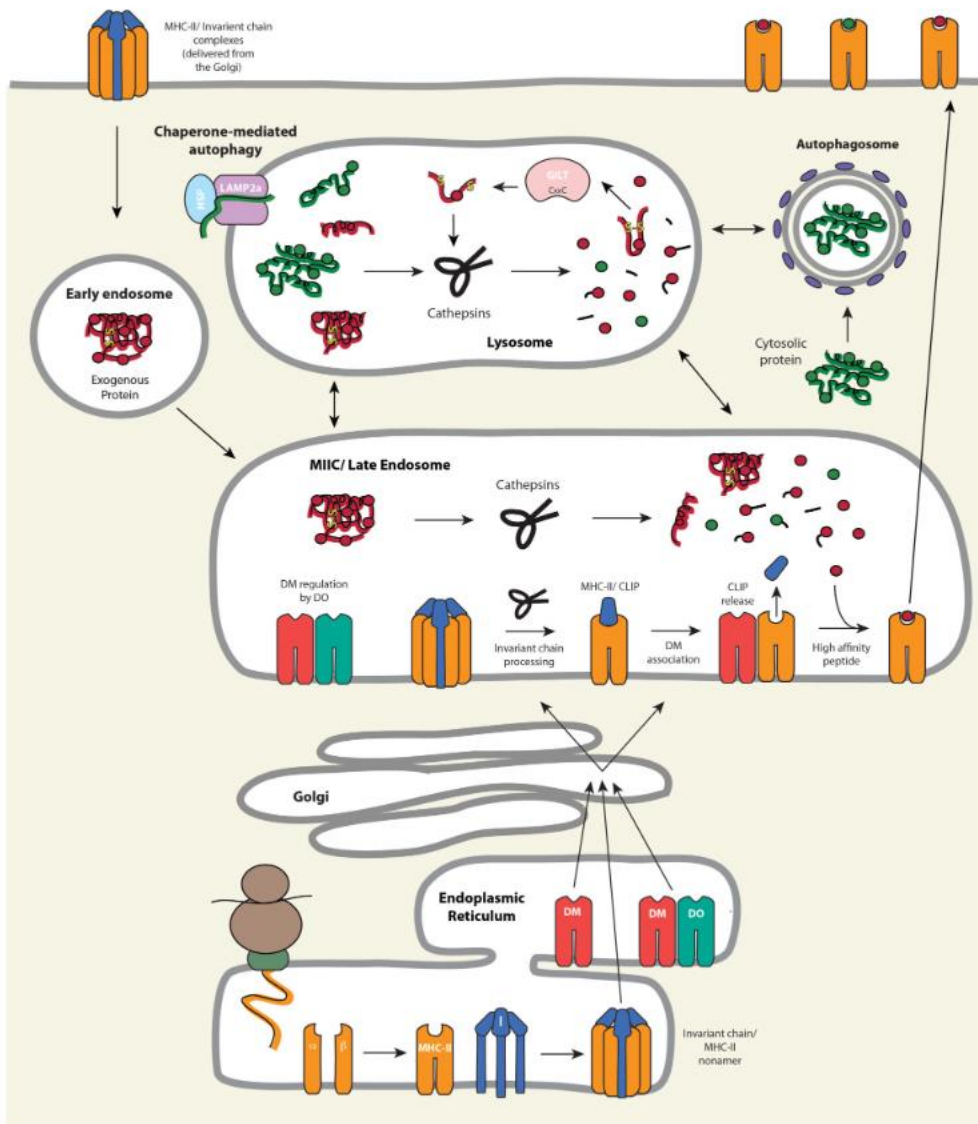


Figure 33. Voie de présentation des antigènes de classe II (figure adaptée de Blum et al 2013). Les protéines α et β du CMH-II s'associent aux trimères de la chaîne invariante Li (I) pour former des nonamères. Ces complexes transitent vers les endosomes tardifs/MIIC soit via l'appareil de Golgi, soit par recyclage à partir de la surface cellulaire. Dans le MIIC, la chaîne Li est protéolysée séquentiellement pour donner le fragment résiduel de la chaîne Li, CLIP. Le déplacement de CLIP du sillon pour un ligand affiné du CMH-II est médié par HLA-DM, lui-même régulé par HLA-DO. Les antigènes sont livrés aux endosomes tardifs/MIIC par phagocytose, pinocytose, endocytose et autophagie puis traités par les cathepsines. Les complexes CMH-II-peptides sont ensuite transportés à la surface des cellules pour être reconnus par les cellules T CD4+.

à la partie Fc des immunoglobulines de type G ($Fc\gamma R$) qui capturent les complexes immuns. Ces différents récepteurs de surface ciblent les antigènes dans les compartiments endolysosomiaux puis dans le MIIC pour leur présentation aux LTs CD4+.

La lectine « DC-SIGN » est notamment connue dans le contexte du VIH-1 pour capturer et promouvoir la présentation de peptides viraux sur les molécules du CMH-II et initier une réponse LT CD4+ spécifique (Moris *et al.*, 2006a). DC-SIGN est un récepteur présent sur les macrophages et les DCs, en condition inflammatoire, qui fixe les pathogènes par la reconnaissance de glycoprotéines riches en mannose. L'expression de DC-SIGN est délétère car elle favorise l'attachement du virus à la cellule cible lors de l'infection. Bien que l'infection des DCs soient généralement peu productive (10 à 100 fois moins efficace *in vivo* que

l'infection des TCD4⁺), elles agissent néanmoins comme des vecteurs de transport pour le virus. Après sa reconnaissance par DC-SIGN, le virus est séquestré dans des compartiments appelés corps multivésiculaires (MVB). Si le MVB fusionne avec un lysosome, le virus est dégradé et présenté sur les molécules du CMH-II. Sinon, le virus est transporté par l'intermédiaire de la DC vers les ganglions lymphatiques où il sera libéré par synapse immunologique ou exocytose. Ce processus est appelé trans-infection (Figure 34) (Martín-Moreno and Muñoz-Fernández, 2019).

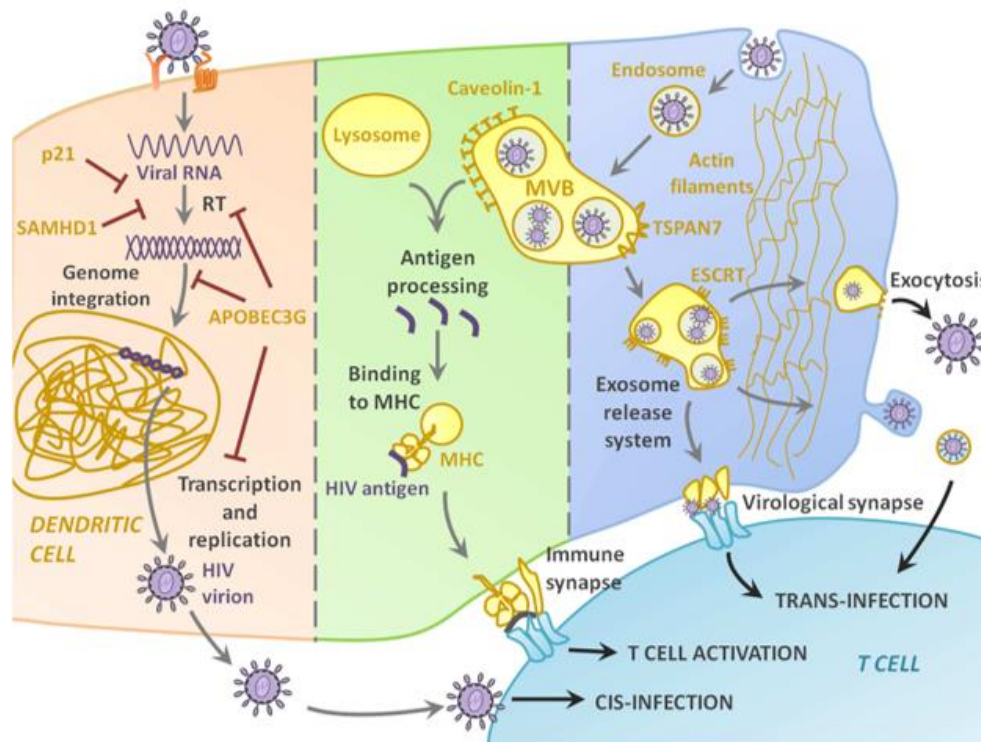


Figure 34. Les différentes voies et acteurs moléculaires utilisés par le VIH-1 à l'intérieur de la DC (figure originale de Martín-Moreno and Muñoz-Fernández 2019). Grâce à l'intervention de facteurs de restriction cellulaires tels que p21, SAMHD1, et APOBEC3G, l'infection des DCs par le VIH-1 est peu efficace (panel de gauche). Néanmoins le VIH-1 peut pénétrer dans la DC par endocytose et s'accumuler dans les MVB. La présence du virus dans les MVB peut conduire à l'activation ou à la trans-infection des LTs CD4⁺.

Concernant les LBs, ils internalisent principalement les antigènes exogènes par l'intermédiaire de leur récepteur BCR (« B-cell receptor »). Le BCR est une immunoglobuline qui reconnaît un épitope spécifique. Les LBs dotés de BCR spécifiques d'antigènes sont 1 000 fois plus efficaces dans le traitement et la présentation de l'antigène aux LTs apparentés que les LBs dotés de BCR aspécifiques (Rock, Benacerraf and Abbas, 1984).

Le MIIC

Le MIIC est un compartiment vésiculaire spécialisé dans la présentation des antigènes de classe II. C'est un endosome tardif qui a un pH acide et où les molécules HLA-DR, -DM, -DO, et les cathepsines B, H, L et S sont présentes (Lautwein *et al.*, 2002; Hartman *et al.*, 2010). Après leur synthèse dans le RE, les molécules du CMH-II sont couplées à la chaîne invariante Li, aussi appelée CD74. CD74 est une glycoprotéine transmembranaire qui stabilise et protège le sillon peptidique des molécules du CMH-II par son domaine CLIP, inhibant ainsi la fixation des épitopes de faible affinité (Roche and Cresswell, 1991; Romagnoli and Germain, 1994; Sadegh-Nasseri *et al.*, 1994; Busch *et al.*, 1996; Blum, Wearsch and Cresswell, 2013; Lindner, 2017). Les

molécules du CMH-II présentes dans le MIIC proviennent soit de la ré-internalisation des molécules du CMH-II exposées à la membrane plasmique soit d'une néosynthèse dans le RE (Roche *et al.*, 1993; Raiborg and Stenmark, 2009). A ce stade, les molécules du CMH-II du MIIC sont immatures.

L'acidification du MIIC mène à l'activation des protéases endolysosomales représentées principalement par la famille des cathepsines. La protéolyse de CD74, essentielle au chargement de peptides de haute affinité, est finement régulée par HLA-DM et les cathepsines. Plusieurs cathepsines et endo-peptidases ont été rapportées pour jouer un rôle dans la dégradation de CD74 mais les cathepsines S et L semblent jouer un rôle prépondérant (Manoury *et al.*, 2003; Blum, Wearsch and Cresswell, 2013). La chaîne invariante CD74 est donc entièrement dégradée à l'exception du peptide CLIP qui reste intégré dans le sillon peptidique. C'est finalement la molécule HLA-DM qui catalyse la dissociation de CLIP pour un peptide de plus haute affinité (Denzin and Cresswell, 1995; Sanderson *et al.*, 1996). La molécule HLA-DO est exprimée par les APCs et permet une régulation négative de HLA-DM (Guce *et al.*, 2013).

Le processus de transport des molécules du CMH-II chargées du MIIC à la surface de la cellule n'est pas encore bien caractérisé. Deux voies de transport ont été proposées : l'une impliquant un réseau d'actine et l'autre impliquant le transport du MIIC le long des microtubules (Neeffes *et al.*, 2011). Le mouvement du MIIC et son extension vers la surface de la cellule sont directement liés à la capacité des DC à présenter des antigènes aux LT CD4⁺ (Boes *et al.*, 2002).

2.4.2.2 Les voies alternatives de présentation antigénique classe II

A l'image de la présentation croisée où des peptides de source exogène sont présentés par les molécules du CMH-I, des peptides issus de la synthèse protéique peuvent être présentés par les molécules du CMH-II. Ces voies alternatives de présentation antigénique sont capables d'initier une réponse LT CD4⁺ spécifique de peptides issus de protéines néosynthétisées. Ce phénomène a été initialement décrit dans les années 1980, notamment quand Jacobson et al ont mis en évidence la reconnaissance de peptides cytosoliques dérivés du virus de la rougeole par des LTs CD4⁺ (Jacobson *et al.*, 1988). Par la suite, de nombreuses études ont également démontré que certaines protéines virales pouvaient être présentées par les molécules du CMH-II selon la voie de présentation endogène (Leung, 2015). L'ultime démonstration est faite suite à l'analyse du répertoire de peptides présentés par les molécules du CMH-II (appelé ligandome ou immunopeptidome) qui révèle qu'une fraction des peptides présentés est issue de protéines intracellulaires (Chicz *et al.*, 1993; H. Rammensee *et al.*, 1999; Dongre *et al.*, 2001; Muntasell *et al.*, 2002). Depuis, de nombreuses voies alternatives ont été identifiées, en particulier grâce à l'étude des antigènes viraux (Purnima, Ganesan and Eisenlohr, 2017). Différentes voies d'apprêtement des antigènes endogènes restreints aux molécules de classe II sont proposées, notamment la voie vésiculaire, la voie cytosolique, et l'autophagie (Figure 35).

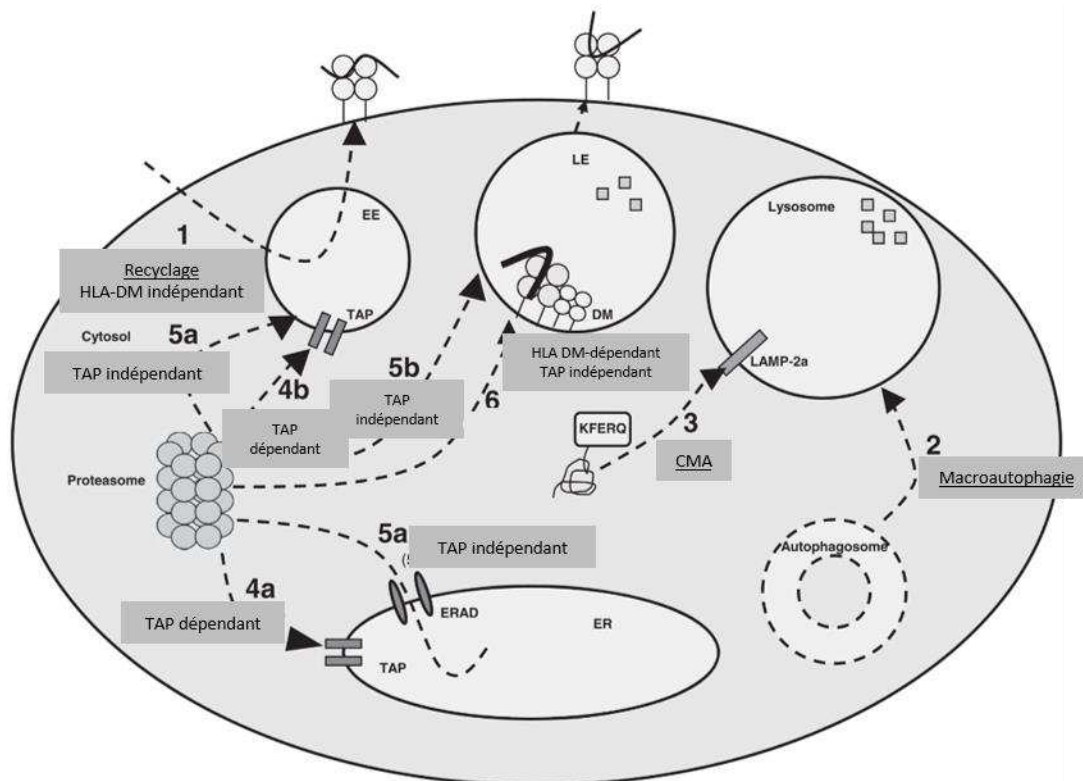


Figure 35. Les différentes voies non canoniques de présentation des antigènes intracellulaires par les molécules du CMH-II (figure adaptée de Purnima et al. 2017 et Pereira, 2019). Les différents phénomènes menant à la présentation d'antigènes endogènes sur les molécules du CMH-II sont représentés. 1) La voie de recyclage indépendante de HLA-DM ; 2) La macroautophagie ; 3) L'autophagie modulée par des chaperonnes (CMA) ; 4a et 4b) les voies d'apprêtement dépendante de TAP ; 5a et 5b) les voies d'apprêtement indépendante de TAP ; 6) L'acheminement par un mécanisme TAP indépendant et HLA-DM dépendant. EE : endosome précoce ; LE : late endosome ; ER : reticulum endoplasmique ; ERAD : complexe de dégradation associé au RE.

La voie vésiculaire : l'exemple du recyclage des molécules du CMH-II

Dans ce modèle, les peptides sont chargés sur les molécules du CMH-II matures suite à leur internalisation par la voie de recyclage. Le chargement des peptides sur les molécules du CMH-II est indépendant de HLA-DM et de CD74 (Figure 35, voie n°1). Après leur recyclage depuis la membrane plasmique, les molécules du CMH-II se situent dans les endosomes précoces. Il a été mis en évidence que suite à leur absorption, certains peptides de la grippe, résistants aux protéases des endosomes précoces, sont chargés sur les molécules de classe II recyclées (Pinet *et al.*, 1995; Sinnathamby and Eisenlohr, 2003). Ayant une conformation ouverte, le sillon du CMH-II présents dans les endosomes précoces peut fixer de grands peptides qui ne nécessitent pas de traitement protéolytique comme il peut y avoir lieu dans les endosomes tardifs. Dans ce cas de figure, les endosomes précoces et tardifs (voie exogène canonique) génèrent des complexes pMHC-II différents (Griffin, Chu and Harding, 1997). L'étude de Tewari et al montre que la génération et l'acheminement de peptides de la grippe dans les endosomes précoces où se situent les molécules du CMH-II recyclées sont dépendants du protéasome et de la protéine TAP (Tewari *et al.*, 2005) (Figure 35, voie n°4b). En utilisant un modèle murin, Miller et al ont montré que la majorité des réponses CD4⁺ dirigées contre les protéines du virus de la grippe sont générées suite au traitement de l'antigène par la voie de présentation endogène dans les APCs infectées (Miller *et al.*, 2015) (Figure 35, voie n°5a). Ce processus est décrit comme dépendant du protéasome mais indépendant des pompes TAP.

La voie cytosolique

Chargement dans les endosomes tardifs ou le RE

Les mécanismes permettant de charger les précurseurs peptidiques dans les endosomes tardifs ou dans le RE peuvent dépendre séparément de HLA-DM et de TAP, ou être totalement indépendant de ces éléments (Purnima, Ganesan and Eisenlohr, 2017). Dani et al, montrent que des peptides cytosoliques sont présents dans les endosomes tardifs et que la formation du complexe pMHC est dépendant de la chaîne invariante et de la chaperonne HLA-DM. Néanmoins, l'acheminement du précurseur dans l'endosome tardif ne nécessite pas l'implication de TAP (Figure 35, voie n°6) (Dani *et al.*, 2004). Malnati et al mettent en évidence l'implication de la protéine TAP dans la présentation des peptides dérivés du virus de la grippe restreints par HLA-DR1 (Malnati *et al.*, 1992) (Figure 35, voie n°4a). Une autre étude vient démontrer que l'acheminement des peptides endogènes dans le RE peut être TAP indépendant (Miller *et al.*, 2015) (Figure 35, voie n°5a). Dans le même esprit, l'antigène tumoral NY-ESPO-1 est apprêté par un processus TAP-indépendant sur les molécules du CMH-II (Tsuji *et al.*, 2012; Matsuzaki *et al.*, 2014). Son apprêtement nécessite l'activité de la protéine HSP90. Dans le modèle du mélanome, l'antigène gp100 est chargé sur les molécules du CMH-II par le biais d'une séquence signal d'adressage au MIIC. Ces séquences optimisent la réponse LT CD4⁺ spécifiques de gp100 (Lepage and Lapointe, 2006).

L'autophagie et l'implication de ses différents effecteurs moléculaires

Plusieurs études ont démontré l'implication de l'autophagie dans la présentation des peptides endogènes. Par l'utilisation de molécules inhibitrices et de ARN interférant, deux études montrent que l'antigène nucléaire EBNA1 de l'EBV active une réponse LT CD4⁺ via la macroautophagie (Münz *et al.*, 2000; Paludan *et al.*, 2005; Leung *et al.*, 2010) (Figure 35, voie n°2). L'analyse de l'immunopeptidome des molécules du CMH-II révèle que l'autophagie participe au traitement des protéines cytosoliques et nucléaires (Dengjel *et al.*, 2005).

Toutefois, un autre mécanisme de présentation non canonique des antigènes de classe II indépendant de la voie autophagique vient d'être mise en évidence au laboratoire. Ce mécanisme repose sur l'implication d'un seul effecteur moléculaire de l'autophagie : le récepteur autophagique TAX1BP1. Lors de précédents travaux réalisés au laboratoire, Coulon et al, ont mis en évidence que des peptides dérivés des protéines endogènes Gag et Env du VIH-1 sont chargées sur les molécules du CMH-II et initient une réponse des LTs CD4⁺ (Coulon *et al.*, 2016). Le traitement endogène de ces protéines virales est démontré comme étant un mécanisme indépendant de l'autophagie. Néanmoins, on observe une potentialisation de la réponse CD4⁺ quand les antigènes sont directement adressés à l'autophagosome. D'autres exemples où le processus de l'autophagie (formation de l'autophagosome puis fusion avec une vésicule lysosomale) n'intervient pas dans la présentation des antigènes endogènes ont été décrits (Paludan *et al.*, 2005; Leung, 2015). Une littérature de plus en plus fournie indique que certains effecteurs moléculaires de l'autophagie peuvent affecter directement ou indirectement la présentation des antigènes restreints par les molécules du CMH-II (Fletcher *et al.*, 2018; Sarango *et al.*, 2022). Les récepteurs autophagiques sont impliqués dans de multiples processus cellulaires tels que la xénophagie, la mitophagie, la dégradation des bactéries

internalisées etc. Notre équipe s'est donc intéressé à étudier le rôle des récepteurs autophagiques dans la présentation des antigènes endogènes par les molécules du CMH-II.

L'étude menée par Sarango et al démontrent que seule la déplétion de TAX1BP1 entraîne une diminution drastique de l'activation des LTs CD4⁺ spécifiques de peptides dérivés du VIH-1 et du HCMV (Sarango *et al.*, 2022). L'analyse de l'immunopeptidome montre que la déplétion de TAX1BP1 influence drastiquement la nature des peptides présentés par les molécules du CMH-II. Les auteurs montrent une relocalisation périnucléaire des molécules du CMH-II en l'absence de TAX1BP1. Ils observent également que le récepteur autophagique stabilise et protège la chaîne invariante CD74 d'une dégradation ce qui permet le maintien du contrôle qualité des peptides présentés. L'analyse de l'interactome de TAX1BP1 montre une interaction avec la calnexine, chaperonne présente dans le RE qui assure la stabilité de la chaîne invariante CD74. Cette étude démontre donc le rôle majeur du récepteur autophagique TAX1BP1 dans la présentation des antigènes viraux endogènes par les molécules du CMH-II. La voie d'apprêtement par laquelle les peptides sont acheminés au MIIC reste à élucider.

2.5. Les sources d'antigènes présentés par les molécules du CMH

Nous venons de voir qu'une partie importante des peptides présentés par les molécules du CMH de classe I et II sont issus de protéines endogènes cytosoliques ou nucléaires, soulignant le lien direct entre la traduction et la présentation des antigènes. Dans cette sous-partie de mon manuscrit, je vais détailler le rôle de la traduction canonique et non canonique en tant que sources antigéniques, qui constituent un axe important de mon sujet de thèse.

2.5.1 La traduction canonique comme source antigénique

2.5.1.1 Les protéines retraitées VS les DRiPs

Les protéines néosynthétisées à partir d'un ORF sont stables, assurent leurs fonctions, puis sont dégradées selon une cinétique de premier ordre avec une demi-vie caractéristique. Lorsque la protéine a assuré ses fonctions et a atteint son temps de demi-vie, elle est alors appelée « retraitée » (« retires » est le terme employé en anglais) et elle est ensuite dégradée par le protéasome puis présentée à la surface des cellules par les molécules du CMH. La dégradation par le protéasome des protéines mal repliées suite à leur traduction ou des protéines stables ayant assurées leurs fonctions est essentielle au maintien de l'homéostasie cellulaire en condition physiologique (Kniepert and Groettrup, 2014; Rock *et al.*, 2014). Historiquement, le renouvellement protéique régulier constitue la source majoritaire d'antigène, que les protéines soient défectueuses ou exemptes d'erreurs (Antón and Yewdell, 2014; Rock *et al.*, 2014).

Néanmoins, en 1992, le D^r Esquivel fait la découverte que les LTs CD8⁺ sont capables de reconnaître des peptides viraux présentés à la surface des molécules du CMH-I de cellules infectées bien avant que les protéines virales puissent être détectées (Esquivel, Yewdell and Bennink, 1992). Cela sous-entend l'existence d'un mécanisme alternatif, plus rapide que la traduction canonique, menant à la production de peptides immunogènes. Suite à cette étude les D^{rs} Yewdell, Antón et Bennink introduisent le modèle des DRiPs (« Defective ribosomal products ») (Yewdell, Antón and Bennink, 1996). A l'origine, les DRiPs sont définis comme des produits défectueux de traduction qui n'atteignent pas une conformation stable et qui sont

dégradés pendant ou rapidement après leur traduction. A cette époque, le D^r Yewdell se focalise sur la traduction des ORFs et conclut que les DRiPs proviennent d'une terminaison prématurée ou d'un mauvais repliement de la chaîne naissante. La caractéristique des DRiPs est leur rapidité : dans leur apparition et leur dégradation. En utilisant la méthode de pulse-chase, Schubert et al, sous la direction des D^{rs} Yewdell et Bennink, démontrent que la dégradation des DRiPs par le protéasome s'effectue dans les 30 minutes maximum après leur synthèse (Schubert *et al.*, 2000). Les auteurs montrent également dans plusieurs types cellulaires que les DRiPs représentent environ 30% des produits nouvellement synthétisés. Les DRiPs sont néanmoins comptabilisés avec les protéines stables mais à courtes durée de vie que l'on appelle les SLiPs (« Short-life proteins ») due à la difficulté technique pour les distinguer. Environ 70% des protéines synthétisées sont stables avec une demi-vie d'en moyenne 3 000 minutes, alors que le reste, les DRiPs et les SLiPs ont une demi-vie d'en moyenne 10 minutes avant leur dégradation par le protéasome (Schubert *et al.*, 2000; Yewdell, 2003). De plus, en transfectant des cellules HeLa avec le clone moléculaire infectieux du VIH-1 (pNL4-3) et en utilisant des inhibiteurs du protéasome, ils détectent par pulse-chase des DRiPs ubiquitinés issus du précurseur Gag Pr55. La conclusion de leurs nombreux travaux dans ce domaine est que les DRiPs représenteraient la source majeure d'antigènes associés aux molécules de classe I et participent à l'immunosurveillance des cellules T (Antón and Yewdell, 2014).

Plus récemment, des travaux ont caractérisé les DRiPs à un niveau moléculaire. Une étude analysant l'immunopeptidome des molécules du CMH-I montre que la majorité des peptides viraux sont détectés avant la détection des peptides viraux issus de protéines canoniques suite à l'infection par le virus de la grippe A (IAV) (Wu *et al.*, 2019). La dégradation des DRiPs par le protéasome a été montrée comme dépendante et indépendante de l'ubiquitination en fonction du modèle cellulaire utilisé et des circonstances exactes de la génération des peptides (Wei *et al.*, 2017; Yewdell and Holly, 2020). Beaucoup de questions fondamentales sur la synthèse des DRiPs, leur nature, leur dégradation restent ouvertes.

Il convient de souligner que les DRiPs ont été initialement définis comme des produits défectueux issus de la traduction des ORFs (Yewdell, Antón and Bennink, 1996). Cependant, suite à la publication de nombreuses études mettant en évidence les différents rôles biologiques de la traduction des sORFs, la définition des DRiPs s'est étendue aux produits issus de la traduction non canonique, et plus largement, à toutes les erreurs qui se produisent lors de la conversion de l'information génétique en protéines (Antón and Yewdell, 2014). Ainsi, à ce jour, les DRiPs sont définis comme l'ensemble des produits de traduction défectueux, qu'ils proviennent de régions canoniques ou non canoniques. Par conséquent, les DRiPs peuvent également être inclus dans la section suivante du manuscrit, intitulée « La traduction non canonique comme source antigénique ».

2.5.1.2 Le peptide splicing

L'épissage du protéasome, ou « peptide splicing », est un mécanisme par lequel le protéasome génère des peptides composés de deux fragments initialement éloignés dans la protéine parentale (Vigneron *et al.*, 2019). En 2004, la découverte de peptides épissés issus de cellules cancéreuses a remis en question le concept général selon lequel les LTs CD8+ ne reconnaissent

que des peptides dérivés de fragments linéaires de protéines (Hanada, Yewdell and Yang, 2004; Vigneron *et al.*, 2019). Deux équipes de recherche ont identifié chacune un peptide épissé par le protéasome, présenté à la surface des cellules tumorales et reconnu par les LTs CD8⁺ (Hanada, Yewdell and Yang, 2004; Vigneron *et al.*, 2004). Le phénomène par lequel les deux fragments peptidiques sont assemblés par le protéasome est appelé transpeptidation. Plus la séquence séparant les deux fragments est courte, plus l'épissage par le protéasome est efficace. Quelques années plus tard, l'existence d'un épissage inverse, où les deux fragments peptidiques sont assemblés dans un ordre inverse est rapporté (Vigneron *et al.*, 2019). Le protéasome comme l'immunoprotéasome sont les deux médiateurs de l'épissage peptidique. Des peptides épissés immunogènes ont été rapportés dans le VIH-1, la *Listeria*, et le diabète de type 1 (DeLong *et al.*, 2016; Platteel *et al.*, 2016; Paes *et al.*, 2019). Le modèle du diabète de type 1 est d'ailleurs la première et unique démonstration de peptides épissés restreint aux molécules du CMH-II. Le peptide, reconnu par les clones LTs CD4⁺ diabétogènes, est issu de l'épissage de fragments peptidiques provenant de deux protéines différentes présentes dans les granules sécrétoires des cellules β (DeLong *et al.*, 2016).

Néanmoins, l'abondance des peptides épissés et leur relevance immunologique est très controversée dans la communauté scientifique (Admon, 2021; Lichti *et al.*, 2022). Certains chercheurs s'accordent sur le fait que les peptides épissés ne composent pas une large fraction de l'immunopeptidome, voir même qu'ils en sont totalement absents. Leurs affirmations sont fondées sur les méthodologies utilisées, biochimiques et bioinformatiques. Par exemple, l'eau, qui est un inhibiteur compétitif de la transpeptidation, est trop abondante dans le milieu intracellulaire rendant la transpeptidation alors inefficace voir peu probable. L'annotation bioinformatique des peptides épissés serait incorrecte. Il serait davantage probable que les peptides épissés soient finalement des peptides non épissés résultant de transcrits non canoniques pas encore identifiés (et donc pas répertoriés dans la base de données). De plus, pour certaines études la disponibilité des données brutes n'est pas toujours évidente, et la validation technique et biologique des peptides identifiés sont souvent manquantes.

2.5.1.3 Les néoantigènes

Dans une cellule tumorale, les multiples altérations de l'ADN peuvent conduire à la formation de séquences en acides aminés nouvelles appelées néoantigènes. A la différence des antigènes associés aux tumeurs et aux antigènes de la lignée germinale du cancer (« C/G antigens »), les néoantigènes sont totalement absents des tissus sains car ils dérivent de mutations spécifiques à la tumeur. C'est une caractéristique majeure pour la pertinence immunologique et thérapeutique des néoantigènes. Cela garantit un répertoire de cellules T disponibles qui n'est pas déplété par la tolérance centrale des cellules T et évite l'activation d'une réponse auto-immune (Schumacher and Schreiber, 2015).

Les néoantigènes peuvent dériver de mutations non synonymes, d'insertions et de délétions, mais aussi de changement de phase de lecture (FS), etc. Les néoantigènes qui dérivent d'évènements non canoniques de traduction seront abordés dans la section suivante. Dans le cas des tumeurs viro-induites telles que le cancer du col de l'utérus lié au HPV, le cancer de la

nuque lié à l'EBV, tout épitope qui dérive d'ORFs viraux est une source potentielle de néoantigènes (Jiang *et al.*, 2019).

En 1988, De Plaen et al sont les premiers à mettre en évidence la reconnaissance d'un peptide muté par les cellules T spécifiques de tumeur (De Plaen *et al.*, 1988). S'en suit plusieurs découvertes charnières dans le domaine notamment en 1995 et en 1996 où des néoantigènes sont identifiés pour être communs à plusieurs patients et où la réactivité des lymphocytes T infiltrant la tumeur est spécifique de néoantigènes (Wölfel *et al.*, 1995; Robbins *et al.*, 1996). De nombreux néoantigènes ont ensuite été identifiés dans le mélanome et le carcinome rénal, et plus largement dans les tumeurs malignes. Plus tard, les groupes de recherche dirigés par les D^{rs} Schreiber et Sahin, ont utilisé des modèles de souris pour montrer que les données génomiques du cancer peuvent être utilisées pour identifier des ensembles potentiels de néoantigènes qui peuvent ensuite être évalués pour leur réactivité aux cellules T (Castle *et al.*, 2012; Matsuzaki *et al.*, 2014). D'autres travaux basés sur la reconnaissance des cellules T ont ensuite révélé que des réponses CD8⁺ mais aussi CD4⁺ dirigées contre des néoantigènes sont fréquemment observées au sein des lymphocytes T infiltrant les tumeurs (TILs) chez les patients atteints de mélanome (Robbins *et al.*, 2013; van Rooij *et al.*, 2013; Linnemann *et al.*, 2015; Verdegaal *et al.*, 2016; Tran, Robbins and Rosenberg, 2017). Cependant, peu d'équipes de recherches étudient l'implication de la réponse immunitaire médiée par les LTs CD4⁺ contre les néoantigènes (Tran *et al.*, 2014; Schumacher and Schreiber, 2015)

2.5.2 La traduction non canonique comme source antigénique

2.5.2.1 La traduction cytosolique

Le laboratoire du D^r Boon est le premier à mettre en évidence le potentiel de la traduction non canonique comme substrat antigénique. En 1989, son équipe montre que des cellules transfectées avec des fragments d'ADN tumoraux déplétés de leurs séquences promotrices ou de motifs d'initiation pouvaient être malgré tout reconnus par des LTs CD8⁺ spécifiques de la tumeur (Boon and Van Pel, 1989). Le laboratoire du D^r Nilabh Shastri, novateur dans le domaine de la présentation antigénique non canonique, montre en 1993, que l'épitope « SIINFEKL » dérivé de l'ovalbumine provient d'un cadre de lecture incorrect (Shastri and Gonzalez, 1993; Apcher *et al.*, 2022). Le D^r Boon et son équipe font la découverte du premier néoantigène dérivé d'intron en 1995 (Coulie *et al.*, 1995). De manière captivante, en 1996, le groupe du D^r Steven A. Rosenberg fait la toute première démonstration que les TILs peuvent reconnaître un peptide issu d'un ARF. Le peptide alternatif « MSLQRQFLR », issu du gène gp75 du mélanome, est codé par un codon AUG qui est présent immédiatement en amont de l'ARF. Les auteurs observent également une corrélation entre la présence des LTs CD8⁺ spécifiques du peptide alternatif (considéré comme néoantigène) et la régression tumorale (Wang *et al.*, 1996). En 1999, un ARF issu d'un FS en +1 du gène de la carboxylestérase intestinale est décrit pour générer un épitope reconnu par les TILs de carcinome rénal *in situ*. Aucune des trois mutations ponctuelles sur les trois codons AUG de l'ARF n'abolit la reconnaissance par les cellules T. En revanche, la mutation du premier codon non-AUG « ACG » en amont abolit la reconnaissance par les LTs CD8⁺ (Ronsin *et al.*, 1999). La même année, le laboratoire du D^r Nilabh Shastri montre par spectrométrie de masse et des tests d'activation des cellules T que les produits issus d'une

initiation non-AUG peuvent être une source de substrats peptidique pour les molécules du CMH-I (Malarkannan *et al.*, 1999). En effet, les peptides traduits à partir d'une initiation « CUG » sont des sources antigénique pour activer la réponse CD8⁺ (Dolstra *et al.*, 1999; Starck *et al.*, 2008, 2012). C'est notamment le cas dans le carcinome rénal où le peptide issu de l'initiation CUG du VEGF active la réponse T CD8⁺ (Weinzierl *et al.*, 2008). De façon complémentaire, il a été reporté que la sécrétion de cytokines pro-inflammatoires lors d'une infection virale ou microbienne favorise la présentation d'épitopes issus d'une initiation « CUG » par les molécules du CMH-I (Prasad, Starck and Shastri, 2016). Ce mécanisme pourrait être une stratégie de l'hôte pour générer des épitopes suite à un stress cellulaire. La reconnaissance de néoantigènes dérivés d'ARFs ou issus d'une traduction dépendante d'IRES par les TILs ont également été mis en évidence (Rosenberg *et al.*, 2002; Carbonnelle *et al.*, 2013).

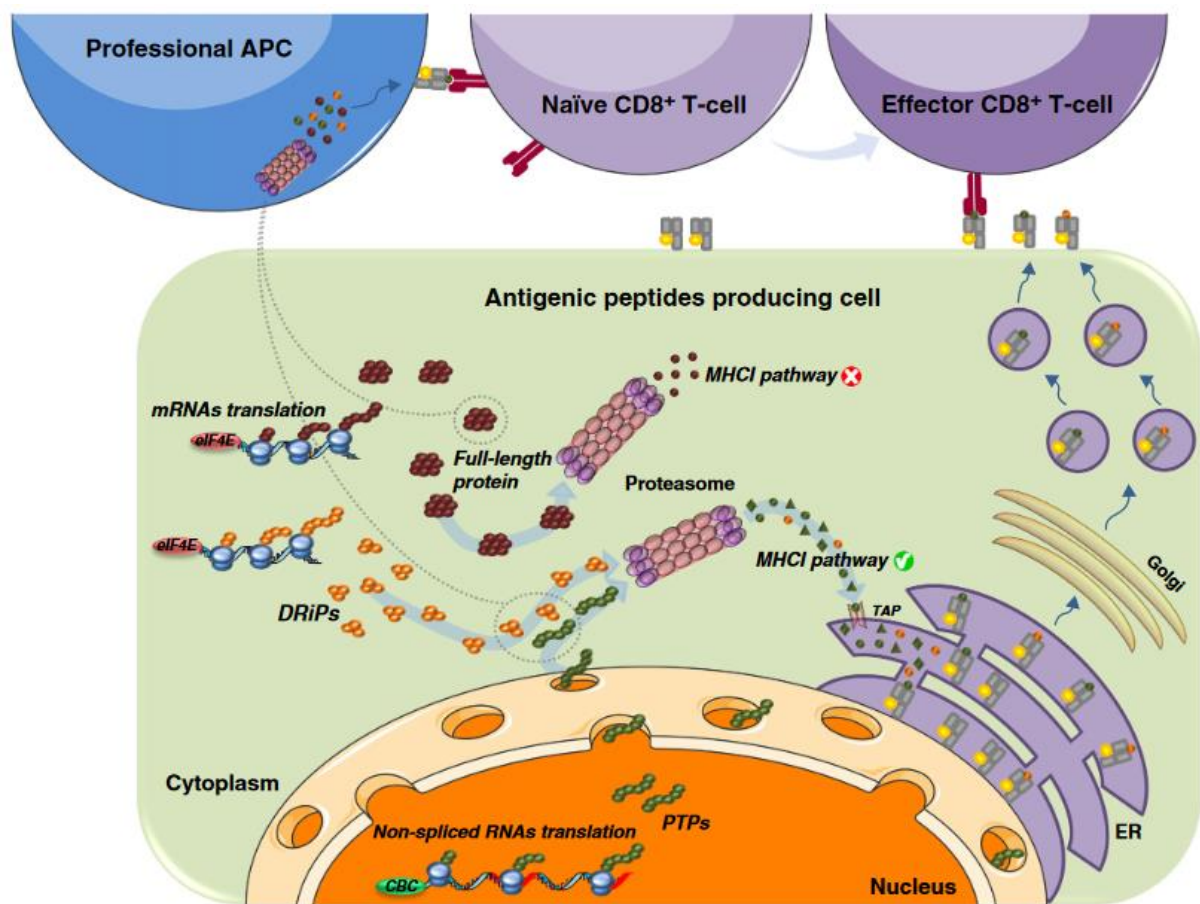
Parallèlement, l'idée que les substrats des molécules du CMH-I puissent dériver d'évènements non conventionnels de traduction fait aussi son chemin sur le modèle viral. Les observations du D^r Fetten ont montré que l'induction d'un décalage de cadre de lecture ne perturbent pas la présentation par les molécules du CMH-I de l'épitope non canonique dérivé de la nucléoprotéine du virus de la grippe (Fetten, Roy and Gilboa, 1991). Et pour cause, Bullock et al ont ensuite démontré que la traduction de ce peptide alternatif est issu d'un leaky scanning (Bullock and Eisenlohr, 1996). La production et la reconnaissance par les LTs CD8⁺ spécifiques de ce peptide est confirmée *in vivo*. Un autre exemple d'épitope dérivé de la nucléoprotéine de la grippe est démontré pour être codé par un ARF (Elliott, Bodmer and Townsend, 1996). Mayrand et al ont fait la démonstration que l'épitope immunodominant « SYNTGRFPP » reconnu par les LTs CD8⁺ est codé par un ARF du rétrovirus murin LP-BM5 (Mayrand, Schwarz and Green, 1998).

Notre équipe de recherche a été l'une des premières à étudier la traduction non canonique du VIH-1 en tant que source antigénique. En 2004, Cardinaud et al mettent en évidence des réponses T CD8⁺ dirigées contre six peptides dérivés d'ARFs chez les individus infectés asymptomatiques et traités (Cardinaud *et al.*, 2004). Plus tard, les études se multiplient dans le modèle simien et humain du virus de l'immunodéficience où une réponse T CD8⁺ contre des peptides dérivés d'ARFs est observée en phase aiguë et chronique d'infection bien que ces ARFs n'aient pas été identifiés par ribosome profiling (Maness *et al.*, 2007, 2010; Garrison *et al.*, 2009; Bansal *et al.*, 2010; Berger *et al.*, 2010; Champiat *et al.*, 2012; Peng *et al.*, 2018).

2.5.2.2 La traduction nucléaire

Malgré le fait que les études pionnières mettent en évidence dans les années 60 la présence de ribosome en cours de traduction dans le noyau par des marquages radioactifs et de la microscopie électronique, le dogme a rapidement dicté que la synthèse protéique n'a lieu exclusivement que dans le cytoplasme (Allfrey, Mirsky and Osawa, 1957; Carneiro and Leblond, 1959; Birnstiel and Hyde, 1963; Zimmerman *et al.*, 1969; Wu and Warner, 1971). Les recherches sur la traduction nucléaire reprennent une trentaine d'années plus tard avec les travaux de Iborra et al (Iborra, Jackson and Cook, 2004). L'équipe du D^r Yewdell rapporte en 2010 que l'inhibition de l'export nucléaire du transcrit de la neuraminidase s'accompagne d'une réduction de seulement 5 fois l'expression du peptide correspondant à la surface cellulaire

suggérant ainsi que les sources antigéniques peuvent être compartimentées (Dolan *et al.*, 2010). Malgré leur effort ainsi que ceux de deux autres équipes, le scepticisme sur la traduction nucléaire persiste (Antón and Yewdell, 2014). En 2011, les travaux du D^r Apcher apportent un soutien à la traduction nucléaire puisque les auteurs montrent que les cellules T CD8⁺ OT-1 reconnaissent aussi bien le peptide « SIINFEKL » qu'il soit issu d'intron ou d'exon (Apcher *et al.*, 2011). L'épissage alternatif a lieu dans le noyau et permet l'élimination des introns, étape qui est nécessaire à la maturation des transcrits avant leur export du noyau vers le cytoplasme. L'introduction d'un changement de cadre de lecture dans le peptide issu de l'intron abolit sa reconnaissance par les cellules T. En bloquant l'export nucléaire ils observent une augmentation de la synthèse du peptide qu'il soit généré à partir de l'intron ou de l'exon. Pour finir, les auteurs mettent en évidence par ribopuromycylation, une technique d'immunofluorescence qui permet de cartographier les sites de traduction à des niveaux subcellulaires, une colocalisation dans le noyau de l'épitope avec les ribosomes en cours de traduction (Apcher *et al.*, 2011; Bastide, Yewdell and David, 2018).



Current Opinion in Immunology

Figure 36. Les diverses sources non canoniques d'antigènes (figure originale de Apcher et al 2016).

Comme nous l'avons vu dans le chapitre 1 (section 1.10.4.2), la traduction de régions génomiques non canoniques telles que les ARFs et les uORFs peut entraîner la synthèse des polypeptides fonctionnels. Par exemple, certains peptides issus de uORFs appelés « peptidoswitch » inhibent la traduction de la séquence codante située en aval. La protéine MYC traduite à partir de l'uORF contrôle la croissance cellulaire. Les protéines « AltFUS » et

« AltRPL36 » issues d'ARFs, inhibent l'autophagie et régulent négativement les voies de signalisations impliquées dans la synthèse protéique et la croissance cellulaire, respectivement. Malgré l'existence de quelques exemples, les démonstrations fonctionnelles manquent cruellement pour la plupart des régions non annotées identifiées (sORFs). Néanmoins, l'ensemble des études discutées dans cette section soulignent la capacité des régions non annotées à coder pour des peptides immunogènes et leur importance en tant que réservoir immunologique (Figure 36). Les régions non annotées sont une source d'antigènes, qu'elles codent ou non pour une protéine fonctionnelle. Le développement de la spectrométrie de masse appliquée à la présentation antigénique a été une avancée majeure dans l'étude des peptides présentés à la surface des molécules du CMH que l'on définit de ligandome ou d'immunopeptidome. En effet, au fil des ans, les scientifiques ont consacré de nombreux efforts pour identifier les antigènes tumoraux qui pourraient être ciblés par les thérapies, visant à stimuler une réponse immunitaire contre la tumeur. Dans la section suivante de mon manuscrit, j'exposerai en détail comment la spectrométrie de masse peut être utilisée pour définir l'immunopeptidome. De façon non exhaustive, je conclurai ce chapitre en présentant quelques-unes de ses utilisations.

2.6. La spectrométrie de masse en tandem appliquée à l'immunopeptidome

2.6.1 Définition de l'immunopeptidomique

L'immunopeptidomique est une nouvelle discipline biotechnologique émergente qui utilise une approche de biologie computationnelle pour définir l'ensemble des peptides présentés à la surface des cellules par les molécules du CMH, également appelé immunopeptidome. En d'autres termes, l'immunopeptidomique est l'identification et la quantification des répertoires de peptides présentés sur les molécules du CMH d'un échantillon après purification par immunoprécipitation et traitement des épitopes élués par chromatographie liquide couplée à la spectrométrie de masse (Caron *et al.*, 2015; Purcell, Ramarathinam and Ternette, 2019). C'est une technique permettant d'interroger de manière exhaustive le répertoire des complexes pCMH présentés naturellement *in vivo* (Bassani-Sternberg and Coukos, 2016).

L'immunopeptidomique repose sur quatre grandes étapes : 1) la préparation de l'échantillon, 2) l'immunoprécipitation des molécules du CMH, 3) l'identification des épitopes par analyse des données de spectrométrie de masse, et 4) la validation du résultat par des analyses complémentaires (Figure 37).

La première étape de l'immunopeptidome repose donc sur l'isolation des épitopes liés aux molécules du CMH à partir de lignées cellulaires ou de tissus. Historiquement, plusieurs approches ont été envisagées comme l'utilisation de tampon acide pour détacher les complexes pCMH à la surface cellulaire, ou encore la génération de molécules du CMH solubles plus faciles à purifier à partir du milieu extracellulaire. Ces approches n'ont pas été concluantes notamment à cause de leur taux élevé de contamination et de leur difficulté d'applications à des échantillons clinique, respectivement (Purcell, Ramarathinam and Ternette, 2019). En utilisant un anticorps spécifique des molécules du CMH, il a été possible d'isoler directement les complexes pCMH solubilisés à partir de lysats cellulaires et tissulaires. Les équipes des D^{rs}

Engelhard et Hunt, Natheson, et Rammensee sont des pionniers dans le développement et l'utilisation de l'immunoprécipitation combinée à la spectrométrie de masse (Rötzschke *et al.*, 1990; Van Bleek and Nathenson, 1990; Falk *et al.*, 1991; Hunt *et al.*, 1992; Engelhard *et al.*, 1993). Cette technique implique l'immunoprécipitation des molécules du CMH et la dissociation ultérieure du peptide lié. Depuis, de nombreux ajustements de protocole ont été effectués, allant de l'immunoprécipitation aux choix de la chromatographie d'affinités, afin d'améliorer le rendement en peptides.

Typiquement, deux types d'anticorps monoclonaux sont utilisés pour purifiés les molécules HLA: W6.32 et TU36 qui reconnaissent les molécules HLA de classe I et classe II, respectivement. L'anticorps W6.32 reconnaît un déterminant antigénique partagés au sein des différents locus HLA-A, -B, et -C. L'anticorps TU36 lui est spécifique du locus HLA-DR. D'autres anticorps, spécifiques de certains allèles HLA peuvent être utilisés.

Une fois l'immunoprécipitation effectuée, les peptides sont dissociés des molécules HLA par une élution en tampon acide, et les peptides élués sont séparés en fonction de leur taille, de leur polarité, et de leur hydrophobicité par chromatographie liquide à haute pression avant d'être directement injectés dans un spectromètre de masse.

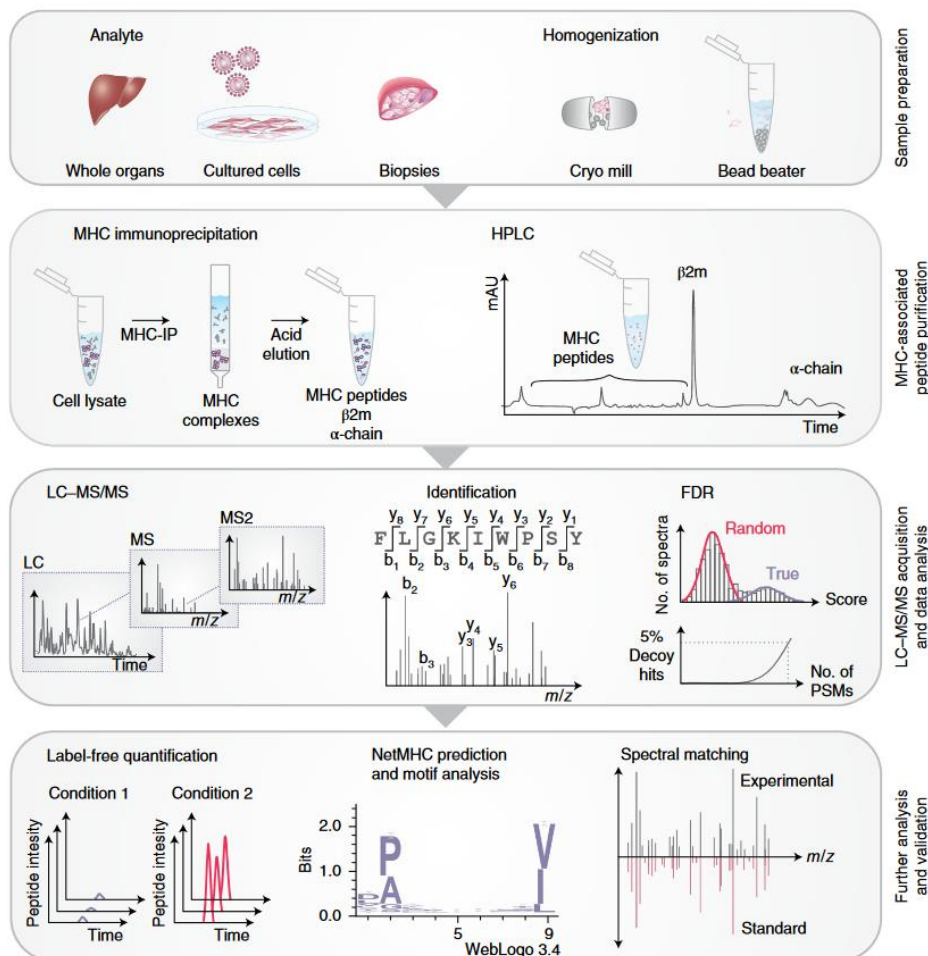


Figure 37. Etapes clés de l'immunopeptidomique (figure originale de Purcell et al 2019).

2.6.2 Principe de la spectrométrie de masse

La spectrométrie de masse est une technique physique d'analyse permettant de caractériser la structure moléculaire et chimique d'un composé. L'identification des molécules repose sur leur séparation en phase gazeuse en fonction de leur rapport masse/charge (m/z ratio). Le spectromètre de masse est souvent associé à un système de chromatographie, et à des méthodes de séparation et d'identification (Figure 38). Le principe de base de séparation de la spectrométrie de masse repose sur l'ionisation. Chaque peptide ou protéine complète ionisée se fragmente en ions chargés avec des rapports m/z différents. Les molécules ionisées sont ensuite soumises à un champ électrique dans une chambre de collision sous vide extrêmement poussé ce qui les fait « voler » jusqu'à l'analyseur de masse. Par des champs électriques et magnétiques, l'analyseur de masse va séparer les ions en fonction de leur m/z . Simultanément, la distribution et l'abondance des ions est enregistrée par le détecteur de masse qui convertit le courant ionique en courant électrique. La distribution et l'abondance des ions sont représentées sous la forme de spectres de masse. L'interprétation des spectres, et donc

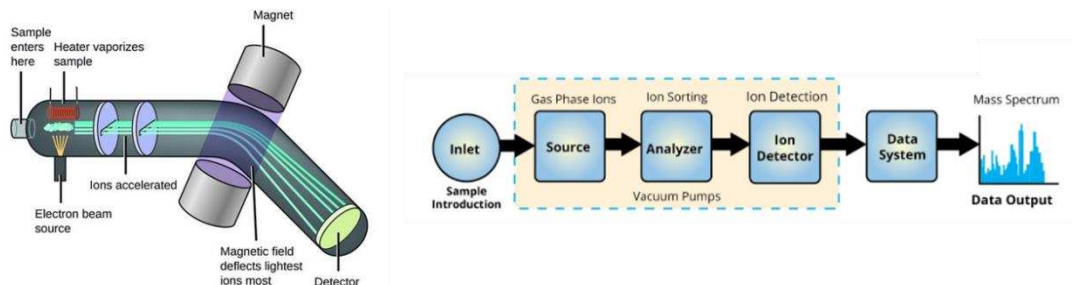


Figure 38. Représentation schématique d'un spectromètre de masse (figures adaptées de Bronkhorst et Vedantu.com).

Le spectromètre de masse se compose principalement de 4 éléments : la source d'ionisation, l'analyseur, le détecteur et la partie qui traite le signal pour générer le spectre de masse.

l'identification des séquences peptidiques peut se faire selon plusieurs approches, par exemple : 1) les données spectrales récoltées peuvent être mises en correspondance avec des spectres de références répertoriés dans la base de données, on cherche alors une similarité spectrale, 2) les données spectrales récoltées permettent l'identification de la masse du peptide qui est ensuite mise en correspondance avec des masses théoriques. Dans tous les cas, l'interprétation des spectres est faite à l'aide de logiciels tels que Mascot, MaxQuant, PEAKS studio, etc.

La spectrométrie de masse en tandem (MS/MS) est une méthode utilisée pour déterminer la séquence des peptides. Elle couple minimum deux cycles de spectrométrie de masse. Le deuxième cycle consiste à sélectionner et isoler un ion fragmenté issu d'une première spectrométrie de masse. Cet ion est appelé l'ion précurseur. Il est sélectionné par le premier analyseur noté MS1. L'ion sélectionné est ensuite fragmenté par une deuxième spectrométrie de masse et les fragments générés sont collectés par l'analyseur MS2 (Figure 39).

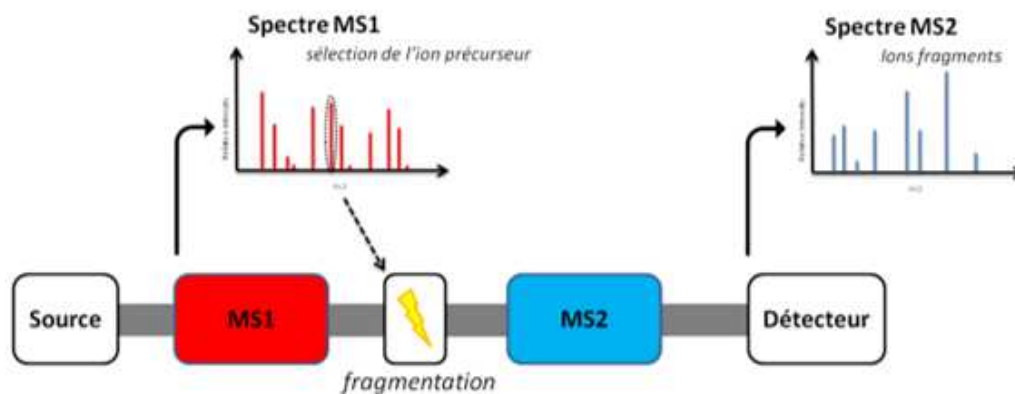


Figure 39. Représentation schématique d'un spectromètre de masse en tandem MS-MS (figure originale de Bilgraer 2015). L'ion précurseur est sélectionné par le premier analyseur MS1 puis fragmenté dans la cellule à collision. Les ions fragmentés résultants sont analysés par le second analyseur MS2.

2.6.3 Identification des peptides

Les logiciels de protéomique conçus pour l'analyse des données LC-MS/MS doivent être préalablement implémentés avec des données d'immunopeptidomique publiées ou avec un fichier FASTA contenant un génome de référence. L'avantage d'utiliser un génome de référence est que ça facilite la détection de peptides non canoniques tels que ceux porteurs de mutations, qui ne seraient pas forcément répertoriés dans une base de données publique. Cela sous-entend que la base de données doit être personnalisée, la plus riche possible, mais aussi la plus représentative de ce qui se passe *in vivo*. L'implémentation des bases de données passe par la protéogénomique, le plus souvent en déterminant l'exosome et/ou le transcriptome des tissus tumoraux et des tissus appariés afin d'identifier les variations somatiques non synonymes (Figure 40). La protéogénomique, typiquement le RNA-seq, va permettre l'identification de toutes les potentielles sources peptidiques. Cependant, cette approche est permissive et présente un taux élevé de faux positifs, ce qui signifie qu'elle peut prédire des peptides comme étant des fixateurs, alors qu'en réalité ils ne sont pas naturellement présentés. Cela gonfle la base de données protéique et entrave le score de correspondance spectre-peptide qui est alors biaisé. Pour déterminer si le spectre correspond réellement au peptide, on définit une valeur seuil de score appelée taux de fausses découvertes « FDR » (« False Discovery Rate ») qui permet de séparer les vraies des fausses identifications. Un FDR à 1% garantit une attribution précise et fiable des séquences aux spectres MS/MS (Bassani-Sternberg and Coukos, 2016). Dans l'étude de la présentation des antigènes viraux non canoniques, la base de données se limite le plus souvent au génome viral traduit dans les six phases de lecture possibles.

L'identification des peptides est souvent couplée à plusieurs étapes finales de validation (Figure 41). La première étape consiste généralement à vérifier la capacité du peptide identifié à fixer les molécules du CMH par l'utilisation des algorithmes de prédictions de fixations (SYFPEITHI, NetMHC, EpiToolKit). Le plus souvent, elle est également accompagnée de la synthèse du peptide et d'une vérification de son spectre de masse par MS. L'évaluation de son affinité est également informative. Sette et al définissent qu'un IC50 inférieur à 500nM est un seuil universel correct pour confirmer l'identification d'un épitope et son affinité (Sette *et al.*, 1994; Paul *et al.*, 2013). La validation de l'épitope identifié par LC-MS/MS peut s'accompagner de

l'évaluation de son immunogénicité par des tests de réactivités des cellules T. Cela consiste à charger des APCs autologues avec l'épitope et à les cocultiver avec des cellules mononuclées sanguines périphériques (PBMCs). L'activation des cellules T est le plus souvent mesurée par ELISPOT IFN γ ou par marquage cytokinique intracellulaire (ICS).

Cette approche apparait comme l'étape ultime de la caractérisation d'un nouvel épitope T immunogène suite à son identification par LC-MS/MS (Zhang *et al.*, 2019; Chong, Coukos and Bassani-Sternberg, 2022).

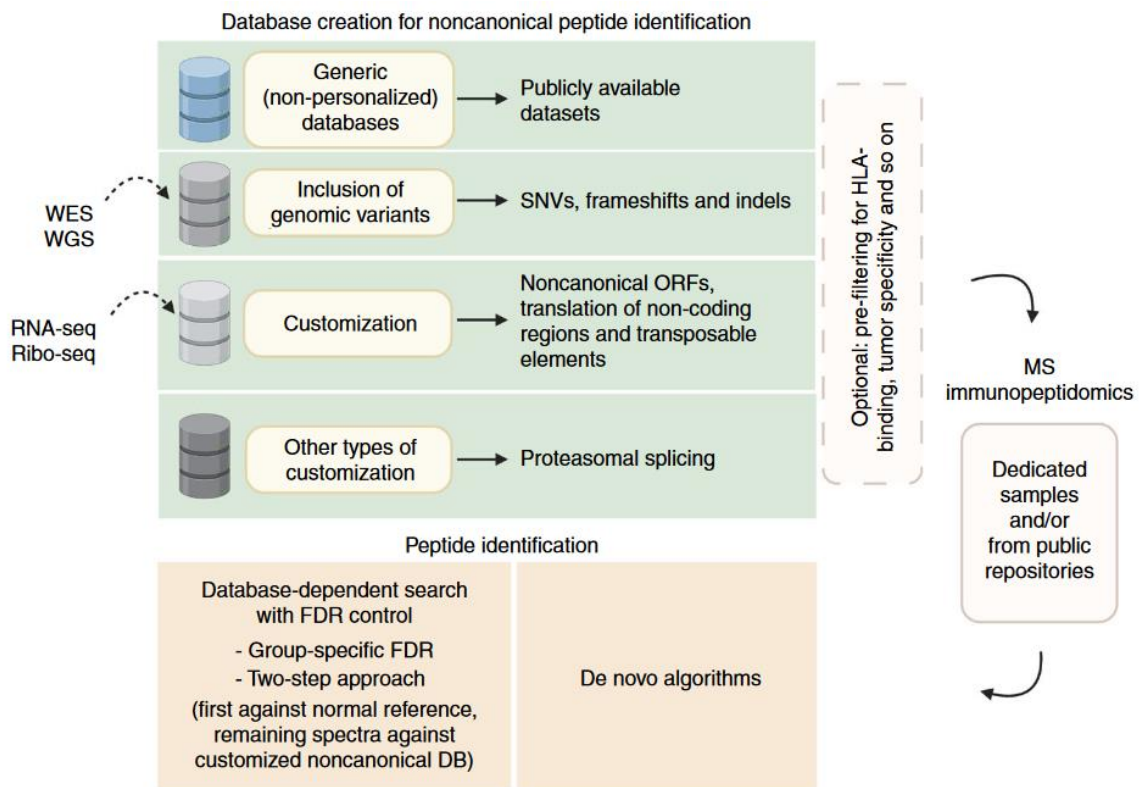


Figure 40. Les différentes bases de données utilisées pour l'identification de peptides non canoniques par LC-MS/MS (figure originale de Chong et al 2022). La création de bases de données est une étape clé de l'identification de peptides par spectrométrie de masse. Les bases de données peuvent être génériques établies à partir de données publiées ou personnalisées avec les données de séquençages disponibles. Des filtres supplémentaires peuvent être ajoutés en fonction de la question scientifique posée. L'identification des peptides est combinée à un contrôle FDR. D'autres outils de recherche peuvent être utilisés pour l'identification des peptides, tel que les algorithmes *de novo* qui n'utilisent pas de base de données préexistantes pour identifier les peptides.

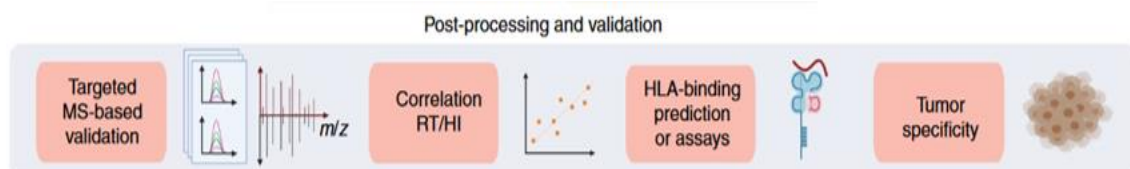


Figure 41. Les différentes procédures de validation de l'immunopeptidomique (figure originale de Chong et al 2022). La validation des peptides identifiés peut s'effectuer par le biais d'une validation ciblée par MS, d'une évaluation des caractéristiques, de la prédictions de liaison aux molécules HLA, et d'une évaluation de l'immunogénicité.

2.6.4 Applications de l'immunopeptidomique à l'identification de peptides non canoniques

Dans cette section, je vais me concentrer sur l'utilité de l'immunopeptidomique pour l'identification des peptides non canoniques, qui sont des peptides qui ne dérivent pas de séquences codantes annotées (CDS ou ORFs) mais de sORFs, de lncRNA ou de pseudogènes.

2.6.4.1 Application dans le cancer

Depuis l'avènement de l'immunopeptidomique, une dizaine d'études remarquables ont réussi à identifier de nombreux peptides non canoniques chez l'homme dans le contexte du cancer. Ce qui ressort de ces études est que les sources alternatives de peptides sont multiples. On retrouve : les pseudogènes, les lncRNA, les régions introniques et intergéniques non codantes, les UTR, l'épissage du protéasome, et des nouveaux ORFs non annotés. En fonction du modèle cellulaire étudié et des allèles HLA, le nombre de peptides non canoniques identifiés par LC-MS/MS représente entre 1,5 et 15% du total des peptides identifiés. De façon surprenante, deux études rapportent une fréquence très élevée (30% à 44%, respectivement) de peptides issus de l'épissage du protéasome présentés sur les molécules HLA de classe I (Liepe *et al.*, 2016; Faridi *et al.*, 2018).

Jusqu'à présent seules quelques études ont rapporté l'existence de peptides non canoniques pour lesquels des cellules T spécifiques ont été trouvées (voir Annexe n°1). Dans un modèle de cellules B transformées par l'EBV, Laumont *et al.* ont identifié par LC-MS/MS 168 peptides non canoniques. Les spectres MS/MS sont comparés à deux bases de données construites à partir des données de RNA-seq obtenues à partir d'un des échantillons analysés. La première base de données aligne les reads sur le génome de référence alors que la deuxième aligne les reads sur l'ensemble des six phases de lectures. La fréquence des peptides non canoniques chargés sur les molécules du CMH de classe I est de 6,5% à 13% selon les critères de sélection appliqués. Ils ont mis en évidence que les peptides issus de régions non canoniques sont plus courts, avec des préférences de liaison à l'allotype différentes, et présentent plus de polymorphisme génétique. Les auteurs ont terminé leur étude avec l'évaluation de l'immunogénicité de quatre peptides non canoniques choisis au hasard, dont la séquence a été validée par LC-MS/MS. Les PBMCs sont cocultivés avec des DCs dérivés de monocytes autologues chargés avec un des quatre peptides. Après douze jours de culture avec différentes cytokines, les LTs CD8⁺ sont triés et leur réponse contre le peptide est testée par ELISPOT IFN γ . Seuls les peptides non canoniques les plus polymorphes sont reconnus par les LTs CD8⁺ (Laumont *et al.*, 2016). Dans une seconde étude et de manière rigoureuse, ils démontrent par LC-MS/MS que la majorité des néoantigènes présentés à la surface des molécules du CMH-I dérivent de régions non codantes (Laumont *et al.*, 2018). Ils font la démonstration *ex vivo* et *in vivo* que certains néoantigènes stimulent une réponse T spécifique.

2.6.4.2 Application dans les maladies infectieuses

Malgré l'apport de connaissances qu'offre la MS sur la présentation des antigènes et leurs sources, peu d'études se sont penchées sur la caractérisation de l'immunopeptidome viral. Les premières caractérisations de peptides viraux par LC-MS/MS ont été faites sur les virus EBV et BVDV (« bovine viral diarrhea virus ») où les auteurs ont identifié des peptides canoniques

présentés sur les molécules du CMH et reconnus par des LTs CD8⁺ (Herr *et al.*, 1999; Planz *et al.*, 2001). Van Els et al ont réalisé le premier immunopeptidome du virus de la rougeole. Ils identifient trois peptides viraux présentés sur les molécules HLA-A*0201 d'une lignée lymphoblastoïde B. De façon remarquable, un peptide dérivé de la protéine canonique C est exprimé à une très forte densité (>10⁵ copies par cellule) et est immunogène chez les souris transgéniques HLA-A2/Kb. Les auteurs ont également démontré l'existence d'une pression de sélection immunitaire sur cet épitope puisque les souches virales persistantes montrent des mutations au niveau des sites d'ancrages à HLA-A*0201 (van Els *et al.*, 2000).

Plus récemment, une étude révèle des informations sur la présentation de peptides issus de la présentation croisée par les DCs (Wu *et al.*, 2019). Les auteurs montrent que le peptide « PA224-232 » du virus la grippe A, malgré qu'il soit faiblement présenté à la surface des cellules infectées, figure parmi les peptides issus de la présentation croisée les plus abondants. Ils démontrent également que ce peptide initie une réponse immunodominante des LTs CD8⁺.

L'immunopeptidome du VIH-1

En 2003 Hickman et al caractérisent l'impact de l'infection du VIH-1 sur la présentation des peptides de l'hôte (Hickman *et al.*, 2003). En infectant la lignée cellulaire CD4⁺ SupT1 préalablement transfectée pour exprimer des molécules HLA-B*0702 solubles, ils identifient 15 peptides du soi qui sont présentés de manière différentielle à la surface des cellules infectées. Les peptides surexprimés au cours de l'infection par le VIH-1 dérivent de protéines de l'hôte impliquées dans de multiples voies cellulaires telles que la régulation du cycle cellulaire, la transcription et la traduction. Ils établissent un lien entre la dégradation des protéines sources de l'hôte suite à l'infection et la présentation sur les molécules HLA. Pour finir, les auteurs émettent l'hypothèse que l'augmentation de la présentation des peptides du soi à la surface des cellules infectées est peut-être la cause de l'apparition de réactions auto-immunes fréquemment observées chez les personnes vivant avec le VIH (PVVIH).

A ma connaissance Nicola Ternette et al sont les premiers à caractériser l'immunopeptidome des molécules classe I du VIH-1 (Ternette *et al.*, 2016). L'immunoprécipitation est réalisée à partir de cellules T CD4⁺ primaires et d'une lignée cellulaire T humaine C8166 préalablement infectées. L'interprétation des spectres de MS/MS est effectuée à l'aide d'une base de données contenant toutes les séquences humaines annotées, y compris la traduction des six phases de lectures du génome de référence HIV-III_B séquencé, ainsi que toutes les séquences canoniques connues du VIH-1. Les auteurs identifient entre 2416 et 6795 peptides présentés par les molécules du CMH à la surface des cellules primaires et des cellules C8166. Seulement 75 peptides dérivent du VIH-1 ce qui représente 1,1% du total des peptides identifiés. 9 peptides présentés par les molécules du CMH des cellules primaires sont identifiés. L'analyse de prédiction de fixation confirme la bonne position des résidus d'ancrage des peptides identifiés par LC-MS/MS. Les peptides viraux élus dérivent de l'ensemble des régions du génome viral à l'exception des gènes *vpr*, *vif*, et *tat*. Dans l'ensemble, un tiers des peptides n'avaient jamais été décrits comme des fixateurs des molécules du CMH-I. Plus de 82% des séquences identifiées proviennent de régions protéiques pour lesquelles des réponses T avaient déjà été signalées sans connaître précisément l'épitope reconnu. De façon intéressante, ils révèlent la

présentation d'un peptide dérivé d'ARF à la surface des cellules C8166. SR11 code pour la séquence suivante « SYALASDAQNR » et chevauche le gène *env* en antisens.

Dans une autre étude, le même groupe de recherche isole 105 peptides à la surface de cellules infectées par le candidat vaccin « MVA.HIVcons » (Ternette *et al.*, 2015). La majorité des peptides proviennent du vecteur mais 7 peptides sont dérivés du VIH. Les PBMCs des personnes vaccinées réagissent aux séquences peptidiques identifiées par LC-MS/MS. Les réponses contre les peptides dérivés des régions conservées du VIH sont détectables rapidement après l'infection et leur abondance corrèle avec les niveaux d'expression protéique. Dans une troisième étude, ils développent une approche computationnelle basée sur le séquençage *de novo* afin d'identifier des peptides de classe I épissés par le protéasome (Paes *et al.*, 2019). Sur un total de 198 peptides viraux identifiés seulement 5 sont définis comme des épitopes épissés. Ils concluent que les peptides épissés par le protéasome constituent une composante mineure de l'immunopeptidome de classe I des cellules infectées par le VIH-1.

Marcilla *et al.* ont défini l'immunopeptidome de l'hôte présenté par les molécules HLA associées au contrôle de l'infection à VIH (Marcilla *et al.*, 2016). Les allèles Mamu-B*08 et HLA-B*27 ou – B*57 sont respectivement surreprésentés chez les macaques et individus qui contrôlent la réplication virale. Les auteurs ont donc cherché à caractériser et comparer les motifs d'ancrage des peptides fixés sur les molécules HLA protectrices simiennes et humaines. Ils démontrent que les peptides ont des similarités dans leur motif de liaison notamment à la position 2 où une arginine est présente dans la majorité des ligands. La position 9 d'ancrage se montre moins restrictive pour les ligands HLA-B*27. Il est étonnant de constater qu'ils n'ont pas étendu cette approche aux peptides dérivés du VIH-1.

D'autres études d'immunopeptidomiques sur le VIH-1 ont permis de révéler 1) l'impact de l'infection sur le répertoire de peptides présentés par les molécules HLA-C*03 :04 et les conséquences sur l'engagement des récepteurs inhibiteurs KIRs des cellules NK, 2) un nouveau mécanisme d'échappement dépendant de l'extension en N-ter des ligands d'HLA-B*57, 3) de nouveaux épitopes immunodominants du VIH-1 présentés par les molécules HLA-C, et 4) de nouveaux épitopes d'HLA-B*57 (Pymm *et al.*, 2017; Ramarathinam *et al.*, 2018; Chikata *et al.*, 2019; Ziegler *et al.*, 2020). A ma connaissance, aucune de ces études ne démontrent la présentation par les molécules du CMH de peptides dérivés de sORFs (ARFs, uORFs, dORF etc.).

2.6.7 Du ribosome profiling à l'immunopeptidomique : la combinaison d'outils puissants

Le translatome, qui constitue l'ensemble des ARNs activement traduits, est la base de données idéale pour l'interprétation des spectres obtenus par MS puisqu'il réduit l'espace de recherche à des événements se produisant *in vivo* et de ce fait optimise le taux de FDR. De plus, en détectant une signature traductionnelle au niveau d'une séquence non annotée et en identifiant un peptide dérivé de cette séquence présentés par les molécules du CMH, nous avons une preuve incontestable de l'existence des sORFs en tant que réservoir immunologique. Cependant à ce jour, seules quelques études ont établi un lien entre le translatome et l'immunopeptidome.

Dans le contexte d'un carcinome rénal à cellules claires, Smith et al établissent un lien entre la traduction des rétrovirus endogènes humains (hERV) et la génération d'épitopes immunogènes (Smith *et al.*, 2018). En réanalysant des données publiées de RNA-seq et ribosome profiling, les auteurs mettent en évidence une signature traductionnelle de certaines séquences d'hERV qui codent pour de potentiels épitopes spécifiques à la tumeur. Ils évaluent ensuite *in vitro* la capacité des épitopes à se fixer sur les molécules HLA par des expériences d'échange de peptides (« UV exchange assay »). Une réponse T spécifique de *gag* et *pol* est détectée par marquage tétramère dans les PBCMs infiltrant la tumeur. De plus, les auteurs montrent que la présence de ces épitopes dérivés de hERV au sein des tumeurs corrèle avec une meilleure évolution de la maladie, ce qui en fait de nouvelles cibles d'immunothérapie dans le carcinome rénal.

La même année, Erhard et al développent un nouvel outil d'analyse des données de ribosome profiling : PRICE qu'ils utilisent pour définir de nouveaux sORFs à partir de bases de données publiées sur les génomes du HSV et du HCMV (Erhard *et al.*, 2018). Parallèlement, ils définissent l'immunopeptidome des molécules du CMH-I infectées par le HCMV. Ils mettent alors en évidence que certains produits de traduction non canoniques sont la source de 112 peptides retrouvés dans leur immunopeptidome. Leurs travaux révèlent également 500 nouvelles protéines non canoniques du HCMV qui n'étaient pas référencées.

L'équipe du Dr Bassani-Sternberg a réalisé une étude unique en son genre en caractérisant pour la première fois simultanément le translatome et l'immunopeptidome à partir d'un même échantillon (Chong *et al.*, 2020). L'immunopeptidome des molécules de classe I et classe II a été réalisé sur sept lignées cellulaires de mélanome dérivées de patients et deux échantillons tissulaires de cancer du poumon. Une centaine de peptides dérivés de séquences présumées non codantes, d'éléments transposables, et d'ARFs est identifiée sur les molécules HLA-I par patient. Au total, cela correspond à l'identification de 452 peptides non canoniques uniques. L'interprétation des spectres de MS obtenus à partir de l'immunoprécipitation des molécules du CMH-II n'est réalisée que sur 4 des échantillons traités. Seulement quatre peptides non canoniques sont retrouvés à la surface des molécules HLA-II sur les quatre échantillons analysés. Les auteurs évaluent ensuite l'implication de ces peptides dans la reconnaissance immunitaire de la tumeur en mesurant la libération d'IFN γ lors de la stimulation par le peptide des lymphocytes T infiltrant la tumeur ou des PBMCs autologues. Seul le peptide non canonique « KYKDRTNILF » issu d'un sORF localisé en 3'UTR (dORF) active une réponse T CD8+. Les auteurs finissent leur étude en démontrant que certains peptides non canoniques spécifiques à la tumeur sont partagés entre les patients ce qui confère un avantage thérapeutique à ces peptides non canoniques.

Dans la même dynamique, l'équipe du Dr Yewdell établissent une corrélation entre le translatome, l'immunopeptidome, et le protéome (Ruiz Cuevas *et al.*, 2021). L'analyse de trois lignées cellulaires lymphoblastiques B montrent que 85% des produits de traduction ne sont pas annotés dans la base de données du génome humain et que 50% d'entre eux correspondraient à des produits alternatifs ne dérivant pas de régions génomiques codant les protéines. Par une approche computationnelle, ils définissent qu'environ trois quarts des régions non canoniques sont initiées avec un codon non-AUG, le plus souvent CUG, bien que

tous les codons near-cognate soient bien représentés. L'immunopeptidome révèle que 7,5% des peptides associés aux molécules du CMH-I dérivent de protéines non canoniques avec une répartition égalitaire entre celles provenant de nouvelles isoformes ou celles provenant d'ARFs, uORFs etc. L'analyse du protéome montre que sur 14 498 protéines identifiées 17% sont non canoniques (Ruiz Cuevas *et al.*, 2021; Yewdell, 2022).

En combinant des données de transcriptomique et de Riboseq dans un flux de travail innovant, Martinez *et al.* identifient dans le génome humaine un millier de sORFs (Martinez *et al.*, 2020). Les séquences sont conservées chez au moins une autre espèce de mammifère, ce qui suggère que les protéines codées par ces sORFs pourraient être fonctionnelles. Ils étudient ensuite l'impact d'un stress cellulaire sur l'efficacité de traduction (TE) des sORFs. L'induction du stress, provoque une augmentation du TE pour un certain nombre de sORFs. En utilisant l'harringtonine ils montrent que 40% des sORFs n'ont pas de codons AUG dans leur cadre de lecture, et de ce fait, sont initiés par d'autres mécanismes non canoniques. La ré analyse de données de MS/MS publiées montre que des peptides dérivés des sORFs sont retrouvés à la surface des molécules du CMH, validant le fait que ces régions non annotées du génome sont une source antigénique.

Ouspenskaia *et al.* définissent le translatome pour un nombre remarquable de cellules : 29 cellules humaines primaires et cancéreuses. Cette étude identifie plus de 200 000 nouveaux sORFs non annotés. Comme leurs confrères, ils démontrent que les protéines non annotées ne représentent que 3,3% de l'immunopeptidome et 0.1% du protéome. (Ouspenskaia *et al.*, 2022; Yewdell, 2022).

Récemment, deux papiers publiés par la même équipe mettent en lumière le fait que des sORFs du SARS-CoV-2 sont traduits et que certains polypeptides générés sont présentés à la surface des cellules et capables d'induire une réponse T spécifique (Finkel *et al.*, 2021; Weingarten-Gabbay *et al.*, 2021). L'immunopeptidome de cellules infectées par le SARS-CoV-2 a permis de mettre en évidence que 9 peptides présentés à la surface des molécules du CMH-I sont issus d'évènements non canonique de traduction. Dans leur étude de 2021, les auteurs avaient montré par ribosome profiling la traduction de 6 uORFs, 11 ARFs, et 6 sORFs codant des isoformes. Au total, l'immunoprécipitation des molécules du CMH-I suivie de l'analyse par spectrométrie de masse détecte uniquement 28 peptides qui dérivent d'ORF canoniques et 9 qui dérivent de trois ARFs : deux chevauchant la protéine spike (ORFs.iORF1 et -iORF2) et l'autre chevauchant la nucléoprotéine (ORF9b). La majorité des peptides identifiés sont des fixateurs d'HLA-A02:01 qui est l'allèle le plus représenté dans la population. Ils étudient également la capacité des peptides non canoniques à initier une réponse T chez des souris HLA-A02 immunisée et chez des individus HLA-A02 convalescents. Sur six individus testés, deux montrent une sécrétion significative d'IFN γ par les lymphocytes T contre le pool de peptides non canoniques identifiés par LC-MS/MS. Le peptide « ELPDEFVVTV » dérivé de ORF9b stimule une réponse CD8+ avec une magnitude plus forte que celles retrouvées contre tous les peptides SARS-Cov-2 de classe I testés canoniques et non canoniques (Weingarten-Gabbay *et al.*, 2021).

En associant l'immunopeptidomique à la protéogénomique tels que le RNA-Seq et le RiboSeq, il est possible d'évaluer de manière exhaustive la relation entre la transcription, la traduction et la présentation antigénique. Cette approche novatrice et intégrative peut être appliquée à de nombreux modèles permettant la compréhension des mécanismes moléculaires impliqués dans la pathogenèse de nombreuses maladies. Les études discutées dans ce chapitre confirment le rôle des événements non conventionnels de traduction dans la présentation antigénique et leur potentiel thérapeutique.

Globalement, les réponses immunes restreintes aux peptides de classe II, qu'ils soient canoniques ou non canoniques, n'ont été que très peu étudiées dans le domaine du cancer ou des maladies infectieuses en comparaison avec les réponses immunes de classe I. A ce jour, aucune étude spécifique n'est publiée sur l'immunopeptidome viral de classe II. Aux vues de l'implication des voies non canoniques d'apprêtement des antigènes de classe II, il reste encore beaucoup à découvrir sur le répertoire des antigènes viraux et cellulaires de classe II.

CHAPITRE 3 : Le VIH-1

3.1 Contexte épidémiologique de l'infection à VIH

Tout a commencé en 1981 quand les premiers signes de l'épidémie apparaissent en Occident. À Los Angeles, cinq individus présentant une pneumonie soudaine à *Pneumocystis Pneumonia* sont admis à l'hôpital (*Pneumocystis Pneumonia --- Los Angeles*, 1981). Il s'agira en réalité des premiers cas de VIH-1 aux Etats-Unis. Les individus infectés présentent des symptômes de type lymphadénopathie qui précèdent souvent le syndrome de l'immunodéficience acquise, aujourd'hui connu sous le nom de SIDA. De manière exceptionnelle, les cliniciens, les épidémiologistes, et les chercheurs, dans un premier temps Américain et Français, unissent rapidement leurs forces pour identifier l'agent pathogène et secourir un grand nombre de malades, puisque le virus se propage de l'autre côté de l'Atlantique, marquant alors le début d'une épidémie émergente dramatique. Les cliniciens considèrent initialement la possibilité d'une infection par le virus T-lymphotrope humain (HTLV). Cette hypothèse est rapidement écartée par les chercheurs qui savent que le HTLV immortalise les lymphocytes T, tandis que les cliniciens observent plutôt une diminution de ces derniers. Dans les laboratoires de l'Institut Pasteur, les D^s Françoise Barré-Sinoussi et Luc Montagnier mettent en culture les cellules ganglionnaires du patient BRU, et observent la production de la transcriptase inverse par les lymphocytes T, ce qui est caractéristique de la réplication d'un rétrovirus. Ils observent également pour la première fois le virus par microscopie électronique. Ce travail est publié le 20 mai 1983 dans la prestigieuse revue Science et sera récompensé par l'obtention du prix Nobel de Médecine et de Physiologie en 2008 (Barré-Sinoussi *et al.*, 1983). Les équipes de Pasteur s'intéressent ensuite au tropisme du VIH. Le chercheur David Klatzmann identifie le récepteur CD4 des lymphocytes T comme le récepteur majeur du virus (Klatzmann, Barré-Sinoussi, *et al.*, 1984; Klatzmann, Champagne, *et al.*, 1984). En 1985, le génome du VIH est séquencé révélant un génome complexe de 9749 nucléotides et distinct du HTLV-1 même si l'organisation génomique est similaire (Alizon *et al.*, 1984; Hahn *et al.*, 1984; Luciw *et al.*, 1984; Rabson and Martin, 1985; Sanchez-Pescador *et al.*, 1985; Wain-Hobson *et al.*, 1985). L'ensemble de ces travaux signent la découverte d'un nouveau rétrovirus humain appelé le virus de l'immunodéficience humaine (VIH).

Depuis le début de l'épidémie, 84,2 millions de personnes ont été infectées par le VIH coutant la vie à plus de 40 millions d'entre elles (*World Health Organization (WHO)*, no date). Malgré la mobilisation internationale des différents corps de la santé communautaire depuis plus de 40 ans, la pandémie du VIH perdure. En 2021, le nombre de personnes vivant avec le VIH (PVVIH) était de 38,4 millions, et 650 000 décès liés au VIH ont été enregistrés à travers le monde (*World Health Organization (WHO)*, no date). L'Organisation Mondiale de la Santé (OMS) tire la sonnette d'alarme. En effet, les infections augmentent là où elles diminuaient (Asie et Pacifique) et le taux de nouvelles infections n'a diminué que de 3,6% entre 2020 et 2021, ce qui constitue la plus faible baisse annuelle depuis 2016. De plus, seulement 73% des personnes qui connaissent leur statut sérologique reçoivent un traitement d'après le rapport du programme commun des Nations unies sur le VIH/SIDA (UNAIDS) ('IN DANGER: UNAIDS Global AIDS Update 2022', 2022). L'objectif « 90-90-90 : un traitement pour tous », fixé par

l'OMS pour l'année 2020, n'a pas été atteint. Cet objectif visait à ce que 90% des personnes infectées connaissent leur statut, que 90% d'entre elles reçoivent un traitement antirétroviral, et que 90% des individus infectés aient une charge virale indétectable sous traitement. Les deux derniers objectifs avaient pour but de lutter contre les inégalités d'accès aux soins et les problèmes d'observance principalement rencontrés dans les pays du Sud. En 2020, les taux de réalisation étaient de 84%, 73%, et 66% respectivement ('IN DANGER: UNAIDS Global AIDS Update 2022', 2022). Les Nations unies ont adopté une nouvelle cible « 95-95-95 » qui doit être atteinte d'ici 2030 pour mettre fin aux inégalités et vaincre le VIH.

3.2 Origine et diversité génétique du VIH

Historiquement, le foyer épidémique ne naît pas aux Etats-Unis mais en Afrique centrale, plus précisément à Kinshasa en 1900 (Peeters, D'Arc and Delaporte, 2014). Il faut attendre environ 80 ans pour que l'épidémie se dissémine aux Etats-Unis suite à l'importation de litres de sérum contaminés en provenance des Caraïbes (Products *et al.*, 1995; Vermund and Leigh-Brown, 2012). En Afrique, l'épidémie n'est connue qu'en 1984 lorsqu'elle est révélée dans la population hétérosexuelle. Les chercheurs identifient en 1986, un deuxième virus responsable du SIDA (Clavel *et al.*, 1986; Rey *et al.*, 1986). Ce virus est nommé VIH-2 puisqu'il s'agit du deuxième agent pathogène responsable du SIDA, le premier agent pathogène identifié en 1983 étant appelé VIH-1. Le VIH-2 reste à ce jour une épidémie localisée à l'Afrique de l'Ouest. Bien que morphologiquement similaire au VIH-1, le VIH-2 est caractérisé par une charge virale plus faible, une progression plus lente vers le stade SIDA et un profil antigénique distinct (Clavel *et al.*, 1986).

La diversité du VIH est représentée par l'existence des deux types du VIH (VIH-1 et -2), ainsi que de quatre groupes distincts du VIH-1, et de multiples sous-types et recombinants. Au cours des années 1990, il est devenu évident que le VIH est une zoonose et que le primate africain est son hôte naturel (Sharp and Hahn, 2011; Peeters, D'Arc and Delaporte, 2014). Beaucoup d'espèces différentes de singes sont infectées par le virus de l'immunodéficience simienne (SIV) mais certaines espèces ne développent pas le stade SIDA. Ces espèces sont alors définies d'hôtes naturels du virus. La diversité du SIV est complexe et implique une transmission inter-espèces ainsi que des recombinaisons entre lignées éloignées. En 1989, les chercheurs identifient les premières souches de SIV chez le chimpanzé au Gabon. Le chimpanzé, contrairement au macaque rhésus, développe le stade SIDA suite à l'infection par le SIV. De nombreuses études épidémiologiques à large échelle révéleront par la suite que les SIV les plus proches du VIH-1 sont le SIV des chimpanzés (SIVcpz) et le SIV des gorilles (SIVgor). Au total, il y a eu quatre transmissions inter-espèces du singe à l'homme donnant naissance aux groupes M, N, O et P du VIH-1 (Peeters, D'Arc and Delaporte, 2014). Les virus du groupe M sont responsables de la pandémie mondiale du VIH-1. Le groupe M est lui-même divisé en neuf clades (A, B, C, D, F, G, H, J, et K) en fonction de leur polymorphisme dans les gènes *gag* et *env*. Le sous-groupe B est le plus répandu en Europe de l'Ouest, en Amérique du Nord et du Sud (Figure 42). Le sous-groupe C est très fortement représenté en Afrique du sud et Asie centrale. Le sous-groupe A est majoritaire en Afrique de l'Est, en Europe de l'Est et également en Asie centrale (*Los Alamos HIV Databases*, no date). La diversité du VIH-1 est augmentée par

la présence de formes recombinantes (CRF), plus d'une centaine de CRF ont été identifiées jusqu'à présent.

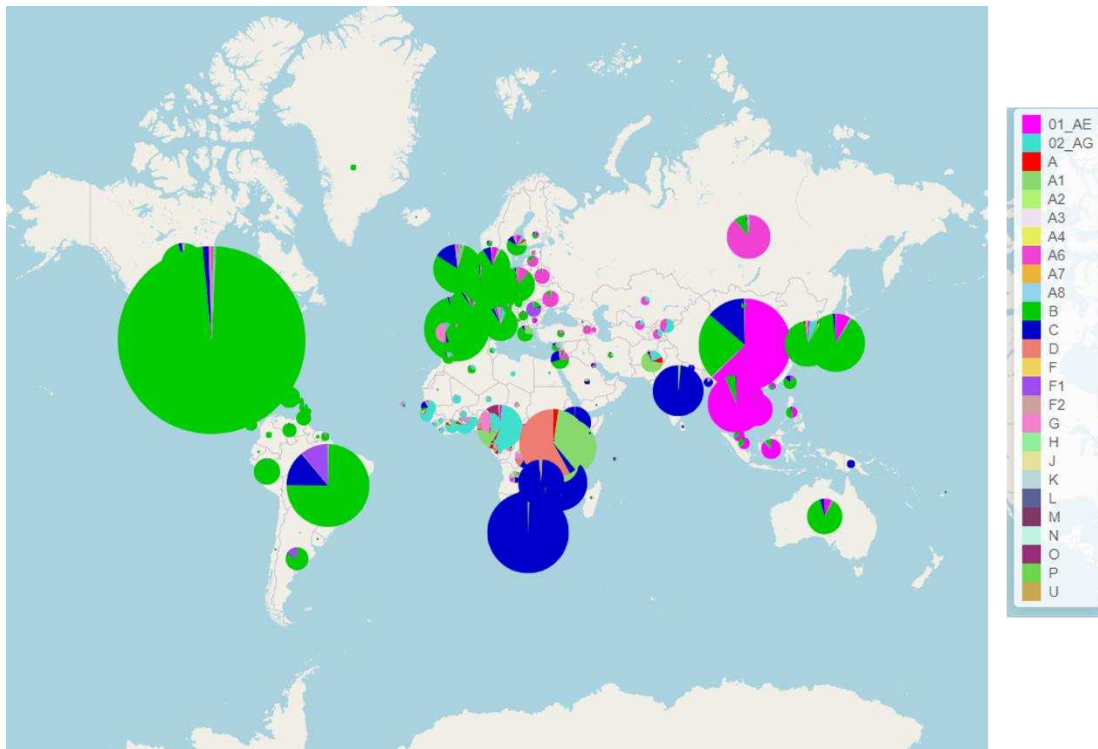


Figure 42. Distribution géographique des sous-types du VIH-1 (figure adaptée de Fischer et al. 2021)
Les diagrammes circulaires représentent la distribution des sous-types et recombinants (CFR01_AE, CFR02_AG) du VIH-1 dans chaque région du monde entre 1959 et 2022 selon la base de données du VIH (LANL database). La surface du diagramme est proportionnelle au nombre de séquences référencées pour cette région du monde.

Les SIV présents chez le singe vert (Sooty Mangabey) sont les plus proches parents du VIH-2. Les neuf sous-groupes de VIH-2 sont le résultat de neuf transmissions inter-espèces (Peeters, D'Arc and Delaporte, 2014).

Pour plus de simplicité, sauf indication contraire, chaque fois que le VIH est mentionné dans la suite de ce manuscrit, cela fera référence au sous-type M du VIH-1.

3.3 Structure du virion et du génome viral

Le VIH est un *Lentivirus* appartenant à la famille des *Retroviridae*. C'est un virus sphérique ayant un diamètre d'environ 145 nanomètres (Briggs *et al.*, 2003). Le VIH est un virus enveloppé à ARN monocaténaire de polarité positive (Figure 43). L'enveloppe du VIH est constituée d'une bicouche lipidique provenant de l'hôte et d'environ 14 exemplaires de la glycoprotéine d'enveloppe Env (Env) (Zhu *et al.*, 2006). Dans la membrane, les glycoprotéines de l'enveloppe sont assemblées en trimère composé chacun de deux sous-unités gp120 et d'une sous-unité gp41. La glycoprotéine gp120 comporte 5 domaines variables annotés V1 à V5, les boucles V1 et V2 étant les plus diversifiées en termes de séquence et de taille (Modrow *et al.*, 1987; Zolla-Pazner and Cardozo, 2010; Pan *et al.*, 2015). La matrice protéique (MA) composée de la protéine p17 s'associe à la face interne de l'enveloppe. A l'intérieur de la matrice sont présentes la protéine accessoire Vpr, l'enzyme protéase (PR) et la capsid (CA) (van Heuvel *et al.*, 2022). La capsid du VIH est formée de la protéine p24 assemblée de façon icosaédrique qui protège le

génomique viral. La capside contient le génome viral, constitué de deux copies d'ARN simple brin, complexées aux protéines de la nucléocapside (NC), aux intégrases (IN), et aux transcriptases inverses (RT).

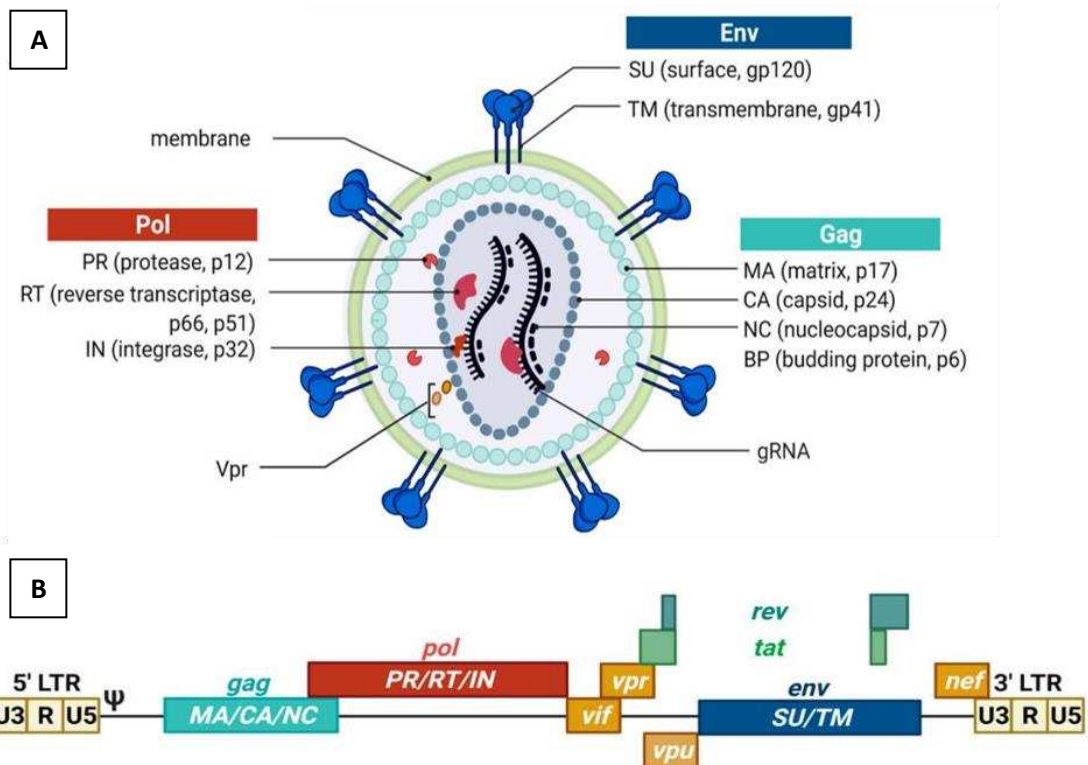


Figure 43. Structure et organisation génomique du VIH-1 (figure adaptée de Heuvel et al. 2022). (A) Représentation schématique d'un virion mature. Le VIH est une particule sphérique enveloppée d'une bicouche lipidique qui contient les protéines Env. Les protéines p17 (MA) forment la matrice située à la face interne de l'enveloppe. Les protéines p24 (CA) forment la capside qui contient l'ARN génomique complexé aux protéines p32 (IN), p7 (NC) et p66 et p51 (RT). La protéine accessoire Vpr ainsi que la protéine p21 (PR) sont présentes à l'intérieur de la matrice. (B) Représentation schématique d'un génome du VIH. L'ARN génomique monocaténaire d'environ 9 kb est flanqué aux extrémités 5' et 3' par des régions LTRs, elles-mêmes composées de trois régions distinctes : U3-R-U5. Les gènes du VIH sont répartis et exprimés sur 3 phases de lectures distinctes. La phase de lecture dans laquelle sont exprimés les gènes varie en fonction de la souche virale.

Le génome du VIH, d'une longueur d'environ 9kb, est composé de 9 ORFs codant 15 protéines (Frankel and Young, 1998; Jäger *et al.*, 2011; Sertznig *et al.*, 2018). L'ARN génomique monocaténaire est flanqué aux extrémités 5' et 3' de séquences nucléotidiques répétées appelées « Long Terminal Repeat sequence » (LTR). Les LTRs sont subdivisés en trois régions : U3, R, et U5. Les régions R sont identiques aux chaque extrémités de l'ARN et jouent un rôle important dans l'étape de transcription inverse du cycle réplcatif. Les gènes *gag*, *pol* et *env* sont communs aux rétrovirus. L'ORF *gag* code pour les protéines MA, CA, NC et p6. L'ORF *pol* permet la synthèse des protéines IN, PR, et RT. Son expression est régulée par un signal alternatif de traduction situé au sein de l'ORF *gag*, et résulte en un précurseur protéique Gag-pol appelé Pr55gag. L'ORF *env* code également pour un précurseur protéique : la glycoprotéine gp160. Dans le cytoplasme, le précurseur Gag-pol est clivé par la PR tandis que gp160 est clivé par une furine (Hallenberger *et al.*, 1992). En plus des gènes communs aux rétrovirus, le VIH code pour des protéines dites accessoires à savoir les protéines Vif, Vpr, Vpu, Tat, Rev, et Nef. Les protéines accessoires interviennent dans différentes étapes du cycle viral. Les protéines Tat

et Rev sont impliquées dans la phase précoce du cycle viral en favorisant la transcription et la traduction des gènes. Vif et Vpr inhibent les facteurs de restrictions cellulaires qui empêchent la réplication du virus tandis que Vpu et Nef interagissent avec les facteurs de restrictions qui inhibent entre autres l'étape de bourgeonnement (Roy *et al.*, 2014).

3.4 Pathogénèse de l'infection à VIH

Le VIH se transmet par passage à travers les muqueuses lors de rapport sexuel non protégés, l'utilisation de matériel injectable contaminé, mais aussi de la mère à l'enfant au cours de la grossesse, de l'accouchement et de l'allaitement.

Suite à la transmission, la pathogénèse du VIH se décompose en trois étapes : la primo-infection, la phase chronique, et la phase SIDA (Figure 44).

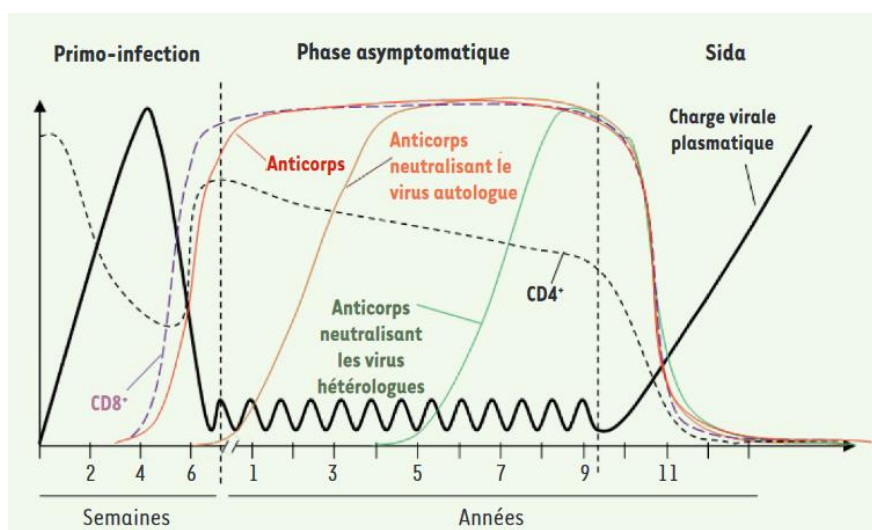


Figure 44. Evolution naturelle de l'infection à VIH (figure originale de Peressin et al 2014). Représentation schématique des trois différentes phases de l'infection à VIH. Les différents acteurs de l'immunité ainsi que la charge virale sont représentés par des lignes distinctives.

3.4.1 La primo-infection

La primo-infection débute par une phase d'éclipse d'une durée d'une à deux semaines. Durant la phase d'éclipse la virémie est indétectable et la réponse immunitaire inexistante. C'est une phase asymptomatique qui permet la réplication du virus et sa propagation du site initial de l'infection aux nombreux autres sites de réplication (Coffin and Swanstrom, 2013). Selon le mode de transmission les premiers foyers de réplication peuvent se situer au niveau des muqueuses oropharyngées, gastro-intestinales ou génitales (Belyakov and Berzofsky, 2004).

3.4.1.1 La phase d'éclipse

Dans les muqueuses résident les différentes cibles du VIH qui sont : les LTs CD4⁺, les macrophages, les DCs, et les cellules de Langerhans (LCs). Toutes expriment le récepteur CD4 et le corécepteur du virus CCR5 (Hladik and McElrath, 2008). En fonction de leur sous-type cellulaire, les DCs et les LTs CD4⁺ expriment de manière différentielle le deuxième corécepteur du virus CXCR4. Les souches virales qui utilisent CCR5 comme corécepteur ont un tropisme appelé X5, ou X4 si elles utilisent le corécepteur CXCR4. La majorité des cellules TCD4⁺ résidentes de la lamina propria sont des cellules mémoires qui expriment le récepteur CCR5 à

un niveau plus élevé que les TCD4⁺ circulant dans le sang (Belyakov and Berzofsky, 2004; Hladik and McElrath, 2008). A l'inverse, les LTs CD4⁺ circulant expriment préférentiellement le corécepteur CXCR4. Dans 80% des cas d'une transmission par les muqueuses, un seul variant se propage chez l'individu receveur malgré une grande diversité de souches virales chez le donneur. Le variant unique qui établit l'infection est appelé « transmitted founder virus » (TF) et a un tropisme X5 (Wolinsky *et al.*, 1992; Zhu *et al.*, 1993; Salazar-Gonzalez *et al.*, 2008, 2009; Fischer *et al.*, 2010; Ochsensbauer *et al.*, 2012). Les LTs CD4⁺ CCR5⁺ constituent donc une cible privilégiée du virus et soutiennent la majorité de la réplication virale dans les premiers jours qui suivent l'infection. Comme abordé dans la section 2.4.2.1 du chapitre 2, les DCs ne sont que rarement infectées de manière productive par le VIH mais potentialisent sa dissémination par le phénomène de trans-infection (Pope *et al.*, 1995; Martín-Moreno and Muñoz-Fernández, 2019). Par la formation d'une synapse virologique ou suite à l'exocytose du virus capturé, les DCs favorisent l'infection des LTs CD4⁺ CCR5⁺ environnants. Les DCs sont également un vecteur de transport du virus jusqu'aux ganglions lymphatiques généralisant l'infection.

3.4.1.2 Propagation et prolifération du VIH au GALT

La phase d'éclipse se termine lorsque la virémie est suffisamment élevée pour que le virus soit détectable dans le sang, marquant ainsi le début de la primo-infection, également connue sous le nom de phase aiguë. La phase aiguë est caractérisée par la propagation et la prolifération du virus au tissu lymphoïde du tube digestif (GALT pour « gut-associated lymphoid tissue » en anglais) (Veazey *et al.*, 1998; Brenchley *et al.*, 2004). Le GALT est une cible privilégiée du VIH puisqu'il contient environ 60% de cellules T et que les LTs CD4⁺ CCR5⁺ mémoires activés y sont très abondants (Veazey *et al.*, 1998; Meng *et al.*, 2000). Ainsi, le GALT favorise une réplication virale massive, ce qui conduit à une déplétion sévère des LTs CD4⁺. Cette déplétion est associée à une charge virale élevée, atteignant environ 10⁷ copies d'ARN viral par millilitre de sang. Il est important de souligner que la déplétion des LTs CD4⁺ n'est pas limitée au GALT, elle affecte également les LTs CD4⁺ périphériques (Moir, Chun and Fauci, 2011; Coffin and Swanstrom, 2013). La déplétion drastique des LTs CD4⁺ a longtemps été considérée comme la conséquence d'une infection productive. Or, en 1995, Finkel et al ont démontré que la majorité des LTs CD4⁺ déplétés ne sont pas infectés (Finkel *et al.*, 1995). La mort cellulaire est induite par divers mécanismes : l'apoptose, l'autophagie, et la pyroptose (Grivel, Malkevitch and Margolis, 2000; Jekle *et al.*, 2003; Espert *et al.*, 2006; Espert and Biard-Piechaczyk, 2009; Doitsh *et al.*, 2014).

3.4.1.3 Le viral set-point

La dissémination du virus présent dans le GALT aux autres tissus lymphoïdes tels que les ganglions lymphatiques permet l'établissement de réservoirs dans les LTs CD4⁺ latents de différents tissus. Le GALT et les ganglions lymphatiques sont des réservoirs persistants du VIH (Moir, Chun and Fauci, 2011; Vanhamel, Bruggemans and Debyser, 2019; Chaillon *et al.*, 2020). Cette hausse de la charge virale et la diminution des LTs CD4⁺ est souvent accompagnée de l'apparition des premiers symptômes cliniques assimilés à ceux d'une grippe. Trois à quatre semaines après l'infection, la virémie plasmatique atteint son pic et décline ensuite progressivement jusqu'à atteindre un point d'équilibre appelé « viral set-point ». Le viral set-

point est un facteur déterminant pour prédire la progression de la maladie chez les individus infectés par le VIH (Mellors *et al.*, 1996). Parallèlement au viral set-point, le système immunitaire se met en place et se manifeste par l'apparition de deux types de réponses : la réponse humorale et une réponse cytotoxique. Cette dernière, médiée par les LTs CD8⁺, corrèle directement avec la diminution de la charge virale. La pression immunitaire exercée par les LTs CD8⁺ entraîne une diminution de la charge virale d'en moyenne d'un facteur 100 (Coffin and Swanstrom, 2013). Cependant, le virus mute dans la région codant les épitopes reconnus par les LTs CD8⁺ pour échapper à cette pression de sélection. La réponse T CD8⁺ et l'évolution virale apparaissent très tôt au cours de l'infection, bien avant l'apparition d'anticorps neutralisants (Koup *et al.*, 1994; Goonetilleke *et al.*, 2009). Les premières IgM et IgG ciblant des antigènes de l'enveloppe sont détectables 13 et 28 jours après l'infection, respectivement (Tomaras *et al.*, 2011). Malheureusement, cette réponse anticorps n'est pas capable de neutraliser la souche virale autologue. Il faudra attendre plusieurs mois avant l'apparition d'anticorps neutralisants (voir section 3.7.1).

3.4.2 La phase chronique

La phase chronique est la phase de l'infection la plus longue pouvant durer de 1 à 20 ans sans aucun symptôme. La phase chronique repose sur le contrôle de la charge virale par la réponse immunitaire. Malgré un contrôle de la virémie, l'exposition des antigènes du VIH persiste à bas bruit provoquant une activation chronique du système immunitaire. L'activation systémique des cellules de l'immunité innée et adaptative favorise la sécrétion de cytokines pro-inflammatoires et augmente le renouvellement cellulaire (Appay and Sauce, 2008). Au niveau du GALT, la libération des facteurs solubles induite par le VIH, entraîne la perturbation de la barrière de la muqueuse intestinale et provoque une translocation de produits microbiens dans la circulation, ce qui favorise l'activation globale du système immunitaire. Ces phénomènes ont pour conséquences l'apparition de troubles liés à l'inflammation et une convergence des lymphocytes T CD8⁺ et CD4⁺ vers des sous-populations plus activées et différenciées. L'activation du système immunitaire se traduit par une fréquence accrue des marqueurs CD38 et HLA-DM sur les LTs CD8⁺ et CD4⁺, et l'accumulation de monocytes CD14⁺ CD16⁺ pro-inflammatoires (Sauce, Elbim and Appay, 2013). Les cellules T expriment des molécules de costimulation inhibitrices telles que PD-1, CTLA-4, TIM-3 etc, qui résultent en une perte de fonctionnalité, un état d'épuisement et une progression de la maladie (Day *et al.*, 2006; Petrovas *et al.*, 2006; Trautmann *et al.*, 2006; Kaufmann *et al.*, 2007; Jones *et al.*, 2008). Le traitement antirétroviral (TARV) réduit que partiellement l'expression à la hausse des ces molécules inhibitrices exprimées par les cellules T (Blackburn *et al.*, 2009; Youngblood *et al.*, 2013). Les cellules T chroniquement activés ont des capacités prolifératives diminuées et une réduction de leurs fonctions effectrices.

3.4.3 La phase SIDA

L'immunodéficience acquise, caractéristique de la phase SIDA, est la conséquence de la déplétion constante des LTs CD4⁺ et du déclin progressif de l'immunocompétence. On considère que lorsque la concentration de LTs CD4⁺ diminue à 200 cellules/mL, la charge virale n'est plus contrôlée. Le système immunitaire alors en état d'épuisement, n'est plus capable de

faire face aux maladies opportunistes. Sans prise en charge, le SIDA est l'une des maladies infectieuses les plus mortelles avec un taux de mortalité supérieur à 95% (Coffin and Swanstrom, 2013).

3.5 Le cycle viral du VIH-1

Le VIH se réplique en un à deux jours et le cycle répliatif se décompose en 11 étapes. On distingue dans l'ordre : l'attachement, la fusion, la transcription inverse, la décapsidation, l'intégration, la transcription, la traduction, l'assemblage, le bourgeonnement, la libération et la maturation (Figure 45).

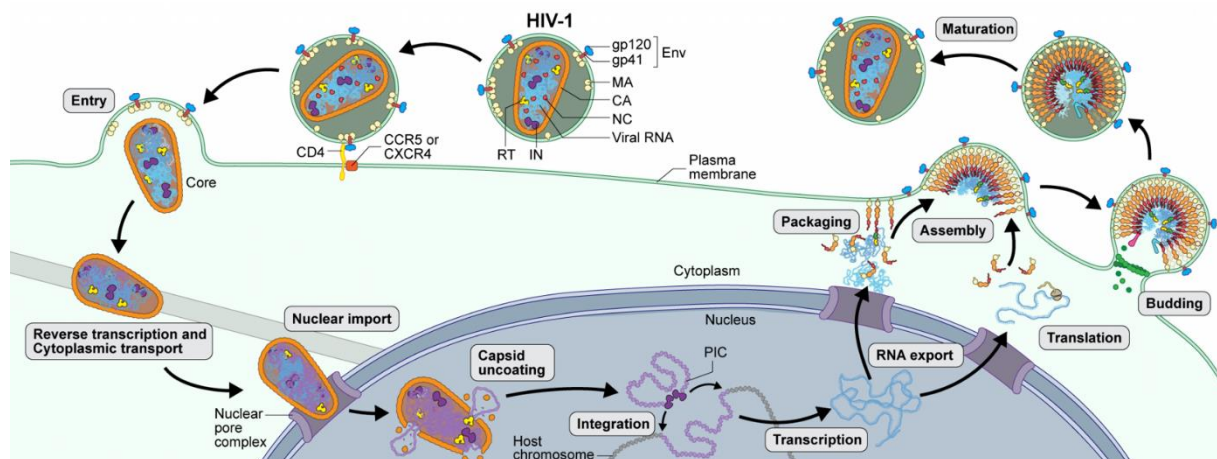


Figure 45. Cycle répliatif du VIH (figure originale de sciences of hiv.org). 11 étapes clés composent le cycle répliatif du VIH. Premièrement, le virus se fixe à la cellule cible par la fixation du récepteur CD4 et corécepteur CCR5. Il y a ensuite fusion des membranes et libération de la capsid dans le cytoplasme. La transcription inverse a lieu à l'intérieur de la capsid qui est acheminée au noyau à travers les pores nucléaires. La décapsidation, l'intégration du génome viral à celui de la cellule hôte, et la transcription se réalisent dans le noyau. Dans le cytoplasme, l'ARNm viral est traduit en protéines. Enfin, les protéines néo-synthétisées sont assemblées à la membrane plasmique pour former un nouveau virion qui sera libéré par bourgeonnement. La maturation de la particule virale est la dernière étape du cycle répliatif.

3.5.1 Attachement et fusion

L'attachement du virus à la cellule cible est facilitée par l'interaction de la protéine Env avec des récepteurs de surface tels que l'intégrine $\alpha 4\beta 7$, DC-SIGN, les glycolipides, etc. (Chen, 2019). Ces molécules de surface ne sont pas des corécepteurs d'entrée du VIH mais favorisent l'infection, en permettant le rapprochement de Env à son récepteur d'entrée : le récepteur CD4 (Cicala *et al.*, 2009; Hijazi *et al.*, 2011; Martín-Moreno and Muñoz-Fernández, 2019). Plusieurs études *in vitro* et *ex vivo* ont montré que les LTs CD4⁺ exprimant l'intégrine $\alpha 4\beta 7$ sont préférentiellement infectés (Martinelli *et al.*, 2013; Joag *et al.*, 2016; Sivro *et al.*, 2018). Néanmoins, la fixation du virus à ces récepteurs de surface n'est pas essentielle pour une infection productive.

La fixation de la glycoprotéine gp120 au récepteur CD4 joue un rôle central dans le processus d'entrée du virus dans la cellule cible (Chen, 2019; Murakami and Ono, 2021). La liaison de la gp120 au récepteur CD4 provoque un réarrangement conformationnel de la protéine Env ce qui permet l'extension de la gp41 et l'insertion du peptide de fusion à la membrane plasmique. Ce phénomène entraîne un repliement dynamique de la gp41 qui provoque le rapprochement

et la fusion des membranes du virus et de la cellule cible (Lin and Da, 2020). La capside est alors libérée dans le cytoplasme de la cellule hôte.

Les corécepteurs CCR5 et CXCR4 interviennent dans l'étape de fusion en permettant le rapprochement et la stabilisation du complexe gp120-CD4 afin de favoriser l'insertion du peptide de fusion à la membrane plasmique (Chen, 2019).

3.5.2 La transcription inverse

La transcription inverse est le mécanisme selon lequel un ARN génomique simple brin est transcrit en ADN complémentaire double brin. La transcription inverse a lieu dans la capside et fait appel aux deux activités enzymatiques de la transcriptase inverse (RT). L'activité polymérase de la RT va permettre la synthèse d'un brin d'ADN complémentaire à partir d'une séquence d'ADN ou d'ARN, tandis que l'activité ribonucléase H hydrolyse l'ARN uniquement si ce dernier est sous la forme d'un double hybride ADN/ARN (Hu and Hughes, 2012).

La transcription inverse débute par la synthèse du brin d'ADN (-) à partir d'une matrice qui n'est autre que la région 5'LTR de l'ARN viral génomique. La RT utilise l'ARN de transfert (ARNt) Lys3 de l'hôte comme amorce, qui en s'appariant avec un codon présent au site de liaison PBS va permettre l'initiation de la polymérisation du brin d'ADN complémentaire (Figure 46 étape A). La molécule d'ARN composant le double hybride naissant ADN/ARN est dégradé par l'activité ribonucléase H de la RT (Etape B). Les séquences répétées « R » situées aux deux extrémités de l'ARN viral servent d'amorces en transférant l'ADN (-) de l'extrémité 5' à l'extrémité 3' (Etape C). La synthèse du brin d'ADN (-) se poursuit le long du génome en même temps que la dégradation du brin d'ARN (Etape D). Cependant, deux régions de l'ARN génomique sont résistantes à l'activité RNase H. Ces séquences, appelées « ppt », sont enrichies en purine et servent d'amorces pour la synthèse du brin d'ADN (+). En plus de synthétiser le brin d'ADN (+), la RT copie également les 18 premiers nucléotides de l'amorce de l'ARNt Lys3 (Etape E). Ces 18 nucléotides sont donc complémentaires à la séquence PBS de l'extrémité 5' du brin d'ADN (-). Les deux extrémités s'hybrident et l'élongation des brins d'ADN (+) et (-) se poursuit (Etape F). Le produit final est un ADN double brin qui contient à chacune de ses extrémités les séquences « U3-R-U5 », essentielles à l'étape d'intégration (Etape G) (Hu and Hughes, 2012). L'ADN viral en association avec les protéines virales Vpr, MA, IN, et la protéine cellulaire nommée « facteur barrière à l'auto-intégration » (BAF1) forment le complexe de pré-intégration PIC que j'appellerai ici « PICv » pour « PIC viral » afin de ne pas induire de confusion avec le complexe PIC de l'initiation eucaryote énoncé dans le chapitre 1.

L'immense diversité génétique du VIH observé au cours de l'infection est le plus souvent attribuée à la RT puisqu'elle possède un taux élevé d'erreur expliqué par l'absence de fonction de relecture. Néanmoins, d'autres étapes du cycle répliatif peuvent être à la source d'erreurs à savoir : la transcription de l'ADN viral en ARNm par l'ARN polymérase II de l'hôte et les hypermutations G-A médiées par le facteur de restriction cellulaire APOBEC3G (Hu and Hughes, 2012; Coffin and Swanstrom, 2013). Peu importe la source d'erreur, le taux global de mutations pour un cycle répliatif est de 3.10^{-5} mutations par base (Coffin and Swanstrom, 2013).

Plusieurs études ont récemment révolutionné le domaine du VIH en démontrant que la libération du matériel génétique par le désassemblage de la capsid (ou décapsidation) n'avait pas lieu dans le cytoplasme mais à l'intérieur du noyau (Li *et al.*, 2021; Scoca and Di Nunzio, 2021). Li et al ont récemment montré que la capsid passe les pores nucléaires, et se désassemble près du site d'intégration une à deux heures avant l'intégration (Li *et al.*, 2021).

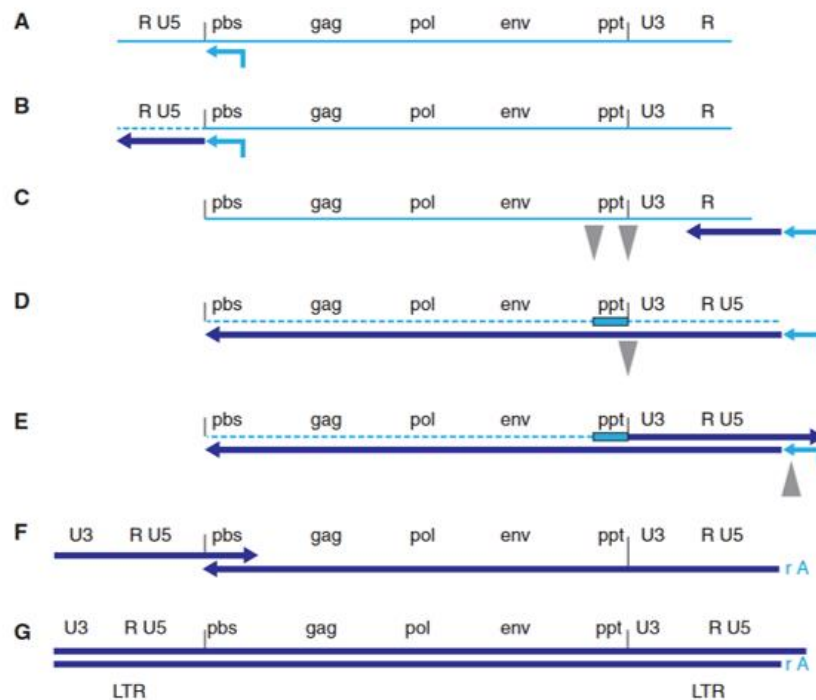


Figure 46. Etapes clés de la transcription inverse du VIH (figure originale de Hu and Hughes, 2012).
Le génome viral est représenté par les lignes bleu clair. L'amorce cellulaire (Lys3) est indiquée par les flèches épaisses bleu clair, et la polymérisation des brins complémentaires correspond aux flèches bleu foncé. La dégradation de l'ARN génomique par la RNase H est représentée par les lignes en pointillés.

3.5.3 L'intégration

Une fois libéré de la capsid, l'intégrase se tétramérise au sein du PICv, en se fixant aux extrémités de l'ADN viral au niveau des LTRs. Ce processus forme un complexe appelé « intasome » (Craigie and Bushman, 2012; Engelman and Singh, 2018). L'intégrase élimine ensuite un dinucléotide GT à chaque extrémités 3' de l'ADN viral. Les groupements hydroxyles libres en 3' servent alors de nucléophiles pour catalyser le transfert de brin (Engelman, Mizuuchi and Craigie, 1991). L'extrémité 3'OH de l'ADN viral forme alors une liaison covalente avec l'extrémité 5'phosphate de l'ADN cellulaire. L'intégration s'effectue préférentiellement dans des régions génomiques activement transcrites ou riches en gènes (Berry *et al.*, 2018). Une fois l'intégration terminée, l'ADN viral est stable et qualifié de provirus.

Suite à l'intégration, le provirus peut être éteint transcriptionnellement ce qui entraîne la formation de réservoirs viraux, ou transcrit puis traduit en protéine afin de produire de nouvelles particules virales. Le provirus est transcrit en ARN viral par le processus de transcription cellulaire faisant intervenir l'ARN polymérase II et des facteurs de transcription mais implique également les protéines virales accessoires Tat et Vpr. Comme pour les ARN

cellulaires, l'ARN viral subit également un processus d'épissage alternatif pour éliminer les introns et générer des isoformes différentes. Une fois épissé, l'ARN viral est exporté dans le cytoplasme, où il est traduit en protéines par les ribosomes de la cellule hôte.

3.5.4 La traduction

3.5.4.1 Les différents transcrits

L'ADN du provirus est transcrit en ARN pré-messager de 9kb qui contient tous les ORFs du VIH ainsi que tous les sites d'épissage alternatif. Dans le noyau, l'épissage alternatif permet la génération de plus de 50 isoformes d'ARN viral (Schwartz, Felber, Benko, *et al.*, 1990; Purcell and Martin, 1993; Ocwieja *et al.*, 2012; Sertznig *et al.*, 2018). Par conséquent, les ARNm du VIH sont classés en fonction de leur taille. On retrouve : 1) l'ARNm génomique non épissé de 9kb dit de pleine longueur, 2) les ARNm partiellement épissés de 4kb et 3) les ARNm de 2kb multi épissés (Figure 47) (de Breyne and Ohlmann, 2018; Sertznig *et al.*, 2018; Nguyen Quang *et al.*, 2020). Chacune de ces formes joue un rôle différent dans la production de nouvelles particules virales. L'ARNm génomique non épissé peut être utilisé comme matériel génétique et être encapsidé lors de l'assemblage des nouvelles particules virales ou servir de substrat pour la traduction des protéines Gag et des précurseurs protéiques Gag-pol. Les ARNm de 4kb codent

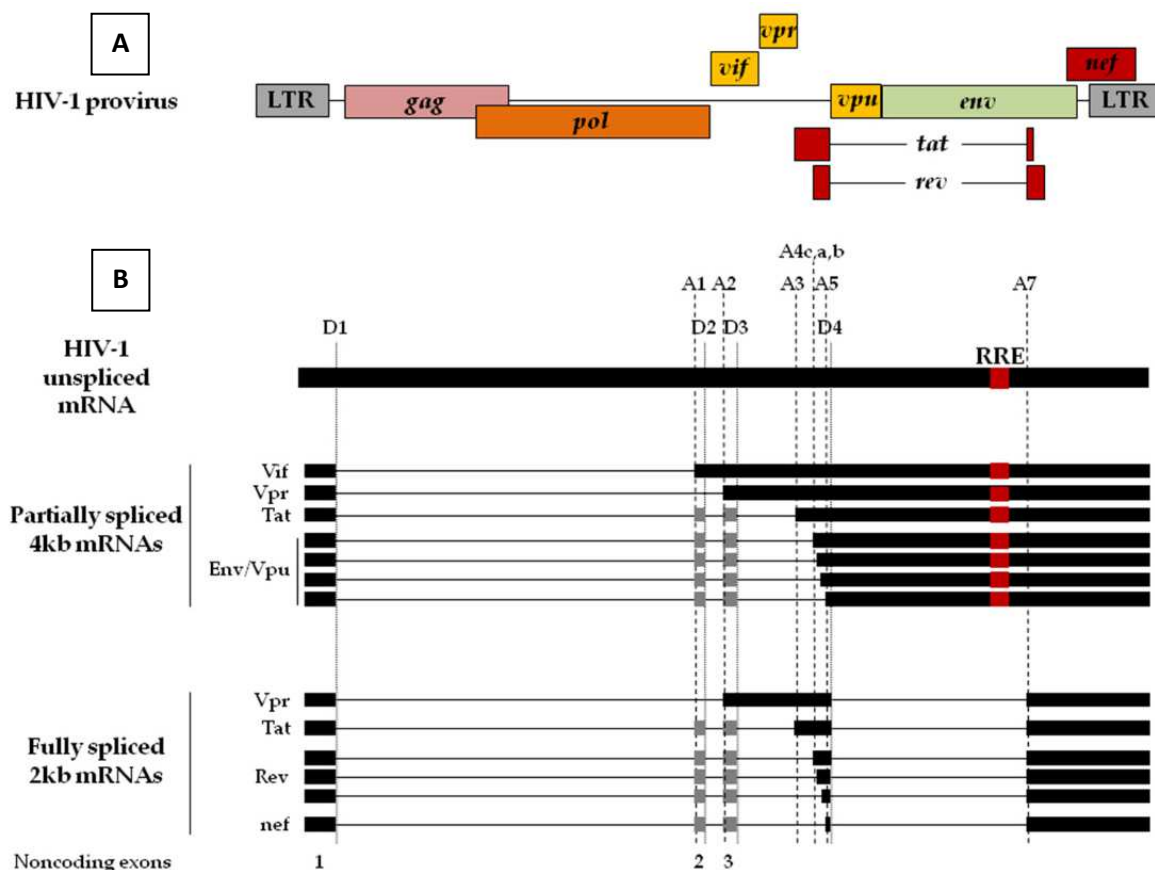


Figure 47. Les différents transcrits du VIH (figure originale de De Breyne and Ohlmann, 2018). (A) Représentation schématique du génome du VIH. (B) Représentation schématique des différents types de transcrits. Les positions des sites accepteurs (A1 à A7) et donneurs (D1 à D4) de l'épissage alternatifs sont indiqués. L'ARN génomique non épissé de pleine longueur permet la synthèse des protéines Gag et Gag-pol. Le site RRE, essentiel à la fixation de Rev pour l'export nucléaire des ARN partiellement et non épissés, est indiqué par un carré rouge. Les exons des transcrits partiellement et multi épissés sont représentés par des rectangles noirs. L'ensemble de ces transcrits possèdent un exon non codant en 5' (exon1). Deux autres exons non codants (carrés gris) peuvent être inclus.

pour les protéines Vif, Vpr, un exon de Tat, et l'ARNm bicistronique *env/vpu*. Les ARNm de 2kb codent pour les protéines Tat, Rev, et Nef. L'épissage alternatif permet de favoriser l'initiation de la traduction des protéines codées en 3' en réarrangeant les ORFs en aval de l'extrémité 5' (Sertznig *et al.*, 2018).

Les ARN sont exprimés de manière différentielle. Les ARN entièrement épissés sont les premiers transcrits générés lors de l'infection puisque leur export nucléaire est indépendant de la protéine Rev (Kim *et al.*, 1989). La traduction des ARN partiellement et non épissés est dépendante de la protéine Rev, qui suite à sa traduction, permet l'export nucléaire des transcrits par sa fixation à la protéine cellulaire CRM1 et à l'élément réactif de Rev « Rev-responsive element » (RRE) localisé dans le gène Env (Malim *et al.*, 1989; Cochrane, Chen and Rosen, 1990; Fernandes, Booth and Frankel, 2016). L'expression des ARN est également différentielle d'un point de vue quantitatif. Par exemple, la quantité des transcrits de 2kb codant pour Tat est cinq fois inférieure aux transcrits codant pour Rev et Nef. Les transcrits codant pour Env représentent environ 90% des ARNm partiellement épissés (de Breyne and Ohlmann, 2018).

3.5.4.2 La traduction dépendante de la coiffe

Comme les ARNm cellulaires, les ARNm viraux possèdent une coiffe m7G à l'extrémité 5' qui permet le recrutement de complexe de pré-initiation PIC. Pour rappel, lors de l'initiation de la traduction des ARNm cellulaires, le complexe PIC composé de la petite sous unité 40S du ribosome, l'ARNt méthionine, et les facteurs d'initiation eIF2, -1, -1A, -3, et -5 se lie au complexe eIF4F préalablement fixé à la coiffe. Le complexe d'initiation devient alors mature et débute le phénomène de scanning jusqu'à la rencontre du codon AUG. La reconnaissance du codon AUG induit le recrutement de la large sous unité du ribosome 60S au complexe d'initiation. Le ribosome devient mature et démarre l'élongation. Néanmoins, le processus d'initiation de la traduction des ARNm viraux est plus compliqué en raison de la présence de nombreuses structures d'ARNm, notamment la structure TAR, au sein des LTRs. TAR est une structure essentielle à la régulation de la transcription puisqu'elle se situe dans la région « R » de l'extrémité 5' et fixe la protéine virale Tat qui stimule la transcription (Rice, 2017). Cependant, la structure TAR constitue un obstacle majeur au recrutement du complexe PIC à la coiffe et au scanning (Parkin *et al.*, 1988; SenGupta *et al.*, 1990; Svitkin, Pause and Sonenberg, 1994; de Breyne and Ohlmann, 2018). Afin de permettre la traduction de ses protéines virales le VIH recrute des protéines cellulaires pour contourner la structure TAR. On distingue : les protéines La, Staufen 1, TRBP et les protéines DEAD-Box hélicase. L'immunoprécipitation de l'ARNm de lymphocytes T infectés par le VIH a révélé une interaction entre la protéine La et l'ARNm viral (Chang *et al.*, 1995). Quelques années plus tard Hühn *et al.* ont montré que La possède une activité hélicase, et favorise le déroulement des structures des régions LTRs (Hühn *et al.*, 1997). La protéine TRBP favorise l'initiation de la traduction par sa liaison à la structure TAR et par l'inhibition de la protéine kinase PKR, qui induite par l'IFN γ , bloque l'initiation de la traduction par phosphorylation du facteur d'initiation eIF2 α (de Breyne and Ohlmann, 2018). La protéine Staufen 1 lie les ARN double brins et intervient dans de nombreux processus cellulaires. En combinant l'immunoprécipitation des ARN à du Ribo-Seq et du RNA-seq, l'équipe du Dr Ricci a montré que Staufen 1 lie les ARNm en cours de traduction et préférentiellement les ARNm

enrichis en structures secondaires (Ricci *et al.*, 2014). D'autres études ont également montré le rôle de Staufen 1 dans le recrutement des polysomes et le déroulement des structures secondaires des LTRs (Villacé, Marión and Ortín, 2004; Dugré-Brisson *et al.*, 2005; Hartman *et al.*, 2006). Pour finir, l'hélicase DDX3 dépendante de l'ATP déroule les structures secondaires de l'ARNm (voir chapitre 1 section 1.4.2). L'expression de DDX3 est stimulée par la protéine Tat (Yedavalli *et al.*, 2004). En effet, l'analyse du méta-gène humain a montré des modifications des génétiques traduisant une augmentation de l'expression de DDX3 lors du cycle viral (Bushman *et al.*, 2009). L'ensemble de ces facteurs cellulaires jouent un rôle important dans l'initiation de la traduction du VIH dépendante de la coiffe. Malgré un grand nombre d'études beaucoup de questions mécanistiques restent encore ouvertes.

3.5.4.3 Le frameshifting

L'expression des gènes du VIH est également modulée au niveau traductionnel. Les protéines Gag sont dix à vingt fois plus exprimées que les précurseurs protéiques Gag-pol alors qu'ils sont issus du même transcrit (Shehu-Xhilaga, Crowe and Mak, 2001). La traduction de *pol* n'est pas initiée avec un codon start AUG en début de l'ORF mais est dépendante d'un mécanisme alternatif de traduction appelé le « ribosome frameshifting » (FS) défini dans le chapitre 1 en section 1.9.6. L'ORF codant pour *pol* est alors traduit uniquement si le ribosome engagé dans l'élongation de *gag* poursuit la traduction en aval du codon stop. Le FS gag/pol est un phénomène qui induit un changement de cadre de lecture du ribosome qui se décale d'un nucléotide en 5'. Le changement de cadre de lecture en -1 du ribosome est provoqué par la présence de deux séquences nucléotidiques spécifiques situées 200 nucléotides en amont du codon stop de *gag* (Brierley and Dos Ramos, 2006; Guerrero *et al.*, 2015) (Figure 48).

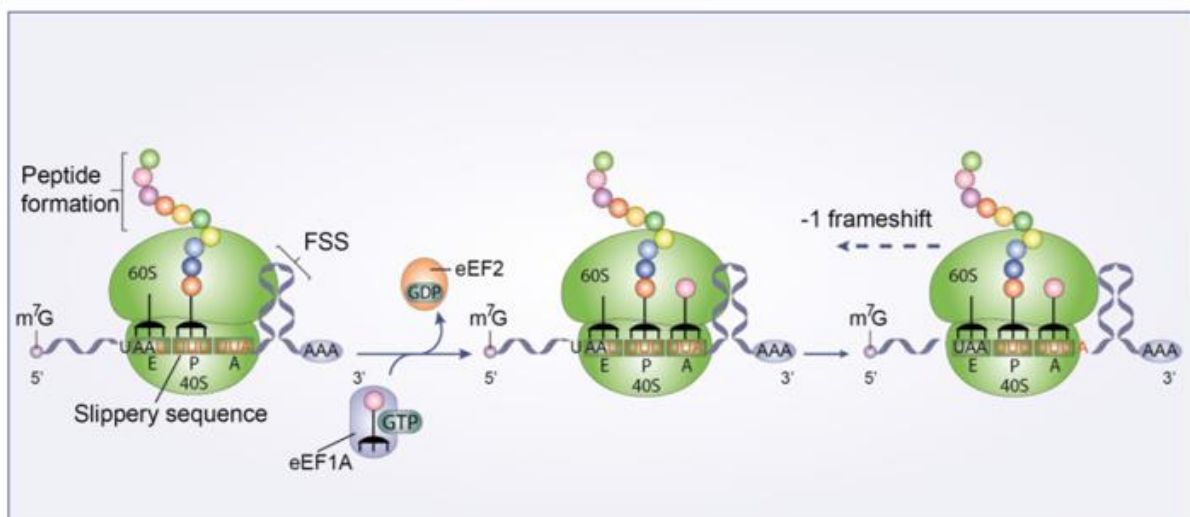


Figure 48. Le -1 ribosomal frameshifting du VIH (figure originale de Guerrero et al, 2015).

Le -1 FS a lieu environ 200 nucléotides en amont du codon stop de *gag*, au niveau de la séquence heptanucléotidique glissante UUUUUUA. L'élément d'ARN en tige boucle en aval de la séquence glissante contrôle le décalage de phase.

Depuis sa première mise en évidence en 1988 par Jacks et al, la structure du signal de FS et son rôle dans la réplication ont été l'objet de nombreuses études (Jacks *et al.*, 1988; Brierley and Dos Ramos, 2006). Le FS est dépendant d'une séquence signal « UUUUUUA » appelée séquence glissante. Cette séquence est conservée parmi les isolats du VIH et sa mutation

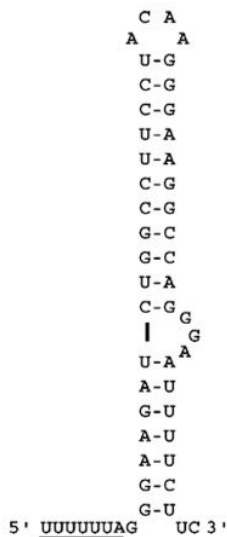


Figure 49. Représentation schématique de la structure d'ARN du -1 FS (figure originale de Brierley et al, 2006).
Modèle de structure d'ARN proposé par Dulude et al 2002.

entraîne une réduction l'efficacité du FS et diminution drastique de l'infectivité (Biswas *et al.*, 2004). Cependant, l'unique présence de la séquence glissante n'est pas suffisante pour induire le FS qui implique également la présence d'une structure en tige boucle. L'efficacité du FS est d'ailleurs corrélée à la stabilité de cette boucle. Plusieurs modèles de tige boucle sont proposés mais le plus décrit est le modèle défini par le laboratoire du D^r Brakier-Gingras (Figure 49) (Dulude, Baril and Brakier-Gingras, 2002; Baril *et al.*, 2003; Gaudin *et al.*, 2005). La présence de ces deux éléments structuraux déclenche une pause du ribosome au niveau de la tige boucle. Cette pause entraîne le décalage de cadre de lecture en -1 (Namy *et al.*, 2006).

L'efficacité du FS -1 du VIH-1 est de 5% à 10% (Jacks *et al.*, 1988). Le maintien de ce ratio est primordial puisque la traduction unique de *gag* ou de *gag-pol* mènent à la production de particules virales non infectieuses et à l'inhibition du clivage des

précurseurs protéiques par la protéase (Park and Morrow, 1991; Karacostas *et al.*, 1993; Shehu-Xhilaga, Crowe and Mak, 2001; Cen *et al.*, 2004).

La traduction de l'ARNm génomique non épissé est initiée selon 2 mécanismes majeurs : l'initiation dépendante de la coiffe et le recrutement du complexe de pré-initiation PIC par les IRES.

3.5.4.4 La traduction dépendante d'IRES

Au sein de l'ARNm génomique du VIH, trois séquences IRES ont été caractérisées pour initier la traduction indépendamment de la coiffe (Figure 50). La première séquence IRES se situe dans la région 5'UTR au niveau du site PBS et a été mise en évidence dans les cellules T et les monocytes infectées (Gendron *et al.*, 2011; Plank, Whitehurst and Kieft, 2013; Plank *et al.*, 2014).

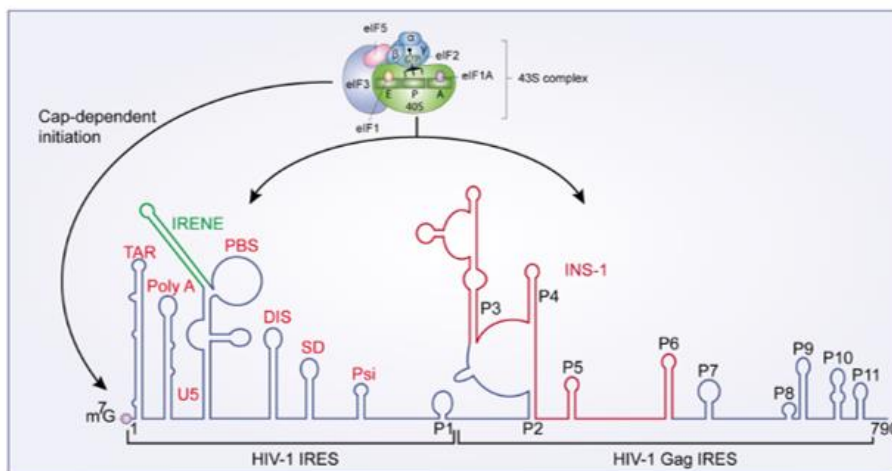


Figure 50. Les IRES du VIH (figure originale de Guerrero et al, 2015). Les trois IRES du VIH sont représentés : celui localisé dans la région 5' (HIV-1 IRES) et les deux IRES successifs situés dans l'ORF de gag (HIV-1 Gag IRES). Les différentes structures des IRES sont représentées et annotées. PBS : Primer binding site, est le site de liaison de l'amorce Lys3 intervenant dans la transcription inverse. IRENE est une région qui régule négativement l'activité de l'IRES.

Le facteur eIF1A permet le recrutement du complexe PIC à l'IRES. Cependant, l'identification précise des acteurs cellulaires impliqués dans ce mécanisme reste incomplète. Deux IRES supplémentaires ont été décrits dans l'ORF de gag (Buck *et al.*, 2001; Deforges *et al.*, 2017). L'interaction de ces IRES avec les sous-unités 40S des ribosomes ont été décrites comme directes et indépendantes des facteurs cellulaires (Deforges *et al.*, 2017). Seul le facteur eIF4A est nécessaire dans la reconnaissance du codon AUG (de Breyne *et al.*, 2012).

3.5.4.5 Impact de l'infection virale sur les protéines cellulaires

Tout au long du cycle réplcatif, les protéines du VIH inhibent différents facteurs cellulaires de la traduction afin de moduler l'expression des protéines virales. Par exemple, la protéase du VIH est capable de cliver les facteurs eIF4GI, PABP, et eIF3d entraînant ainsi une inhibition de l'initiation dépendante de la coiffe (Ventoso *et al.*, 2001; Alvarez *et al.*, 2006; Castelló *et al.*, 2009). La protéolyse de ces facteurs d'initiation induit une diminution drastique de la traduction coiffe dépendante (coiffe-eIF4E-eIF4G-PABP) des ARNm cellulaires et viraux (de Breyne and Ohlmann, 2018). Il a été rapporté depuis 1995 que l'expression de la protéine Vpr entraîne un arrêt du cycle cellulaire en phase G2. Les mécanismes impliqués dans l'inhibition du cycle cellulaire ne sont pas tous élucidés. Vpr semble activer la voie de signalisation des dommages à l'ADN. L'arrêt du cycle en phase G2 corrèle avec une forte diminution de la traduction coiffe dépendante.

Lors de l'inhibition de la traduction cellulaire par l'action des protéines virales, la traduction du VIH dépendante des IRES a été démontrée fonctionnelle, ce qui permet la poursuite du cycle viral et la production de nouveaux virions (Vagner, Galy and Pyronnet, 2001; de Breyne and Ohlmann, 2018). La cinétique de traduction est la suivante, les transcrits codants les protéines accessoires sont traduits rapidement par une traduction dépendante de la coiffe, à l'inverse de l'ARNm génomique non épissé de 9kb dont la traduction est majoritairement initiée à un stade plus tardif par les IRES.

Exemple d'APOBEC3G

Certaines étapes du cycle viral sont inhibées par les facteurs de restriction cellulaires APOBEC3G, SAMHD1, Tethrin, TRIM5 α (Colomer-Lluch *et al.*, 2018). APOBEC3G est une cytidine désaminases qui catalyse la conversion de la cytidine en uridine dans le brin d'ADN (-) entravant ainsi la transcription inverse et l'intégration du VIH (Sheehy *et al.*, 2002). Néanmoins, l'action d'APOBEC3G est contrecarrée par la protéine virale Vif selon deux mécanismes. Premièrement, la liaison de Vif à APOBEC3G provoque le recrutement de l'ubiquitine ligase E3 qui poly-ubiquitine APOBEC3G alors dégradée par le protéasome. Deuxièmement, Vif perturbe la traduction d'APOBEC3G. Mariani et al ont été les premiers a montré le rôle de Vif sur la traduction d'APOBEC3G (Mariani *et al.*, 2003). En utilisant la méthode du pulse-chase, ils démontrent que Vif provoque la diminution de la synthèse d'APOBEC3G d'environ un facteur 5. Stopak et al confirment ce résultat en montrant que Vif inhibe la traduction d'APOBEC3G de 30 à 40% (Stopak *et al.*, 2003). En absence de Vif, les ribosomes qui traduisent APOBEC3G sont recrutés au codon AUG de l'ORF par le mécanisme de ré-initiation à partir de l'uORF (chapitre 1 section 1.9.4). Lors de l'infection à VIH, Vif inhibe ce mécanisme de ré-initiation par la

séquestration des ribosomes au niveau de l'uORF d'APOBEC3G (Libre *et al.*, 2021). Les auteurs montrent également une relocalisation cellulaire dans les granules de stress corrélée à la fonctionnalité de l'uORF et dépendante de Vif. Le rôle précis des facteurs cellulaires dans le mécanisme d'inhibition d'APOBEC3G via l'uORF médié par Vif, reste à déterminer.

La figure 51 représente les différents éléments viraux et cellulaires impliqués dans le contrôle de la traduction du VIH.

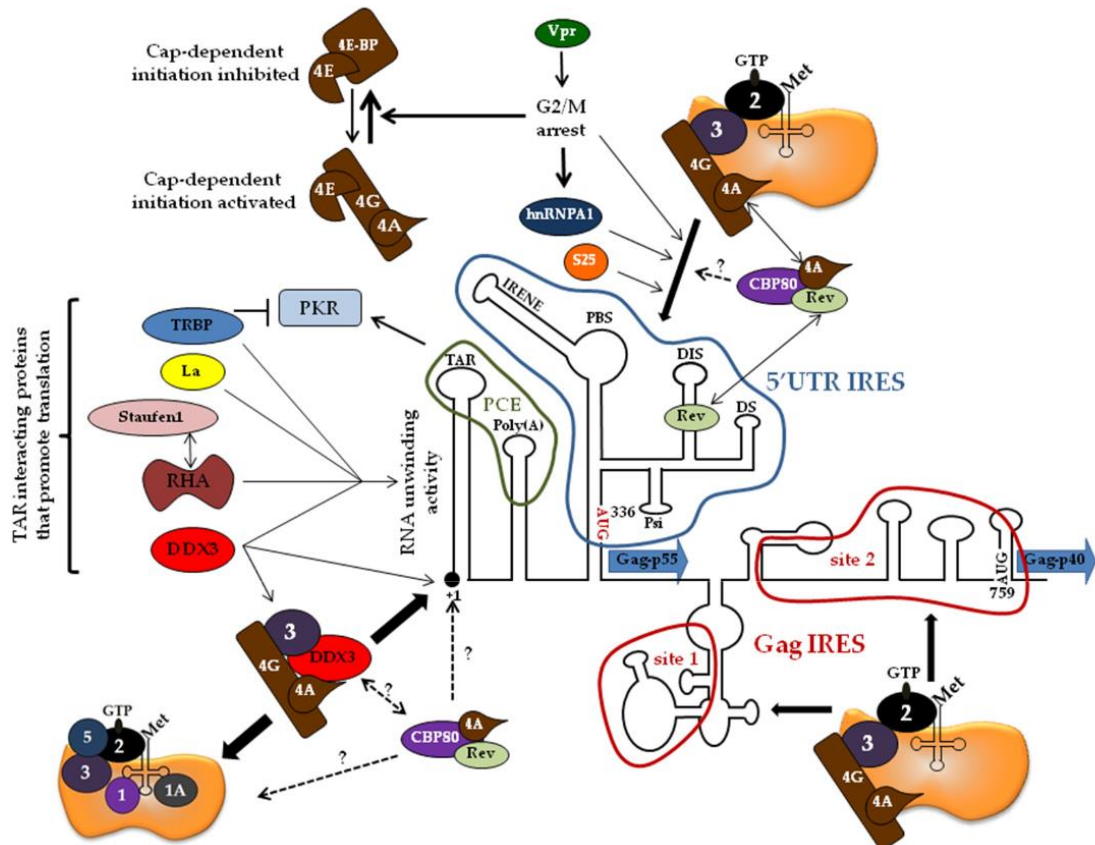


Figure 51. Les différents facteurs viraux et cellulaires impliqués dans le contrôle de la traduction du VIH (figure originale de Breyne and Ohlmann 2018).

3.5.5 Assemblage, bourgeonnement et maturation

Dans les cellules T CD4⁺ infectées, les protéines virales néosynthétisées sont assemblées à la membrane plasmique pour former de nouveaux virions (Figure 52). La protéine Env est acheminée à la membrane plasmique par la voie sécrétoire, transitant du RE au Golgi puis dans des vésicules sécrétoires jusqu'à rejoindre la membrane. Les protéines Gag et Gag-pol se multimérisent dans le cytoplasme tout en recrutant l'ARNm génomique puis rejoignent la membrane plasmique. Le complexe d'assemblage Env-Gag-Gag-pol-ARNm génomique recrute les facteurs cellulaires ESCRT qui interviennent dans le bourgeonnement et la scission de la membrane plasmique. Le virion libéré devient mature suite à la protéolyse du précurseur Gag-pol par la protéase PR ce qui permet la formation de la capsid et donc la maturation du virion (Freed, 2015).

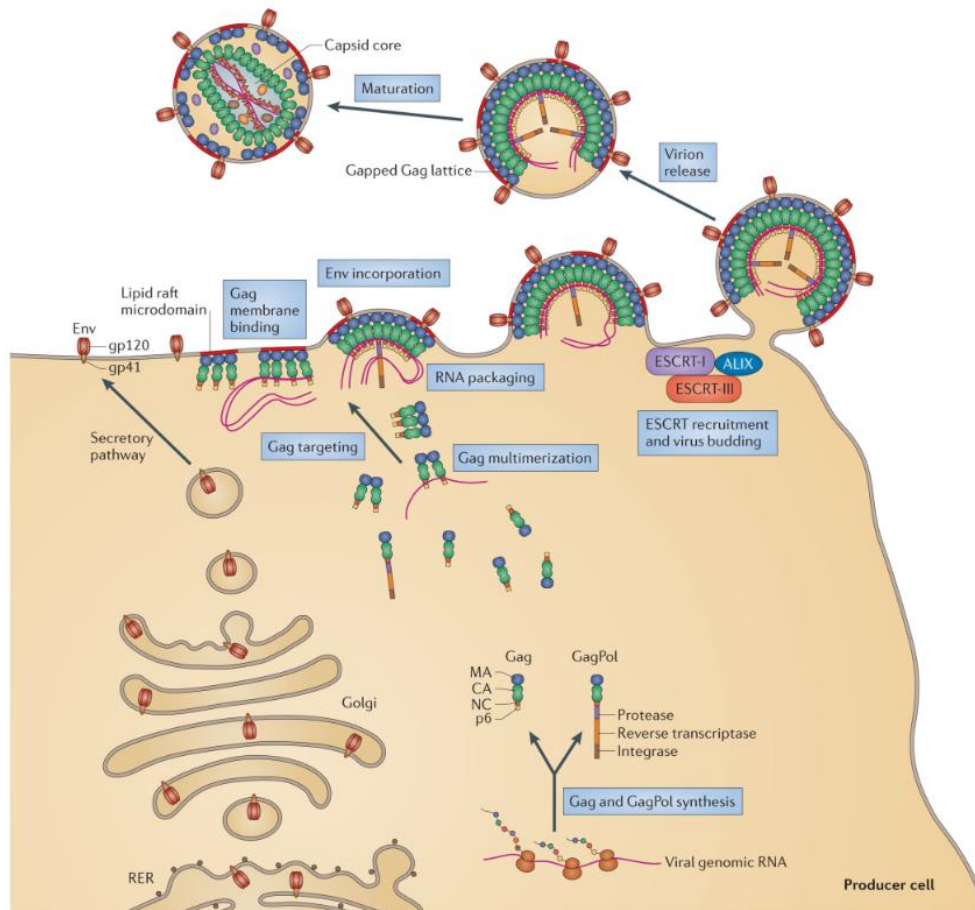


Figure 52. Assemblage, bourgeonnement et maturation des virions (figure originale de Freed 2015).

3.6 Les réservoir viraux

Le réservoir du VIH est un ensemble de cellules infectées de manière latente (White, 2022). En virologie, la latence fait référence à un stade non productif d'infection dans un ensemble de cellules transcriptionnellement silencieuses, mais qui conservent la capacité à produire des particules virales lors de stimulation (Vanhamel, Bruggemans and Debyser, 2019). C'est donc une infection non productive mais réversible qui persiste malgré de longues périodes de virémie supprimée par les TARV. Dans le contexte du VIH, la latence se caractérise par l'intégration d'un génome viral inactif transcriptionnellement mais compétent pour la réplication (Sloan *et al.*, 2015).

Plusieurs études ont démontré que le réservoir s'établit dans les premiers jours suivant l'infection, pendant la phase d'éclipse (Chun *et al.*, 1998; Vanhamel, Bruggemans and Debyser, 2019). Chez les individus asymptomatique le réservoir est constitué de 10^6 - 10^7 cellules. On sait d'ailleurs que l'initiation précoce de TARV diminue la taille des réservoirs et est associée à une meilleure évolution de la maladie (Vanhamel, Bruggemans and Debyser, 2019). La localisation du provirus dans les sanctuaires et la longue demi-vie des cellules quiescentes constituent des obstacles majeurs à l'éradication du VIH.

Les cellules mémoires sont localisées dans un grand nombre de tissus, élargissant alors le réservoir du VIH à l'intestin, le cerveau, les poumons, les organes lymphoïdes, la peau, aux

organes génitaux etc. (Mueller and Mackay, 2016). Certains de ces tissus comme le cerveau, la moëlle osseuse, les tissus adipeux, le tractus génital sont considérés comme des sanctuaires où l'accès des drogues y est encore plus difficile. Parmi les lymphocytes T CD4+ mémoires, la sous-population mémoire centrale est considérée comme la cible préférentielle du virus. L'implication des macrophages, des fibroblastes, des cellules microgliales et des podocytes dans les réservoirs est débattue malgré l'identification de provirus dans ces cellules (Vanhamel, Bruggemans and Debyser, 2019). Les DCs folliculaires présentent dans les follicules B des ganglions lymphatiques contribueraient aux réservoirs en présentant à leur surface et de manière constante un ensemble de virions (Spiegel *et al.*, 1992).

3.7 La réponse lymphocytaire anti-VIH

Etant au centre de ma thèse, je développerai principalement dans cette section les réponses lymphocytaires T CD4+ et CD8+ dirigées contre le VIH. Les grandes lignes de la réponse humorale seront abordées puisqu'elle participe également à la reconnaissance et à l'élimination des cellules infectées.

3.7.1 La réponse humorale

Dans le cadre d'une infection à VIH, la réponse humorale est altérée. En effet, les premières immunoglobulines plasmatiques produites en primo-infection, à savoir des IgM et des IgG spécifiques de gp41 et gp120, respectivement, ne sont pas capables de neutraliser la souche virale autologue (Tomaras *et al.*, 2008). Les anticorps capables de neutraliser la souche virale autologue n'apparaissent qu'en phase chronique d'infection. Un anticorps est dit neutralisant quand il est capable d'interférer avec une ou plusieurs étapes du cycle viral (Zolla-Pazner, 2004). Les anticorps neutralisants du VIH ciblent différents sites sur les protéines gp120 et gp41 de Env. Ces sites sont appelés sites de vulnérabilité (Mouquet, 2014). Les épitopes de la gp120 qui induisent une réponse humorale neutralisante sont majoritairement situés dans la boucle V2 et V3 et dans le domaine de fixation au récepteur CD4. Au sein de la gp41 se sont les épitopes issus du peptide de fusion, du domaine transmembranaire plus précisément le « membrane-proximal external region », de sa région de répétition heptadique, et de l'interface gp120-gp41 qui stimulent la synthèse d'anticorps neutralisants (Caillat *et al.*, 2020). La réponse immune exerce une pression de sélection sur les virus conduisant à l'émergence de mutations d'échappement dans les épitopes de l'enveloppe reconnus par les anticorps neutralisants. Ce phénomène explique entre autres la grande diversité des séquences codantes pour Env intra et inter-individus.

Dans la population de PVVIH, une proportion équivalente à 10-25% génère des anticorps capables de neutraliser plusieurs souches virales, communément appelés anticorps cross-neutralisants, tandis que le reste de la population génère des anticorps qui neutralisent uniquement la souche autologue (Gray *et al.*, 2011; Hraber *et al.*, 2014). En 2009, de nouveaux anticorps neutralisants à large spectre (bNAbs) sont découverts (Scheid *et al.*, 2009). Parmi les 25% d'individus générant des anticorps cross-neutralisants, 1 à 2% développent des bNAbs capables de neutraliser *in vitro* un large spectre de souches virales à travers les clades (Simek *et al.*, 2009; Hraber *et al.*, 2014; Moore, Williamson and Morris, 2015; Landais and Moore, 2018; Moris, Pereira and Chakrabarti, 2019a). Les individus qui produisent les bNAbs sont appelés

« Elite Neutralizer » (EN). Néanmoins, le profil EN n'est pas corrélé au profil des « Elite Controllers » (EC). Les EN ont une charge virale élevée, un taux élevé de Tfh circulants, et des populations de LTs régulateurs et natural killer altérées (Moore, Williamson and Morris, 2015; Haynes *et al.*, 2023).

Les bNAbs sont des IgG qui reconnaissent différents sites de vulnérabilité de Env (Haynes *et al.*, 2023). Les bNAbs neutralisent directement les virions et stimulent l'ADCC dépendante du FcγR (Julg and Barouch, 2021). Certains bNAbs peuvent neutraliser jusqu'à 95% des souches de VIH (Klein *et al.*, 2013).

3.7.2 La réponse lymphocytaire T

3.7.2.1 La réponse lymphocytaire T CD4⁺ anti-VIH

Les lymphocytes T CD4⁺ (LTs CD4⁺) sont les médiateurs de la réponse immune. On distingue plusieurs sous-populations : les Th1, Th2, Th9, Th17, Th22, les Tfh, et les T régulateurs induits (Eyerich *et al.*, 2009; Zhu, Yamane and Paul, 2010). Chaque population est caractérisée par un facteur de transcription distinct, associée à des fonctions effectrices spécifiques. La différenciation d'un LT CD4⁺ naïf en l'une de ces sous-populations est conditionnée par l'APC et l'environnement cytokinique (Rouers, Jeger-Madiot, *et al.*, 2017).

La réponse CD4 auxiliaire

Les LTs CD4⁺ spécifiques du VIH sont détectables en phase aiguë et chronique d'infection (Betts *et al.*, 2001; Palmer *et al.*, 2002). Néanmoins, leur fréquence est faible (0,2 à 2,94%) et bien inférieure à celle des LTs CD8⁺ (1,6 à 18,4%) (Betts *et al.*, 2001). Younes et al mettent en évidence que cette baisse de fréquence est causée par un défaut de sécrétion d'IL2 engendrant une absence de prolifération des LTs CD4⁺ spécifiques du VIH (Younes *et al.*, 2003). L'absence de LTs CD4⁺ spécifiques du VIH ayant des capacités prolifératives est une caractéristique de l'infection progressive à VIH (Wahren *et al.*, 1987; Berzofsky *et al.*, 1988; Krowka *et al.*, 1989; Porichis and Kaufmann, 2011). Un LT CD4⁺ activé exprime les marqueurs moléculaires d'activation CCR5 et CD38, ainsi que le marqueur de prolifération cellulaire Ki-67 (Zaunders *et al.*, 2005; Sauce, Elbim and Appay, 2013). Sans traitement, la fréquence des LTs CD4⁺ activés détectés en phase aiguë d'infection diminue (Oxenius *et al.*, 2000; Rosenberg *et al.*, 2000). Boritz et al ont mis en évidence une corrélation négative entre la charge virale et les capacités prolifératives des LTs CD4⁺ spécifiques du VIH (Boritz, Palmer and Wilson, 2004). Le traitement antirétroviral permet le maintien et la restauration des capacités prolifératives des LTs CD4⁺ (Autran *et al.*, 1997; Weiss *et al.*, 1999).

La présence des LTs CD4⁺ spécifiques du VIH est essentielle pour la génération de réponses T CD8⁺ effectrices et mémoires ainsi que pour la génération de réponses lymphocytaires B (Janssen *et al.*, 2003; Juno *et al.*, 2017). Dans le modèle simien, Li et al ont montré que le recrutement des TCD8⁺ à la muqueuse est corrélé à la proportion de LTs CD4⁺ et que le ratio CD8/CD4 représente un bon pronostic de l'infection (Li *et al.*, 2009). La présence de Tfh spécifiques du VIH est essentielle au développement d'anticorps. En effet, la proportion de Tfh dans les organes lymphoïdes secondaires est corrélée avec la production d'anticorps neutralisants à large spectre (Locci *et al.*, 2013; Martin-Gayo *et al.*, 2017). Chez le singe,

Havenar-Daughton et al ont également montré que le développement d'anticorps neutralisants est lié à la quantité de Tfh dans les centres germinatifs (Havenar-Daughton *et al.*, 2016). Néanmoins, la balance bénéfiques/risques de la forte proportion de Tfh est sensible car les Tfh sont une cible privilégiée de l'infection chez le macaques et l'humain. L'infection à VIH provoque une modification de l'architecture des ganglions lymphatiques et un dysfonctionnement des Tfh dont les mécanismes restent à élucider (Tangye *et al.*, 2013; Rouers, Jeger-Madiot, *et al.*, 2017; Jeger-Madiot, Heredia and Graff-Dubois, 2019).

La réponse CD4 cytotoxique

Mis en évidence dans les années 1980, les LTs CD4⁺ cytotoxiques (CTLs) sont une sous-population de LTs CD4⁺ caractérisée par la sécrétion de perforines et granzyme B, et par la répression et l'expression des facteurs de transcription ThPOK et Runx3, respectivement (Appay *et al.*, 2002; Mucida *et al.*, 2013; Reis *et al.*, 2013; Juno *et al.*, 2017). ThPOK est un facteur de transcription exprimé continuellement dans les LTs CD4⁺ du thymus puis dans les LTs CD4⁺ matures pour inhiber les gènes cytotoxiques de la lignée CD8⁺, eux-mêmes régulés par le facteur de transcription Runx3 (Cheroutre and Husain, 2013; Mucida *et al.*, 2013; Reis *et al.*, 2013). La population T CD4⁺ cytotoxique a d'abord été considérée par la communauté scientifique comme un artefact causé par la culture *in vitro* des LTs, et/ou à une mauvaise identification de LTs CD8⁺ (Juno *et al.*, 2017). Les D^{rs} Heinkelein, Suni et Appay ont été les premiers à mettre en évidence l'existence *ex vivo* de LTs CD4⁺ cytotoxiques spécifiques du VIH, du HCMV et de l'EBV (Heinkelein *et al.*, 1996; Suni *et al.*, 2001; Appay *et al.*, 2002). Appay et al ont mis en évidence une fréquence des LTs CD4⁺ CTLs plus élevée chez les individus infectés que chez les individus séronégatifs. Les LTs CD4⁺ CTLs spécifiques du VIH sont détectés à tous les stades d'infection mais semblent plus abondants chez les individus infectés chroniquement et non traités. Les LTs CD4⁺ CTLs identifiés expriment la perforine, le granzyme A, le TNF α , l'IFN γ , et GMP-17/TIA-1, et semblent avoir un phénotype mémoires (CD27 et CD28 négatives). En concordance avec ces résultats, Zaunders et al ont montré un phénotype cytotoxique des LTs CD4⁺ CCR5⁺ chez les individus non progresseurs infectés par le VIH ou le HCMV (Zaunders *et al.*, 2004). Un an plus tard, la même équipe a détecté chez des individus non traités en tout début de primo-infection, une fréquence élevée de LTs CD4⁺ fortement activés (CCR5⁺, Ki-67⁺, et CD38⁺) qui expriment également la perforine et le granzyme B (Zaunders *et al.*, 2005). Les LTs CD4⁺ CTLs sont aussi caractérisés par l'expression du marqueur de dégranulation CD107a et de LAMP-1 (Juno *et al.*, 2017). L'activité cytotoxique des LTs repose sur la libération des granules lytiques cytoplasmiques, composées de perforines et de granzymes, de manière polarisée vers la cellule cible ; ce mécanisme est appelé dégranulation (Betts and Koup, 2004). Récemment, Collora et al ont mis en évidence, par combinaison de diverses techniques en « single-cell », que sous traitement antirétroviral, le clone de LTs CD4⁺ infecté le plus représenté possède un phénotype CTL mémoire avec l'expression du granzyme B (Collora *et al.*, 2022). L'ensemble de ces études soulignent l'importance de la réponse cytotoxique médiée par les LTs CD4⁺ dans la lutte contre l'infection à VIH.

Malgré l'ensemble de ces travaux, les phénotypes et la dynamique de différenciation des LTs CD4⁺ au cours de l'infection à VIH restent peu documentés. Les cellules T CD4⁺ spécifiques du VIH semblent avoir majoritairement un profil Th1 avec la sécrétion d'IFN γ , un profil Tfh, et un

profil CTL. Plusieurs études montrent également l'induction d'un profil Th17 en phase aiguë d'infection qui se perd en phase chronique (Brenchley *et al.*, 2008; MacParland *et al.*, 2016). Les réponses T CD4⁺ spécifiques du VIH ciblent des épitopes dérivés de l'ensemble des gènes du VIH (Ranasinghe *et al.*, 2012).

3.7.2.2 La réponse lymphocytaire T CD8⁺ anti-VIH

En primo-infection, l'apparition des LTs CD8⁺ corrèle directement avec la diminution de la charge virale (Borrow *et al.*, 1994; Koup *et al.*, 1994; Yang *et al.*, 1997; Chen *et al.*, 2009; Ndhlovu *et al.*, 2015). Il a été observé que la pression immunitaire exercée par les LTs CD8⁺ entraîne une diminution de la charge virale d'en moyenne d'un facteur 100 (Coffin and Swanstrom, 2013). La conséquence de cette pression de sélection est l'apparition très rapide de variants d'échappement situés au sein des épitopes reconnus par les LTs CD8⁺ lors de la phase aiguë (Goonetilleke *et al.*, 2009; Liu *et al.*, 2013). L'étude de la réponse T CD8⁺ dans le modèle primate non humain a montré une très forte augmentation de la virémie suite à la déplétion des LTs CD8⁺ en phase aiguë et chronique d'infection (Schmitz *et al.*, 1999).

Lors de la phase chronique on observe chez les individus non traités, une diminution de la fréquence des LTs CD8⁺ liée à une réduction de l'expression des antigènes circulants du VIH (Jones and Walker, 2016). L'exposition continue des antigènes résiduels du VIH durant la phase chronique provoque une altération et l'épuisement de la réponse T CD8⁺.

Durant les deux premières années qui suivent la mise sous traitement des personnes infectées, la fréquence des LTs CD8⁺ spécifiques du VIH diminue (Gray *et al.*, 1999; Kalams *et al.*, 1999; Ogg *et al.*, 1999; Casazza *et al.*, 2001; Addo *et al.*, 2003). Les réponses restent néanmoins détectables chez la majorité des individus même après plusieurs années de traitement. Le TARV permet une diminution de l'expression des récepteurs d'inhibition exprimés par les LTs CD8⁺ et une réduction de la fréquence des LTs CD8⁺ pro-apoptotiques (Rinaldo *et al.*, 2000; Rehr *et al.*, 2008). Néanmoins, Migueles et al ont montré que les capacités prolifératives et effectrices ne sont pas restaurées après l'initiation du traitement (Migueles *et al.*, 2002, 2008). Plusieurs études ont démontré que seule une initiation précoce en phase aiguë d'infection permet de restaurer les capacités prolifératives et une activité cytotoxique *in vitro* (Oxenius *et al.*, 2000; Takata *et al.*, 2022). D'autres études observent également une restauration de la polyfonctionnalité et une diminution partielle de l'expression des molécules associées à l'épuisement immunitaire après une longue période de traitement (Rehr *et al.*, 2008; Grabmeier-Pfistershammer *et al.*, 2011).

3.8 Les différents statuts cliniques

Les prédispositions génétiques, la capacité du système immunitaire à contrôler la réplication virale, et l'aptitude virale sont d'autant plus de facteurs qui influencent l'évolution de la maladie. Plusieurs statuts cliniques sont observés chez les individus infectés. On distingue : les progressseurs qui, sans traitement antirétroviral (TARV), progressent dans la maladie, et les non progressseurs qui ont une charge virale faible et un taux de LTs CD4⁺ élevé et stable pendant une longue période sans TARV. Parmi ce dernier groupe, les « Elite Controllers » (EC) sont les

individus qui maintiennent une charge virale en dessous des limites de détection et qui ne progressent pas vers le stade SIDA.

3.8.1 Les Elite Controllers

Découvert en 2004, les Elite Controllers (EC) forment un groupe restreint qui représente 1% des individus infectés par le VIH (Lambotte *et al.*, 2005; Deeks and Walker, 2007; Okulicz *et al.*, 2009). Ces individus sont cliniquement caractérisés par : une infection à VIH connue depuis plus de dix ans, une charge virale indétectable équivalente à < 400 copies/ml dans 90 % des mesures, et l'absence de TARV (Lambotte *et al.*, 2005; Lambotte, 2012). Les EC sont un groupe d'individus qui s'approche le plus d'une guérison fonctionnelle. Dans les stratégies de thérapies curatives ou prophylactiques, de nombreux efforts ont été déployés pour identifier les facteurs et déterminants liés au contrôle du VIH. Les différentes caractéristiques virologiques, génétiques et immunologiques des EC sont listées ci-dessous.

3.8.1.1 Caractéristiques virologiques

La plupart des EC sont infectés par des virus compétents pour la réplication mais les individus parviennent sans TARV à maintenir une virémie indétectable et stable au cours du temps (Lambotte *et al.*, 2005; Grabar *et al.*, 2009). Malgré une charge virale indétectable, il a été démontré par l'utilisation d'une méthode ultrasensible de quantification, une réplication virale à bas bruit chez les EC (Pereyra *et al.*, 2009). La réplication virale n'est donc pas inexistante chez les EC même si la charge virale est indétectable. Des études ont observé chez certains individus un rebond viral temporaire, appelé *blips*, associé à une diminution du taux de LTs CD4⁺ (Boufassa *et al.*, 2011; Rosás-Umbert *et al.*, 2019). L'évolution immunovirologique à long terme des EC est donc incertaine (Lambotte, 2012). La taille du réservoir est plus faible chez les EC que chez les individus infectés virémiques (Lambotte *et al.*, 2005).

3.8.1.2 Caractéristiques génétiques

Il a été rapidement mis en évidence que certains allèles HLA-I sont associés à des évolutions cliniques distinctes. On distingue les HLA-I protecteurs associés à une évolution de la maladie plus lente et les HLA-I associés à la progression. Parmi les EC, les allèles protecteurs HLA-B*2705, -B*5701, et B*5703 sont surreprésentés (Miguelles *et al.*, 2000; Lambotte *et al.*, 2005). Les allèles HLA-B*27 et -B*57 ont été associés à un contrôle de la réplication virale et à une persistance des fonctions effectrices des lymphocytes T au cours du temps (Pereyra *et al.*, 2010; Thèze *et al.*, 2011; Jones and Walker, 2016). L'expression de l'allèle HLA-B*57 a également été corrélée à la présence de cellules B sécrétrices d'IgG1 et à une activité neutralisante des anticorps (Rouers, Klingler, *et al.*, 2017).

3.8.1.3 Caractéristiques immunologiques

La réponse lymphocytaire T

La réponse T CD8⁺ des EC présente plusieurs caractéristiques associées au contrôle viral. Tout d'abord, suite à l'exposition à l'antigène les cellules T CD8⁺ spécifiques du VIH ont des capacités prolifératives et effectrices accrues en comparaison à la réponse T CD8⁺ observée chez les individus sous traitement (Sáez-Ciri6n *et al.*, 2007). En effet, les LTs CD8⁺ des EC ont une capacité supérieure à lyser les cellules infectées *in vitro* lors de coculture avec des LTs CD4⁺

infectés autologues. Cette capacité de lyse est dépendante de la sécrétion élevée de perforine et granzyme, dont l'expression est régulée par le facteur de transcription T-bet (Hersperger *et al.*, 2010). Deuxièmement, la réponse T CD8⁺ chez les EC a été démontrée ayant une polyfonctionnalité supérieure à celle des individus progressseurs (Betts *et al.*, 2006; Sáez-Ciri6n *et al.*, 2007; Migueles *et al.*, 2008; Ferre *et al.*, 2009; Hersperger *et al.*, 2010). La polyfonctionnalité est la capacité d'un lymphocyte T à produire plusieurs cytokines ou chimiokines simultanément telles que l'IL2, l'IFN γ , le TNF α , Mip1 β et CD107a. La polyfonctionnalité est un critère clé dans la pr6diction de l'6volution de la maladie puisque la fr6quence des r6ponses T CD8⁺ ayant cinq fonctions (= s6cr6tions de 5 cytokines) est inversement corr6l6e avec la charge virale chez les individus progressseurs (Betts *et al.*, 2006). En d'autres termes, la capacit6 6 supprimer la r6plication *in vitro* et la pr6sence d'une r6ponse T CD8⁺ polyfonctionnelle a 6t6 d6montr6 comme 6tant associ6e 6 un contr6le de l'infection.

Des exp6riences *in vitro* et *ex vivo* visant 6 caract6riser la r6ponse T CD8⁺ sp6cifique de l'6pitope immunodominant KK10 chez les EC ont montr6 une corr6lation entre l'avidit6 du TCR et la polyfonctionnalité (Almeida *et al.*, 2007, 2009; Iglesias *et al.*, 2011). L'avidit6 du TCR refl6te une efficacit6 fonctionnelle des clones LTs CD8⁺ en pr6sence d'une faible concentration d'antig6ne. Des exp6riences sur le SIV ont 6galement confirm6 l'importance de l'avidit6 du TCR dans le contr6le de l'infection (Osuna *et al.*, 2014). L'avidit6 du TCR est donc 6galement un crit6re de pr6diction de la qualit6 fonctionnelle des LTs (Lissina *et al.*, 2016).

De fa6on int6ressante, le r6pertoire d'6pitope reconnu par les TCRs semblent 6galement varier entre les EC et les progressseurs. La r6ponse T CD8⁺ des EC ciblent pr6f6rentiellement des 6pitopes d6riv6s de Gag et ont une r6activit6 crois6e accrue, c'est-6-dire qu'un seul TCR est capable de reconna4tre plusieurs 6pitopes issus de mutations d'6chappement (Turnbull *et al.*, 2006; Pereyra *et al.*, 2008; Miura *et al.*, 2009; S6ez-Ciri6n *et al.*, 2009; Chen *et al.*, 2012; Mothe *et al.*, 2012; Lissina *et al.*, 2016).

Les efforts mis en 6uvre pour caract6riser la r6ponse lymphocytaire T chez les EC ont permis de renforcer les preuves que la r6ponse T CD4⁺ joue un r6le fondamental dans la r6ponse immunitaire anti-VIH. En effet, la pr6sence de LTs CD4⁺ est associ6e 6 un contr6le de la r6plication virale et 6 une meilleure progression de la maladie (Rosenberg *et al.*, 1997; Pitcher *et al.*, 1999; Chevalier *et al.*, 2011; Schieffer *et al.*, 2014). 6 l'image de la r6ponse T CD8⁺, la r6ponse T CD4⁺ chez les EC pr6sente une polyfonctionnalité plus 6lev6e, une avidit6 des TCRs plus forte et cible majoritairement des 6pitopes issus de Gag (Rosenberg *et al.*, 1997; Betts *et al.*, 2001; Dyer *et al.*, 2008; Ferre *et al.*, 2010; Vingert *et al.*, 2010; Schieffer *et al.*, 2014; Benati *et al.*, 2016; Lissina *et al.*, 2016; Galperin *et al.*, 2018; Kervevan and Chakrabarti, 2021). De plus, les LTs CD4⁺ retrouv6s chez les EC ont des capacit6s prolif6ratives accrues et une diminution de l'expression des r6cepteurs de co-inhibition (Kervevan and Chakrabarti, 2021). Les LTs CD4⁺ des EC produisent de l'IL2, ce qui leur permet un maintien de leur capacit6 prolif6rative n6cessaire au contr6le de la r6plication (Younes *et al.*, 2003; Harari *et al.*, 2004; Potter *et al.*, 2007).

Dans leur 6tude Ferre et al mettent en 6vidence que les EC porteurs des all6les HLA-DQB1*13 et -DQB1*06 poss6dent des r6ponses T CD4⁺ polyfonctionnelles au niveau des muqueuses (Ferre *et al.*, 2010). En accord avec ces observations, l'6quipe du Dr Chakrabarti met 6galement

en évidence que la réponse T CD4⁺ des EC dirigée contre Gag est caractérisée par une fréquence plus élevée de sécrétion d'IFN γ et de perforine, d'expression de CD107a, et par une polyfonctionnalité (Vingert *et al.*, 2012). Ce phénomène est observé chez les EC malgré une exposition antigénique très faible tandis que les LTs CD4⁺ d'individus virémiques montrent une perte de fonction. En effet, la réponse T CD4⁺ CTL CD107a⁺ est plus fréquente chez les EC, et elle est corrèle avec un viral set point plus bas (Johnson *et al.*, 2015).

L'avidité des TCRs pourrait expliquer la magnitude plus élevée des réponses T CD4⁺ chez les EC (Vingert *et al.*, 2010). Ranasinghe et al mettent en évidence que la source protéique d'épitopes ciblée par les LTs CD4⁺ diffère selon le statut clinique (Ranasinghe *et al.*, 2012). Les EC ont une fréquence environ six fois plus élevée de LTs CD4⁺ spécifiques de peptides issus de Gag comparée aux patients très virémiques. A l'inverse, une fréquence élevée de réponses T CD4⁺ dirigées contre des peptides issus de Env est associée à un profil d'individus très virémiques.

Récemment, l'immunométabolisme a été montré comme un déterminant dans la pathogénèse du VIH. La plasticité du métabolisme des LTs CD4⁺ et macrophages leur confère une restriction d'infection. Dans la population LTs CD8⁺ mémoires, la plasticité du métabolisme permet une activité soutenue des cellules par l'adaptation aux conditions restrictives de réplication telles que le stress métabolique, une stimulation antigénique persistante, et des signaux négatifs tels que les espèces réactives de l'oxygène (ROS) (Sáez-Cirión and Sereti, 2021).

La réponse lymphocytaire B

La réponse humorale est hétérogène parmi les EC (Moris, Pereira and Chakrabarti, 2019b). Certains individus possèdent des titres faibles d'anticorps neutralisants et une baisse de la neutralisation croisée en comparaison aux individus virémiques (Bailey *et al.*, 2006; Pereyra *et al.*, 2008, 2009; Lambotte *et al.*, 2009). A l'inverse, la présence d'anticorps ayant une forte capacité de neutralisation croisée ou d'anticorps neutralisants à large spectre (bNAbs) ont également été observées chez des EC (Scheid *et al.*, 2009, 2011).

Récemment, les travaux des D^{rs} Rouers et Buckner ont mutuellement démontré que l'une des caractéristiques de la réponse humorale des EC est le maintien des différentes sous-population de LTs B mémoires spécifiques du VIH malgré une stimulation antigénique limitée (Buckner *et al.*, 2016; Rouers, Klingler, *et al.*, 2017). Dans le modèle simien, la réplication virale à bas bruit observée chez les animaux controllers a été décrite comme ayant lieu dans les centres germinatifs des organes lymphoïdes (Fukazawa *et al.*, 2015). Il est donc possible que la persistance des sous-populations B repose sur la proximité de l'antigène aux sites de production des anticorps. Le maintien des sous-populations a été montré dépendant des fonctions des Tfh, qui sont préservées chez les EC (Buranapraditkun *et al.*, 2017). Dans leur étude de 2017, Rouers et al mettent en évidence chez des individus porteurs de l'allèle protecteur HLA-B*57, une corrélation entre la fréquence des cellules sécrétrices d'anticorps et la capacité à neutraliser différents variant, y compris les TF (Rouers, Klingler, *et al.*, 2017).

D'autres études ont ensuite mis en évidence une différence dans les isotypes des IgG sécrétées entre les progressseurs et les EC, chez qui le profil d'anticorps est enrichi en IgG2 et IgG3

(Sadanand *et al.*, 2018). Il a également été observé que les sérums des EC présentant de fortes activités antivirales polyfonctionnelles sont enrichis en IgG1 et IgG3 (Ackerman *et al.*, 2016). Ces observations suggèrent une forte affinité de liaison des anticorps IgG1 et IgG3 spécifiques de l'enveloppe aux récepteurs Fc. Néanmoins, l'implication du phénomène d'ADCC (« antibody dependent cell cytotoxicity ») dans le contrôle de l'infection reste à ce jour controversé (Moris, Pereira and Chakrabarti, 2019b).

3.9 Les stratégies thérapeutiques

3.9.1 Le traitement antirétroviral

3.9.1.1 Les antirétroviraux

Les TARV ont permis aux PVVIH d'avoir une meilleure espérance de vie tout en restaurant et préservant leur immunité. Les TARV ont transformé l'infection à VIH en une maladie chronique. De plus, les TARV sont également importants pour limiter la diversité génétique du virus. Lorsqu'ils sont combinés à une bonne observance, les TARV permettent de supprimer la charge virale et le risque de transmission du VIH.

En 1985, deux ans après l'identification du virus, le premier agent thérapeutique est testé dans le cadre d'un essai clinique. L'HPA23 est un antirétroviral (ARV) qui agit sur la transcription inverse de rétrovirus murins. Malheureusement, l'essai thérapeutique n'est pas concluant et les études se focalisent alors sur la zidovidine aussi connue sous le nom d'AZT. L'AZT est un inhibiteur nucléosidique de la transcriptase inverse (INTI). Puis vient le tour de la didanosine qui est également un INTI. Avant les années 1990 la stratégie thérapeutique consistait à utiliser des ARV en monothérapie. Toutes fois, même avec l'introduction de la bithérapie en 1994, ces stratégies thérapeutiques n'empêchaient pas la réplication virale puisqu'un rebond est observé après 24 semaines. L'année 1996 fut une année charnière dans le traitement contre le VIH avec la découverte de la trithérapie qui dévoile des résultats très encourageants. La trithérapie combine trois antirétroviraux, dont une classe de molécules innovantes : les inhibiteurs de protéases (IP). Après une quarantaine d'années de progrès constants, une dizaine d'ARVs ont été approuvés par la Food and Drug Administration (HIVinfo.org). On distingue plusieurs types d'ARV : les INTI, les IP, les inhibiteurs de l'intégrase (INI), les inhibiteurs non nucléosidiques de la transcriptase inverse (INNTI), les inhibiteurs de fusion (IE), et les inhibiteurs du CCR5. La trithérapie, aussi appelée thérapie antirétrovirale combinée (cART) est la combinaison de deux INTI avec un IP, ou un INNTI, ou bien un INI. Une charge virale inférieure à 50 copies/mL est le seuil international reconnu d'une trithérapie efficace.

3.9.1.2 L'échec thérapeutique

L'échec thérapeutique se définit par une perte d'efficacité du TARV. L'échec peut être virologique si la charge virale augmente, immunologique si le nombre de LTs CD4⁺ diminue, ou clinique suite à l'apparition de maladies opportunistes.

L'échec virologique est le plus précoce, survenant avant l'échec immunologique et clinique. Il est facilement détectable par une analyse de la charge virale. L'échec virologique est diagnostiqué si le seuil de la charge virale plasmatique est supérieur à 1 000 copies/mL (*World*

Health Organization (WHO), no date). Il existe plusieurs facteurs qui favorise la survenue de l'échec virologique.

L'observance est le premier facteur évalué suite à la découverte d'un échec virologique. En effet, la diminution soudaine de la concentration du TARV dans l'organise favorise l'émergence de variants résistants aux drogues. Des études sociologiques montrent que les facteurs d'inobservance sont multiples et évolutifs. On retrouve : le jeune âge, l'absence de partage du statut sérologique avec le conjoint, l'état dépressif etc. Dans les pays du Sud, il est favorisé par une difficulté d'accès aux soins, l'insécurité alimentaire, et la peur de la stigmatisation.

L'échec virologique peut également être détecté chez des individus observants. Dans ce cas de figure, les cliniciens étudient les caractéristiques du virus et des drogues. Lors de la primo-infection et avant tout traitement, il existe des mutations de résistances aux ARV que l'on appelle résistances primaires. L'analyse de ces résistances au sein des populations permet de faire évoluer le traitement de première ligne en inter-changeant les ARV. La conséquence finale de l'accumulation de résistance est la diminution des options thérapeutiques.

3.9.2 Le traitement prophylactique

L'année 2004 signe la révolution de la prévention du VIH avec l'apparition de PrEP aussi appelée prophylaxie de préexposition (« pre-exposure prophylaxis »). La PrEP est le premier outil de prévention à avoir montré son efficacité. La PrEP consiste en la prise d'ARVs pour prévenir l'infection chez des individus non infectés très exposés. Cette population peut inclure des hommes ayant des rapports sexuels avec des hommes (HSH), des personnes ayant des partenaires infectés (couples sérodiscordants), des travailleurs.ses du sexe et les usagers de drogues injectables. Lorsque l'observance est respectée, la PrEP est extrêmement efficace avec une réduction des risques de 99% chez les hommes et de 90% chez les femmes (Anderson *et al.*, 2012; Baeten *et al.*, 2012). La PrEP, commercialisée sous le nom de Truvada ou Descovy, se présente sous la forme d'un seul comprimé combinant deux INTI : le téfovofir et l'emtricitabine (Nicol, Adams and Kashuba, 2013). Ces INTI ont été choisis pour leur efficacité dans le modèle animal et leur pharmacocinétique favorable (Cranage *et al.*, 2008; Denton *et al.*, 2008; García-Lerma *et al.*, 2008; Parikh *et al.*, 2009; Anderson *et al.*, 2012). Initialement, la PrEP a été recommandé en prise continue. La protection par prise intermittente de la PrEP dans la population HSH a été démontrée en 2016 par l'essai iPergay en France (*ANRS / Résultats finaux de l'essai ANRS IPERGAY*, 2016). Pour les femmes et les femmes trans, le schéma recommandé reste à ce jour une prise en continue.

3.9.3 La vaccination VIH

Plusieurs obstacles sont rencontrés dans l'élaboration d'un vaccin efficace contre le VIH. Premièrement, le VIH possède une grande diversité qui se traduit par un nombre élevé de variants intra et inter-individus. L'établissement de réservoirs viraux et l'échappement au système immunitaire sont également des freins au développement d'un agent thérapeutique efficace. Une deuxième problématique rencontrée lors du développement d'un vaccin VIH est la difficulté d'exécution des tests d'efficacité qui est liée au manque de groupes « contrôle » où les individus ne peuvent pas être sans protection.

Différents types de vaccins ont été testés depuis le début des premiers essais vaccinaux en 1990 : les protéines recombinantes de l'enveloppe, les bNAbs, les vaccins vectorisés, les vaccins à ARN et mosaïques (Kim *et al.*, 2021). Malgré un nombre conséquent d'essais, l'essai RV144 conduit en Thaïlande est la seule étude à avoir démontré une efficacité à 31% à 3,5 ans après immunisation (Haynes *et al.*, 2012). L'essai RV144 fait appel à la combinaison de deux vaccins « ALVAC-HIV » et « Aidsvax » administrés en utilisant le principe du « prime-boost ». Cette stratégie permet de renforcer la réponse immunitaire induite par « ALVAC », vecteur recombinant de la variole du canari, par l'injection de protéines recombinantes de l'enveloppe « Aidsvax ». L'immunisation induit une réponse T CD4⁺ polyfonctionnelle spécifique de l'enveloppe, des anticorps IgG ayant des fonctions d'ADCC (Kim *et al.*, 2021).

3.9.4 Les stratégies curatives

Au vu de la complexité de la pathogénèse du VIH, les stratégies de guérison doivent passer par une double approche : réduire les réservoirs et renforcer les barrières de l'hôte (Deeks *et al.*, 2012). Cette double approche est décomposée en trois grands axes : 1) cibler les réservoirs, 2) induire des mécanismes optimaux de contrôle, et 3) limiter l'inflammation chronique.

3.9.4.1 Cibler les réservoirs

Deux stratégies ont vu le jour pour cibler les réservoirs : le « shock and kill » et le « block and lock ».

Le shock and kill

La stratégie du shock and kill vise à lever la latence du virus dans les cellules quiescentes et à les éliminer. La grande variabilité des sites d'intégration du provirus et la difficulté à définir des marqueurs spécifiques des cellules quiescentes sont les principaux obstacles rencontrés dans cette stratégie. L'utilisation de super-agonistes tels que l'IL15, et le TLR7 a récemment montré des résultats prometteurs *ex vivo* et *in vivo* dans l'induction de la transcription du VIH et l'augmentation de l'expression des gènes viraux ainsi que dans l'activation des acteurs de l'immunité adaptative dans les organes lymphoïdes (Du *et al.*, 2018; Garrido *et al.*, 2018; Webb *et al.*, 2018; Miller *et al.*, 2022). C'est une piste prometteuse dans la stratégie « shock and kill » bien que les marqueurs des cellules quiescentes infectées soient encore à déterminer. Les marqueurs PD1, CD32a, OX40 pourraient être des marqueurs indicateurs, tout comme les marqueurs anti apoptotique des cellules infectées.

Le block and lock

A l'inverse du shock and kill, le block and lock a pour but de verrouiller la latence (Ahlenstiel *et al.*, 2020). Là encore, deux approches sont utilisées pour empêcher la réactivation du provirus intégré : la répression du provirus par des agents bloquant la transcription et l'élimination du virus par la modification des gènes. La stratégie du block and lock apparaît comme davantage prometteuse puisqu'elle permet une inhibition de la réplication virale mais des zones d'ombres subsistent concernant la durée du traitement et la génotoxicité. Cette stratégie est également beaucoup mieux acceptée chez les PVVIH que le shock and kill.

Dans le cadre du "shock and kill" ou du "block and lock", les mêmes problèmes persistent : la difficulté à identifier les cellules du réservoir, à accéder aux sanctuaires et la présence de nombreux sites d'intégration.

3.9.4.2 Renforcer les barrières de l'hôte

Diverses stratégies sont utilisées pour renforcer les barrières de l'hôte telles que : l'introduction de résistance intrinsèque, les vaccins thérapeutiques, les modulateurs immuns, et la thérapie cellulaire adoptive.

Le développement d'anticorps à double affinité (« Dual-Affinity Re-targeting Molecules », DART) permet une cytolysse des cellules infectées par le VIH. Les DARTs sont des anticorps bispécifiques qui combinent une chaîne spécifique aux antigènes de la protéine d'enveloppe et une chaîne spécifique de la molécule CD3 (Ferrari *et al.*, 2017). Cette double spécificité permet la reconnaissance de l'antigène Env présenté à la surface des cellules infectées et l'engagement simultané des LTs CD8⁺ par le récepteur CD3 de façon indépendante des molécules du CMH. Plusieurs études *in vitro* et *in vivo* ont montré l'efficacité des DARTs dans l'élimination des LTs CD4⁺ infectés médiée par les LTs CD8⁺ (Sloan *et al.*, 2015; Tuyishime *et al.*, 2021).

La découverte des anticorps neutralisants à large spectre (bNAbs) en 2009 a ouvert de nouvelles perspectives et entraîné des progrès significatifs dans la lutte contre l'infection à VIH. D'un point de vue prophylactique, l'immunothérapie passive par les bNAbs a démontré son efficacité dans des modèles murins et simiens par un contrôle de la charge virale à des niveaux indétectables et par une potentialisation de la réponse LTs CD8⁺ et une amélioration des réponses LTs CD4⁺ spécifiques du virus. Plusieurs bNAbs ont fait preuve de leur efficacité en essais cliniques thérapeutiques chez l'Homme. C'est notamment le cas des bNAbs 3BNC117, 10-1074, et N6 (Julg and Barouch, 2021). L'utilisation de bNAbs en monothérapie a souvent conduit à l'émergence de variants résistants, ce qui a suscité l'idée de combiner des bNAbs ciblant différents sites de vulnérabilité de l'enveloppe pour réduire les risques d'échappement. Plusieurs études démontrent la nécessité de combiner l'administration de deux ou plusieurs bNAbs afin d'étendre le contrôle de la charge virale à plusieurs mois et d'éviter le rebond viral lié à l'émergence de virus résistants (Shingai *et al.*, 2013; Bar-On *et al.*, 2018; Mendoza *et al.*, 2018). Plusieurs bNAbs à triple spécificité sont actuellement en cours d'évaluation chez l'Homme. La diversité des séquences de l'enveloppe parmi les différents clades du VIH reste l'obstacle majeur de la thérapie par les bNAbs. En plus d'exister entre les individus, l'hétérogénéité virale existe également entre les tissus complexifiant l'utilisation des bNAbs (Chaillon *et al.*, 2020). La stratégie des bNAbs représente une piste très prometteuse pour l'établissement d'un vaccin curatif contre le VIH. Il reste à déterminer la combinaison idéale et le nombre d'épitopes cibles requis pour éviter l'échappement viral (Julg and Barouch, 2021).

3.9.4.3 « HIV cure » et « Cancer cure » : des divergences et similitudes

Le VIH et le cancer présentent des caractéristiques communes telles que l'exposition chronique aux antigènes, l'épuisement des lymphocytes T, l'échappement au système immunitaire, la latence, et la mortalité en absence de traitement (Vigano *et al.*, 2020).

Les similitudes entre le cancer et l'infection à VIH ont permis de décliner les thérapies prometteuses du cancer au VIH. La stratégie « shock and kill » vient d'ailleurs de la cancérologie. Les oncologues ont vite utilisé des approches combinatoires pour atteindre la rémission. Les inhibiteurs de points de contrôle PD-1, PDL-1 et CTLA-4 ont été une révolution en cancérologie avec des premiers essais cliniques en 2010 très concluants. Ces inhibiteurs restaurent la réponse immune en bloquant la liaison du récepteur PD-1 exprimé par les lymphocytes T à leur ligand PDL-1 exprimé par les cellules tumorales. Les anticorps monoclonaux ciblant les points de contrôle ont un taux de réponses élevé en situation de maladie disséminée. L'utilisation d'un inhibiteur de point de contrôle PD-1 dans le VIH se montre moins efficace que dans le cancer, avec une plus forte toxicité et un effet modéré mais reste une stratégie prometteuse à combiner avec d'autres (Boyer and Palmer, 2018).

L'utilisation des thérapies cellulaires ont également inspiré le domaine du VIH avec le développement de la reprogrammation des cellules T CD8⁺ en T CD8⁺ ayant des propriétés associées au contrôle naturel de l'infection, ou en utilisant des CAR-T cells (Qi *et al.*, 2020; Perdomo-Celis *et al.*, 2022).

Il est très probable que l'efficacité de guérison du VIH découle de la combinaison de plusieurs stratégies thérapeutiques, tant dans l'approche utilisée que dans les acteurs de l'immunité ciblés (cellulaire et humoral). C'est le cas notamment des travaux des D^{rs} Walker-Sperling et Barouch qui montrent un contrôle virologique suite à l'injection de la triple combinaison de ligand de TLR7, de bNAbs, et de vaccin (Moldt *et al.*, 2022; Walker-Sperling *et al.*, 2022).

La découverte des neoantigènes a permis l'émergence de nouvelles immunothérapies du cancer (Brennick *et al.*, 2017; Verdegaal and van der Burg, 2017; Chen *et al.*, 2021). Récemment, les neoantigènes ont été utilisés sous forme de vaccins peptidiques, ARN, ADN, de vaccin basé sur les DCs et d'immunothérapies adoptives à base de TCR transgéniques (Bassani-Sternberg and Coukos, 2016; Laumont and Perreault, 2018; Arnaud *et al.*, 2021; Chen *et al.*, 2021). Dans le cancer, la majorité du protéome dérive de régions non codantes et ces produits alternatifs de traduction ont un rôle en tant que source antigénique (Ruiz Cuevas *et al.*, 2021). Une étude publiée en 2021 démontre même un rôle fonctionnel des ORFs non canoniques dans la survie des cellules cancéreuses. Dans le cancer du sein, l'ORF alternatif GREP1 code pour une protéine sécrétée qui favorise l'expression de la cytokine oncogène GDF15 (Prensner *et al.*, 2021). De plus, 99% des mutations spécifiques aux cellules cancéreuses ont lieu dans les régions non codantes (Khurana *et al.*, 2016).

Toutes ces études justifient l'intérêt d'étudier le génome non codant et les événements de traduction non canoniques afin de mieux comprendre le développement des pathologies et pour mieux appréhender la conception des stratégies thérapeutiques. Dans la prochaine et dernière section de ce manuscrit, je détaillerai les données de la littérature sur l'existence des régions non codantes du VIH appelées cadres alternatifs de lectures.

3.10 Les ARFs du VIH

Jusqu'à présent, la découverte des cadres alternatifs de lecture (ARFs) du VIH a reposé sur des méthodes d'identification indirectes telles que l'immunologie inverse et la méthode HLA footprint.

La méthode HLA footprint consiste en une corrélation entre les mutations d'échappement et les molécules HLA de classe I. Les cellules infectées présentent par les molécules du CMH-I des peptides viraux aux LTs CD8⁺. La reconnaissance des peptides par les LTs CD8⁺ est restreinte par les molécules HLA. L'activation d'une réponse T CD8⁺ spécifique induit une pression de sélection qui provoque des mutations d'échappement au sein de l'épitope reconnu. Les mutations qui entraînent une absence de reconnaissance par les LTs CD8⁺ sont sélectionnées par le virus dans le cas où elles ne sont pas délétères à la réplication. Ces mutations sont caractérisées d'« empreintes » HLA (Goulder *et al.*, 1997; McMichael and Klenerman, 2002; Moore *et al.*, 2002; Martinez-Picado *et al.*, 2006; Schneidewind *et al.*, 2007; Matthews *et al.*, 2008, 2009). Les mutations d'échappement sont donc associées aux molécules HLA et certaines sont conservées au sein des souches circulantes, ce qui permet de prédire les mutations de certains épitopes (Leslie *et al.*, 2004; Schneidewind *et al.*, 2007; Matthews *et al.*, 2008).

Au total, neuf études ont identifié des ARFs dans le génome du VIH-1, et deux études se sont intéressées aux ARFs exprimés dans le génome du SIV (voir Annexe n°2) (Cardinaud *et al.*, 2004; Maness *et al.*, 2007, 2010; Garrison *et al.*, 2009; Bansal *et al.*, 2010, 2015; Berger *et al.*, 2010, 2015; Champiat *et al.*, 2012; Ternette *et al.*, 2016; Peng *et al.*, 2018).

3.10.1 Les ARFs identifiés au laboratoire

Notre équipe de recherche a été précurseur dans l'étude des ARFs du VIH. De précédentes études extérieures au laboratoire sur les virus, les maladies auto-immunes, et les cancers ont démontré l'existence d'ARFs ayant la capacité à coder des polypeptides immunogènes reconnus par les LTs CD8⁺. Des ARFs dérivés de la nucléoprotéine de la grippe ont été identifiés par les D^{rs} Bullock, Fetten, et Elliot, tandis que le peptide "SYNTGRFPPL", codé par un ARF du rétrovirus murin LP-BMP5, a été identifié par le D^r Mayrand (Fetten, Roy and Gilboa, 1991; Bullock and Eisenlohr, 1996; Elliott, Bodmer and Townsend, 1996; Mayrand, Schwarz and Green, 1998) (chapitre 2 section 2.5.2.1). Dans ce contexte, notre équipe a précédemment étudié si des polypeptides dérivés d'ARFs pouvaient être générés lors de l'infection à VIH et stimuler une réponse T. Les ARFs candidats ont été sélectionnés après analyse des six phases de lecture du génome viral et en fonction de la présence de codons start AUG et de motifs Kozak, qui suggéraient une expression potentielle des ARFs (Cardinaud *et al.*, 2004). Selon ces critères de sélection Cardinaud *et al.* ont identifié 3 ARFs du VIH localisés dans les gènes *gag*, *pol* et *env*. Les auteurs ont ensuite évalué la capacité des peptides issus des ARFs à fixer les molécules HLA-B*0702 en utilisant l'algorithme SYFPEITHI. Au total, ils ont défini 6 peptides potentiellement présentés par les molécules HLA-B*0702. Parmi les peptides, un retiendra particulièrement leur attention par sa forte capacité de liaison, c'est le peptide « QPRSDTHVF » codé par un ARF localisé dans *gag*, qu'ils nommeront Q9VF. Les auteurs ont ensuite évalué la capacité des peptides à stimuler une réponse T CD8⁺ spécifiques *in vivo* en immunisant des souris transgéniques HLA-B*07. Les 6 peptides sélectionnés sont immunogéniques et induisent

une réponse T CD8⁺. Cette observation a été confirmée chez les PBMCs d'individus infectés, puisqu'une réponse T CD8⁺ a été mise en évidence contre chacun des peptides. L'introduction d'un codon stop au sein de l'ARF codant Q9VF abolit sa reconnaissance par les LTs CD8⁺. Pour conclure, les auteurs ont démontré que les ARFs identifiés sont conservés parmi les isolats du clade B de 24 à 80%, Q9VF étant conservé à 68%.

De façon très intéressante, une seconde étude menée au laboratoire a mis en évidence une mutation d'échappement dans l'épitope Q9VF (Cardinaud *et al.*, 2011). Un polymorphisme a été observé à la 5^{ème} position de l'épitope où l'asparagine est substituée par un résidu aspartate (Q9VF5D->Q9VF5N). Les deux peptides sont reconnus *ex vivo* par les LTs CD8⁺ d'individus infectés HLA-B*07. Environ 16% des individus testés ont présenté une réponse T CD8⁺ contre les épitopes Q9VF. Notre équipe a cherché à comprendre les mécanismes d'expression du variant Q9VF5N. En infectant des cellules T avec des isolats de laboratoire portant ou non la mutation, il a été démontré que l'épitope Q9VF5N n'était reconnu ni par les LTs CD8⁺ spécifiques de Q9VF5N, ni par ceux spécifiques de Q9VF5D, tandis que l'épitope Q9VF5D était bien reconnu par les clones T spécifiques Q9VF5D. Des expériences de liaisons aux protéines TAP et de spectrométrie de masse MS/MS ont été réalisées sur des immunoprotéasomes purifiés, permettant de mettre en évidence que l'introduction du résidu aspartate induisait un clivage par le protéasome de l'épitope, abolissant ainsi sa production et sa présentation par les molécules HLA.

3.10.2 Les ARFs publiés par les autres laboratoires

Suite aux travaux de l'équipe du Dr Moris, huit études ont démontré l'existence de réponses T CD8⁺ dirigées contre des peptides dérivés d'ARFs du VIH (Garrison *et al.*, 2009; Bansal *et al.*, 2010, 2015; Berger *et al.*, 2010, 2015; Champiat *et al.*, 2012; Ternette *et al.*, 2016; Peng *et al.*, 2018). Sept études ont travaillé avec des isolats du clade B (Garrison *et al.*, 2009; Berger *et al.*, 2010, 2015; Champiat *et al.*, 2012; Bansal *et al.*, 2015; Ternette *et al.*, 2016; Peng *et al.*, 2018), tandis qu'une seule étude a étudié les ARFs exprimés dans les isolats du clade C (Bansal *et al.*, 2010).

3.10.2.1 Identification par approche computationnelle

Quatre études ont utilisé une approche computationnelle pour identifier les ARFs du VIH (Garrison *et al.*, 2009; Champiat *et al.*, 2012; Bansal *et al.*, 2015; Berger *et al.*, 2015).

Pour identifier les ARFs, les chercheurs ont analysé les six phases de lecture du génome viral et ont recherché la présence de codons start AUG en amont des séquences (Champiat *et al.*, 2012; Berger *et al.*, 2015). Berger et al ont identifié 5 ARFs *antisens* pour lesquels ils observent une réponse T spécifique chez des individus chroniquement infectés (Berger *et al.*, 2015). Dans une autre étude menée par Champiat et al, 13 ARFs en *sens* et 70 en *antisens* ont été identifiés, chevauchant principalement la région Env (Champiat *et al.*, 2012). Les auteurs ont évalué les réponses T spécifiques des peptides dérivés d'ARFs par ELISPOT IFN γ , chez des individus en phase aiguë et chronique d'infection avant et après TARV. Des pools de peptides ont été utilisés pour faciliter le criblage. L'objectif de cette étude étant de caractériser l'impact du TARV sur les réponses T spécifiques des ARFs au cours de l'infection, aucune déconvolution des pools

positifs n'a été réalisée. Les pools 2 et 3 contenant les peptides dérivés d'ARFs en *sens* ont été reconnus par les PBMCs d'individus en phase aiguë et chronique d'infection. Les chercheurs ont montré que le TARV semblait favoriser l'émergence de réponses T spécifiques des ARFs, et ont émis l'hypothèse que la pression de sélection médiée par le TARV exercée sur les ORFs pourrait également impacter l'expression des ARFs.

Par une approche similaire de prédiction *in silico*, Garrison et al ont mis en évidence 4 peptides dérivés d'ARFs pour lesquels ils détectent une réponse T chez des individus infectés (Garrison *et al.*, 2009). Les peptides « SRDIHHYRFR » et « RIFTIIVSD » dérivent des sites d'épissage alternatif des exons de Tat et Rev, respectivement. De façon intéressante, les auteurs ont également montré que la réponse T dirigée contre les peptides dérivés d'ARFs est plus fréquente chez les individus chroniquement infectés traités et les EC que les individus virémiques. Ces observations rejoignent celles de Champiat et al (Champiat *et al.*, 2012).

Dans cette même ligne, les recherches menées par le Dr Bansal ont révélé que les réponses T spécifiques de peptides dérivés d'ARFs sont plus fréquentes chez les individus porteurs des allèles HLA protecteurs HLA-B*27, -B57, et -B58 (Bansal *et al.*, 2015). L'étude a identifié plusieurs peptides dérivés d'ARFs localisés principalement sur le brin *antisens*, bien que deux peptides dérivés d'ARFs situés sur le brin *sens* aient également été identifiés et reconnus par les PBMCs d'un donneur HLA-B27.

Identification des ARFs du SIV par approche computationnelle

Selon un protocole similaire d'identification, c'est-à-dire une analyse du génome viral suivie d'une prédiction de fixation des peptides dérivés d'ARFs sur les molécules du CMH Mamu-B*17, Maness et al ont identifié un ARF codé par Env (Maness *et al.*, 2007). Parmi 20 macaques testés, le peptide prédit nommé cRW9 est reconnu par les LTs CD8⁺ de 13 macaques chroniquement infectés dont deux sont des macaques contrôlant naturellement l'infection. Sur ces deux animaux controllers les auteurs ont évalué la fréquence des LTs CD8⁺ spécifique de cRW9, en comparaison avec celles dirigées contre des peptides canoniques en utilisant des tétramères Mamu-B*17 couplés au peptide cWR9. Une réponse spécifique à cRW9 a été détectée chez les deux animaux mais était particulièrement forte chez l'un d'eux, avec une fréquence représentant 41% des réponses T CD8⁺ totales. L'analyse du séquençage du génome viral au moment de la mort des macaques révéla la présence d'une mutation identique au sein de cRW9 chez 5 animaux. La substitution d'un tryptophane en arginine à la position d'ancrage C-terminale provoque une diminution drastique de la fixation de l'épitope aux molécules Mamu-B*17. Enfin, les auteurs ont démontré que cette mutation est une mutation d'échappement à la pression exercée *in vivo* par les LTs CD8⁺.

3.10.2.2 Identification par HLA footprint

Concernant les ARFs du VIH, l'utilisation de la méthode HLA footprint par trois équipes de recherche a permis d'identifier des ARFs du VIH exclusivement localisés au sein des gènes *gag*, *pol*, et *nef* (Bansal *et al.*, 2010; Berger *et al.*, 2010; Peng *et al.*, 2018). Le fait qu'aucune de ces études ne mettent en évidence des ARFs issus de Env est un biais de la méthode HLA footprint. En effet, les réponses T associées à un contrôle viral ciblent principalement des épitopes issus de Gag, Pol puis plus modestement Nef, motivant ainsi les recherches de mutation

d'échappement dans ces régions spécifiques du génome (Carlson *et al.*, 2008; Matthews *et al.*, 2008).

Au total ce sont 70 et 13 ARFs du VIH qui ont été décrits en *sense* dans les études de Berger et al, et de Peng et al, respectivement (Berger *et al.*, 2010; Peng *et al.*, 2018). L'étude de Bansal et al qui, comme celle de Peng et al, s'intéressent aux ARFs du VIH situés en *antisense*, a révélé 391 ARFs répartis sur les deux brins du génome viral (Bansal *et al.*, 2010). En se basant sur le polymorphisme associé aux molécules HLA de classe I et l'utilisation de l'algorithme EpiPred pour définir l'épitope le plus probable de la séquence candidate, les auteurs ont sélectionné et synthétisé 24 peptides dérivés d'ARFs *antisens* pour lesquels ils ont testé leur capacité à stimuler une réponse T chez les individus infectés (Bansal *et al.*, 2010). Les peptides sont reconnus par les cellules T de donneurs en phase aiguë et chronique d'infection. Dans l'ensemble, 48% des peptides testés sont reconnus par les cellules T d'au moins un donneur. L'étude longitudinale effectuée sur 37 individus infectés a permis de mettre en évidence que les peptides dérivés d'ARF en *antisens* sont rapidement mutés après le début de l'infection, à l'instar des épitopes issus des protéines canoniques, connus pour être soumis à une pression du système immunitaire dès la première année d'infection (Brumme *et al.*, 2008). L'étude de Peng et al précisent ces précédentes observations en démontrant que les peptides dérivés d'ARFs *antisens* sont bien immunogéniques en phase aiguë et chronique d'infection, mais que les mutations observées aux stades précoces d'infection, ne sont globalement pas associées à la réponse immunitaire dirigée contre les peptides dérivés d'ARFs (Peng *et al.*, 2018).

Dans leur étude Berger et al ont sélectionné et synthétisé 22 peptides dérivés d'ARFs du VIH en *sens* pour évaluer les réponses T spécifiques chez des donneurs infectés chroniquement (Berger *et al.*, 2010). Dix peptides dérivés d'ARFs ont été reconnus par les lymphocytes T des donneurs, certaines séquences peptidiques étant chevauchantes, cela signifie que 8 ARFs ont la capacité à coder des polypeptides viraux. Le peptide « RTSKASLER » (RR9) qui est codé par un ARF localisé au sein du gène *pol* et présenté par les molécules HLA-A*A03, est celui qui induit le plus de réponses T. Un polymorphisme associé aux molécules HLA-A*03 statistiquement significatif a été observé au niveau du 6^{ème} résidu de l'épitope. La mutation de la sérine en une proline abolit la reconnaissance par les LTs *in vitro*. Le mécanisme d'initiation de la traduction du peptide RR9 a été ensuite investigué. La leucine présente en amont du peptide RR9 a été mutée en une méthionine ou valine et l'impact de cette mutation sur l'expression du peptide a été évaluée par des tests d'inhibition de la réplication virale par des clones T spécifiques. Les auteurs ont montré que le variant possédant la méthionine est fortement inhibé par les clones à la différence du variant portant la valine pour lequel aucune inhibition n'a été observée. Ces données suggèrent un rôle de la leucine en tant que codon start alternatif du peptide RR9.

3.10.2.3 Identification par immunopeptidomique

A ma connaissance, l'étude de Ternette et al publiée en 2016 est l'unique démonstration directe et non biaisée de l'existence des ARFs en tant que source antigénique grâce à la mise en évidence d'un peptide dérivé d'ARF présenté par les molécules du CMH-I par immunopeptidomique (Ternette *et al.*, 2016). Dans cet article, les auteurs ont cherché à

caractériser le répertoire de peptides viraux associé aux molécules du CMH de classe I, à partir de cellules humaines infectées. La lignée cellulaire C8166 et des cellules T CD4⁺ primaires provenant de 3 donneurs ont été infectées par l'isolat de laboratoire HIV III B. Une lyse suivie d'une immunoprécipitation des molécules du CMH de classe I ont été effectuées. L'analyse de la spectrométrie de masse LC-MS/MS réalisée sur les épitopes élués a permis de mettre en évidence un total de 75 peptides uniques provenant du génome viral. Le nombre total de peptides identifiés variait de 2 416 à 6 795, en fonction de l'échantillon, avec un enrichissement des 9-mers. Pour l'ensemble des cellules, les auteurs ont observé un enrichissement des acides aminés prédits aux résidus d'ancrage pour le génotype de l'allèle HLA correspondant. Les peptides viraux identifiés sont majoritairement issus des protéines Gag, Pol, et Env ; des peptides issus de Nef, Rev, et Vpu sont identifiés, bien que minoritaires. L'ensemble de ces peptides dérivent donc de protéine canonique, à l'exception d'un, nommé SR11, qui est issu d'un ARF. SR11 a été isolé dans les cellules C8166 et dérive d'un ARF lu en *antisense* qui chevauche le gène *env*. La capacité des peptides viraux à stimuler une réponse T chez des individus infectés a été ensuite évaluée par ELISPOT IFN γ . Parmi 24 donneurs testés, les PBMCs d'un seul donneur reconnaissent le peptide dérivé de l'ARF SR11. Dans cette étude les auteurs identifient un peptide dérivé d'ARF du VIH par immunopeptidomique, mais ne discutent pas la nature et les causes du différentiel obtenu entre l'immunopeptidome viral et cellulaire.

Les ARFs du VIH localisés sur le brin *sens* (5'-3') ont été mis en évidence dans la littérature par des méthodes indirectes, et leur nombre s'élève à 103 d'après les données disponibles. Toutefois, seule une petite partie de ces ARFs ont été validés pour leur expression par la détection d'une réponse T spécifique ou par immunopeptidomique. Contrairement aux virus HCMV, KSV, SARS-CoV-2, etc., pour lesquels l'ensemble des événements de traduction a été caractérisé *in vivo* par ribosome profiling, aucune démonstration n'a été faite sur le génome du VIH. Par conséquent, il est primordial d'établir une liste exhaustive des ARFs du VIH de manière non biaisée.

**MISE EN CONTEXTE,
OBJECTIFS DE THÈSE,**

&

**RÉSUMÉ DES
RÉSULTATS**

Mise en contexte

Les avancées récentes dans les méthodes de détection à haut débit ont révélé l'existence de petits cadres ouverts de lecture (sORFs) qui codent pour des polypeptides ou micro-protéines de moins de 100 acides aminés. Ces sORFs sont largement présents dans le génome humain, principalement dans les régions non traduites 5' UTR des gènes connus. Bien que certaines de ces micro-protéines aient été montrées comme ayant des rôles fondamentaux, par exemple dans la régulation de l'expression des gènes, le développement embryonnaire, les fonctions mitochondriales etc, les fonctions biologiques de la plupart d'entre elles restent énigmatiques. La caractérisation des sORFs repose principalement sur la technique du ribosome profiling (RiboSeq) qui permet une évaluation impartiale des séquences d'ARNm activement traduites. En combinant le RiboSeq et la spectrométrie de masse LC-MS/MS, plusieurs études notamment dans le cancer, ont mis en évidence que de nombreux peptides présentés par les molécules du CMH dérivent de sORFs, ce qui valide leur rôle en tant que source antigénique. Des études ont également démontré que les lymphocytes T infiltrant les tumeurs reconnaissent des peptides tumoraux traduits à partir d'une sous-population de sORFs appelée cadre alternatif de lecture (ARFs). Le microenvironnement immunitaire de la tumeur ainsi que les conditions de stress favorisent les évènements non canoniques de traduction des sORFs.

Le VIH-1 et le cancer présentent des caractéristiques communes telles que l'induction d'un stress majeur pour la cellule, l'exposition chronique aux antigènes, l'épuisement des lymphocytes T et la résistance. Dans ce contexte, notre équipe de recherche a cherché à avoir une approche similaire, à grande échelle et non biaisée, appliquée à l'infection à VIH-1. De façon non surprenante, des évènements non canoniques de traduction tels que les ARFs ont été décrits par RiboSeq dans plusieurs génomes viraux tels que HCMV, KSV, MLV, et SARS-CoV-2, mais aucune démonstration n'a été faite à ce jour sur le génome du VIH. Notre équipe avait déjà montré que plusieurs ARFs du VIH codent pour des petits polypeptides reconnus par les lymphocytes T de donneurs infectés par le VIH. D'autres équipes ont également démontré l'existence de peptides dérivés d'ARFs, mais toutes ces observations, y compris les nôtres, étaient basées sur l'immunologie inverse ou l'HLA footprint, autrement dit sur des approches indirectes. L'intégralité de ces études se sont concentrées sur la présentation de ces peptides non canoniques par les molécules du CMH-I aux lymphocytes T CD8⁺, et ont démontré qu'une réponse T spécifique est induite tout au long de l'infection à VIH et semble préférentiellement abondante chez les EC.

Notre objectif était de mettre en évidence le potentiel des régions présumées non codantes du VIH en définissant tous les ARFs et ORFs activement traduits et d'évaluer si les ARFs du VIH codent pour des polypeptides viraux. Pour répondre aux questions de ce sujet passionnant, plusieurs objectifs ont été soigneusement considérés :

Objectif 1 : Caractérisation du translatome du VIH-1

La traduction des ARFs et des ORFs a été identifiée par ribosome profiling. Les ARFs qui présentaient une signature traductionnelle significative par rapport au bruit de fond ont été évalués quant à leur conservation au sein des clades B et C du VIH.

Objectif 2 : Définition de l'immuno-peptidome du VIH-1

Le répertoire de peptides présentés par les molécules du CMH de classe I et II des cellules infectées a été établi par immuno-peptidomique afin de confirmer le rôle des ARFs en tant que source antigénique.

Objectif 3 : Détermination du potentiel immunogène des ARFs

La capacité des ARFs à coder des polypeptides viraux immunogènes a également été évaluée par la détection de réponses T spécifiques dans les PBMCs de donneurs infectés selon un protocole d'expansion *in vitro* qui précédait des tests ELISPOT IFN γ .

Objectif 4 : Evaluation des propriétés phénotypiques et fonctionnelles des lymphocytes T spécifiques des peptides dérivés d'ARFs

Les réponses T spécifiques des peptides dérivés d'ARFs ont été caractérisées par cytométrie en flux pour évaluer leurs phénotypes et polyfonctionnalité au travers de l'expression des marqueurs CD4, CD8, Mip1 β , IFN γ , IL2, CD107a, et TNF α .

Résumé des résultats

Dans une lignée T CD4⁺ infectée par l'isolat de laboratoire NL4-3, le ribosome profiling a révélé la traduction de 98 ARFs répartis sur l'ensemble du génome viral. Il s'agit du premier transcriptome complet du VIH-1. Un second ribosome profiling en présence d'harringtonine nous a permis de prédire des codons AUG ou non-AUG comme sites d'initiation de certains ARFs. En générant une base de données interne au laboratoire du VIH-1, nous avons démontré que les ARFs sont hautement conservés parmi les isolats de clade B et de clade C du VIH-1 avec une conservation médiane en acides aminés de 88% et 89%, respectivement. En analysant la conservation des régions chevauchantes des ORFs et de la troisième phase, nous avons observé que les ARFs ne semblent pas soumis à une sélection positive, la phase de l'ORF étant significativement plus conservée. Cependant, 8 ARFs sont plus conservés que la région chevauchante de l'ORF. Sur la base de leur conservation et de la prédiction de leur fixation aux molécules du CMH, nous avons synthétisé et testé par ELISPOT IFN γ un total de 96 peptides dérivés des ARFs sur les PBMCs de 3 individus traités et 4 individus contrôlant naturellement l'infection. Parmi ces peptides, 46 ont été reconnus par les lymphocytes T, démontrant ainsi qu'au moins 42 ARFs codent pour des polypeptides viraux. Les peptides reconnus dérivent de l'ensemble des régions génomiques, y compris la région 5'UTR. Certains peptides dérivés d'ARFs ont induit une sécrétion d'IFN γ aussi forte que celle obtenue après stimulation avec des peptides dérivés d'ORFs. Parallèlement, l'analyse de l'immuno-peptidome a permis d'identifier, dans des cellules T CD4⁺ primaires infectées, un ligand de HLA-A*0201 dérivé d'un ARF identifié par notre second ribosome profiling. De plus, nous avons également étudié les caractéristiques phénotypiques et fonctionnelles des cellules T ciblant les peptides dérivés d'ARFs chez les individus infectés par marquage cytokinique intracellulaire. La majorité des réponses observées était médiées par des lymphocytes T CD4⁺ polyfonctionnels sécrétant au moins 3 cytokines simultanément. L'analyse du profil cytokinique a révélé des similitudes entre les réponses T spécifiques des peptides dérivés d'ARFs et d'ORFs.

RÉSULTATS

Unveiling novel conserved HIV-1 reading frames encoding T-cell antigens using ribosome profiling.

Lisa Bertrand^{1,2}, Annika Nelde^{3*}, Bertha Cecilia Ramirez^{1*}, Isabelle Hatin^{1*}, Hugo Arbes¹, Pauline François¹, Emiliano Ricci⁴, Emmanuel Labaronne⁴, Sylvie Grégoire^{1,2}, Anne Bet², Guillaume Beauclair¹, Antoine Gross⁵, Maja C. Ziegler⁶, Mathias Pereira^{1,2}, Yann Verdier⁷, Joelle Vinh⁷, Stéphanie Graff-Dubois², Audrey Esclatine¹, Cécile Gouttefangeas³, Marcus Altfeld⁶, Laurent Hocqueloux⁸, Assia Samri², Brigitte Autran², Olivier Lambotte⁹, Hans-Georg Rammensee³, Juliane Walz³, Olivier Namy¹ and Arnaud Moris^{1,2#}.

¹ Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91190, Gif-sur-Yvette, France.

² Sorbonne Université, INSERM, CNRS, Center for Immunology and Microbial Infections (CIMI-Paris), 75013, Paris, France.

³ Department of Immunology, Institute for Cell Biology, University of Tübingen, 72076, Tübingen, Germany.

⁴ Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, 46 allée d'Italie F-69364 Lyon, France

⁵IRIM, CNRS, Université de Montpellier, France.

⁶ Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg

⁷ESPCI Paris, PSL University, Spectrométrie de Masse Biologique et Protéomique, CNRS UMR8249, Paris, France.

⁸ Centre Hospitalier Universitaire d'Orléans, Orléans, France.

⁹ Université Paris Saclay, AP-HP, Department of Internal Medicine & Clinical Immunology, Bicêtre Hospital, Inserm, CEA, UMR1184, Le Kremlin-Bicêtre, France.

*: These authors contributed equally to this work.

#: To whom correspondence should be addressed:

Arnaud Moris, arnaud.moris@i2bc.paris-saclay.fr

Present address:

SGD: Sorbonne Université, INSERM U959, Immunology-Immunopathology-Immunotherapy (i3), Paris, France.

BCR: INSERM U1016, CNRS UMR8104, Université de Paris Cité, Paris, France.

Running Title: HIV translome and T cell immunity.

Abstract

Alternative reading frames (ARFs) were described in several viral genomes using ribosome profiling such as HCMV, KSV, and SARS-CoV-2 but no demonstration has been made on HIV genome. Using ribosome profiling, we uncovered 98 conserved HIV ARFs distributed across the genome, with high conservation among HIV clade B and C isolates. Our analysis revealed that at least 42 ARFs encode viral polypeptides, as demonstrated by T-cell responses targeting 45 ARF-derived peptides in patients under treatment or naturally controlling the infection. At the same time, we identified a ligand of HLA-A*0201 derived from an identified ARF on primary infected cells. These responses were mediated by polyfunctional CD4⁺ T-cells secreting at least 3 cytokines simultaneously. Our discovery expands the list of conserved viral polypeptides that might be potential targets for vaccination strategies.

Key Words:

Virus / HIV-1 translation / ribosome profiling / Immunopeptidome / CD4⁺ and CD8⁺ T cells / polyfunctionality

Introduction

Open reading frames (ORF) of the human genome have been long annotated using restricted criteria including the presence of classical start (AUG) and stop codons, and the potential to encode protein longer than 100 amino acids (aa) (Andrews and Rothnagel, 2014). Recent advances in detection methods challenged this definition revealing that thousands of small ORF (sORF) encode polypeptides or microproteins shorter than 100 aa (Martinez et al., 2020). These sORF are widely spread along the human genome, some overlap, on a different frame, classical ORF but the majority is found within 5' untranslated regions (UTR) of known genes (Ji et al., 2015). Although some microproteins encoded by sORF have been shown to play fundamental roles, for instance, in DNA repair, mitochondrial functions, RNA regulation, etc (Schlesinger and Elsässer, 2022; Wright et al., 2022), the biological functions of most remain enigmatic. Nonetheless, the thorough characterization of sORF greatly diversified the translation landscape in healthy and malignant cells expanding in particular the sources of cancer antigens.

To date the characterization of sORF mostly relies on the ribosomal profiling technic (Riboseq) that allows unbiased assessment of actively translated mRNA sequences. However, only a limited number of studies correlated Riboseq data with the generation of peptides (Yewdell, 2022). Combining Riboseq and liquid chromatography–tandem mass spectrometry (LC-MS/MS) several studies revealed that a fraction of peptides displayed by MHC molecules (so-called immunopeptidome) is derived from sORF that are specific or overrepresented in tumor cells (Chong et al., 2022; Erhard et al., 2020; Ouspenskaia et al., 2022). Other studies demonstrated that the tumour immune microenvironment is likely to favour the expression of human endogenous retroviruses (hERVs) that are normally silent in healthy tissues and of out-of-frame polypeptides that both constitutes antigens sources for T cell immunity (Bartok et al., 2021; Smith et al., 2018). Remarkably, these findings and others (Starck et al., 2016) highlighted that noncanonical translation events from sORF, in 5'UTR or out-of-frame, and internal start codons are likely favoured in the context of stress.

Viral infections represent a major stress to the cell whose translation machinery is mobilized in the translation of viral proteins (Stern-Ginossar et al., 2019). Viruses acquired specific features to control viral protein expressing such as sequences favouring frame-shifting and internal entry of ribosomes (Firth and Brierley, 2012; Sorokin et al., 2021). Thereafter not surprisingly peptide derived from alternative translation events were described during the last decades in influenza (Bullock and Eisenlohr, 1996; Yang et al., 2016; Zanker et al., 2019), murine leukemia virus (MLV) (Gaur and Green, 2005) and human immunodeficiency virus (HIV-1) (Bansal et al., 2010; Berger et al., 2010; Cardinaud et al., 2004) infected cells.

The expression of these non-canonical viral peptides often referred to as cryptic epitopes (CE), was mainly studied in the context of peptide presentation by MHC class-I molecules to cytotoxic CD8⁺ T cells (CTLs), probably due to the fact that sORF are more likely to generate MHC-associated peptide than transcripts from canonical ORF (Ruiz Cuevas et al., 2021). To date, only a limited number of studies combined Riboseq and immunopeptidomic approaches to characterize the landscape and origin of viral peptides presented by MHC-I molecules in the context of viral infection. These studies confirmed that MHC-I molecules present peptides from un-annotated ORF within alternative reading frames (ARF) of the human cytomegalovirus

(HCMV) (Erhard et al., 2018) and of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Weingarten-Gabbay et al., 2021).

In the context case of HIV-1 infection, the existence of peptides derived from ARF (ARFP) was highlighted using T cell-based assays (Cardinaud et al., 2011, 2004; Champiat et al., 2012). CTLs specific to peptides derived from ARF were indeed detected during the acute and chronic phase of infection and seem preferentially abundant in patients with favourable clinical outcome expressing protective HLA alleles (Bansal et al., 2015). Underlining the contribution of ARFP-specific CTL in the control of viral replication, several labs have shown that HIV adapts to and escapes from ARFP-specific T cell responses by introducing mutations within ARF (Bansal et al., 2010; Berger et al., 2010; Cardinaud et al., 2011, 2004). Remarkably, in the macaque model of SIV infection, a single mutation within an ARFP-derive epitope was strongly associated with viral rebound (Maness et al., 2010, 2007). Bioinformatics approaches analysing the association between HLA polymorphisms and HIV sequence variations (HLA footprint) revealed that viral RNA might encode ARF-derived peptides from both forward and reverse viral ARF (Bansal et al., 2010; Berger et al., 2010; Peng et al., 2018).

In the present study, we used Riboseq to delineate the first translome of HIV-1 in infected CD4⁺ T cells. In addition to canonical viral ORF, we identify 98 ARF, corresponding to sORF, that are distributed across HIV genome including the 5'UTR region. Using a database of complete and unique HIV genomes, we show that most aa sequences of ARF are highly conserved among HIV-1 clade B and C strains, with 8 ARF aa sequences being more conserved than the overlapping ORF region. Using T-cell based assays and immunopeptidomic, we demonstrate that at least 42 ARF encode viral polypeptides. In PBMCs of HIV-infected individuals, ARFP elicit potent and poly-functionnal T cells responses mediated by both CD4⁺ and CD8⁺ T cells. Our findings broaden the spectrum of HIV immunogen that might help in the design of efficacious vaccines and might reveal the existence of viral microproteins or pseudogenes.

Results

HIV Ribosome Profiling reveals 98 Alternative Reading Frames

To date, the existence of ARF in the HIV-1 genome has been highlighted using indirect approaches such as T cell-based assays and HLA footprint analyses. We thus intended to provide an unbiased assessment of actively translated viral mRNA sequences in HIV-infected cells. To this end, we infected the CD4⁺ T cell line, SupT1, with HIV_{NL4-3} and analysed viral mRNA translation using ribosome profiling. 24h post-infection, the viability of the cells and the infection rates were analysed using a viability dye and a combination of anti-HIV-Gag and anti-CD4 antibodies, respectively (Supplementary Fig 1A and 1B). Combining Gag and CD4 stainings allowed the direct detection of cells expressing HIV-Gag protein and the indirect assessment of cells expressing viral proteins that down modulate CD4 expression such HIV-Nef, -Vpu or -Env, respectively (Tanaka et al., 2003). Among the three biological replicates, the viability of the cells ranged from 44 to 71% and from 54 to 73% of Gag⁺CD4⁻ and Gag⁺CD4⁺ infected cells (Supplementary Fig 1B). The cells were then harvested, polysome extracted, ribosome-protected fragments (RPF) isolated and sequenced (Supplementary Fig 1A). Data were analysed using the RiboDoc tool (François et al., 2021) and some inhouse biological pipelines. The length of RPFs ranged between 25–35 nts (not shown). We analysed the periodicity of the reads by aligning the RPF to annotated start codons of human coding sequences (CDS). The reads, in particular the 28- and 29-mer, uniformly aligned along the human CDS with a periodicity of three nucleotides corresponding to codon triplets, thus strongly suggesting that the RPF correspond to bona fide translation events (Supplementary Fig 1C). The data from the three biological replicates were then combined and viral RPF aligned to the HIV_{NL4-3} genome (Figure 1A). The ribosome footprints aligned only on the forward strand throughout the viral genome. On the Figure 1, note that each reading frame is highlighted by a specific colour: green, red and blue for Frame 0 (F0), 1 (F1) and 2 (F2), respectively (Figure 1A). The number and density of RPF was particularly high in 5' UTR and 3' UTR regions of the HIV genome (Figure 1A). This is coherent with the described pausing of ribosomes just before starting and after terminating translation of viral CDS in the 5' and 3' UTR regions, respectively (Han et al., 2014; Ingolia et al., 2009; Wolin and Walter, 1988). As expected, we observed patches of RPF in frame and uniformly covering the CDS of known viral proteins such as HIV-Gag and -Env (Figure 1A, in green and red, respectively). The decrease in density and number of RPF covering and aligning on the CDS of HIV-Gag and -Pol also highlighted the specific feature of the HIV-GagPol poly-protein expression, which is controlled by a ribosomal frameshifting during translation (Brierley and Dos Ramos, 2006; Jacks et al., 1988; Namy et al., 2006). Based on the number of RPF and the length of the CDS, we calculated the relative expression of each CDS (Figure 1C). Although we did not observe a correlation between the length of the CDS and the relative expression, the relative expression of HIV-GagPol polyprotein was about 6% of HIV-Gag, which is consistent with the literature (Chen et al., 1995; Jacks et al., 1988). Thereafter, our ribosome profiling of HIV translation identifies all known HIV CDS and confirms specific features of the regulation of HIV translation. Strikingly, we also observed patches of RPF overlapping known CDS but aligning on other reading frames (e.g. patches of RPF in red overlapping HIV-Gag CDS or RPF in blue overlapping HIV-Vif CDS) suggesting the existence of permissive translation events of ARF

in HIV genome (Figure 1A). Having no information, thus far, on potential start codons, we then defined an ARF as a sequence between two stop codons with minimal size of 10 codons (shorter sequences being unlikely to be processed to produce MHC-ligands). In order to discriminate between background/technical events and *bona fide* translation events, we then determined the number of reads in frame within each potential ARF, which allowed calculating a mean of RPF/codon for each ARF. We observed that 75% of the ARF had less than 8 RPF/codon and decided to use this value as minimum threshold to define *bona fide* translation of ARF. Based on these criteria, we identified 98 HIV ARF sequences distributed across the viral genome (Figure 1A bottom panel, 1D and Table 1). Examples of potential ARFs that failed or succeeded to pass our selection criteria are presented in Supplementary Fig 1A, right panel. Note that ARF were named based on the CDS that they overlap plus a specific number (e.g. HIV-GAG15). Figure 1B, four examples of selected ARF overlapping HIV-Gag, -Vif, -Env and -Nef CDS are highlighted, named ARF-GAG15, ARF-VIF3, ARF-ENV4 and ARF-NEF7, respectively. Interestingly, nine of the ARF were previously described in the literature using indirect approaches (Table 1) (Bansal et al., 2010; Berger et al., 2010; Champiat et al., 2012; Garrison et al., 2009; Peng et al., 2018). We also identified 9 ARF in the 5' UTR region of HIV genome, commonly referred as uORF, which is consistent with previous studies describing translation events in the 5' UTR regions of different viral mRNA (Arias et al., 2014; Bencun et al., 2018; Finkel et al., 2021; Stern-Ginossar et al., 2012; Whisnant et al., 2020; Yang et al., 2015; Zong et al., 2017). Note that we sequenced a large number of RPFs on the 5' UTR region which might be due to arrest of the ribosomes before initiation and to the fact that all the viral transcripts carry this 5' UTR region (de Breyne and Ohlmann, 2018; Han et al., 2014; Sertznig et al., 2018). In a second set of Riboseq experiments, using Harringtonine (HAR), a drug that stops the elongation step of translation and thus leads to the accumulation of ribosomes at starts sites, we intended to define the start codons of each CDS and ARF. Unfortunately, the number of reads was limited and did not allow the thorough and systematic identification of start codons with the exception of the start codons of *env* and *vpr* CDS (Supplementary Fig 2A and B) and four ARFs (Supplementary Fig 2C to F). We calculated the amino acid (AA) length of the selected ARF, keeping in mind that this is an overestimate since we did not readily define the start sites of the ARFs. The AA length of ARF ranges from 10 to 101 AA with a mean of 23,5 AA (Supplementary Fig 1D).

Using ribosome profiling, we therefore reveal the existence of at least 98 ARFs distributed across the HIV genome that are actively translated in infected CD4+ T cells. We confirm and extend previous work showing that other region than annotated HIV CDS are indeed translated.

Identified ARFs are conserved among HIV clade B and C isolates

Assuming that these ARFs might be a reservoir of genetic novelty and a source of T-cell antigen, we analysed their aa conservation among clade B and C HIV isolates. To this end, using the Los Alamos HIV Sequence Database, we created in house databases of clade B and C HIV sequences containing: complete HIV genome sequences without aberrant mutations within known CDS and isolated from different individuals. We obtained 1609 and 411 clade B and C sequences, respectively. Using UGENE, we then aligned the ARF that we identified in HIV_{NL4-3} genome to the databases (Figure 2). The ARF AA conservation ranged from 48% to 98% with a median of 88% and 42% to 98% with a median of 89% among the clade B and C,

respectively (Figure 2A and 2B). We did not notice a particular trend in AA conservation when comparing the reading frames in which the ARF are located (Supplementary Figure 3A). In contrast, a statistically significant difference in AA conservation can be observed depending on the overlapping ORF (Supplementary Figure 3B). The most significant difference in AA conservation is found between ARFs overlapping *pol* and *env* genomic regions, which is consistent with the fact that *env* ORF is the most variable region (Stefic et al., 2019). There is no difference in ARF AA conservation between clade B and C regardless of the reading frame in which the ARF are translated (Supplementary Figure 3D). Finally, we noticed no correlation between ARF AA conservation and ARF relative expression (Supplementary Figure 3E).

We then investigated whether it is the conservation of the ORF that imposes the conservation of the ARF. To this end, we analyzed the percentage of AA conservation of the overlapping ORF sequence and third frame (i.e., corresponding neither ORF nor ARF). To simplify the analysis, ARFs that overlap two ORFs were excluded. We observed a significantly higher conservation of ORF regions compared to the ARF and third frame, with 93, 88 and 88% AA conservation, respectively (Figure 2C and Supplementary Figure 3C). It is interesting to note that the most variable ARF among clade B and C, ARF-ENV3, overlaps the highly variable V1-loop of HIV-Env glycoprotein (Figure 2 and Supplementary Figure 3C) (Curlin et al., 2010; Hamoudi et al., 2013; Palmer et al., 1996; Pan et al., 2015). These observations further highlight that ORF encoding the known viral proteins required for viral replication are under positive selection and also influenced by the pressure of humoral immune responses (Ruiz-Orera and Albà, 2019; Stefic et al., 2019). In addition, there is no statistical difference in AA sequence conservation between the ARF frame and the third one (88.25 and 88% respectively, Supplementary Figure 3C), which might imply that there is a low or none selection pressure on ARFs. However, it is particularly interesting to note that 8 ARFs sequences are more conserved than their overlapping ORF: ARF-GAG9, -VIF5, -VPU1, -ENV4, -ENV9, -ENV20, -ENV21, and -NEF3 (Figure 2C).

Overall, we demonstrated that 98 non-canonical transcripts identified by ribosome profiling are highly conserved among HIV clade B and C clinical isolates. Although, most ARF are probably not under selective evolutionary pressure, eight ARF are more conserved than their overlapping ORFs. These ARF might encode polypeptides implicated in the viral life cycle in infected cells and/or with the potential to elicit antiviral immune responses.

To determine whether all these ARF readily encode viral polypeptides, we then used two complementary approaches, monitoring on one hand immune responses to ARF-derived polypeptide in the PBMCs of HIV-infected individual and on the second hand, using a biochemical approaches to isolate directly, in infected cells, ARF-derived peptides.

At least 42 ARFs encode viral polypeptides eliciting T cell responses

We first designed a set of peptides to be synthesized and tested their capacity to elicit T-cell responses in the PBMCs of HIV-infected individuals, using a cultured IFN γ -Elispot assay. ARF-derived peptides were selected based on their AA conservation among HIV isolates and their predicted capacity to bind HLA alleles exhibiting high frequencies in the general population (Table 2). For ARF with a long potential AA sequences, several peptides were synthesized and tested (Table 2). We then monitored T cell-responses to potential ARF-derived

peptides, using blood samples from individuals under anti-retroviral treatment (ART) and individuals who naturally control viral replication without treatment, so-called elite controllers (EC) (Table 3). As control, we also included peptides derived from canonical HIV proteins (HIV) and when possible, from non-HIV common viral peptides (CEF). In total, 96 ARF-derived peptides were synthesized ranging from 11 to 24 AA in length, and distributed in 9 different pools (Table 2). The HIV-classical peptide pool was composed of previously characterized T-cell epitopes (Berzofsky et al., 1991; Fonseca et al., 2006; Geretti et al., 1994; Malhotra et al., 2003; Mirano-Bascos et al., 2008; Ranasinghe et al., 2012; Schrier et al., 1989; Wahren et al., 1989) (Supplementary Table 1).

PBMCs from HIV-infected individuals were seeded in wells of p24-well plates and loaded with the different peptides pools in the presence of various cytokines. 7-days later, a fraction of the cells was counted and loaded with or without the peptide pool used for the initial culture and T-cell reactivity, and monitored using an IFN γ -Elispot assay (Supplementary 4A and B). Although with different magnitude, the PBMCs from all the donors reacted in a specific manner to the pool of peptides derived from classical HIV-proteins (HIV+) (Supplementary 4B). Out of the 7 samples from infected individuals 6 reacted to at least one peptide pool (Supplementary 4B). Cells from donor EC-1 reacted to up to 8 peptide-pools. In contrast using cells from donor ART-3, we could not identify peptide positive response due to the high IFN γ -secretion observed in wells without peptide re-stimulation (-). The response to the pool of classical HIV peptides was also particularly low in this donor (HIV+) (Supplementary 4B). Overall, several peptide-pools (including Pool-1, -2, -3, -4 and -6) induced very strong and specific IFN γ -secretions in PBMCs from several HIV-positive donors (Supplementary 4B). The cells were then allowed to expand further by renewing the cytokine cocktail and tested on day 12, after the initial culture, for their capacity to react to individual peptides of the pools tested positive at day-7 (Figure-3 and Supplementary 5). Note that, despite the high background of IFN γ -secretion observed at day 7, we tested the capacity of the cells from donor ART-3 to react to individual peptides of the pool 4, 6 and 8 that showed a tendency of specific responses at day-7 (see Supplementary 4B). In addition, responses where the spots appeared visually particularly large at day-7, even if not exceeding our selection threshold in terms of spot numbers, were also tested at day 12. The reactivity of each individual to the peptides is summarized and exhaustively presented, with pictures of wells of the IFN γ -Elispot plates, in Figure-3 and Supplementary 5, respectively. At day-12, all samples from HIV-infected donors exhibited very strong and saturated responses to classical HIV-peptide in the IFN γ -Elispot assay. Remarkably, the PBMCs of all donors reacted to at least one individual ARF-derived peptide (Figure-3). Among the donors, the median of T-cell responses against ARF-derived peptides was 6. However, responses were very heterogeneity with donor EC-1 reacting 33 ARF-derived peptides and EC-2 only to a single one (Figure 3A). A total 60 ARF-derived peptide-specific T cell responses were identified across HIV infected donor samples (Figure 3A). 14 ARF-derived peptides elicited T-cell responses in samples from two different individuals (Figure 3B). Using the exact same protocol, the entire set of ARF-derived peptide pools was also tested, at day-7, on PBMCs from non-infected donors (n=3) and on day-12, peptide deconvolution performed on pools stimulating IFN γ -responses (Supplementary 6). Cross-reactivity to one peptide pool was observed for EFS 1619 and EFS 5174 samples (Supplementary 6) but the peptide specificity could not be

confirmed when testing individual peptides (not shown). Donor EFS 0685 exhibited a high background on all tested peptide pools (Supplementary 6). Peptide deconvolution showed a slight cross-reactivity to 3 peptides that were then excluded from the study (Supplementary 6). Thereafter, the results using samples from HIV-infected individuals reveal 46 unique ARF-derived peptide specific T-cell responses.

These responses target peptides encoded by ARF distributed along HIV genome sequence from the 5' UTR to HIV-*Nef* CDS (Figure 3E). The magnitude of T cell responses was heterogeneous between and among HIV-infected donor samples, with a frequency ranging from above 1250 (saturated) to 13 SFU/10⁶ (Figure 3C) with a median of 311 SFU/10⁶ PBMCs. For instance, despite exhibiting a broad ARF-specific T-cell response, the intensity of most responses from donor EC-1 was low as compared to other donors (Supplementary 7). Interestingly, although donor EC-3 exhibited the highest magnitude of ARF-specific T-cell responses with a median at 569 SFU/10⁶ PBMCs (Supplementary 7), we did not observe a significant difference in the magnitude of T-cell responses between ART and EC individuals (Supplementary 7).

Overall, we demonstrate here that T-cell responses from HIV-1 infected donors targeted 46 ARF-derived peptides encoded by 42 different ARF. We thus reveal that at least 42 ARF described in our ribosome profiling of infected cells encode viral polypeptides and elicit T-cell responses *in vivo* in the course of infection.

ARF-specific T cell responses are mediated by CD4⁺ and CD8⁺ T cells with a polyfunctional profile.

We then asked whether CD8⁺ and/or CD4⁺ T cells mediate ARF-specific T cell responses and studied the quality of these responses. The quality of HIV-specific T cell activation, defined as the capacity to produce multiple antiviral cytokines/chemokines, and not the magnitude of T-cell responses has been linked to disease outcome (Sauce et al., 2013). Thereafter, we characterized the capacity of ARF-specific T-cells to produce MIP1 β , IFN γ , IL2, CD107a, and TFN α using an Intra-Cellular Cytokine Staining (ICS) (Figure 4). As before PBMCs from the 2 ART and 4 EC individuals (ART-2 was not tested in ICS due to a lack of sample availability) were seeded in wells of p24-well plates and loaded only with the respective peptides that induced T-cell responses in the previous cultured IFN γ -Elispot assay (one ARF-derived peptide per well). In total for all donors, we tested 48 ARF-derived peptides, 46 peptides which induced IFN γ -responses cultured IFN γ -Elispot assay plus 2 potential candidates (Supplementary Table 2). In addition, to the HIV⁺ peptide pool (HIV⁺), we included immunodominant peptides derived from classical HIV proteins with known HLA binding capacity (the Los Alamos National Laboratory (Frahm et al., 2008). The pool of the immunodominant peptides was adapted to the HLA allotype expressed by each individual (Supplementary Table 3). As previously, at day-7 post culture, a fraction of the cells was used in an IFN γ -Elispot assay (not shown). At day 12, peptides inducing an IFN γ secretion were then used to monitor T-cell activation using ICS combined with CD4, CD8 and CD3 extracellular stainings (Figure 4).

As illustrated with the ARF-derived peptide G05 and HIV-Env derived peptide env01, peptide loading led to strong secretions of MIP1 β , IFN γ , IL2, CD107a, and TFN α by CD4⁺ T cells as compared to the negative controls (NS-ARF and NS_ORF, respectively) (Figure 4A). Overall, 28 peptides induced the secretion of at least one cytokine which allowed to discriminate

whether T cells responses were mediated by CD4⁺ and/or CD8⁺ T cells (Figure 4B). Remarkably, most T-cell responses, 86%, were mediated by CD4⁺ T cells (Figure 4B and C). The ARF-derived peptide H04 and A05, G02 and G05 led to the stimulation of CD8⁺ and both CD4⁺ and CD8⁺ T cells respectively (Figure 4B).

ORF-specific CD4⁺ T cell responses are characterized by a significantly higher proportion of cells producing five individual cytokines as compare to ARF-specific CD4⁺ T-cells (Supplemental Figure 8A) which is not the case for CD8⁺ T cell responses (Supplemental Figure 8B). However, although the frequency of responses against ARF-derived peptides is lower than against ORF-derived peptides, ARF-derived peptides stimulate a polyfunctional response in both CD4⁺ and CD8⁺ T-cells (Figure 4D and 4E). As tri-functional T-cell responses are often considered polyfunctional and might be associated with protection from disease progression (Larsen et al., 2012), we focused our analysis on activated T-cells with a polyfunctional profile with at least 3 functions. Clustering the data, we observed that 80% of CD4⁺ T cells against ARF-derived peptides secrete at least 3 cytokines simultaneously compared to 93,3% for CD4⁺ T cells responding to HIV-classical peptide stimulation (Figure 4D). 90.9% of CD8⁺ T cell responses directed against HIV immunodominant peptides expressed 5 cytokines at the same time upon peptide stimulation, whereas 80% of ARF-specific CD8⁺ T cell responses harbour a tri-functional phenotype (Figure 4E). The profiles and frequencies of cytokines responses induced by each ARF- and ORF-derived peptide using cells from the 6 HIV-infected individuals are presented as heatmaps (Supplemental Figure 8B and C). It was interesting to note that upon G02 peptide stimulation, the proportion of activated T-cells secreting 5 cytokines simultaneously was higher than for some HIV-classical immunodominant epitopes or even non-HIV common cold epitopes stimulation (see donor EC-3 donor). Overall, taking into account cells producing simultaneously at least 3 cytokines, in ART and EC individuals, we did not observed a significant difference in the capacity of ARF- and ORF-encoded peptides to induce polyfunctional cytokine secretions in CD4⁺ or CD8⁺ T cell (Figure 4F and G). There is also no obvious influence of the clinical status on the polyfunctional profile of ARF- and ORF-specific T-cell responses (Figure 4G).

Finally, since the cytokine combination secreted by T cells might influence disease outcome (Almeida et al., 2009; Benati et al., 2016; Betts et al., 2006; Ferre et al., 2010; Vingert et al., 2012), we dissected ARF- and ORF-specific T-cell cytokine secretion patterns. To this end, using the software SPICE, we analyzed the 32 possible cytokine combinations (out of the 5 cytokines detected: MIP1 β , IL-2, INF γ , CD107a, and TNF α) secreted by CD4⁺ and CD8⁺ T-cells upon ARF- or ORF-derived peptide stimulation. Overall, although the percentage of ARF-specific CD4⁺ T cell activation was lower, activated CD4⁺ T cells exhibited similar cytokine secretion patterns whether they target ARF- or ORF-derived peptides (Supplemental Figure 8D). The majority of polyfunctional ARF-specific CD4⁺ T cells secreted TNF α γ , CD107a, IFN γ , and Mip1 β simultaneously. Among ARF- and ORF-specific CD4⁺ T cells, the most prevalent monofunctional categories were TNF α ⁺ and Mip1 β ⁺ cells, respectively (Supplemental Figure 8D). ARF- and ORF-specific CD8⁺ T-cell responses displayed similar pentafunctional profiles (Supplemental Figure 8D).

Taken together, our multiparametric study showed that ARF-derived peptides are recognized predominantly by CD4⁺ T-cells from ART and EC donors and to a lesser extent by CD8⁺ population. The percentage of ARF-specific T cell activation is lower than ORF-specific T cell responses. Nonetheless, some ARF-derived peptide induced a high magnitude of T cell activation than ORF-derived peptides. ARF- or ORF-derived peptide stimulation both induce a polyfunctional T cell activation.

Identification of an ARF-derived peptide presented by HLA-A*0201 molecules on the surface of infected primary CD4⁺ T cells.

Recent studies combined Riboseq and immunopeptidomic approaches to demonstrate that MHC molecules also display peptides derived from sORF that are specific or overrepresented in tumor cells (Chong et al., 2022; Erhard et al., 2020; Ouspenskaia et al., 2022). In the context of viral infection, two studies also highlighted that MHC-I molecules present peptide from ARF of HCMV and SARS-CoV-2 viruses (Erhard et al., 2018; Weingarten-Gabbay et al., 2021). We thus intended to identify ARF-derived peptides presented by MHC-I and MHC-II molecules on the surface of HIV-infected cells. We decided to perform *de novo* experiments but also to reanalyze LC-MS/MS data from previously published immunopeptidomic experiments that used HIV-infected cells. In both cases, we used as matrix to identify HIV-derived peptides, the genome of HIV_{NL4-3}, taking into account all potential peptides encompassed between 2 stop codons on the 6 potential reading frames (3 forward and 3 backwards) of the HIV genome. For our own experiments we used the CD4⁺ T cells lines SupT1, that was used in the Riboseq experiments and C8166 which express MHC-I molecules and both MHC-I and MHC-II molecules, respectively. In order to by-pass the action of HIV-Nef on the modulation of MHC-I and MHC-II molecule expressions (Pawlak and Dikeakos, 2015; Stumptner-Cuvelette et al., 2001), we used a *nef*-deficient HIV_{NL4-3} isolate (HIV_{NL4-3ΔNef}) bearing a stop codon in the *nef* CDS (Schwartz et al., 1996). Cells were readily infected, with more than 50 and 20 percent of HIV-gag⁺ SupT1 and C8166 cells, respectively (not shown and Supplementary Figure 9A), lysed and then submitted to MHC-I molecule immunoprecipitations using W632 antibody. Peptide-ligands were identified using LC-MS/MS. For both cell lines, although we obtained a large number of MHC-I ligand and we could detect 5 known HIV-derived MHC-I ligands, we could not identify any ARF-derived peptides. As most T cell recognizing ARF-derived peptides were CD4⁺ T cells, we also submitted HIV_{NL4-3ΔNef}-infected C8166 cells to MHC-II molecules immunoprecipitation using the TŪ39 antibody (specific to HLA-DP, DQ, and DR) (Supplementary Figure 9B). Using LC-MS/MS, we identify a total of 4 834 peptides (not shown) from which 13 were derived from the HIV-Gag protein (not shown and Supplementary Figure 9C). To our knowledge, these HIV-derived peptides are the first natural MHC-II ligands identified in infected cells. Nonetheless, unfortunately, we could not detect any ARF-derived peptide. We thus re-analyze the LC-MS/MS data from previously published work in particular work from Ziegler et al where primary CD4⁺ T cells infected HIV_{NL4-3} were submitted to W632 immunoprecipitation and MHC-I ligand identified (Ziegler et al., 2020). This work originally focused on the characterization of the cellular peptides repertoire in infected cells. As previously using as matrix the HIV_{NL4-3} sequence but taking into account all potential peptides encompassed between 2 stop codons on the 6 potential reading frames, we identified 7 HIV-derived peptides (Figure 5A). Remarkably, among these peptides, one was derived from an

ARF overlapping the *pol* gene, ILINGQYSL, that we named YSL. The LC-MS/MS peptide profile was confirmed using synthetic a peptide (Supplemental Figure 9D). The YSL peptide is a likely a strong binder of the HLA-A*0201 molecules expressed by the primary CD4⁺ T cells used for the infection, as estimated using the NetMHCpan 4.0 algorithm (Jurtz et al., 2017). The YSL-ARF is highly conserved (91% of aa conservation) among HIV clade B and C which is probably explained by the fact that it overlaps the sequence encoding the catalytic site of the HIV reverse transcriptase (RT) which is among the most conserved region of HIV crucial for its replication (Jacobo-Molina et al., 1993). Remarkably, the YSL-ARF did not pass through the selection criteria of our initial Riboseq profiling with cycloheximide because of its low coverage per codon (2,5 RPF/codon). However, it was readily identified in our second Riboseq performed in the presence of HAR (10,1 RPF/codon). The HAR treatment also leads to the accumulation of RFP at the leucine position 3223, suggesting that it might be the alternative start codon of YSL-ARF (Supplemental Figure 2F).

We then tested the capacity of the YSL peptide (YSL-I) and an elongated form, LHSFGWVMNSILING (YSL-II), with a high prediction score to bind HLA-DR molecules, to activate T cells response in the PBMCs of HIV-infected donors. Using our cultured IFN γ -Elipsot assay, we observed in two individuals intermediate to very strong T cell responses targeting the YSL-I peptide, in PBMCs from EC-1 and EC-3, respectively (Figure 5B). Remarkably, ICS revealed that the YSL-I peptide is targeted by CD8⁺ and CD4⁺ T cells with a polyfunctional profile equivalent to responses specific to immunodominant ORF-derived peptides (Figure 5C and D). Responses to the YLS-I peptide were particularly strong among CD4⁺ T cells from the EC-3 donor (up to 30% of CD4⁺ T cells) and were also detected, but to a weaker extend, in the cells from donors EC-1, EC-2 and EC-4 (Figure 5E). Two donors also exhibited weak specific CD4⁺ T cell responses to the YLS-II peptide (Figure 5F). Overall, 5 donors reacted to either YSL-I and YSL-II peptides (Figure 5G) demonstrating that the YSL-ARF readily encodes viral polypeptides.

Thereafter, using two complementary and independent approaches, detection of ARF-derived peptide-specific T cells in the PBMCs of HIV-infected individuals and direct isolation of MHC-bound ARF-derived peptide using biochemical methods, we readily demonstrate that HIV ARF identify by ribosome profiling encode viral polypeptides with potential to elicit broad T-cell responses.

Discussion

We provide here the first characterization of HIV translome in infected CD4⁺ T cells. To this end, we used an unbiased approach based on the identification of actively translated viral mRNA fragments. The ribosome footprints aligned on the three forward reading frames throughout the viral genome. As expected, we observed patches of RPF in frame with and covering the CDS of classical viral proteins. The density and number of reads allowed a clear identification of the ribosomal frame shifting required for HIV-GagPol poly-protein expression. Thereafter, we revealed the existence of at least 98 ARFs distributed across HIV genome, including the 5'UTR, that are actively translated in infected CD4⁺ T cells. These ARFs are highly conserved among HIV clade B and C clinical isolates. We demonstrate that at least 42 ARFs readily encode viral polypeptides, since ARF-derived polypeptides are targeted by CD4⁺ and CD8⁺ T cells from HIV-infected individuals in the course of infection. Remarkably, ARF-derived peptides induce a polyfunctional T cell activation that is reminiscent of T cell responses targeting classical ORF-derived immunodominant epitopes. Finally, using a biochemical approach, we identified a conserved ARF-derived epitope, YSL, exposed at the surface of HIV-infected cells by the HLA-A*0201 molecules. YSL is translated from a highly conserved ARF which also encodes a conserved MHC-II ligand inducing polyfunctional T cell activation. Using two complementary approaches, we thus demonstrate that the ARFs identify using ribosome profiling readily encode viral polypeptides with the capacity to induce potent T cell immunity.

In human cells, recent advances in ribosome profiling, RNAseq, and bioinformatics have revealed that sORFs are widely spread throughout the human genome, with some overlapping with different frames of classical ORFs or located within the 5' UTR of known genes in human cells (Ji et al., 2015). Remarkably these sORF encode polypeptides or microproteins shorter than 100 aa (Martinez et al., 2020; Schlesinger and Elsässer, 2022; Wright et al., 2022). In fact, it has been estimated that 85% of the translation products are translated from nonannotated regions of the human genome and mostly from out-of-frame sequences from CDS (Ingolia et al., 2019; Ruiz Cuevas et al., 2021). We identified 98 ARF in HIV-1 genome, ranging from the 5' UTR region to the *nef* gene. 9 are located in the 5' UTR region. In viral genomes, 5' UTR was until recently considered as non-coding, but required for viral replication to recruit and assemble translation initiation factors (Araujo et al., 2012; Leppek et al., 2018; Zhang and Crumpacker, 2022). Several studies using ribosome profiling recently demonstrated that polypeptides might be encoded from in the 5' UTR regions of viral and human genomes (Arias et al., 2014; Chong et al., 2020; Finkel et al., 2021; Ingolia et al., 2011, 2009; Lee et al., 2012; Ouspenskaia et al., 2022; Ruiz Cuevas et al., 2021; Stern-Ginossar et al., 2012; Whisnant et al., 2020; Yang et al., 2015). These ARF located at the 5' UTR, are probably regulatory elements of the translation of downstream CDS (Brar and Weissman, 2015; Jürgens and Wethmar, 2022; Orr et al., 2020; Wethmar, 2014).

In our dataset, outside the 5' UTR, most ARF are located within *gag*, *pol* and *env* genes which correlates with the size of the ORF. Previous studies using the HLA footprint or predictive ORF approaches, also found ARF overlapping *gag*, *pol*, and *nef* sequences (Bansal et al., 2010; Berger et al., 2015, 2010; Cardinaud et al., 2004; Champiat et al., 2012; Garrison et al., 2009; Peng et al., 2018; Ternette et al., 2016). Only 16 HIV ARF identified in our

ribosomal profiling of infected cells are in common with ARF described in the literature on the forward strand of HIV genome (Berger et al., 2010; Champiat et al., 2012; Garrison et al., 2009; Peng et al., 2018). This relatively low number of shared ARF might be due to an over-estimation of the number of ARF by the various prediction methods, the lack of thorough annotations of some ARF in the published works, but also the rigorous criteria that we applied for the selection of ARF based on the codon coverage by RPF, meaning the third percentile of RFP/codon (8 reads/codon) of the whole dataset. This is illustrated by the ARF overlapping *gag* CDS encoding the Q9VF epitope that we previously characterize in HIV-infected cells (Cardinaud et al., 2011, 2004) whose coverage (5,9 RPF/codon) is just below our selection criteria. However, we previously shown that HIV_{NL4-3}-infected cells are not recognized by Q9VF-specific CTL is due to a proteasomal cleavage of the Q9VF epitope (Cardinaud et al., 2011, 2004). Our ribosome profiling data also suggest at the absence of Q9VF epitope presentation might be a combination of factors including a lower translation rate of this ARF, as observed in here HIV_{NL4-3}-infected SupT1 CD4⁺ T cells. Remarkably, we also did not identify a significant number of RPF aligning in frame on the three minus/backwards strand of HIV. In particular, we could not identify the CDS encoding the antisense protein of HIV-1 (ASP) which overlaps *env* gene, although several lines of evidences suggest that it is expressed *in vivo* e.g. existence in patients of antibodies and CTLs specific to ASP epitopes, (Bet et al., 2015; Vanhée-Brossollet et al., 1995). The lack of ASP CDS detection might in fact be due to a specific regulation of the translation from the 3' LTR that is down modulated in lymphoid cells such as CD4⁺ T cell used in our study (Laverdure et al., 2012; Michael et al., 1994; Savoret et al., 2021). Therefore, the differences in ARF detections might as well be due to the fact that translation events are differently regulated depending on the cell lines and cell types. Nonetheless, we reveal here additional novel ARFs translated from the *vif*, *vpr*, *vpu*, *tat* and *env* sequences as well as an ARF overlapping *tat* and *rev* exons. Importantly, the ribosome profiling technic allows us to obtain an unbiased view of translation events in HIV-infected cells.

Combining ribosome profiling with drugs blocking elongation such as harringtonine (HAR), we intended to define the start codons of the newly described ARF. Unfortunately, we did not obtain a sufficient amount of reads to allow a systematic identification of start codons including start codons from the CDS. We could only identify start codons from *env* and *vpr* CDS and 4 ARF, including YSL-ARF. Among the 4 ARF, three use alternative L start codons and one M. However, we know that although ARF translation can be dependent on the presence of a classical start codon (AUG), a large part of eukaryotic and viral annotated as well as unannotated ORF or ARF initiate translation by non-canonical mechanisms such as leaky scanning, the read-codon through, ribosome shunting, re-initiation, IRES, FSs, and alternative start codons (Andreev et al., 2022; Firth and Brierley, 2012; Malys and McCarthy, 2011; Shatsky et al., 2018; Sorokin et al., 2021; Yang and Wang, 2019). Due to the lack of information on possible start codons, we defined the ARF based on flanking stop codons which also probably has an influence on the threshold of ARF RFP/codon and thus ARF identification. Based on these criteria, the length of the identified ARF ranged from 10 to 101 aa and can therefore be considered as sORF (Finkel et al., 2018). Using an in-house HIV database of clade B and C strains, we demonstrated that the ARF aa sequences are conserved among both clades. Overall, we observed that the overlapping ORF are significantly more conserved than the ARF and the third frame, suggesting that ARF might not be under genomic selective pressure.

Nonetheless, 8 ARFs appear to be more conserved than their overlapping ORF. We cannot exclude thus far that these 8 ARF might encode polypeptides with functional viral properties, however identified ARF might as well be seen as a genomic reservoir of unselected sequences allowing the emergence of *de novo* genes. Indeed, pervasive translation of short peptides derived from presumed non-coding regions might expose these ARF-encoded polypeptides to selection, allowing from time to time, the passage of ARF into to the world of established, regulated, and selected products, as observed for ASP (Carvunis et al., 2012; Cassan et al., 2016; Papadopoulos et al., 2023; Ruiz-Orera et al., 2018).

Remarkably, in infected cells, MHC-I epitopes originate from viral proteins but also from truncated or misfolded viral polypeptides, also called DRiPs (defective ribosomal products) (Dolan et al., 2010; W. Yewdell, 2005; Yewdell et al., 1996). DRiPs are labile products whose processing and presentation by MHC-I molecules proceed the production of epitopes from highly stable viral proteins (Yewdell and Nicchitta, 2006). DRiPs are processed by proteasomes shortly after translation, allowing rapid loading of MHC-I molecules and thus CTL recognition within minutes after viral infections (Dolan et al., 2011, 2010). Remarkably, DRiPs were first identified in HIV-infected cells (Schubert et al., 2000) and we could show that the presentation of truncated viral polypeptides enhances HIV-specific CTL activation (Cardinaud et al., 2017; Casartelli et al., 2010). We and others demonstrated that ARF-encoded polypeptides (ARFP) constitute a source of alternative antigens for the stimulation of HIV-specific CTLs (Bansal et al., 2010; Berger et al., 2015, 2010; Cardinaud et al., 2004; Champiat et al., 2012; Garrison et al., 2009; Maness et al., 2010, 2007; Peng et al., 2018; Ternette et al., 2016). ARFP-specific CTL recognize infected cells (Bansal et al., 2010; Berger et al., 2010; Cardinaud et al., 2004) and exert a selection pressure on the virus *in vivo* (Bansal et al., 2010; Berger et al., 2010; Cardinaud et al., 2011). In the present work, we readily demonstrate that at least 42 ARFs, identified by Riboseq, express viral polypeptides inducing T cell responses in HIV-infected individuals. Some of these ARF-derived peptides are targeted by CTLs.

Using immunopeptidomics, we intended to detect natural ARF-derived ligands of MHC-I molecules in the context of HIV-infected cells. However, we did not find any ARF-derived peptides. The total number of classical HIV-derived peptides was also rather low suggesting that we might reach here technical limitations of the immunopeptidomic approach at least using HIV-infected cells. However, re-analyzing published data using primary CD4⁺ T cells as infected cells, we, readily demonstrate that MHC-I molecules present ARF-derived peptides. We identified one ARF-derived peptide, namely YSL-I. Remarkably, we revealed in PBMCs of one HIV-infected individual polyfunctional CTL responses targeting this YSL-I peptide. In addition, 5 HIV-infected individuals mounted a CD4⁺ T cell responses against YSL-I peptide or the longer YSL-II peptide. Thereafter, we show that alternative translation events may give rise to epitopes recognized independently or concomitantly by CD4⁺ and CD8⁺ T cells and presented on different HLA class I and HLA II alleles. Overall, we actually observed that the majority of T cell responses targeting ARF-derived peptides were mediated by CD4⁺ T cells. Although we cannot exclude here a biased in our experimental protocol favoring the expansion of CD4⁺ T cells, HIV-specific CD4⁺ T cells clearly play an important role in HIV-infection. The breadth and specificity of HIV-specific CD4⁺ T cell responses are associated with improved viral control and low viremia during acute and chronic infection, respectively (Ranasinghe et al., 2012; Schieffer et al., 2014). In particular, the generation of Gag-specific

CD4⁺ T cells and their maintenance correlates with a better disease outcome and viral control, respectively (Ranasinghe et al., 2012; Schieffer et al., 2014). There are also evidences of an association between certain HLA-DRβ1 alleles and Gag-specific CD4⁺ T cell activity with delayed disease progression (Ranasinghe et al., 2013). HS CD4⁺ T cells can also exert direct antiviral functions by eliminating infected cells and/or inhibiting viral replication (Zheng et al., 2009). Remarkably others and we have shown that CD4⁺ T cell recognize peptides derived from newly-synthesized, so-called endogenous antigens in virus infected cells, including HIV-infected (Coulon et al., 2016; Eisenlohr and Hackett, 1989; Jacobson et al., 1988; Jaraquemada et al., 1990; Nuchtern et al., 1990; Schmid et al., 2007; Sekaly et al., 1988; Thiele et al., 2015) or in tumor cells (Fonteneau et al., 2016; Tsuji et al., 2012). In fact, depending on the cellular localization, the trafficking and the nature of the antigen itself, different pathways are involved in the degradation and presentation of endogenous antigens by MHC-II molecules, this includes components of the MHC-I processing pathway such as proteasome and TAP (Lich et al., 2000; Tewari et al., 2005), autophagy (Dengjel et al., 2005) and receptors regulating vesicular trafficking (Sarango et al., 2022). In the field of cancer, it was recently revealed that CD4⁺ T cells also recognize peptide derived from unannotated CDS and neoantigens (Cachot et al., 2021; Hu et al., 2021; Kreiter et al., 2015; Malandro et al., 2016; Ott et al., 2017; Sahin et al., 2017).

To get some insight on the quality of ARF-derived peptide specific T cell responses, we also analyze the polyfunctional profiles of these CTL and CD4⁺ T cell responses in PBMCs of patients. Indeed, the quality of HIV-specific T cell activation, defined as the capacity to produce multiple antiviral cytokines/chemokines, and not the magnitude of T-cell responses has been linked to disease outcome (Sauce et al., 2013). In addition, the composition of the pool of functionally distinct HIV-specific CD4⁺ T cells has been shown to be critical for viral control during acute and chronic stages (Duvall et al., 2008; Harari et al., 2004; Iyasere et al., 2003; Rosenberg et al., 1997; Yamamoto et al., 2009). We show here that CD4⁺ and CD8⁺ T cell responses against ARF-derived peptides exhibit a polyfunctional profile with 80% of cells secreting at least 3 cytokines simultaneously. The pattern of secretion does not differ between previously described T cells responses targeting immunodominant epitopes and ARFP-specific T cells responses, suggesting that the latter are as potent as classical T cell response in controlling viral replication. To this regard, it has been shown the ARFP-specific CTL cells are predominant in patient naturally controlling the infection (Bansal et al., 2015).

Using ribosome profiling, we defined here the translome of HIV in infected CD4⁺ T cells. We demonstrate that HIV genome harbours ARF that are readily translated giving rise to viral polypeptides used by the immune system to initiate potent T cell responses. Our work using a highly pathogenic virus extend current findings demonstrating that in human cells 85% of translation products are not annotated and to a large extend encoded by ARF (Ingolia et al., 2019; Ruiz Cuevas et al., 2021). Both in cancer and HIV-infected cells, ARF-derived polypeptides represent promising targets to promote the control of tumour growth or viral replication. Understanding how the translation of non-canonical genomic regions is regulated and its involvement in various cellular or viral processes will certainly help in fighting cancer and infections.

MATERIAL AND METHODS

Virus and infection

HIV_{NL4-3 XCS} and HIV_{NL4-3ΔNef} were produced by transfection of 293T cells using CAPHOS kit (Sigma-Aldrich). 1×10^9 SupT1 cell line were infected with 4 055 ng of HIV_{NL4-3 XCS} HIV-Gagp24, 3h in IMDM plus 10mM HEPES (Gibco/ThermoFisher Scientific) supplemented with 2 μg/ml of DEAE-dextran (Sigma). Three biological replicates were performed corresponding to 300 million cells infected at day 0. 20h post-infection 2×10^5 cells were harvested, stained with CD4-Vio770 antibody (Miltenyi) and a Live/Dead fixable violet dye (Invitrogen), fixed with 4% paraformaldehyde (ChemCruz) and permeabilized (PBS1X, BSA 0,5%, saponin 0,05%). Cells were then stained with an HIV-1 Gag-p24 specific antibody (KC57-PE, Beckman Coulter). Sample acquisition and analysis were performed using a BD LSRFortessa and FlowJo v10.8 Software, respectively (both from BD Life Sciences). Mock-infected cells were used as negative control. 24h post-infection, cycloheximide (CHX) was added to cell culture (100 μg/ml, 10 min, 37°C, Sigma). Cells were then washed in cold PBS added with CHX and lysed at 4°C (10 mM Tris-HCl pH7.5, 100 mM KCl, 10 mM Mg²⁺ acetate, 1% Triton X100 and 2 mM DTT). To identify potential start codons, prior harvest and lysis, cells were treated with Harringtonine (without CHX) (7min, 2μg/ml, 37°C). Glass beads were then added to the cell supernatant and incubated on an orbital shaker (5 min, 4°C). Beads were removed by centrifugation and the supernatant was quickly frozen in liquid nitrogen and stored at -80°C.

Ribosome profiling

Cells collected were gently defrosted on ice in presence of EDTA-free complete protease inhibitor cocktail (Roche) and 200U of RNase Murine Inhibitor (NEB) were added. The cells were mechanically broken through a 25 gauge needle and centrifugated 3 min at 800rcf at +4°C. The absorbance of the crude extract obtained was measured at 260 nm. The extracts were digested using 5U/UA260nm RNase I (Ambion) for 1h at 25°C with followed by addition of 500U of SUPERasin RNase inhibitor (Invitrogen). Monosomes generated were loaded on a 24% sucrose cushion and ultracentrifugated at 100 krpm on a TLA110 rotor at +4°C. The concentrated monosomes were resuspended in 600 microl of lysis buffer. RNA were extracted by acid phenol at 65°C, chloroform and precipitated by ethanol with 0.3 M sodium acetate pH5.2. Resuspended RNA were loaded on 17% polyacrylamide (19:1) gel with 7M urea and run in 1xTAE buffer for 6h at 100V. RNA fragments corresponding to 28-34nt were retrieved from gel and precipitated in ethanol with 0.3 M sodium acetate pH5.2 in presence of 100mg glycogen. rRNA were depleted using Ribo-Zero Gold rRNA Removal kit (Illumina). The supernatants containing the ribosome footprints were recovered and RNA were precipitated in ethanol in the presence of glycogen overnight at -20°C. The RNA concentration was measured by Quant-iT microRNA assay kit (Invitrogen) and the RNA integrity and quality was verified using Bioanalyzer small RNA Analysis kit (Agilent). cDNA library from 100 ng RNA was prepared by the High-throughput sequencing facility of I2BC, using the NebNext Small RNA Sample Prep kit with 3' sRNA Adapter (Illumina) according to the manufacturer's protocol with 12 cycles of PCR amplification in the last step followed by DNA purification with AMPpure XP beads cleanup. Library molarity was analyzed using Bioanalyzer DNA Analysis kit (Agilent) and an equimolar pool of libraries was sequenced with NextSeq 500/550 High output kit v2 (75 cycles) (Illumina) with 10% PhiX.

Bioinformatical analysis

The ribosome profiling analysis was made using the RiboDoc tool (v0.9.0) (François et al., 2021). The different main steps with corresponding programs, versions and command lines used in its analysis are described below. The reference genomes are Homo_sapiens.GRCh38.104 for Human (from the Ensembl database (Yates et al., 2020)) and HIV1-pNL4-3 XCS for HIV.

The sequencing adapters were trimmed by cutadapt v4.3 (Martin, 2011) and the lengths of the RPFs was filter to keep reads from 25 to 35 nucleotides long as there expected length is around 30 nucleotides:

```
cutadapt -e 0.125 --max-n=1 -m 25 -M 35 -a ${adapter_sequence} -o ${output.fastq}
${input.fastq}
```

The removal of the rRNA reads was made by an alignment on the rRNA sequences by bowtie2 v2.5.1 (Langmead and Salzberg, 2012) :

```
bowtie2 -x ${index.rRNA} -U ${input.fastq} --un-gz ${output.fastq}
```

The alignment on the genome was made with both hisat2 v2.2.1 (Kim et al., 2019) and bowtie2 v2.5.1 :

```
hisat2 -x ${hisat2_index.genome} --no-softclip -U ${input.fastq} --un-gz ${output.fastq} -S
${output.sam_hisat2}
```

```
bowtie2 -x ${bowtie2_index.genome} --end-to-end -U ${output.fastq} -S
${output.sam_bowtie2}
```

The selection of the reads uniquely mapped on the genome was made with samtools v1.14 (Li and Durbin, 2009) :

```
samtools view -F 3844 -q 1 -h -o ${output.bam}
```

The counting of the reads corresponding to each transcript was done by htseqcount v2.0.2 (Anders et al., 2015) :

```
htseq-count -f bam -t CDS -i Parent --additional-attr Name -m intersection-strict --nonunique
fraction ${input.bam} ${input.gff} > ${output.txt}
```

The differential analysis was made with the bioconductor package DESeq2 v1.34.0 (Love et al., 2014).

The qualitative analysis was made on a transcriptome made from the genome with a selection of transcripts annotated as having a 5'UTR region only (for the human genome). This analysis and the determination of the P-site offset for each read length of every sample was made by the riboWaltz v1.2.0 package (Lauria et al., 2018).

To study the reading frame of the Ribosome Protected Fragments (RPF), each read is represented by the coordinate corresponding to the first base of the associated ribosome's P-site. To determine where the P-site is, a P-site offset has to be defined for every read length. This step is done with the riboWaltz program (Lauria et al., 2018) which looks for the first base of each read beginning of the signal.

Data sets and multiple alignments

Sequences were downloaded from the Los Alamos HIV Sequence Database (HIV-1 clades B and C). Sequences were filtered based on the presence of only one sequence per patient and excluding sequences carrying premature stop codons within ORFs. Very similar sequences and incomplete sequences (more than 10 gaps/unknowns) were additionally discarded. We obtained 1609 and 411 of HIV-1 clade B and clade C, respectively. The AA sequence conservation of

the overlapping ORF and third frame were analyzed using Unipro UGENE toolkit. Multiple alignments were performed using ClustalO. Plots were generated using R (version 4.2.2) with the ggplot2, and rstatix packages.

Immuno-peptidome

Isolation of HLA ligands Koyn

1×10^9 C8166 cells were infected with HIV_{NL4-3ΔNef} at a MOI of 0,05 and HLA class-I and -II molecules of isolated using standard immunoaffinity purification (Falk *et al*, 1991; Nelde *et al*, 2019). Snap-frozen samples were lysed in 10 mM CHAPS/PBS (AppliChem, Gibco) with 1x protease inhibitor (Roche). HLA class-I and -II-associated peptides were isolated using the pan-HLA class I-specific mAb W6/32 and the pan-HLA class II-specific mAb TÛ39 (both in house mouse monoclonal) covalently linked to CNBr-activated Sepharose (GE Healthcare). HLA-peptide complexes were eluted by repeated addition of 0.2% TFA (trifluoroacetic acid, Merck). Eluted HLA ligands were purified by ultrafiltration using centrifugal filter units (Millipore). Peptides were desalted using ZipTip C18 pipette tips (Millipore), eluted in 35 µl 80% acetonitrile (Merck)/0.2% TFA, vacuum-centrifuged and resuspended in 25 µl of 1% acetonitrile/0.05% TFA and samples stored at -20 °C until LC-MS/MS analysis.

Analysis of HLA ligands by LC-MS/MS

Isolated peptides were separated by reversed-phase liquid chromatography (nano-UHPLC, UltiMate 3000 RSLCnano; ThermoFisher) and analysed in an online-coupled Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher). Samples were analysed in five technical replicates and sample shares of 20% trapped on a 75 µm × 2 cm trapping column (Acclaim PepMap RSLC; Thermo Fisher) at 4 µl/min for 5.75 min. Peptide separation was performed at 50 °C and a flow rate of 175 nl/min on a 50 µm × 25 cm separation column (Acclaim PepMap RSLC; Thermo Fisher) applying a gradient ranging from 2.4 to 32.0% of acetonitrile over the course of 90 min. Samples were analysed on the Orbitrap Fusion Lumos implementing a top-speed CID method with survey scans at 120k resolution and fragment detection in the Orbitrap (OTMS2) at 60k resolution. The mass range was limited to 400–650 m/z with precursors of charge states 2+ and 3+ eligible for fragmentation.

Database search and spectral annotation

LC-MS/MS results were processed using Proteome Discoverer (v.1.3; Thermo Fisher) to perform database search using the Sequest search engine (Thermo Fisher) and the HIV_{NL4-3ΔNef} proteome containing all possible polypeptides from the 6 frames of HIV_{NL4-3ΔNef} sequence as reference database. The search-combined data of five technical replicates was not restricted by enzymatic specificity, and oxidation of methionine residues was allowed as dynamic modification. Precursor mass tolerance was set to 5 ppm, and fragment mass tolerance to 0.02 Da. False discovery rate (FDR) was estimated using the Percolator node (Käll *et al.*, 2007) and was limited to 5%. For HLA class-I ligands, peptide lengths were limited to 8–12 amino acids. For HLA class-II, peptides were limited to 12–25 amino acids of length. HLA class-I annotation was performed using NetMHCpan 4.0 (Jurtz *et al.*, 2017) annotating peptides with percentile rank below 2% as previously described (Ghosh *et al.*, 2019).

Patients and Samples

EC (n=4) were recruited from the CO21 CODEX cohort implemented by the ANRS (Agence nationale de recherches sur le SIDA et les hépatites virales). PBMC were cryopreserved at enrolment. ECs were defined as HIV-infected individuals maintaining viral loads (VL) under

400 copies of HIV RNA/mL without cART for more than 5 years. HIV-infected efficiently treated patients (ART) (n=3) were recruited at Kremlin Bicêtre Hospital. They were treated for at least 1 year (mean of 10 years) and had an undetectable viral load using standard assays. HIV-negative individuals (n=3) were anonymous blood donors from EFS (Établissement Français du sang). A detailed description of the patients is provided in Table 3, including the median and interquartile range for age (at the time of the study), CD4 T cell count and RNA load for each group. HIV-RNA loads were measured on site with different real-time PCR-based assays; depending on the date of enrolment in the cohort and the assay routinely used on each site, the VL detection limit varied from 50 to 10 copies/mL.

Ethic statement

All the subjects provided their written informed consent to participate in the study. The CO21 CODEX cohort and this sub-study were funded and sponsored by ANRS and approved by the Ile de France VII Ethics Committee. The study was conducted according to the principles expressed in the Declaration of Helsinki.

Peptides

ARF-derived peptides were chosen based on the aa conservation among the clade B isolates and the predictions to bind frequent HLA alleles. The binding affinity of ARF-derived peptides was evaluated using the NetMHCPan-4.1 and NetMHCIIpan-4.1. For the longest ARF sequences, some overlapping peptides were also designed. HIV-classical epitopes were chosen based on the literature (Berzofsky et al., 1991; Fonseca et al., 2006; Geretti et al., 1994; Malhotra et al., 2003; Mirano-Bascos et al., 2008; Ransinghe et al., 2012; Schrier et al., 1989; Wahren et al., 1989). The peptide AA sequences are presented in Table-2. ARF-derived peptides and HIV-classical peptides were synthesized by Vivitide company and in-house at the Dpt of Immunology of Tübingen University. Non-HIV common peptides were obtained from Mabtech (PepPool: CEF (CD8), human, 3616-1).

12-day in vitro T cell amplification prior ELISPOT-assay

PBMCs were thawed and rested 2-3h in IMDM (Gibco/ThermoFisher Scientific) containing 5% human AB serum (SAB, Institut Jacques Boy), supplemented with recombinant human IL-2 (rhIL2, 10 U/ml, Miltenyi) and Dnase (1 U/ml, New England Biolabs). Cells were washed and $5-9 \times 10^6$ PBMCs were then seeded/well of a 24-well plate in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids and sodium pyruvate (all from Gibco/ThermoFisher Scientific). Nevirapine (NVP, 1,2 μ M, HIV reagent program) was added to inhibit potential viral replication and poly I:C (2 μ g/ml, InvivoGen) to facilitate the presentation of long peptides by antigen presenting cells (Schuhmacher et al., 2020). PBMCs were loaded with peptide pools (each peptide at 10 μ g/ml), except for the CEF pool (Mabtech) used at 5 μ g/ml, and cultured overnight at 37°C, 5% CO₂. HIV-classical peptides and/or CEF peptide pools were used as positive control for the expansion of HIV-specific and non-HIV common anti-virus-specific T cells, respectively. On day 1 and 3, T cell media were then complemented with rhIL2 (10 U/ml) and rhIL-7 (20 ng/ml, Miltenyi) with NVP (1,2 μ M), respectively, and maintained throughout the culture. On days 3, 5, 7, and 9, when the cell layer was > 70% confluent, cells were split in two wells before the addition of rhIL-2, rhIL-7 and NVP. On day 7, cells were harvested, counted and a fraction submitted to IFN- γ ELISPOT assay using the ARF-derived peptide pools and positive controls: HIV classical peptide and/or CEF peptide pools. The remaining cells were maintained in culture and submitted, on day 12, to IFN- γ ELISPOT assay using individual ARF-derived peptides and the controls.

Enzyme-linked immunospot assay - Cultured IFN- γ ELISPOT

From $10-30 \times 10^5$ cells/well were seeded in an Elispot plates (MSIPN4550, Millipore) in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids, sodium pyruvate 10 %, and HEPES 10mM (Gibco/ThermoFisher Scientific), loaded with either 50, 10 or 2 μ g/ml of

ARF-derived peptides, HIV classical peptides and CEF peptides, respectively and incubated at 37°C for 16h. ELISPOT plates were pre-coated with anti-IFN γ primary antibody and after cell-incubation IFN γ revealed using anti-IFN γ secondary antibody conjugated with biotin (both from Mabtech) as described (Cardinaud et al., 2017). The ELISPOT analysis was performed in technical triplicates or duplicates. Spots were counted with the AID ELISPOT Reader according standard protocols. Responses were considered positive when IFN γ production was superior to 20 spots/10⁶ PBMCs and at least twofold higher than background (cells loaded with the DMSO containing medium used to solubilize the peptides).

Intracellular Cytokine Staining (ICS) assay

Cells were cultured as for the 12-day in vitro T cell amplification. One day 12, cells were harvested, washed and counted. 20-100 x 10⁵ PBMCs were then seeded/well of a 96-well U-bottom plate in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids, sodium pyruvate 10% and HEPES 10mM. PBMCs were immediately loaded with individual ARF-derived peptides (50 μ g/ml), individual HIV classical peptides (10 μ g/ml) or CEF peptide pool (2 μ g/ml) and the CD107a-BV786 antibody (BD Biosciences), for degranulation detection, was added in each condition. After 1h at 37°C, brefeldin A was added (5 μ g/ml) and the incubation carried out for additional 5h. PBMCs were washed in PBS and incubated with LVD and human Fc Block (BD Biosciences). Cells were then stained 30 min on ice with a mix of cell surface antibodies: CD3-PE, CD8-SuperBright600 (Invitrogen) and CD4-AF750 (Beckman Coulter). Next, cells were washed, fixed with PFA 4% and permeabilized (PBS + BSA 0,5%, saponin 0,05%). Cytokine production were detected by intracellular staining using IFN- γ -PerCP-Cy-5.5, Mip1 β -FITC, TNF α -PE-Cy7 and IL2-APC antibodies (all from BD Biosciences).

Flow cytometry and data analysis

Sample acquisition and analysis were performed using Cytoflex with CytExpert software (Beckman Coulter). The compensation matrix was calculated automatically by the CytExpert software after measuring negative and single-color controls. Sample analysis was performed using FlowJo v10.8 Software. After elimination of doublets, lymphocytes were gated on side scatter area versus forward scatter area pseudo-colour dot plot and dead cells were removed according to LVD staining. CD4⁺CD3⁺ events were gated versus individual gates made for the detection of Mip1 β , IFN γ , IL2, CD107a and TNF α secretion. All gates were set based on the unstimulated control (cells loaded with the DMSO containing medium used to solubilized the peptides). Individual cytokine secretions were then combined together using Boolean gating to create a full array of possible combinations of response patterns from the CD4⁺CD3⁺-cell subsets. The same procedure was used to CD8⁺CD3⁺ events. Positive responses were reported after background subtraction. Samples were considered positive when the % of activated cells was at least 2 times higher than negative controls.

Statistical analysis

Statistical significances (p-values) were calculated using Prism Software (GraphPad).

Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al*, 2019) partner repository with the dataset identifier - XXXX.

Reviewer account details:

Username:

Password:

Acknowledgements

This work was granted by the Agence Nationale de Recherche sur le SIDA (ANRS-Maladies infectieuses émergentes) and Sidaction for fundings. L.B and A.B. were supported by ANRS. M.P. was supported by Sidaction. We thank Remi Villette for his expertise using R, Bernard Maillere and his team for the access to the ELISPOT reader, Anne Lopes, Paul Roginski, for discussions, Frederic Suba and Clemence Richetta for access to the L3 facility of ENS-Paris-Saclay. We thank all participants of the ANRS CODEX cohort and the NIH AIDS Research and Reference Reagent Program for providing drugs.

Author contributions

A.M. conceived and designed the project together with O.N. for the Riboseq experiments. Design and performed the experiments: L.B., A.N., B.C.R., I.H., S.G., A.B., Y.V. Analysed the data: H.A., P.F., L.B., G.B., A.N., B.C.R., I.H., S.G., Y.V., J.V, M.P., S.G.-D.,J.W. O.N. and A.M. Contributed reagents/materials/analysis tools: H.A., P.F., G.B., I.H., A.G., M.C.Z., C.G., M.A., L.H., E.R., E.L., A.E., A.S., B.A., O.L., H.-G.R., J.W., O.N. and A.M.. Wrote the paper: L.B. and A.M. with contributions from H.A., I.H. and O.N, and all authors.

Conflict of interest

All other authors declare no financial or commercial conflict of interest.

FIGURE LEGENDS & TABLES

Figure 1. The translational landscape of HIV-1: identification of ARFs. (A) Distribution of ribosome protected fragments (RPF) across HIV-1 genome obtained from HIV_{NL4-3}-infected SupT1 cells. Ribosomal footprints are represented according to the position in HIV_{NL4-3} genome. RPF translated in frames 0, +1, and +2 are represented in green, red, and blue, respectively. HIV genomic structure is indicated below; annotated coding sequences are indicated and colored accordingly to their reading frames from the first position of the genome. Note that the classical reference strain, HIV_{HXB2}, owns an extra T (position 5772 within the *vpr* gene) as compared to most clade B strains. As a direct consequence in the HIV_{NL4-3} genome, *env* CDS is not in the same reading frame as *pol* and *tat* CDS. *Rev*, *vpu* and *nef* frames are also different from the HIV_{HXB2} genome. The positioning of ARFs (purple lines) are indicated in the lower genome representation. The selection criteria were: 1) not corresponding to ORF sequences; 2) amino acid length ≥ 10 ; 3) translated between 2 stops codons and 4) reads per codon ≥ 8 . **(B) Ribosome densities reveal novel viral coding regions.** The number of ribosomal footprints (reads) are represented according to the position in HIV_{NL4-3} genome. The color codes indicate the frames of the ARF and the relative ORF (green, 0; red, +1; blue, +2). Filled and open rectangles indicate the relative ORF and the noncanonical ARF, respectively. **a-d**, ARFs within *gag* (a), *vif* (b), *env* (c), and *nef* (d) ORFs. Putative start and stop codons are labelled according to their reading frames with triangles and crosses, respectively. **(C) Relative expression of HIV ORFs.** Relative expressions of HIV_{NL4-3} CDS were obtained by dividing the total read counts of each translated fragment by its length establishing a read/nucleotide value. **(D) Genomic distribution of the 98 identified ARF according to the overlapping CDS.** The number of ARF is indicated in brackets.

Figure 2: Amino acid conservation of ARF sequences among HIV-1 clade B and C strains. Percentage of amino acid (aa) conservation of each identified ARFs among HIV-1 clade B (A) and clade C (B). ARFs are represented according to the reading frame of their overlapping CDS (top legend). ARF expressed in +1 or +2 are represented by a green square and a red triangle, respectively. ARFs overlapping two ORFs are represented by a blue reverted triangle and those that correspond to peptide leaders or that are expressed in a first reading frame in the 5'UTR having no CDS, are represented by a purple circle. The horizontal lines indicate: the global ARF median (black dotted line), the median for ARF in +1 (green dotted line), +2 (red dotted line), ARF in 0 (purple dotted line), and the median for ARF overlapping 2 ORFs (blue dotted line). Genomic regions are indicated by the background color and the ARF name in the x-axis starting from 5'UTR (left) to *nef* region (right). **(C) Percentage of ARF amino acid conservation and overlapping genomic regions, among HIV-1 clade B.** Genomic regions are indicated by the background color and the ARF name in the x-axis starting from 5'UTR (left) to *nef* region (right). The aa conservation of each ARF (red triangle), overlapping CDS (purple circle), and third frame (green square) is presented. The horizontal colored lines indicate the corresponding median of the set.

Figure 3: Detection of T-cell responses specific to ARF-derived peptides in PBMCs of HIV-infected individuals. (A) Overview of ARF-derived peptide-specific T cell responses. The number of ARF-derived peptide-specific T cell responses is represented for each donor. The largest rectangle represent several peptides encoded by the same ARF. The color code, on the top right corner, indicates the overlapping CDS. **(B) The number of responding HIV-infected individuals to each ARF-derived peptide stimulation.** As in (A) the bar colors indicate the overlapping CDS. **(C) ELISPOT IFN γ deconvolution of positive pools at day 12 and for all tested donors.** Only IFN γ -positive T cell responses upon individual ARF-derived

peptide stimulations are presented, in SFU/10⁶ PBMCs. Intrinsic negative controls (POOL (-), cells incubated with DMSO-containing medium) are indicated by a white bar for each tested pool. POOL (+), HIV (+), and CEF (+) correspond to PBMCs cultures loaded with the pool of ARF-derived peptides, HIV-classical peptides, and non-HIV common-derived peptides, respectively. The bar color indicates the overlapping CDS. SAT: saturated well for which the IFN γ signal was so high that it could not be quantified. **(D) Genomic repartition of ARF-derived peptide stimulating T cell responses in the PBMCs of HIV-infected individuals.** The number of peptides is indicated in brackets. The color code is as in (A).

Figure 4: ARF-derived peptide (ARFP) specific T cell responses are mediated by CD4⁺ and CD8⁺ T-cells exhibiting polyfunctional cytokine profiles. (A) Representative gating scheme for the identification of ARFP-specific T cell responses. Gatings for G05-ARF derived peptide-specific CD4⁺ T cell responses. are shown for EC-3 HIV-infected individual. Gates for each cytokines/chemokines were set based on the negative controls (cells loaded with DMSO containing medium) NS_ARF and NS_ORF for the assessment of G05-ARF derived peptide-specific and env01-specific CD4⁺ T cell responses, respectively. Env01 is a previously characterized immunodominant peptide from HIV-Env. CD4⁺ CD3⁺ T cell populations were pre-gated on lived cells and doublets were excluded. **(B) Heatmap of ARF-derived peptides recognized by CD4⁺ and CD8⁺ T cells from PBMCs of HIV-infected donors.** NA = Not applicable, NS = Not significant. **(C) Percentage of CD4⁺ and CD8⁺ T cell responses among ARFP-specific T cells responses after *in vitro* stimulation with ARF-derived peptides at cohort the level.** Frequencies of ARFP- and ORF-specific CD3⁺ CD4⁺ **(D) or CD3⁺ CD8⁺ (E) T-cells secreting 1 to 5 cytokines simultaneously.** The numbers on the top indicate the percentage of polyfunctional T cells secreting at least 3 cytokines. ARFP-specific T-cell responses from all tested donors are combined. **(F) Comparison of the percentage of polyfunctional T cells secreting at least 3 cytokines among ARFP- and ORF-specific CD3⁺ CD4⁺ T-cells in the EC and ART patient groups. (G) same as (F) among ARFP- and ORF-specific CD3⁺ CD8⁺ T-cells in EC group.** Each dot represents one peptide-specific T-cell response (as in B). Kruskal-Wallis test with Dunn's comparison was applied.

Figure 5: Identification of ARF-derived peptide, YSL, presented by the HLA-A*02:01 molecules using immunopeptidomics and YSL-specific polyfunctional T-cells responses in HIV-infected donors. (A) Identification of ORF and ARF derived peptides from HIV_{NL4-3}-infected primary CD4⁺ T cells. LC-MS/MS data from Ziegler *et al.* were reanalyzed using as reference database all potential polypeptides encoded by the 6 frames of HIV_{NL4-3} genome. The sequences of HIV-derived MHC-I ligands identified and their characteristics are presented. The ILINGQYSL peptide, named YSL, derived from on ARF overlapping *pol* CDS exhibit a strong binding score for HLA-A*02:01 as monitored using NetMHCpan4.1. **(B) Detection of T-cell responses specific to YSL ARF-derived peptide in PBMCs of HIV-1 infected donors.** The IFN γ responses of donors EC-1 and EC3 who responded to the YSL-peptide stimulation are presented. Cells loaded with DMSO-containing medium, used as negative controls, (POOL (-)) are indicated by a white bar for each donor. POOL (+), YSL-I, and YSL-II correspond to cells loaded with a pool of peptides containing YSL-I and YSL-II peptides, or only with YSL-I and YSL-II peptides, respectively. SAT: saturated well for which the IFN γ signal was so high that it could not be quantified. **(C) Functional characteristics of YSL-specific CD8⁺ T cell response from EC-3 donor.** Following a 12-day *in vitro* expansion, intracellular staining was performed for Mip1 β , IFN γ , IL-2, CD107a, and TNF α after stimulation with YSL-I (left panel - upper gates) or a Nef-derived peptide (left panel - lower gates) used as positive control. Cells were gated on the CD8⁺ CD3⁺ population and each of the five individual gates for each cytokine/chelokine set based

on the negative controls (NEG, cells loaded with DMSO-containing medium). CD8⁺CD3⁺ cells were pre-gated on lived cells and doublets were excluded. The polyfunctional profile of activated YSL- or Nef- specific T-cells are shown in the right panel, upper and lower lines, respectively. The left pie charts indicated the activated (yellow) and non-activated (blue) T-cells following peptides stimulation. Polyfunctionality of activated T-cells are shown in the right pie charts. Boolean analysis was conducted to determine whether T-cells secreted 1, 2, 3, 4 or 5 cytokines simultaneously (red, green, purple, blue, and orange, respectively). Numbers indicate the percentage of activated cells with one to five cytokines secreted simultaneously (**D**) as in (C), for YSL-specific CD4⁺ T cell response from EC-3 donor. Panel (E) and (F), polyfunctional profiles of CD4⁺ T cell response against YSL-I and YSL-II peptides, respectively, The polyfunctionality of activated T-cells, for all responding donors, is shown as pie charts. Boolean analysis was conducted to determine whether T-cells secreted 1, 2, 3, 4, or 5 cytokines simultaneously (red, green, purple, blue, and orange, respectively). Numbers indicate the percentage of activated cells with one to five cytokines secreted simultaneously (**G**) Heatmap summarizing YSL-I and -II peptide-specific CD4⁺ and CD8⁺ T cell responses in all tested HIV-infected donors. NA = Not applicable, NS = Not significant.

Table 1: List of HIV ARFs identified by Ribosome profiling. Each of the 98 identified ARFs is listed according to their name, frame, length, and position in the HIV_{NL4-3} genome. Their description in previous studies is also informed.

Table 2: List of synthesized ARF-derived peptides. Peptides were named according to the ARF from which there are translated (left column). The short name of the peptides is the peptide name used in the Figures. Peptide sequences and length are indicated.

Table 3: Clinical characteristics of HIV-infected individuals. Age, sex, years of infection, treatments, CD4 T cell counts, viral loads, and HLA types of HIV-infected individuals are indicated.

SUPPLEMENTAL FIGURES & TABLE LEGENDS

Supplemental Figure-1-related to Figure-1. (A) Outline of ribosome profiling protocol. 20h post-infection a fraction of HIV_{NL4-3}-infected CD4⁺ SupT1 cells was collected, the viability and percentage of HIV-Gag⁺ infected cells assessed using flow cytometry (B). 4h later (24h post-infection) infected cells were treated with CHX and lysed. Polysomes were then extracted, treated with RNase and ribosomes protected fragments (RPF) isolated and sequenced. Data were analysed using an inhouse biological pipeline and RPF were aligned on HIV_{NL4-3} genome according to their reading frame. CDS and ARF translated in frame 0, +1, and +2 are represented in green, red, and blue, respectively. The ARF inclusion criteria were: 1) not corresponding to ORF/CDS sequences; 2) amino acid length ≥ 10 ; 3) translated between 2 stops codons and 4) reads per codon ≥ 8 . Green and blue rectangles indicate the canonical gag ORF and the noncanonical ARF-GAG16 that passed the inclusion criteria, respectively. Untagged red rectangle indicates ARF that failed the inclusion criteria. **(B) Gating strategy for the analysis of HIV_{NL4-3}-infected CD4⁺ SupT1 cells.** Three biological replicates were used. Infected cells were labeled with a viable dye and stained with antibodies specific to CD4 and HIV-Gag. In the middle the numbers indicate the percentages of viable cells. In the right panels, the lower and top numbers indicated the percentage of infected CD4⁺HIVGag⁺ and CD4⁺HIVGag⁺ cells, respectively. The same negative control was used for replicate 1 and 2 since they derived from the same experiment. **(C) Trinucleotide periodicity on cellular transcripts/metagene of 29 nt.** Data were analysed using an inhouse biological pipeline. We analysed the periodicity of the reads by aligning the RFP to annotated start codons of human

CDS. The reads, in particular the 28- and 29-mer, uniformly aligned along the human CDS with a periodicity of three nucleotides corresponding to codon triplets, thus strongly suggesting that the RPF correspond to bona fide translation events. **(D) Violin plot comparing amino acid lengths of uORF (in the 5'UTR) and non-uORF ARFs.**

Supplemental Figure-2-related to Figure-1: Identification of ARF and CDS start sites using translation inhibitors harringtonine (HAR) and Riboseq. In a second set of Riboseq experiments, HAR-treatment was used to define the start codons of each CDS and ARF. The number of reads was limited and did not allow the thorough and systematic identification of start codons. However, start codons indicated, with a black triangle, of *vpr* (A) and *env* (B) CDS and ARF-GAG7 (C), ARF-VPR3 (D), ARF-ENV/NEF1 (E) and ARF-YSL (F) could be defined. Elongating ribosomal P-site reads are depicted as bars, aligned to HIV_{NL4-3} genome, and are color coded by reading frame. For *vpr* (A) and *env* (B) CDS and ARF-ENV/NEF1 (E) the three frames are depicted whereas for ARF-GAG7 (C), ARF-VPR3 (D) and ARF-YSL (F) only the ORF and ARF frames are shown. Read coverage peak in HAR-treated cells (grey) confirmed translation initiation at the upstream-most AUG, start codon encoding methionine in ARF-GAG7 (C) and UUG, CUU, CUC alternative start codons encoding leucine in ARF-VPR3 (D), ARF-ENV/NEF1 (E) and ARF-YSL (F), respectively.

Supplemental Figure-3-related to Figure-2: Features of ARF amino acid (AA) conservation among HIV-1 clade B and C inhouse database. (A) Percentage of amino acid conservation of each identified ARFs among HIV-1 clade B, depending on the reading frame. ARFs are represented in relation to the reading frame of the ORF/CDS (x-axis). Kruskal-Wallis test was performed, and the p-value is reported on the graphic. **(B) Percentage of AA conservation of each identified ARFs depending on the overlapping CDS.** Kruskal-Wallis test was conducted followed by a post-hoc Dunn test (Holm correction). P-value: ***=0,00022 and **=0,0011. **(C) Percentage of AA conservation of the overlapping ORFs, ARFs, and third frames.** Kruskal-Wallis test followed by a pairwise Wilcoxon test was achieved, and p-values are shown on the graph. **(D) Comparison of ARF AA sequence conservation between clade B and clade C in-house database.** A pairwise Wilcoxon test was achieved and p-values are shown on the graph. **(E) Correlation between ARF AA conservation and ARF relative expression (expressed as reads/nucleotide).**

Supplemental Figure-4-related to Figure-3: Detection of ARF-derived peptides (ARFP)-specific T-cell responses using cultured IFN γ -Elispot assay. (A) Experimental procedure of the 12-day T-cell amplification. PBMCs from HIV-infected individuals were seeded in wells of p24-well plates and loaded with the different peptide pools in the presence of the indicated cytokines, polyI:C to allow a better presentation of elongated peptides and Nevirapine to block potential viral replication. 7-days later, a fraction of the cells was counted and loaded with or without the peptide pool used for the initial culture and T-cell reactivity monitored using an IFN γ -Elispot assay. The remaining cells were then allowed to expand further by renewing the cytokine cocktail and tested on day 12, after the initial culture, for their capacity to react to individual peptides of the pools tested positive at day-7, using IFN γ -Elispot assay. **(B) Day 7 IFN γ -Elispot responses against all peptide pools for every HIV-infected individual,** presented as SFU/10⁶ PBMCs (y-axis). A response was considered positive when the number of spots was >20 and 2 times higher than the negative control (POOL (-) white bar, cells loaded with DMSO-containing medium). Positive responses are indicated by colored bars depending on the overlapping genomic region. HIV-CDS-derived peptides pools were used as a positive control. SAT: saturated well for which the IFN γ signal was so high that it could not be quantify.

Supplemental Figure-5-related to Figure-3: Detection of T-cell responses specific to unique ARF-derived peptides in the PBMCs of HIV-infected individuals. Day 12, after the initial culture, the capacity of cultured PBMCs from HIV-infected donors to react to individual peptides of the pools, tested positive at day-7, was assessed using IFN γ -Elispot assay. The results are presented as SFU/10⁶ PBMCs. A response was considered positive when the number of spots was >20 and 2 times higher than the negative control (POOL (-) white bar, cells loaded with DMSO-containing medium). Shaded lines represent the threshold of positivity. HIV-classical peptide pool and CEF peptides were used as positive controls. **(A), (B), (C), (D), (E), (F) and (G) results from the cultured IFN γ -Elispot assay of cells from ART-1, ART-2, ART-3, EC-1, EC-2, EC-3 and EC-4 HIV-infected individuals, respectively.** Only deconvolutions for which a peptide pool elicited a significant T response are shown. Pictures of the original rough data from the ELISPOT plates are presented below or on the sides of the corresponding graphic (ARF-derived peptides and controls were tested in duplicates).

Supplemental Figure-6-related to Figure-3: Detection of T-cell responses to unique ARF-derived peptides in the PBMCs of non-infected individuals. As exemplified in panel (A) of the Supplemental Figure-4-related to Figure-3, PBMCs from non-infected individuals (EFS 0685, EFS 1619, and EFS 5174) were seeded and loaded with the different peptide pools in the presence of cytokines, polyI:C and Nevirapine cocktail. 7-days later, T-cell reactivity to the peptide pools used for the initial culture was monitored using an IFN γ -Elispot assay. Positive responses are indicated by colored bars depending on the overlapping genomic region. **(A).** 12 day, after the initial culture, their capacity to react to individual peptides of the pools, tested positive at day-7, was evaluated also, using IFN γ -Elispot assay. Shaded lines represent the threshold of positivity. Only deconvolutions for which a peptide pool elicited a significant T response are shown. **(B).** IFN γ -Elispot are presented as SFU/10⁶ PBMCs (y-axis). A response was considered positive when the number of spots was >20 and 2 times higher than the negative control (POOL (-) white bar, cells loaded with DMSO-containing medium). (A) CEF peptide pool was used as a positive control. SAT: saturated well for which the IFN γ signal was so high that it could not be quantify.

Supplemental Figure-7-related to Figure-3: Comparison of ARF-derived peptide (ARFP) specific T cells responses for all tested HIV-infected donors. Responses to unique ARFP were plotted for each HIV-infected individual as SFU/10⁶ PBMCs. Negative controls were subtracted and saturated responses attributed on arbitrary score of (1250 SFU/10⁶ PBMCs). Kruskal-Wallis test with Dunn's correction was applied. No significant differences could be observed.

Supplemental Figure-8-related to Figure-4: Analysis of the polyfunctional profiles of ARFP-, -ORF and CEF-specific T cell responses in HIV-infected individuals. (A) Percentages of ARF or ORF-specific CD4⁺ (left panel) and CD8⁺ (right panel) T cell responses secreting the individual cytokine/chemokine among total CD4⁺ and CD8⁺ T cells, respectively. CEF-specific CD8⁺ T cell responses are also presented (right panel). In contrast, to CD8⁺ T cells, a higher proportion of ORF-specific CD4⁺ T cells secrete each individual cytokine/chemokine as compared to ARF-specific CD4⁺ T cells. Mann-Whitney U test and p values are shown. **(B) and (C), Heatmaps summarizing the frequency of simultaneously secreted cytokine/chemokine (from 1 to 5) upon activation of CD4⁺ (B) and CD8⁺ (C) T-cells by individual ARF-, ORF- or CEF-peptides for each HIV-infected donor.** The frequency of secreting T-cells is represented by the filling gradient in red. The absence of the secretion is shown in dark grey (i.e "NS= Not Significant"). ARF-derived peptides not tested by ICS (i.e not positive in ELISPOT) are represented in light grey. **(D) Cytokine/chemokine secretion profiles of CD4⁺ (top panel) and CD8⁺ T (bottom panel) cell responses against**

ARF and ORF-derived peptides. ARF- and -ORF-specific T cell profiles are presented in the top and bottom rows, respectively. The 32 possible patterns of cytokine/chemokine secretions are indicated by the grey line below the heatmap and defined in a listed format on the right. (+) and (-) indicated secreted and not secreted respectively.

Supplemental Figure-9-related to Figure-5: Immunoepitidomic of HIV-infected cells. (A), (B) and (C), Defining the ligandome of HIV-derived HLA-class II ligands in HIV_{NL4-3ΔNef}-infected C8166 cells by LC-MS/MS. (A) Briefly, C8166 cells were infected or not (mock) with HIV_{NL4-3ΔNef} at 0,05 MOI. 24 hours post-infection, a fraction of cells was stained with an HIVGag-p24 antibody to estimate the fraction of HIVGag⁺ infected cells. Mock infected cells were used to delineate the HIVGag⁺ population (21% of HIVGag⁺ cells). **(B)** Cells were then lysed, HLA class-II complexes immunoprecipitated using TU39 antibody and HLA-ligands sequenced using LC-MS/MS. **(C) Table summarizing HIV-derived HLA class II ligands isolated from HIV_{NL4-3ΔNef}-infected C8166 cells.** All peptides are derived from classical HIV CDS. **(D) Confirmation of the LC-MS/MS profile of the ILINGQYSL peptide, so-called YSL.** A synthetic ILINGQYSL peptide (bottom spectrum) was used to validate the LC-MS/MS profile of the ILINGQYSL peptide (top spectrum) identified by reprocessing of the data from Ziegler *et al.* obtained using HIV_{NL4-3}-infected primary CD4⁺ T cells.

Supplemental Table 1: List of synthesized classical HIV peptides (pool HIV) derived from viral CDS.

Supplemental Table 2: List of ARF-derived peptides tested in ICS. Peptides were named according to the ARF from which there are translated (left column). The short name of the peptides is the peptide name used in the Figures. Pools where the peptides were included, peptide sequences and lengths are indicated.

Supplemental Table 3: List of classical HIV peptides used in ICS. These HIV ORF-derived peptides were previously described as immunodominant CD8⁺ T-cell epitopes (<https://www.hiv.lanl.gov/content/immunology>). The sequence, ORF and HLA-restriction of each peptide is indicated.

REFERENCES

- Almeida, J.R., Sauce, D., Price, D.A., Papagno, L., Shin, S.Y., Moris, A., Larsen, M., Pancino, G., Douek, D.C., Autran, B., Sáez-Cirión, A., Appay, V., 2009. Antigen sensitivity is a major determinant of CD8⁺ T-cell polyfunctionality and HIV-suppressive activity. *Blood* 113, 6351–6360. <https://doi.org/10.1182/blood-2009-02-206557>
- Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Andreev, D.E., Loughran, G., Fedorova, A.D., Mikhaylova, M.S., Shatsky, I.N., Baranov, P.V., 2022. Non-AUG translation initiation in mammals. *Genome Biol.* 23, 111. <https://doi.org/10.1186/s13059-022-02674-2>
- Andrews, S.J., Rothnagel, J.A., 2014. Emerging evidence for functional peptides encoded by short open reading frames. *Nature Reviews Genetics.* <https://doi.org/10.1038/nrg3520>
- Araujo, P.R., Yoon, K., Ko, D., Smith, A.D., Qiao, M., Suresh, U., Burns, S.C., Penalva, L.O.F., 2012. Before It Gets Started: Regulating Translation at the 5' UTR. *Comp. Funct. Genomics* 2012, 475731. <https://doi.org/10.1155/2012/475731>
- Arias, C., Weisburd, B., Stern-Ginossar, N., Mercier, A., Madrid, A.S., Bellare, P., Holdorf, M., Weissman, J.S., Ganem, D., 2014. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* 10, e1003847. <https://doi.org/10.1371/journal.ppat.1003847>
- Bansal, A., Carlson, J., Yan, J., Akinkunle, O.T., Schaefer, M., Sabbaj, S., Bet, A., Levy, D.N., Heath, S., Tang, J., Kaslow, R.A., Walker, B.D., Ndung'u, T., Goulder, P.J., Heckerman, D., Hunter, E., Goepfert, P.A., 2010. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med.* 207, 51–59. <https://doi.org/10.1084/jem.20092060>

- Bansal, A., Mann, T., Sterrett, S., Peng, B.J., Bet, A., Carlson, J.M., Goepfert, P.A., 2015. Enhanced Recognition of HIV-1 Cryptic Epitopes Restricted by HLA Class I Alleles Associated With a Favorable Clinical Outcome. *J. Acquir. Immune Defic. Syndr.* 1999 70, 1–8. <https://doi.org/10.1097/QAI.0000000000000700>
- Bartok, O., Pataskar, A., Nagel, R., Laos, M., Goldfarb, E., Hayoun, D., Levy, R., Körner, P.-R., Kreuger, I.Z.M., Champagne, J., Zaal, E.A., Bleijerveld, O.B., Huang, X., Kenski, J., Wargo, J., Brandis, A., Levin, Y., Mizrahi, O., Alon, M., Lebon, S., Yang, W., Nielsen, M.M., Stern-Ginossar, N., Altelaar, M., Berkers, C.R., Geiger, T., Peeper, D.S., Olweus, J., Samuels, Y., Agami, R., 2021. Anti-tumour immunity induces aberrant peptide presentation in melanoma. *Nature* 590, 332–337. <https://doi.org/10.1038/s41586-020-03054-1>
- Benati, D., Galperin, M., Lambotte, O., Gras, S., Lim, A., Mukhopadhyay, M., Nouël, A., Campbell, K.-A., Lemercier, B., Claireaux, M., Hendou, S., Lechat, P., de Truchis, P., Boufassa, F., Rossjohn, J., Delfraissy, J.-F., Arenzana-Seisdedos, F., Chakrabarti, L.A., 2016. Public T cell receptors confer high-avidity CD4 responses to HIV controllers. *J. Clin. Invest.* 126, 2093–2108. <https://doi.org/10.1172/JCI83792>
- Bencun, M., Klinke, O., Hotz-Wagenblatt, A., Klaus, S., Tsai, M.-H., Poirey, R., Delecluse, H.-J., 2018. Translational profiling of B cells infected with the Epstein-Barr virus reveals 5' leader ribosome recruitment through upstream open reading frames. *Nucleic Acids Res.* 46, 2802–2819. <https://doi.org/10.1093/nar/gky129>
- Berger, C.T., Carlson, J.M., Brumme, C.J., Hartman, K.L., Brumme, Z.L., Henry, L.M., Rosato, P.C., Piechocka-Trocha, A., Brockman, M.A., Harrigan, P.R., Heckerman, D., Kaufmann, D.E., Brander, C., 2010. Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. *J. Exp. Med.* 207, 61–75. <https://doi.org/10.1084/jem.20091808>
- Berger, C.T., Llano, A., Carlson, J.M., Brumme, Z.L., Brockman, M.A., Cedeño, S., Harrigan, P.R., Kaufmann, D.E., Heckerman, D., Meyerhans, A., Brander, C., 2015. Immune screening identifies novel T cell targets encoded by antisense reading frames of HIV-1. *J. Virol.* 89, 4015–4019. <https://doi.org/10.1128/JVI.03435-14>
- Berzofsky, J.A., Pendleton, C.D., Clerici, M., Ahlers, J., Lucey, D.R., Putney, S.D., Shearer, G.M., 1991. Construction of peptides encompassing multideterminant clusters of human immunodeficiency virus envelope to induce in vitro T cell responses in mice and humans of multiple MHC types. *J. Clin. Invest.* 88, 876–884. <https://doi.org/10.1172/JCI115389>
- Bet, A., Maze, E.A., Bansal, A., Sterrett, S., Gross, A., Graff-Dubois, S., Samri, A., Guihot, A., Katlama, C., Theodorou, I., Mesnard, J.-M., Moris, A., Goepfert, P.A., Cardinaud, S., 2015. The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* 12, 15. <https://doi.org/10.1186/s12977-015-0135-y>
- Betts, M.R., Nason, M.C., West, S.M., De Rosa, S.C., Migueles, S.A., Abraham, J., Lederman, M.M., Benito, J.M., Goepfert, P.A., Connors, M., Roederer, M., Koup, R.A., 2006. HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells. *Blood* 107, 4781–4789. <https://doi.org/10.1182/blood-2005-12-4818>
- Brar, G.A., Weissman, J.S., 2015. Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* 16, 651–664. <https://doi.org/10.1038/nrm4069>
- Brierley, I., Dos Ramos, F.J., 2006. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res., Translational Control During Virus Infections* 119, 29–42. <https://doi.org/10.1016/j.virusres.2005.10.008>
- Bullock, T.N., Eisenlohr, L.C., 1996. Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. *J. Exp. Med.* 184, 1319–1329. <https://doi.org/10.1084/jem.184.4.1319>
- Cachot, A., Bilous, M., Liu, Y.-C., Li, X., Saillard, M., Cenerenti, M., Rockinger, G.A., Wyss, T., Guillaume, P., Schmidt, J., Genolet, R., Ercolano, G., Protti, M.P., Reith, W., Ioannidou, K., de Leval, L., Trapani, J.A., Coukos, G., Harari, A., Speiser, D.E., Mathis, A., Gfeller, D., Altug, H., Romero, P., Jandus, C., 2021. Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. *Sci. Adv.* 7, eabe3348. <https://doi.org/10.1126/sciadv.abe3348>
- Cardinaud, S., Consiglieri, G., Bouziat, R., Urrutia, A., Graff-Dubois, S., Fourati, S., Malet, I., Guergnon, J., Guihot, A., Katlama, C., Autran, B., Endert, P., van, Lemonnier, F.A., Appay, V., Schwartz, O., Kloetzel, P.M., Moris, A., 2011. CTL Escape Mediated by Proteasomal Destruction of an HIV-1 Cryptic Epitope. *PLOS Pathog.* 7, e1002049. <https://doi.org/10.1371/journal.ppat.1002049>
- Cardinaud, S., Moris, A., Février, M., Rohrlisch, P.-S., Weiss, L., Langlade-Demoyen, P., Lemonnier, F.A., Schwartz, O., Habel, A., 2004. Identification of Cryptic MHC I-restricted Epitopes Encoded by HIV-1 Alternative Reading Frames. *J. Exp. Med.* 199, 1053–1063. <https://doi.org/10.1084/jem.20031869>
- Cardinaud, S., Urrutia, A., Rouers, A., Coulon, P.-G., Kervevan, J., Richetta, C., Bet, A., Maze, E.A., Larsen, M., Iglesias, M.-C., Appay, V., Graff-Dubois, S., Moris, A., 2017. Triggering of TLR-3, -4, NOD2, and DC-SIGN reduces viral replication and increases T-cell activation capacity of HIV-infected human dendritic cells. *Eur. J. Immunol.* 47, 818–829. <https://doi.org/10.1002/eji.201646603>
- Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charlotteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., Brar, G.A., Weissman, J.S., Regev, A., Thierry-Mieg, N., Cusick,

- M.E., Vidal, M., 2012. Proto-genes and de novo gene birth. *Nature* 487, 370–374. <https://doi.org/10.1038/nature11184>
- Casartelli, N., Guivel-Benhassine, F., Bouziat, R., Brandler, S., Schwartz, O., Moris, A., 2010. The antiviral factor APOBEC3G improves CTL recognition of cultured HIV-infected T cells. *J. Exp. Med.* 207, 39–49. <https://doi.org/10.1084/jem.20091933>
- Cassan, E., Arigon-Chifolleau, A.-M., Mesnard, J.-M., Gross, A., Gascuel, O., 2016. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. *Proc. Natl. Acad. Sci. U. S. A.* 113, 11537–11542. <https://doi.org/10.1073/pnas.1605739113>
- Champiat, S., Raposo, R.A.S., Maness, N.J., Lehman, J.L., Purtell, S.E., Hasenkrug, A.M., Miller, J.C., Dean, H., Koff, W.C., Hong, M.A., Martin, J.N., Deeks, S.G., Spotts, G.E., Pilcher, C.D., Hecht, F.M., Kallas, E.G., Garrison, K.E., Nixon, D.F., 2012. Influence of HAART on Alternative Reading Frame Immune Responses over the Course of HIV-1 Infection. *PLOS ONE* 7, e39311. <https://doi.org/10.1371/journal.pone.0039311>
- Chen, X., Chamorro, M., Lee, S.I., Shen, L.X., Hines, J.V., Tinoco, I., Varmus, H.E., 1995. Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting. *EMBO J.* 14, 842–852.
- Chong, C., Coukos, G., Bassani-Sternberg, M., 2022. Identification of tumor antigens with immunopeptidomics. *Nat. Biotechnol.* 40, 175–188. <https://doi.org/10.1038/s41587-021-01038-8>
- Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., Michaux, J., Bilic, I., Hirsekorn, A., Calviello, L., Simó-Riudalbas, L., Planet, E., Lubiński, J., Bryśkiewicz, M., Wiznerowicz, M., Xenarios, I., Zhang, L., Trono, D., Harari, A., Ohler, U., Coukos, G., Bassani-Sternberg, M., 2020. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun.* 11, 1293. <https://doi.org/10.1038/s41467-020-14968-9>
- Coulon, P.-G., Richetta, C., Rouers, A., Blanchet, F.P., Urrutia, A., Guerbois, M., Piguet, V., Theodorou, I., Bet, A., Schwartz, O., Tangy, F., Graff-Dubois, S., Cardinaud, S., Moris, A., 2016. HIV-Infected Dendritic Cells Present Endogenous MHC Class II–Restricted Antigens to HIV-Specific CD4+ T Cells. *J. Immunol.* 197, 517–532. <https://doi.org/10.4049/jimmunol.1600286>
- Curlin, M.E., Zioni, R., Hawes, S.E., Liu, Y., Deng, W., Gottlieb, G.S., Zhu, T., Mullins, J.I., 2010. HIV-1 Envelope Subregion Length Variation during Disease Progression. *PLOS Pathog.* 6, e1001228. <https://doi.org/10.1371/journal.ppat.1001228>
- de Breyne, S., Ohlmann, T., 2018. Focus on Translation Initiation of the HIV-1 mRNAs. *Int. J. Mol. Sci.* 20, 101. <https://doi.org/10.3390/ijms20010101>
- Dengjel, J., Schoor, O., Fischer, R., Reich, M., Kraus, M., Müller, M., Kreymborg, K., Altenberend, F., Brandenburg, J., Kalbacher, H., Brock, R., Driessen, C., Rammensee, H.-G., Stevanovic, S., 2005. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7922–7927. <https://doi.org/10.1073/pnas.0501190102>
- Dolan, B.P., Li, L., Takeda, K., Bennink, J.R., Yewdell, J.W., 2010. Defective Ribosomal Products Are the Major Source of Antigenic Peptides Endogenously Generated from Influenza A Virus Neuraminidase. *J. Immunol. Baltim. Md* 1950 184, 1419–1424. <https://doi.org/10.4049/jimmunol.0901907>
- Dolan, B.P., Li, L., Veltri, C.A., Ireland, C.M., Bennink, J.R., Yewdell, J.W., 2011. Distinct Pathways Generate Peptides from Defective Ribosomal Products for CD8+ T Cell Immunosurveillance. *J. Immunol. Baltim. Md* 1950 186, 2065–2072. <https://doi.org/10.4049/jimmunol.1003096>
- Duvall, M.G., Precopio, M.L., Ambrozak, D.A., Jaye, A., McMichael, A.J., Whittle, H.C., Roederer, M., Rowland-Jones, S.L., Koup, R.A., 2008. Polyfunctional T cell responses are a hallmark of HIV-2 infection. *Eur. J. Immunol.* 38, 350–363. <https://doi.org/10.1002/eji.200737768>
- Eisenlohr, L.C., Hackett, C.J., 1989. Class II major histocompatibility complex-restricted T cells specific for a virion structural protein that do not recognize exogenous influenza virus. Evidence that presentation of labile T cell determinants is favored by endogenous antigen synthesis. *J. Exp. Med.* 169, 921–931. <https://doi.org/10.1084/jem.169.3.921>
- Erhard, F., Dölken, L., Schilling, B., Schlosser, A., 2020. Identification of the Cryptic HLA-I Immunopeptidome. *Cancer Immunol. Res.* 8, 1018–1026. <https://doi.org/10.1158/2326-6066.CIR-19-0886>
- Erhard, F., Halenius, A., Zimmermann, C., L'Hernault, A., Kowalewski, D.J., Weekes, M.P., Stevanovic, S., Zimmer, R., Dölken, L., 2018. Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* 15, 363–366. <https://doi.org/10.1038/nmeth.4631>
- Ferre, A.L., Hunt, P.W., McConnell, D.H., Morris, M.M., Garcia, J.C., Pollard, R.B., Yee, H.F., Martin, J.N., Deeks, S.G., Shacklett, B.L., 2010. HIV controllers with HLA-DRB1*13 and HLA-DQB1*06 alleles have strong, polyfunctional mucosal CD4+ T-cell responses. *J. Virol.* 84, 11020–11029. <https://doi.org/10.1128/JVI.00980-10>
- Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M.,

- Stern-Ginossar, N., 2021. The coding capacity of SARS-CoV-2. *Nature* 589, 125–130. <https://doi.org/10.1038/s41586-020-2739-1>
- Finkel, Y., Stern-Ginossar, N., Schwartz, M., 2018. Viral Short ORFs and Their Possible Functions. *Proteomics* 18, 1700255. <https://doi.org/10.1002/pmic.201700255>
- Firth, A.E., Brierley, I., 2012. Non-canonical translation in RNA viruses. *J. Gen. Virol.* 93, 1385–1409. <https://doi.org/10.1099/vir.0.042499-0>
- Fonseca, S.G., Coutinho-Silva, A., Fonseca, L.A.M., Segurado, A.C., Moraes, S.L., Rodrigues, H., Hammer, J., Kallás, E.G., Sidney, J., Sette, A., Kalil, J., Cunha-Neto, E., 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS Lond. Engl.* 20, 2263–2273. <https://doi.org/10.1097/01.aids.0000253353.48331.5f>
- Fonteneau, J.F., Brilot, F., Münz, C., Gannagé, M., 2016. The Tumor Antigen NY-ESO-1 Mediates Direct Recognition of Melanoma Cells by CD4+ T Cells after Intercellular Antigen Transfer. *J. Immunol. Baltim. Md* 1950 196, 64–71. <https://doi.org/10.4049/jimmunol.1402664>
- Frahm, N., Baker, B., Brander, C., 2008. Identification and Optimal Definition of HIV-Derived Cytotoxic T Lymphocyte (CTL) Epitopes for the Study of CTL Escape, Functional Avidity and Viral Evolution. *HIV Molecular Immunology*.
- François, P., Arbes, H., Demais, S., Baudin-Baillieu, A., Namy, O., 2021. RiboDoc: A Docker-based package for ribosome profiling analysis. *Comput. Struct. Biotechnol. J.* 19, 2851–2860. <https://doi.org/10.1016/j.csbj.2021.05.014>
- Garrison, K.E., Champiat, S., York, V.A., Agrawal, A.T., Kallas, E.G., Martin, J.N., Hecht, F.M., Deeks, S.G., Nixon, D.F., 2009. Transcriptional errors in human immunodeficiency virus type 1 generate targets for T-cell responses. *Clin. Vaccine Immunol. CVI* 16, 1369–1371. <https://doi.org/10.1128/CVI.00410-08>
- Gaur, A., Green, W.R., 2005. Role of a cytotoxic-T-lymphocyte epitope-defined, alternative gag open reading frame in the pathogenesis of a murine retrovirus-induced immunodeficiency syndrome. *J. Virol.* 79, 4308–4315. <https://doi.org/10.1128/JVI.79.7.4308-4315.2005>
- Geretti, A.M., Van Baalen, C.A., Borleffs, J.C., Van Els, C.A., Osterhaus, A.D., 1994. Kinetics and specificities of the T helper-cell response to gp120 in the asymptomatic stage of HIV-1 infection. *Scand. J. Immunol.* 39, 355–362. <https://doi.org/10.1111/j.1365-3083.1994.tb03386.x>
- Hamoudi, M., Simon-Loriere, E., Gasser, R., Negroni, M., 2013. Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality. *Retrovirology* 10, 114. <https://doi.org/10.1186/1742-4690-10-114>
- Han, Y., Gao, X., Liu, B., Wan, J., Zhang, X., Qian, S.-B., 2014. Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation. *Cell Res.* 24, 842–851. <https://doi.org/10.1038/cr.2014.74>
- Harari, A., Petitpierre, S., Vallelain, F., Pantaleo, G., 2004. Skewed representation of functionally distinct populations of virus-specific CD4 T cells in HIV-1-infected subjects with progressive disease: changes after antiretroviral therapy. *Blood* 103, 966–972. <https://doi.org/10.1182/blood-2003-04-1203>
- Hu, Z., Leet, D.E., Allesøe, R.L., Oliveira, G., Li, S., Luoma, A.M., Liu, J., Forman, J., Huang, T., Iorgulescu, J.B., Holden, R., Sarkizova, S., Gohil, S.H., Redd, R.A., Sun, J., Elagina, L., Giobbie-Hurder, A., Zhang, W., Peter, L., Ciantra, Z., Rodig, S., Olive, O., Shetty, K., Pyrdol, J., Uduman, M., Lee, P.C., Bachireddy, P., Buchbinder, E.I., Yoon, C.H., Neuberger, D., Pentelute, B.L., Hacohen, N., Livak, K.J., Shukla, S.A., Olsen, L.R., Barouch, D.H., Wucherpennig, K.W., Fritsch, E.F., Keskin, D.B., Wu, C.J., Ott, P.A., 2021. Personal Neoantigen Vaccines Induce Persistent Memory T Cell Responses and Epitope Spreading in Patients with Melanoma. *Nat. Med.* 27, 515–525. <https://doi.org/10.1038/s41591-020-01206-4>
- Ingolia, N.T., Ghaemmghami, S., Newman, J.R.S., Weissman, J.S., 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223. <https://doi.org/10.1126/science.1168978>
- Ingolia, N.T., Hussmann, J.A., Weissman, J.S., 2019. Ribosome Profiling: Global Views of Translation. *Cold Spring Harb. Perspect. Biol.* 11, a032698. <https://doi.org/10.1101/cshperspect.a032698>
- Ingolia, N.T., Lareau, L.F., Weissman, J.S., 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. <https://doi.org/10.1016/j.cell.2011.10.002>
- Iyasere, C., Tilton, J.C., Johnson, A.J., Younes, S., Yassine-Diab, B., Sekaly, R.-P., Kwok, W.W., Migueles, S.A., Laborico, A.C., Shupert, W.L., Hallahan, C.W., Davey, R.T., Dybul, M., Vogel, S., Metcalf, J., Connors, M., 2003. Diminished proliferation of human immunodeficiency virus-specific CD4+ T cells is associated with diminished interleukin-2 (IL-2) production and is recovered by exogenous IL-2. *J. Virol.* 77, 10900–10909. <https://doi.org/10.1128/jvi.77.20.10900-10909.2003>
- Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., Varmus, H.E., 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature* 331, 280–283. <https://doi.org/10.1038/331280a0>
- Jacobo-Molina, A., Ding, J., Nanni, R.G., Clark, A.D., Lu, X., Tantillo, C., Williams, R.L., Kamer, G., Ferris, A.L., Clark, P., 1993. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase

- complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6320–6324.
- Jacobson, S., Sekaly, R.P., Bellini, W.J., Johnson, C.L., McFarland, H.F., Long, E.O., 1988. Recognition of intracellular measles virus antigens by HLA class II restricted measles virus-specific cytotoxic T lymphocytes. *Ann. N. Y. Acad. Sci.* 540, 352–353. <https://doi.org/10.1111/j.1749-6632.1988.tb27096.x>
- Jaraquemada, D., Marti, M., Long, E.O., 1990. An endogenous processing pathway in vaccinia virus-infected cells for presentation of cytoplasmic antigens to class II-restricted T cells. *J. Exp. Med.* 172, 947–954. <https://doi.org/10.1084/jem.172.3.947>
- Ji, Z., Song, R., Regev, A., Struhl, K., 2015. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* 4, e08890. <https://doi.org/10.7554/eLife.08890>
- Jürgens, L., Wethmar, K., 2022. The Emerging Role of uORF-Encoded uPeptides and HLA uLigands in Cellular and Tumor Biology. *Cancers* 14, 6031. <https://doi.org/10.3390/cancers14246031>
- Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., Nielsen, M., 2017. NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* 199, 3360–3368. <https://doi.org/10.4049/jimmunol.1700893>
- Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., MacCoss, M.J., 2007. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4, 923–925. <https://doi.org/10.1038/nmeth1113>
- Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kreiter, S., Vormehr, M., van de Roemer, N., Diken, M., Löwer, M., Diekmann, J., Boegel, S., Schrörs, B., Vascotto, F., Castle, J.C., Tadmor, A.D., Schoenberger, S.P., Huber, C., Türeci, Ö., Sahin, U., 2015. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* 520, 692–696. <https://doi.org/10.1038/nature14426>
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>
- Larsen, M., Sauce, D., Arnaud, L., Fastenackels, S., Appay, V., Gorochev, G., 2012. Evaluating Cellular Polyfunctionality with a Novel Polyfunctionality Index. *PLoS ONE* 7, e42403. <https://doi.org/10.1371/journal.pone.0042403>
- Lauria, F., Tebaldi, T., Bernabò, P., Groen, E.J.N., Gillingwater, T.H., Viero, G., 2018. riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput. Biol.* 14, e1006169. <https://doi.org/10.1371/journal.pcbi.1006169>
- Laverdure, S., Gross, A., Arpin-André, C., Clerc, I., Beaumelle, B., Barbeau, B., Mesnard, J.-M., 2012. HIV-1 antisense transcription is preferentially activated in primary monocyte-derived cells. *J. Virol.* 86, 13785–13789. <https://doi.org/10.1128/JVI.01723-12>
- Lee, Sooncheol, Liu, B., Lee, Soohyun, Huang, S.-X., Shen, B., Qian, S.-B., 2012. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci.* 109, E2424–E2432. <https://doi.org/10.1073/pnas.1207846109>
- Leppek, K., Das, R., Barna, M., 2018. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat. Rev. Mol. Cell Biol.* 19, 158–174. <https://doi.org/10.1038/nrm.2017.103>
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Lich, J.D., Elliott, J.F., Blum, J.S., 2000. Cytoplasmic processing is a prerequisite for presentation of an endogenous antigen by major histocompatibility complex class II proteins. *J. Exp. Med.* 191, 1513–1524. <https://doi.org/10.1084/jem.191.9.1513>
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Malandro, N., Budhu, S., Kuhn, N.F., Liu, C., Murphy, J.T., Cortez, C., Zhong, H., Yang, X., Rizzuto, G., Altan-Bonnet, G., Merghoub, T., Wolchok, J.D., 2016. Clonal Abundance of Tumor-Specific CD4(+) T Cells Potentiates Efficacy and Alters Susceptibility to Exhaustion. *Immunity* 44, 179–193. <https://doi.org/10.1016/j.immuni.2015.12.018>
- Malhotra, U., Holte, S., Zhu, T., Delpit, E., Huntsberry, C., Sette, A., Shankarappa, R., Maenza, J., Corey, L., McElrath, M.J., 2003. Early induction and maintenance of Env-specific T-helper cells following human immunodeficiency virus type 1 infection. *J. Virol.* 77, 2663–2674. <https://doi.org/10.1128/jvi.77.4.2663-2674.2003>
- Malys, N., McCarthy, J.E.G., 2011. Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* 68, 991–1003. <https://doi.org/10.1007/s00018-010-0588-z>
- Maness, N.J., Valentine, L.E., May, G.E., Reed, J., Piaskowski, S.M., Soma, T., Furlott, J., Rakasz, E.G., Friedrich, T.C., Price, D.A., Gostick, E., Hughes, A.L., Sidney, J., Sette, A., Wilson, N.A., Watkins, D.I., 2007. AIDS

- virus specific CD8⁺ T lymphocytes against an immunodominant cryptic epitope select for viral escape. *J. Exp. Med.* 204, 2505–2512. <https://doi.org/10.1084/jem.20071261>
- Maness, N.J., Wilson, N.A., Reed, J.S., Piaskowski, S.M., Sacha, J.B., Walsh, A.D., Thoryk, E., Heidecker, G.J., Citron, M.P., Liang, X., Bett, A.J., Casimiro, D.R., Watkins, D.I., 2010. Robust, vaccine-induced CD8(+) T lymphocyte response against an out-of-frame epitope. *J. Immunol. Baltim. Md 1950* 184, 67–72. <https://doi.org/10.4049/jimmunol.0903118>
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>
- Martinez, T.F., Chu, Q., Donaldson, C., Tan, D., Shokhirev, M.N., Saghatelian, A., 2020. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* 16, 458–468. <https://doi.org/10.1038/s41589-019-0425-0>
- Michael, N.L., Vahey, M.T., d'Arcy, L., Ehrenberg, P.K., Mosca, J.D., Rappaport, J., Redfield, R.R., 1994. Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. *J. Virol.* 68, 979–987.
- Mirano-Bascos, D., Tary-Lehmann, M., Landry, S.J., 2008. Antigen structure influences helper T-cell epitope dominance in the human immune response to HIV envelope glycoprotein gp120. *Eur. J. Immunol.* 38, 1231–1237. <https://doi.org/10.1002/eji.200738011>
- Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J.C., Brierley, I., 2006. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* 441, 244–247. <https://doi.org/10.1038/nature04735>
- Nuchtern, J.G., Biddison, W.E., Klausner, R.D., 1990. Class II MHC molecules can use the endogenous pathway of antigen presentation. *Nature* 343, 74–76. <https://doi.org/10.1038/343074a0>
- Orr, M.W., Mao, Y., Storz, G., Qian, S.-B., 2020. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* 48, 1029–1042. <https://doi.org/10.1093/nar/gkz734>
- Ott, P.A., Hu, Z., Keskin, D.B., Shukla, S.A., Sun, J., Bozym, D.J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., Chen, C., Olive, O., Carter, T.A., Li, S., Lieb, D.J., Eisenhaure, T., Gjini, E., Stevens, J., Lane, W.J., Javeri, I., Nellaiappan, K., Salazar, A.M., Daley, H., Seaman, M., Buchbinder, E.I., Yoon, C.H., Harden, M., Lennon, N., Gabriel, S., Rodig, S.J., Barouch, D.H., Aster, J.C., Getz, G., Wucherpfennig, K., Neuberg, D., Ritz, J., Lander, E.S., Fritsch, E.F., Hacohen, N., Wu, C.J., 2017. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* 547, 217–221. <https://doi.org/10.1038/nature22991>
- Ouspenskaia, T., Law, T., Clauser, K.R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B.A., Le, P.M., Hartigan, C.R., Keshishian, H., Apffel, A., Oliveira, G., Zhang, W., Chen, S., Chow, Y.T., Ji, Z., Jungreis, I., Shukla, S.A., Justesen, S., Bachireddy, P., Kellis, M., Getz, G., Hacohen, N., Keskin, D.B., Carr, S.A., Wu, C.J., Regev, A., 2022. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* 40, 209–217. <https://doi.org/10.1038/s41587-021-01021-3>
- Palmer, C., Balfe, P., Fox, D., May, J.C., Frederiksson, R., Fenyő, E.-M., Mckeating, J.A., 1996. Functional Characterization of the V1V2 Region of Human Immunodeficiency Virus Type I. *Virology* 220, 436–449. <https://doi.org/10.1006/viro.1996.0331>
- Pan, R., Gorny, M.K., Zolla-Pazner, S., Kong, X.-P., 2015. The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel. *J. Virol.* 89, 8003–8010. <https://doi.org/10.1128/JVI.00754-15>
- Papadopoulos, C., Arbes, H., Chevrollier, N., Blanchet, S., Cornu, D., Roginski, P., Rabier, C., Atia, S., Lespinet, O., Namy, O., Lopes, A., 2023. The Ribosome Profiling landscape of yeast reveals a high diversity in pervasive translation. <https://doi.org/10.1101/2023.03.16.532990>
- Pawlak, E.N., Dikeakos, J.D., 2015. HIV-1 Nef: a master manipulator of the membrane trafficking machinery mediating immune evasion. *Biochim. Biophys. Acta* 1850, 733–741. <https://doi.org/10.1016/j.bbagen.2015.01.003>
- Peng, B.J., Carlson, J.M., Liu, M.K.P., Gao, F., Goonetilleke, N., McMichael, A.J., Borrow, P., Gilmour, J., Heath, S.L., Hunter, E., Bansal, A., Goepfert, P.A., 2018. Antisense-Derived HIV-1 Cryptic Epitopes Are Not Major Drivers of Viral Evolution during the Acute Phase of Infection. *J. Virol.* 92, e00711-18. <https://doi.org/10.1128/JVI.00711-18>
- Ranasinghe, S., Cutler, S., Davis, I., Lu, R., Soghoian, D.Z., Qi, Y., Sidney, J., Kranias, G., Flanders, M.D., Lindqvist, M., Kuhl, B., Alter, G., Deeks, S.G., Walker, B.D., Gao, X., Sette, A., Carrington, M., Streeck, H., 2013. Association of HLA-DRB1-restricted CD4⁺ T cell responses with HIV immune control. *Nat. Med.* 19, 930–933. <https://doi.org/10.1038/nm.3229>
- Ranasinghe, S., Flanders, M., Cutler, S., Soghoian, D.Z., Ghebremichael, M., Davis, I., Lindqvist, M., Pereyra, F., Walker, B.D., Heckerman, D., Streeck, H., 2012. HIV-Specific CD4 T Cell Responses to Different Viral Proteins Have Discordant Associations with Viral Load and Clinical Outcome. *J. Virol.* 86, 277–283. <https://doi.org/10.1128/JVI.05577-11>

- Rosenberg, E.S., Billingsley, J.M., Caliendo, A.M., Boswell, S.L., Sax, P.E., Kalams, S.A., Walker, B.D., 1997. Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science* 278, 1447–1450. <https://doi.org/10.1126/science.278.5342.1447>
- Ruiz Cuevas, M.V., Hardy, M.-P., Holly, J., Bonneil, É., Durette, C., Courcelles, M., Lanoix, J., Côté, C., Staudt, L.M., Lemieux, S., Thibault, P., Perreault, C., Yewdell, J.W., 2021. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep.* 34, 108815. <https://doi.org/10.1016/j.celrep.2021.108815>
- Ruiz-Orera, J., Albà, M.M., 2019. Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation. *Trends Genet.* 35, 186–198. <https://doi.org/10.1016/j.tig.2018.12.003>
- Ruiz-Orera, J., Verdaguier-Grau, P., Villanueva-Cañas, J.L., Messeguer, X., Albà, M.M., 2018. Translation of neutrally evolving peptides provides a basis for de novo gene evolution. *Nat. Ecol. Evol.* 2, 890–896. <https://doi.org/10.1038/s41559-018-0506-6>
- Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.-P., Simon, P., Löwer, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrörs, B., Omokoko, T., Vormehr, M., Albrecht, C., Paruzynski, A., Kuhn, A.N., Buck, J., Heesch, S., Schreeb, K.H., Müller, F., Ortseifer, I., Vogler, I., Godehardt, E., Attig, S., Rae, R., Breitkreuz, A., Tolliver, C., Suchan, M., Martic, G., Hohberger, A., Sorn, P., Diekmann, J., Ciesla, J., Waksman, O., Brück, A.-K., Witt, M., Zillgen, M., Rothermel, A., Kasemann, B., Langer, D., Bolte, S., Diken, M., Kreiter, S., Nemecek, R., Gebhardt, C., Grabbe, S., Höller, C., Utikal, J., Huber, C., Loquai, C., Türeci, Ö., 2017. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* 547, 222–226. <https://doi.org/10.1038/nature23003>
- Sarango, G., Richetta, C., Pereira, M., Kumari, A., Ghosh, M., Bertrand, L., Pionneau, C., Le Gall, M., Grégoire, S., Jeger-Madiot, R., Rosoy, E., Subra, F., Delelis, O., Faure, M., Esclatine, A., Graff-Dubois, S., Stevanović, S., Manoury, B., Ramirez, B.C., Moris, A., 2022. The Autophagy Receptor TAX1BP1 (T6BP) improves antigen presentation by MHC-II molecules. *EMBO Rep.* 23, e55470. <https://doi.org/10.15252/embr.202255470>
- Sauce, D., Elbim, C., Appay, V., 2013. Monitoring cellular immune markers in HIV infection: from activation to exhaustion. *Curr. Opin. HIV AIDS* 8, 125–131. <https://doi.org/10.1097/COH.0b013e32835d08a9>
- Savoret, J., Mesnard, J.-M., Gross, A., Chazal, N., 2021. Antisense Transcripts and Antisense Protein: A New Perspective on Human Immunodeficiency Virus Type 1. *Front. Microbiol.* 11.
- Schieffer, M., Jessen, H.K., Oster, A.F., Pissani, F., Soghoian, D.Z., Lu, R., Jessen, A.B., Zedlack, C., Schultz, B.T., Davis, I., Ranasinghe, S., Rosenberg, E.S., Alter, G., Schumann, R.R., Streeck, H., 2014. Induction of Gag-Specific CD4 T Cell Responses during Acute HIV Infection Is Associated with Improved Viral Control. *J. Virol.* 88, 7357–7366. <https://doi.org/10.1128/JVI.00728-14>
- Schlesinger, D., Elsässer, S.J., 2022. Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins. *FEBS J.* 289, 53–74. <https://doi.org/10.1111/febs.15769>
- Schmid, D., Pypaert, M., Münz, C., 2007. Antigen-loading compartments for major histocompatibility complex class II molecules continuously receive input from autophagosomes. *Immunity* 26, 79–92. <https://doi.org/10.1016/j.immuni.2006.10.018>
- Schrier, R.D., Gnann, J.W., Landes, R., Lockshin, C., Richman, D., McCutchan, A., Kennedy, C., Oldstone, M.B., Nelson, J.A., 1989. T cell recognition of HIV synthetic peptides in a natural infection. *J. Immunol. Baltim. Md* 1950 142, 1166–1176.
- Schubert, U., Antón, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., Bennink, J.R., 2000. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* 404, 770–774. <https://doi.org/10.1038/35008096>
- Schuhmacher, J., Heidt, S., Balchen, T., Richardson, J.R., Schmeltz, C., Sonne, J., Schweiker, J., Rammensee, H.-G., Straten, P.T., Røder, M.A., Brasso, K., Gouttefangeas, C., 2020. Vaccination against RhoC induces long-lasting immune responses in patients with prostate cancer: results from a phase I/II clinical trial. *J. Immunother. Cancer* 8, e001157. <https://doi.org/10.1136/jitc-2020-001157>
- Schwartz, O., Maréchal, V., Gall, S.L., Lemonnier, F., Heard, J.-M., 1996. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat. Med.* 2, 338–342. <https://doi.org/10.1038/nm0396-338>
- Sekaly, R.P., Jacobson, S., Richert, J.R., Tonnelie, C., McFarland, H.F., Long, E.O., 1988. Antigen presentation to HLA class II-restricted measles virus-specific T-cell clones can occur in the absence of the invariant chain. *Proc. Natl. Acad. Sci. U. S. A.* 85, 1209–1212. <https://doi.org/10.1073/pnas.85.4.1209>
- Sertznig, H., Hillebrand, F., Erkelenz, S., Schaal, H., Widera, M., 2018. Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet. *Virology* 516, 176–188. <https://doi.org/10.1016/j.virol.2018.01.011>
- Shatsky, I.N., Terenin, I.M., Smirnova, V.V., Andreev, D.E., 2018. Cap-Independent Translation: What's in a Name? *Trends Biochem. Sci.* 43, 882–895. <https://doi.org/10.1016/j.tibs.2018.04.011>

- Smith, C.C., Beckermann, K.E., Bortone, D.S., De Cubas, A.A., Bixby, L.M., Lee, S.J., Panda, A., Ganesan, S., Bhanot, G., Wallen, E.M., Milowsky, M.I., Kim, W.Y., Rathmell, W.K., Swanstrom, R., Parker, J.S., Serody, J.S., Selitsky, S.R., Vincent, B.G., 2018. Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. *J. Clin. Invest.* 128, 4804–4820. <https://doi.org/10.1172/JCI121476>
- Sorokin, I.I., Vassilenko, K.S., Terenin, I.M., Kalinina, N.O., Agol, V.I., Dmitriev, S.E., 2021. Non-Canonical Translation Initiation Mechanisms Employed by Eukaryotic Viral mRNAs. *Biochem. Biokhimiia* 86, 1060–1094. <https://doi.org/10.1134/S0006297921090042>
- Starck, S.R., Tsai, J.C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., Martins-Green, M., Shastri, N., Walter, P., 2016. Translation from the 5' untranslated region shapes the integrated stress response. *Science* 351, aad3867. <https://doi.org/10.1126/science.aad3867>
- Stefic, K., Bouvin-Pley, M., Braibant, M., Barin, F., 2019. Impact of HIV-1 Diversity on Its Sensitivity to Neutralization. *Vaccines* 7, 74. <https://doi.org/10.3390/vaccines7030074>
- Stern-Ginossar, N., Thompson, S.R., Mathews, M.B., Mohr, I., 2019. Translational Control in Virus-Infected Cells. *Cold Spring Harb. Perspect. Biol.* 11, a033001. <https://doi.org/10.1101/cshperspect.a033001>
- Stern-Ginossar, N., Weisburd, B., Michalski, A., Le, V.T.K., Hein, M.Y., Huang, S.-X., Ma, M., Shen, B., Qian, S.-B., Hengel, H., Mann, M., Ingolia, N.T., Weissman, J.S., 2012. Decoding human cytomegalovirus. *Science* 338, 1088–1093. <https://doi.org/10.1126/science.1227919>
- Stumptner-Cuvelette, P., Morchoisne, S., Dugast, M., Le Gall, S., Raposo, G., Schwartz, O., Benaroch, P., 2001. HIV-1 Nef impairs MHC class II antigen presentation and surface expression. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12144–12149. <https://doi.org/10.1073/pnas.221256498>
- Tanaka, M., Ueno, T., Nakahara, T., Sasaki, K., Ishimoto, A., Sakai, H., 2003. Downregulation of CD4 is required for maintenance of viral infectivity of HIV-1. *Virology* 311, 316–325. [https://doi.org/10.1016/S0042-6822\(03\)00126-0](https://doi.org/10.1016/S0042-6822(03)00126-0)
- Ternette, N., Yang, H., Partridge, T., Llano, A., Cedeño, S., Fischer, R., Charles, P.D., Dudek, N.L., Mothe, B., Crespo, M., Fischer, W.M., Korber, B.T.M., Nielsen, M., Borrow, P., Purcell, A.W., Brander, C., Dorrell, L., Kessler, B.M., Hanke, T., 2016. Defining the HLA class I-associated viral antigen repertoire from HIV-1-infected human cells. *Eur. J. Immunol.* 46, 60–69. <https://doi.org/10.1002/eji.201545890>
- Tewari, M.K., Sinnathamby, G., Rajagopal, D., Eisenlohr, L.C., 2005. A cytosolic pathway for MHC class II-restricted antigen processing that is proteasome and TAP dependent. *Nat. Immunol.* 6, 287–294. <https://doi.org/10.1038/ni1171>
- Thiele, F., Tao, S., Zhang, Y., Muschawekch, A., Zollmann, T., Protzer, U., Abele, R., Drexler, I., 2015. Modified vaccinia virus Ankara-infected dendritic cells present CD4+ T-cell epitopes by endogenous major histocompatibility complex class II presentation pathways. *J. Virol.* 89, 2698–2709. <https://doi.org/10.1128/JVI.03244-14>
- Tsuji, T., Matsuzaki, J., Caballero, O.L., Jungbluth, A.A., Ritter, G., Odunsi, K., Old, L.J., Gnjjatic, S., 2012. Heat Shock Protein 90-Mediated Peptide-Selective Presentation of Cytosolic Tumor Antigen for Direct Recognition of Tumors by CD4+ T Cells. *J. Immunol.* 188, 3851–3858. <https://doi.org/10.4049/jimmunol.1103269>
- Vanhée-Brossollet, C., Thoreau, H., Serpente, N., D'Auriol, L., Lévy, J.-P., Vaquero, C., 1995. A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. *Virology* 206, 196–202. [https://doi.org/10.1016/S0042-6822\(95\)80034-4](https://doi.org/10.1016/S0042-6822(95)80034-4)
- Vingert, B., Benati, D., Lambotte, O., de Truchis, P., Slama, L., Jeannin, P., Galperin, M., Perez-Patrigéon, S., Boufassa, F., Kwok, W.W., Lemaître, F., Delfraissy, J.-F., Thèze, J., Chakrabarti, L.A., 2012. HIV controllers maintain a population of highly efficient Th1 effector cells in contrast to patients treated in the long term. *J. Virol.* 86, 10661–10674. <https://doi.org/10.1128/JVI.00056-12>
- W. Yewdell, J., 2005. The seven dirty little secrets of major histocompatibility complex class I antigen processing. *Immunol. Rev.* 207, 8–18. <https://doi.org/10.1111/j.0105-2896.2005.00309.x>
- Wahren, B., Rosen, J., Sandström, E., Mathiesen, T., Modrow, S., Wigzell, H., 1989. HIV-1 peptides induce a proliferative response in lymphocytes from infected persons. *J. Acquir. Immune Defic. Syndr.* 2, 448–456.
- Weingarten-Gabbay, S., Klaeger, S., Sarkizova, S., Pearlman, L.R., Chen, D.-Y., Gallagher, K.M.E., Bauer, M.R., Taylor, H.B., Dunn, W.A., Tarr, C., Sidney, J., Rachimi, S., Conway, H.L., Katsis, K., Wang, Y., Leistriz-Edwards, D., Durkin, M.R., Tomkins-Tinch, C.H., Finkel, Y., Nachshon, A., Gentili, M., Rivera, K.D., Carulli, I.P., Chea, V.A., Chandrashekar, A., Bozkus, C.C., Carrington, M., Lavin-Parsons, K., Parry, B., Lilley, B., Lodenstein, C., McKaig, B., Charland, N., Khanna, H., Margolin, J., Gonye, A., Gushterova, I., Lasalle, T., Sharma, N., Russo, B.C., Rojas-Lopez, M., Sade-Feldman, M., Manakongtreecheep, K., Tantivit, J., Fisher Thomas, M., Bhardwaj, N., Barouch, D.H., Sette, A., Maus, M.V., Rice, C.M., Clauser, K.R., Keskin, D.B., Pregibon, D.C., Hacohen, N., Carr, S.A., Abelin, J.G., Saeed, M., Sabeti, P.C., 2021. Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell* 184, 3962–3980.e17. <https://doi.org/10.1016/j.cell.2021.05.046>
- Wethmar, K., 2014. The regulatory potential of upstream open reading frames in eukaryotic gene expression. *WIREs RNA* 5, 765–768. <https://doi.org/10.1002/wrna.1245>

- Whisnant, A.W., Jürges, C.S., Hennig, T., Wyler, E., Prusty, B., Rutkowski, A.J., L'hernault, A., Djakovic, L., Göbel, M., Döring, K., Menegatti, J., Antrobus, R., Matheson, N.J., Künzig, F.W.H., Mastrobuoni, G., Bielow, C., Kempa, S., Liang, C., Dandekar, T., Zimmer, R., Landthaler, M., Grässer, F., Lehner, P.J., Friedel, C.C., Erhard, F., Dölken, L., 2020. Integrative functional genomics decodes herpes simplex virus 1. *Nat. Commun.* 11, 2038. <https://doi.org/10.1038/s41467-020-15992-5>
- Wolin, S.L., Walter, P., 1988. Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J.* 7, 3559–3569. <https://doi.org/10.1002/j.1460-2075.1988.tb03233.x>
- Wright, B.W., Yi, Z., Weissman, J.S., Chen, J., 2022. The dark proteome: translation from noncanonical open reading frames. *Trends Cell Biol.* 32, 243–258. <https://doi.org/10.1016/j.tcb.2021.10.010>
- Yamamoto, T., Iwamoto, N., Yamamoto, H., Tsukamoto, T., Kuwano, T., Takeda, A., Kawada, M., Tsunetsugu-Yokota, Y., Matano, T., 2009. Polyfunctional CD4+ T-Cell Induction in Neutralizing Antibody-Triggered Control of Simian Immunodeficiency Virus Infection. *J. Virol.* 83, 5514–5524. <https://doi.org/10.1128/JVI.00145-09>
- Yang, N., Gibbs, J.S., Hickman, H.D., Reynoso, G.V., Ghosh, A.K., Bennink, J.R., Yewdell, J.W., 2016. Defining Viral Defective Ribosomal Products: Standard and Alternative Translation Initiation Events Generate a Common Peptide from Influenza A Virus M2 and M1 mRNAs. *J. Immunol. Baltim. Md 1950* 196, 3608–3617. <https://doi.org/10.4049/jimmunol.1502303>
- Yang, Y., Wang, Z., 2019. IRES-mediated cap-independent translation, a path leading to hidden proteome. *J. Mol. Cell Biol.* 11, 911–919. <https://doi.org/10.1093/jmcb/mjz091>
- Yang, Z., Cao, S., Martens, C.A., Porcella, S.F., Xie, Z., Ma, M., Shen, B., Moss, B., 2015. Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling. *J. Virol.* 89, 6874–6886. <https://doi.org/10.1128/JVI.00528-15>
- Yates, A.D., Achuthan, P., Akanni, W., Allen, James, Allen, Jamie, Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J.C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Lavidas, I., Le, T., Lemos, D., Martinez, J.G., Maurel, T., McDowall, M., McMahan, A., Mohanan, S., Moore, B., Nuhn, M., Oheh, D.N., Parker, A., Parton, A., Patricio, M., Sakthivel, M.P., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sycheva, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Flint, B., Frankish, A., Hunt, S.E., Ilesley, G., Kostadima, M., Langridge, N., Loveland, J.E., Martin, F.J., Morales, J., Mudge, J.M., Muffato, M., Perry, E., Ruffier, M., Trevanion, S.J., Cunningham, F., Howe, K.L., Zerbino, D.R., Flicek, P., 2020. Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688. <https://doi.org/10.1093/nar/gkz966>
- Yewdell, J.W., 2022. MHC Class I Immunopeptidome: Past, Present, and Future. *Mol. Cell. Proteomics MCP* 21, 100230. <https://doi.org/10.1016/j.mcpro.2022.100230>
- Yewdell, J.W., Antón, L.C., Bennink, J.R., 1996. Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? *J. Immunol. Baltim. Md 1950* 157, 1823–1826.
- Yewdell, J.W., Nichchitta, C.V., 2006. The DRiP hypothesis decennial: support, controversy, refinement and extension. *Trends Immunol.* 27, 368–373. <https://doi.org/10.1016/j.it.2006.06.008>
- Zanker, D.J., Oveissi, S., Tschärke, D.C., Duan, M., Wan, S., Zhang, X., Xiao, K., Mifsud, N.A., Gibbs, J., Izzard, L., Dlugolenski, D., Faou, P., Laurie, K.L., Vigneron, N., Barr, I.G., Stambas, J., Van den Eynde, B.J., Bennink, J.R., Yewdell, J.W., Chen, W., 2019. Influenza A Virus Infection Induces Viral and Cellular Defective Ribosomal Products Encoded by Alternative Reading Frames. *J. Immunol. Baltim. Md 1950* 202, 3370–3380. <https://doi.org/10.4049/jimmunol.1900070>
- Zhang, J., Crumpacker, C., 2022. HIV UTR, LTR, and Epigenetic Immunity. *Viruses* 14, 1084. <https://doi.org/10.3390/v14051084>
- Zheng, N., Fujiwara, M., Ueno, T., Oka, S., Takiguchi, M., 2009. Strong ability of Nef-specific CD4+ cytotoxic T cells to suppress human immunodeficiency virus type 1 (HIV-1) replication in HIV-1-infected CD4+ T cells and macrophages. *J. Virol.* 83, 7668–7677. <https://doi.org/10.1128/JVI.00513-09>
- Ziegler, M.C., Nelde, A., Weber, J.K., Schreitmüller, C.M., Martrus, G., Huynh, T., Bunders, M.J., Lunemann, S., Stevanovic, S., Zhou, R., Altfeld, M., 2020. HIV-1 induced changes in HLA-C*03:04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS Lond. Engl.* 34, 1713–1723. <https://doi.org/10.1097/QAD.0000000000002596>
- Zong, L., Qin, Y., Jia, H., Ye, L., Wang, Y., Zhang, J., Wands, J.R., Tong, S., Li, J., 2017. Differential regulation of hepatitis B virus core protein expression and genome replication by a small upstream open reading frame and naturally occurring mutations in the precore region. *Virology* 505, 155–161. <https://doi.org/10.1016/j.virol.2017.02.020>

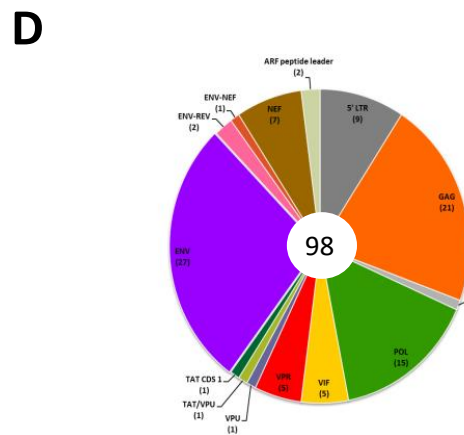
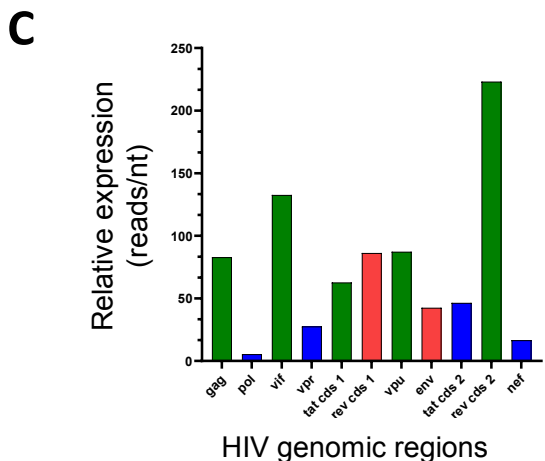
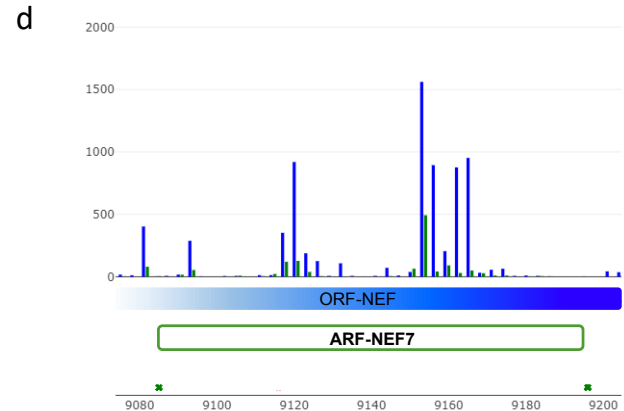
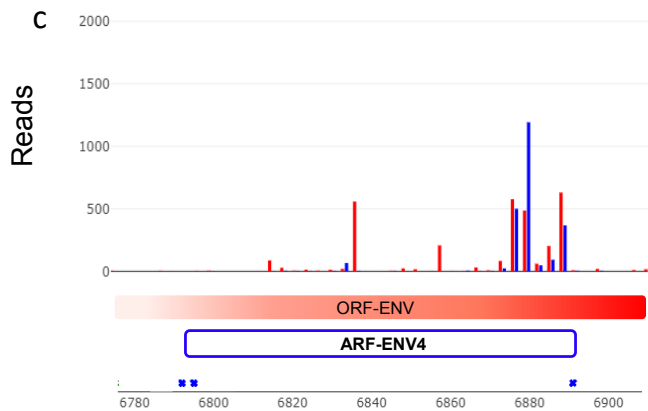
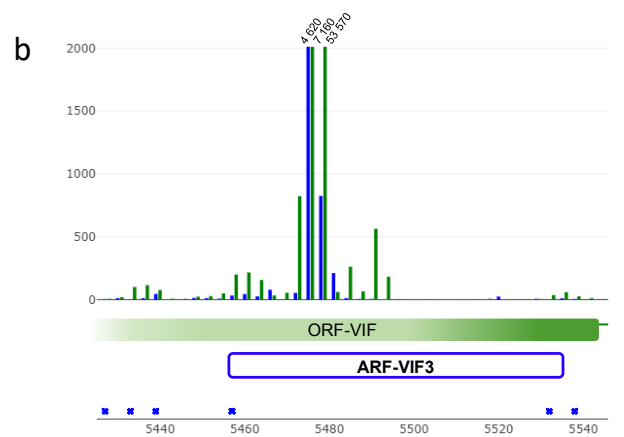
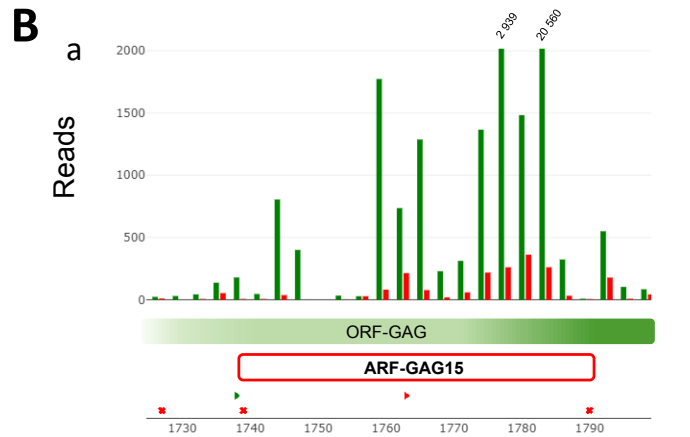
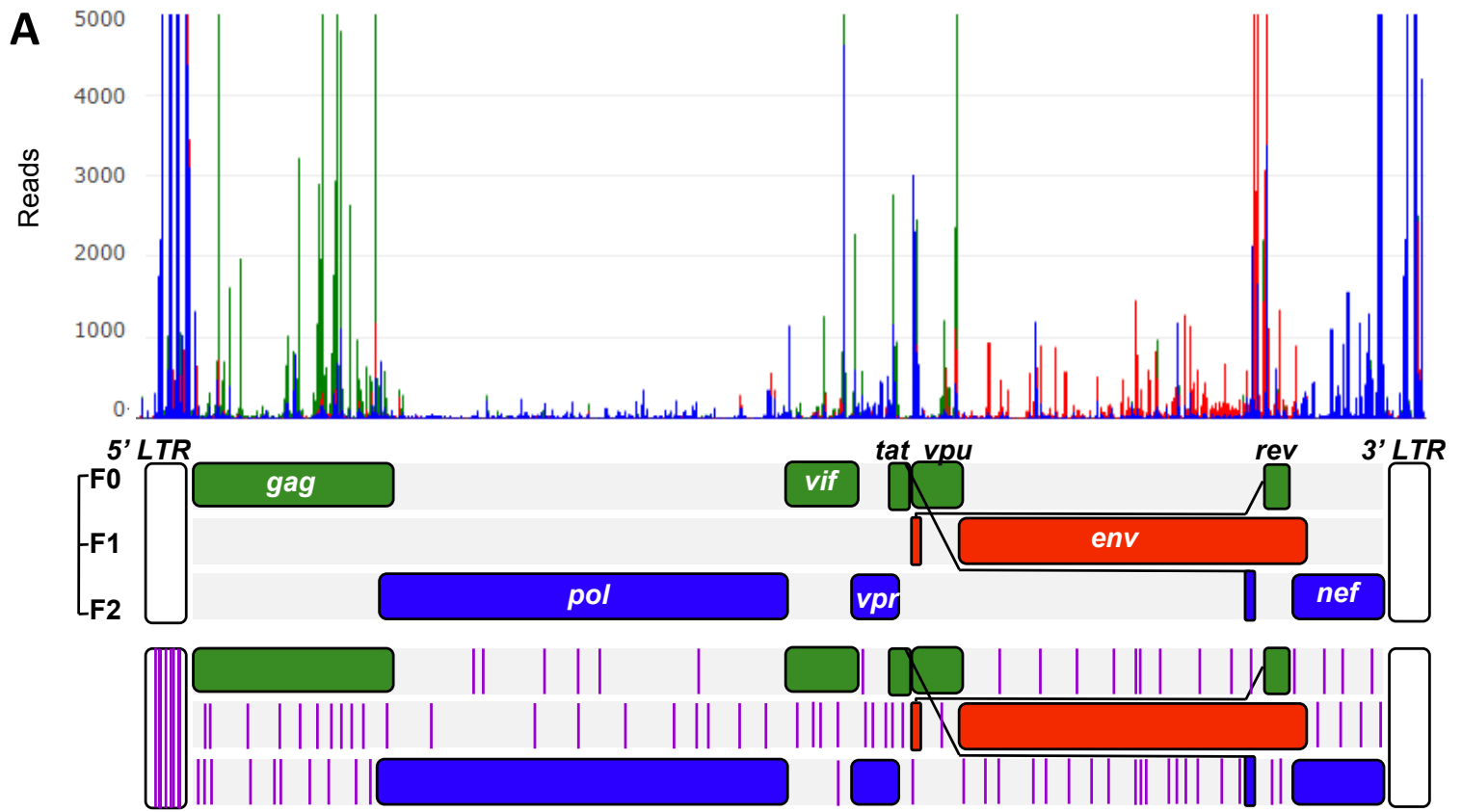


Figure 1

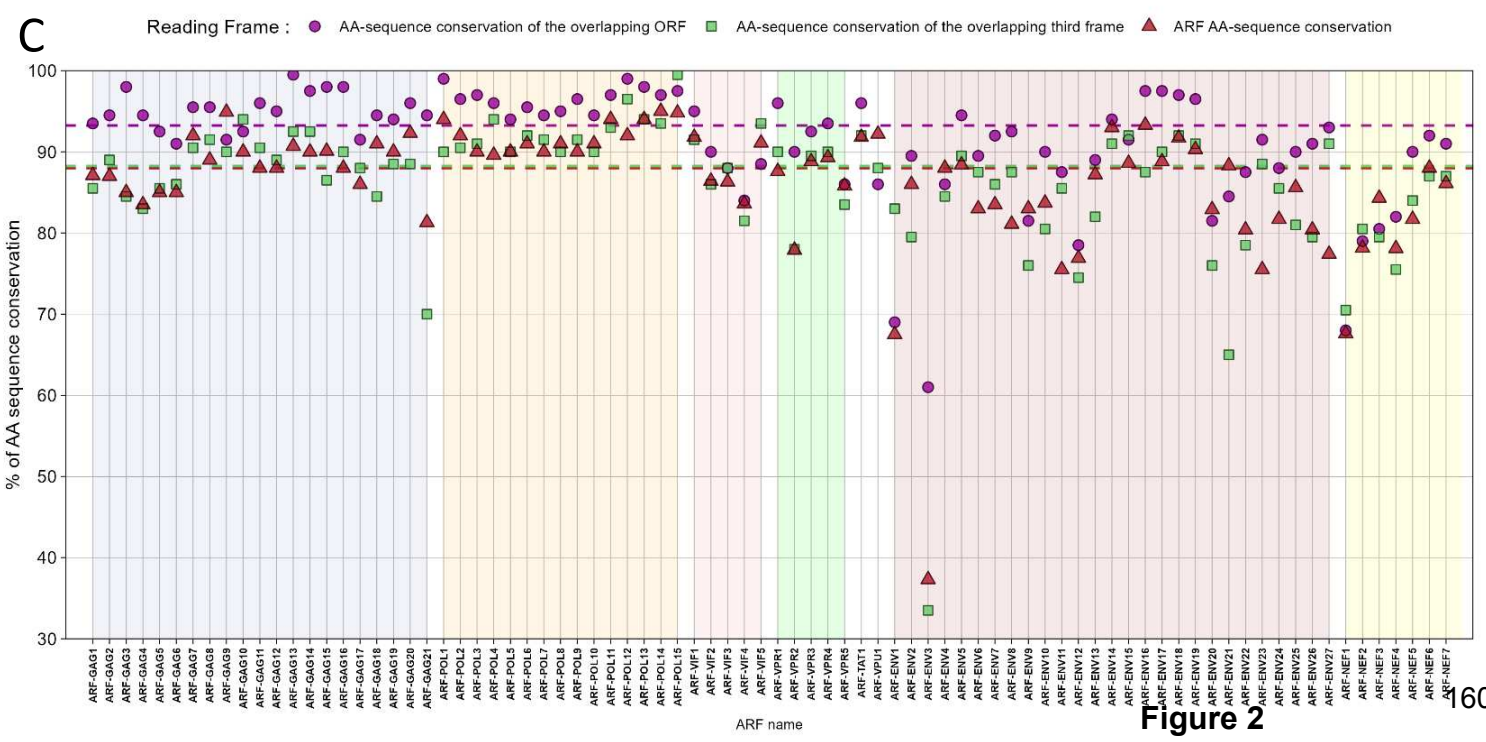
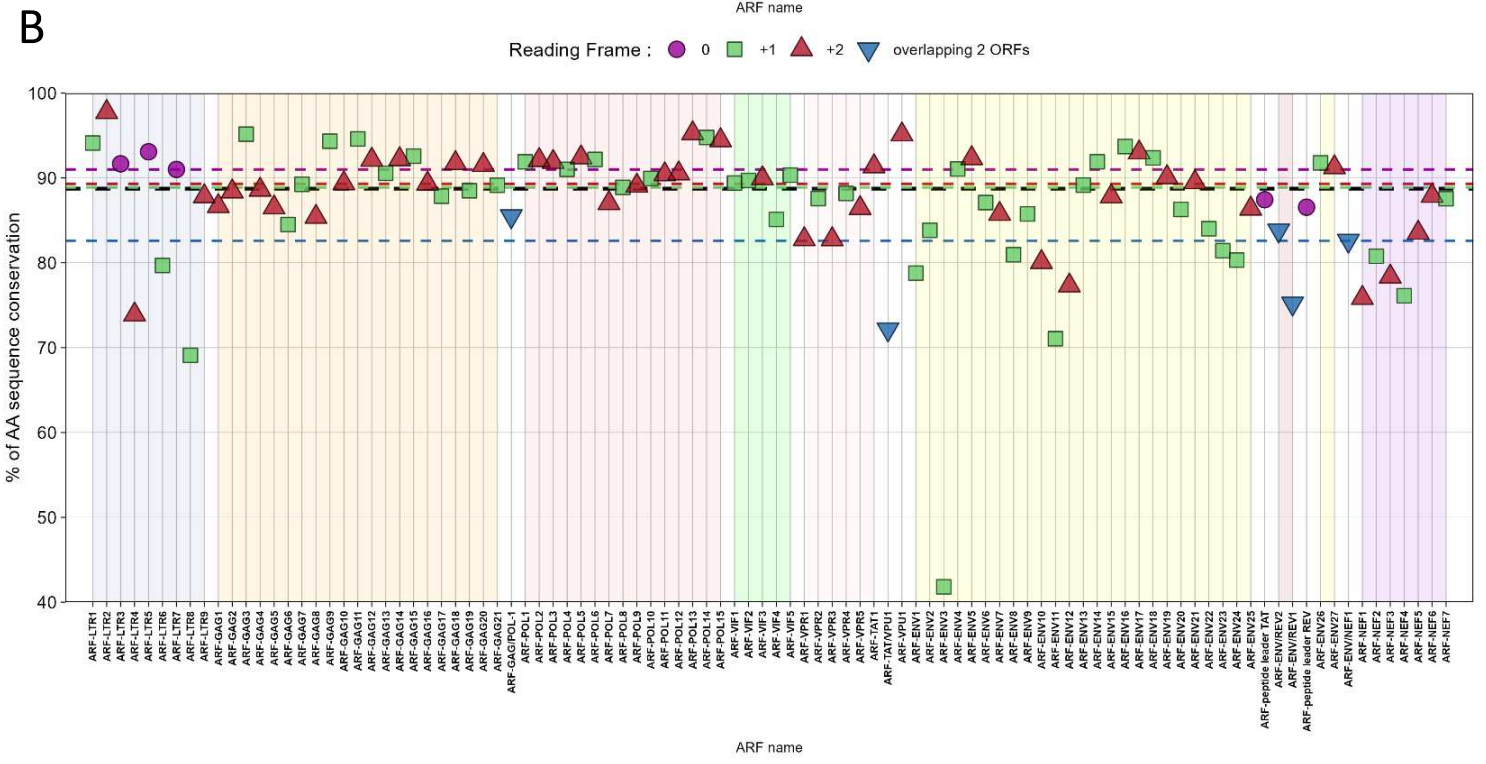
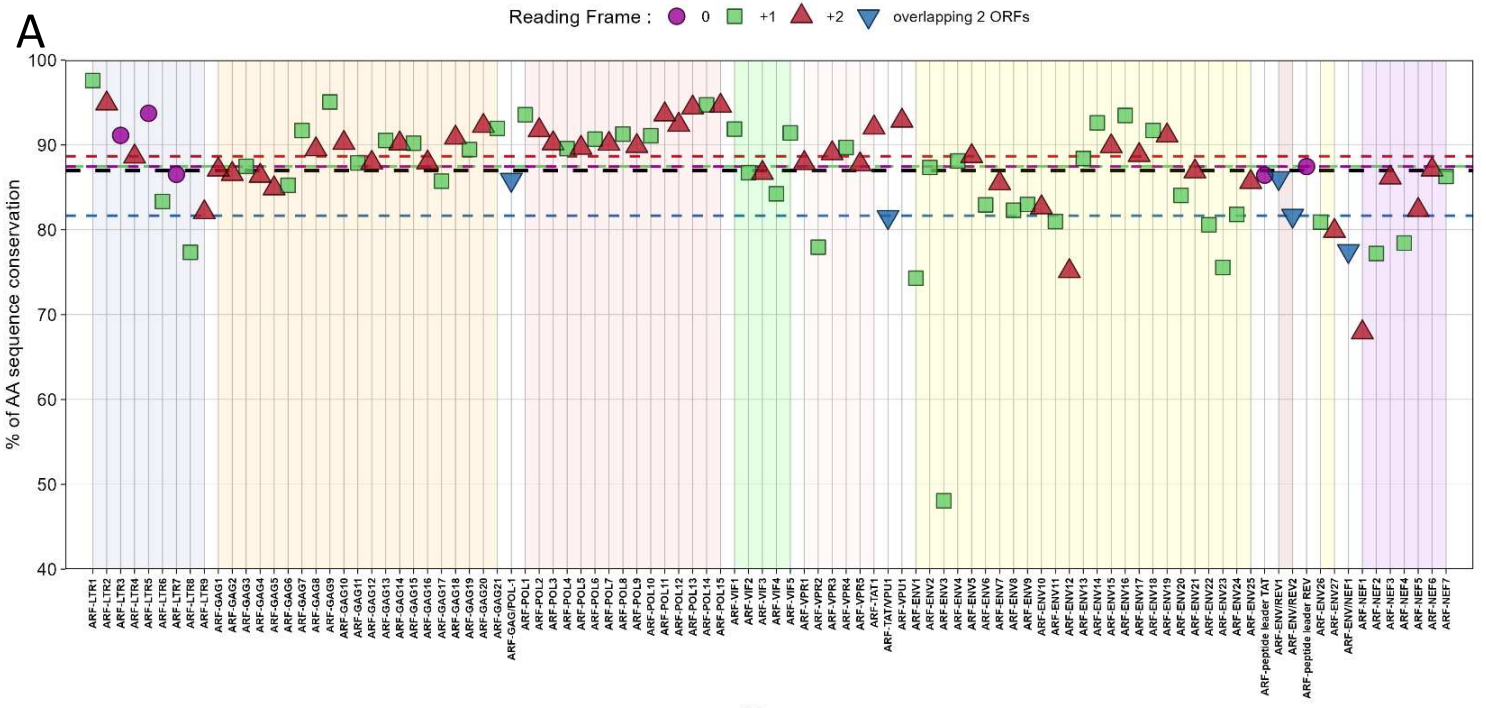


Figure 2 160

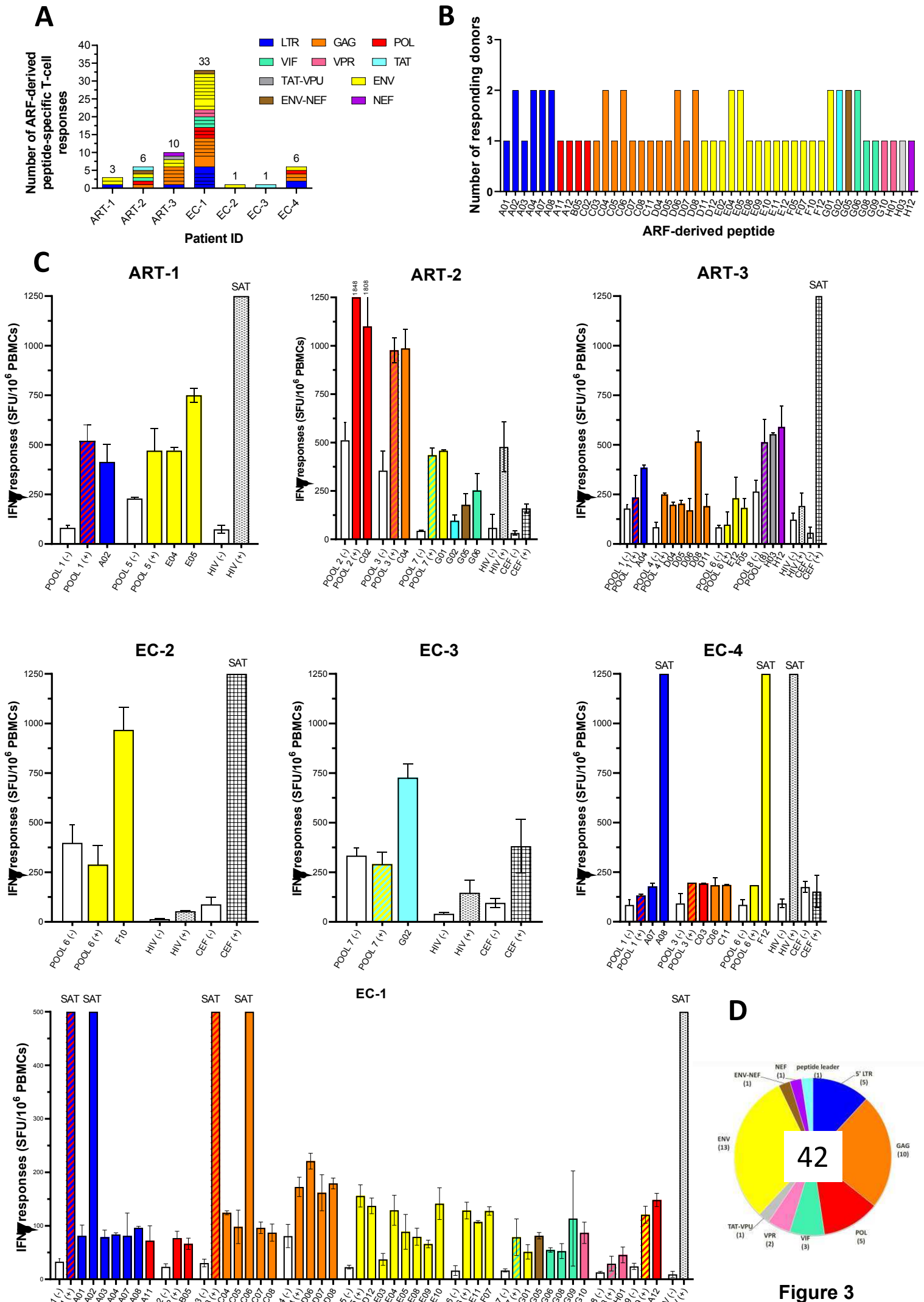


Figure 3

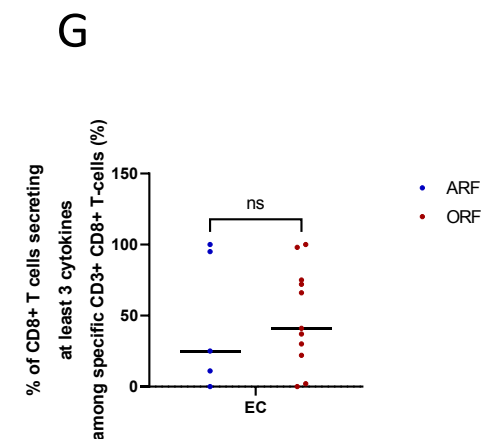
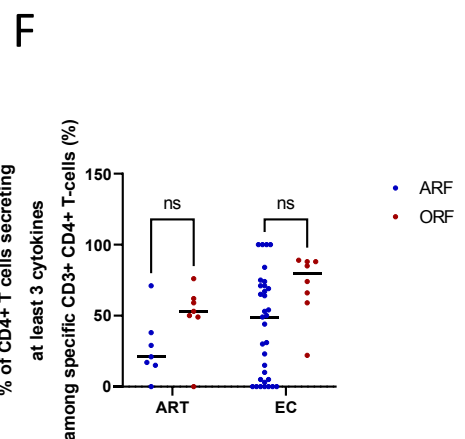
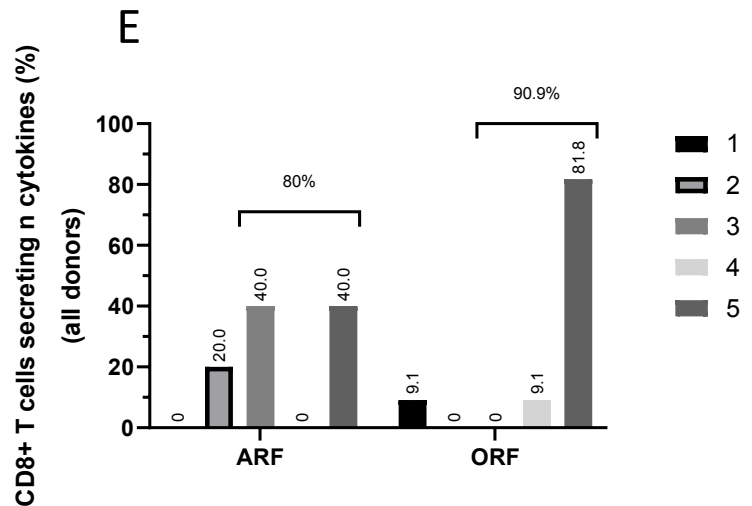
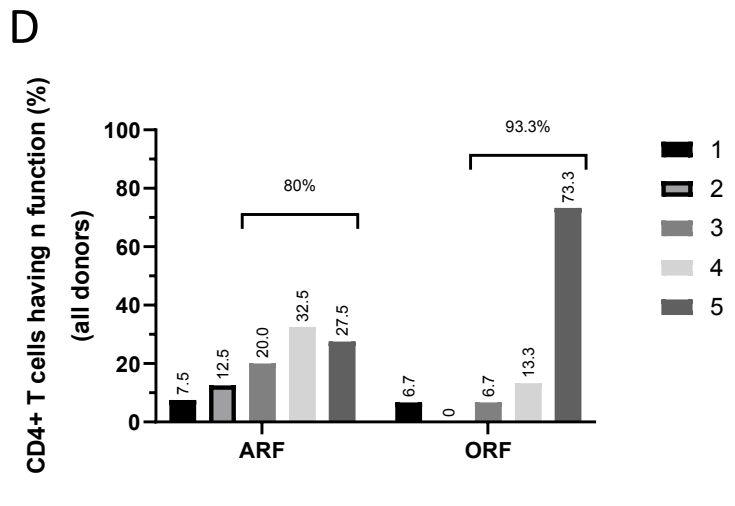
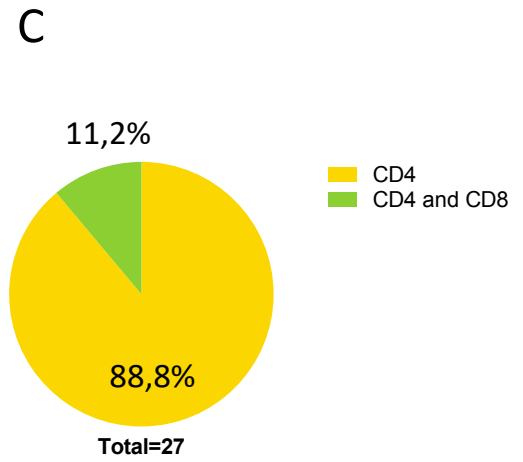
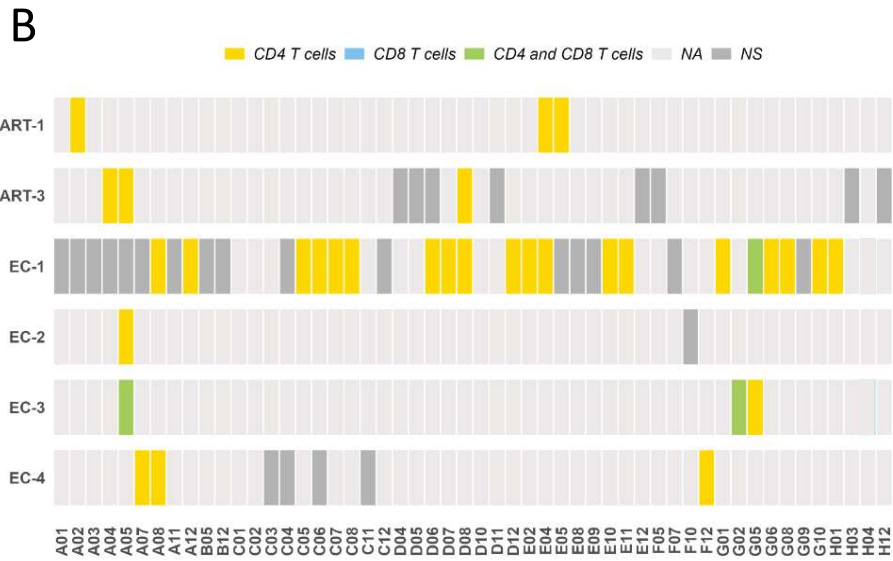
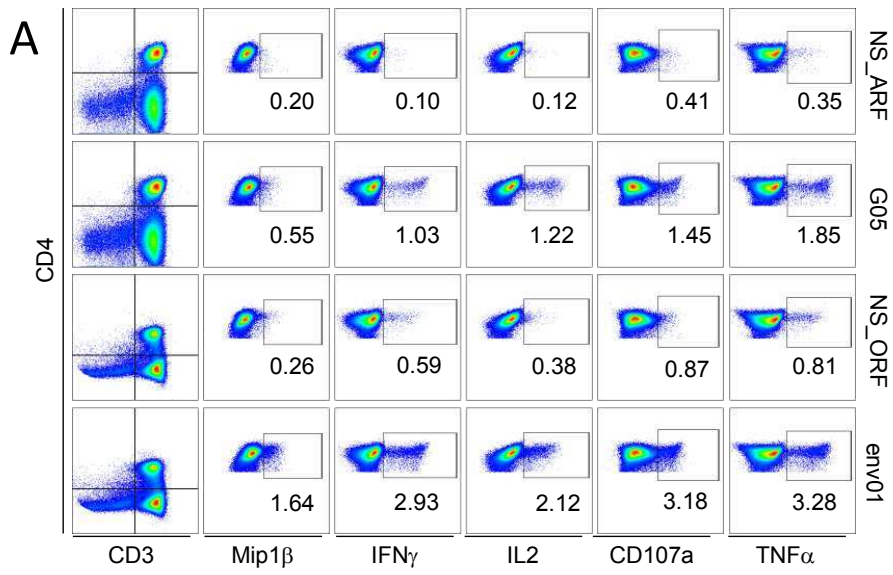


Figure 4

A

Sample	HIV strain	HLA type	HIV ligands	Frame	5' Position (nt)
Primary CD4+ T-cells	NL4-3	A*02:01; B*27:05; B*40:01; C*02:02; C*03:04	ILINGQYSL	ARF	3253
			AVLSIVNR	<i>env</i> CDS	8312
			SESAIRNTL	<i>vif</i> CDS	5386
			SESAIRNTIL	<i>vif</i> CDS	5386
			KELYPLASL	<i>gag</i> CDS	2230
			VLAEAMSQV	<i>gag</i> CDS	1873
			GENTSLHPVSL	<i>nef</i> CDS	9261

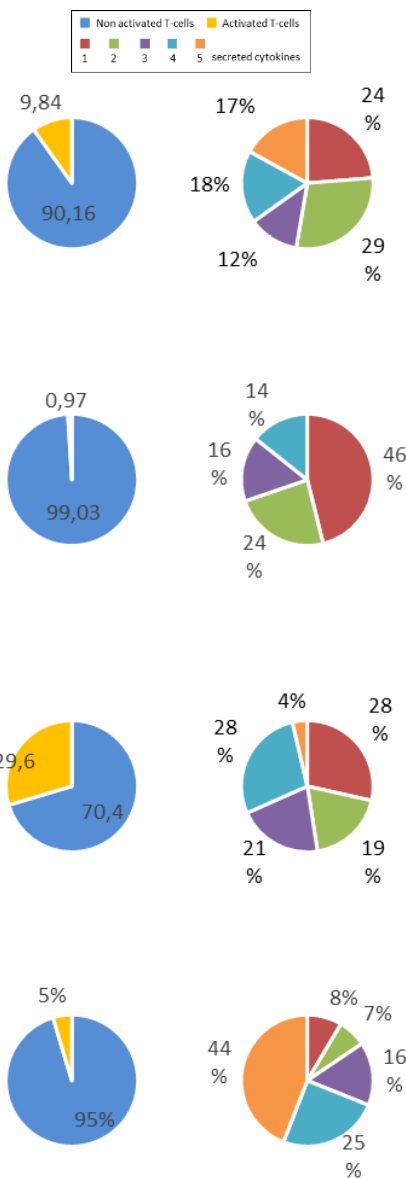
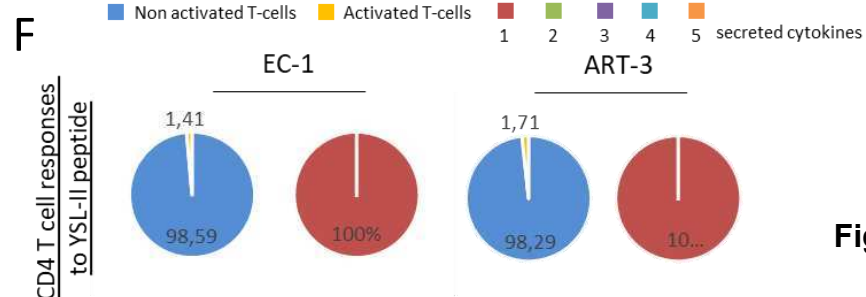
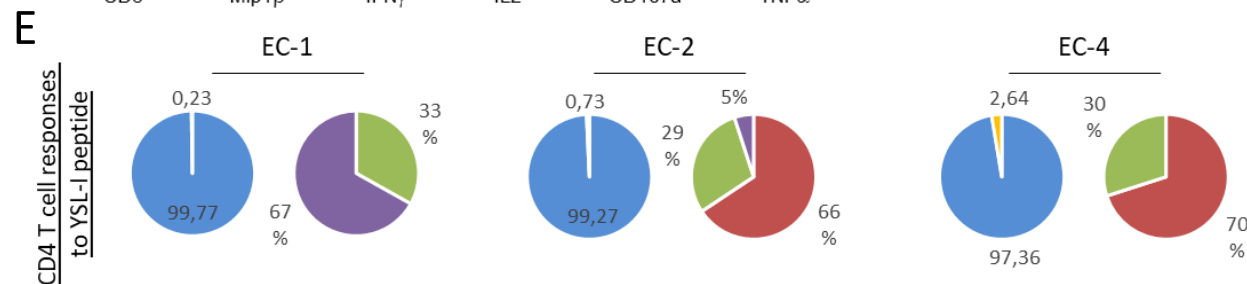
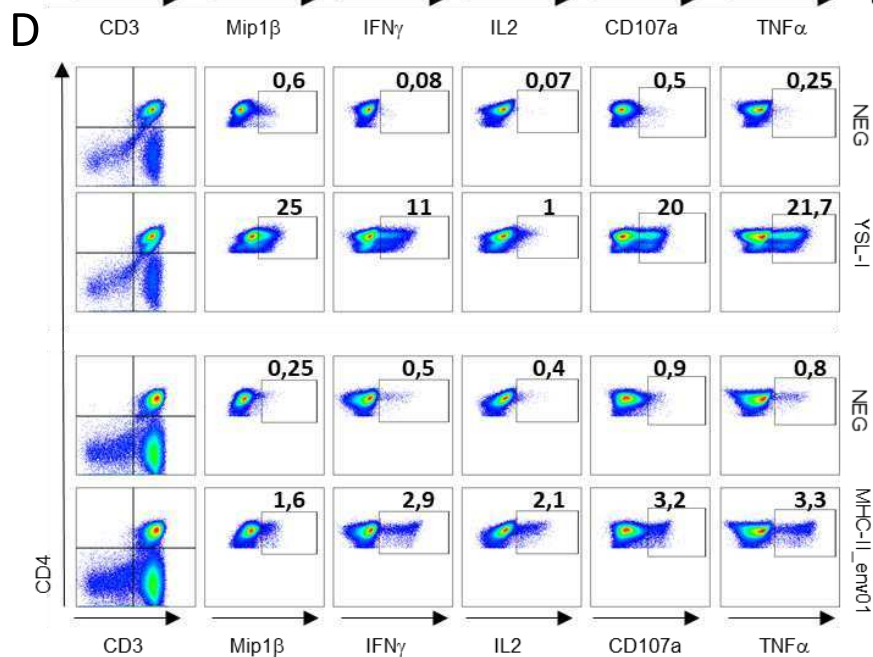
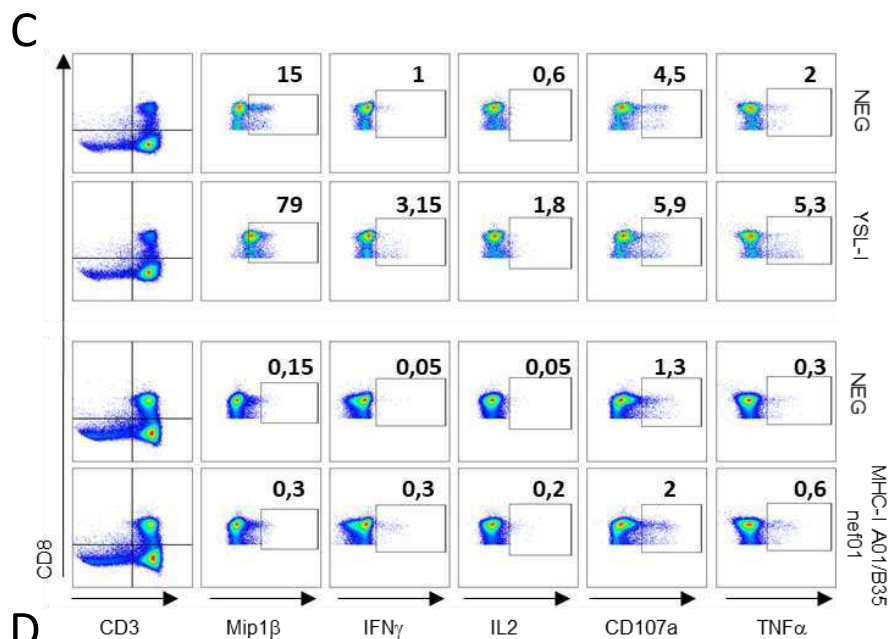
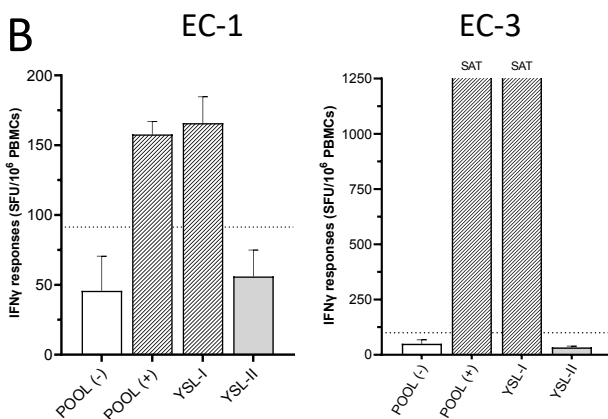


Figure 5

ARF name	sequence (ribosome profiling)	frame	length (nt)	length (AA)	start position	stop position	previously described
ARF-LTR1	GGAAACCCTGCTTAAGCCTCAATAAAGCTTGCCTTGAG	1	51	17	506	556	NO
ARF-LTR2	GCCTCAATAAAGCTTGCCTTGAAGTCTCAAGTAGTGTG	2	57	19	522	578	NO
ARF-LTR3	GTGCTCAAAGTAGTGTGCGCCCTGTTGTGTGACTCT	0	48	16	544	591	NO
ARF-LTR4	TCAGAGATCCCTCAGACCTTTTGTAGTGTGAAAAAT	2	45	15	588	632	NO
ARF-LTR5	TCAGTGTGAAAACTCTAGCAGTGGCCCGAACAGG	0	45	15	613	657	NO
ARF-LTR6	TGTGTGCGCCCTGTTGTGTGACTCGGTAAGTAGAGAT	1	153	51	557	709	NO
ARF-LTR7	AGCCAGAGGAGATCTCTCAGCAGGACTCGGCTTGTCT	0	78	26	670	747	NO
ARF-LTR8	AGCCGCACGGCAAGAGGCGAGGGCGGCGACTGGTC	1	60	20	710	769	NO
ARF-LTR9	CAGTGGCCCGAACAGGACTTGAAGCGAAAGTAAV	2	147	49	633	779	NO
ARF-GAG1	ATGGGAAAAAATTCGTTAAGCCAGGGGGAAGAAA	2	45	15	834	878	NO
ARF-GAG2	ACTAAAAATATAGTATGGGCAAGCAGGGAGCTAGAAC	2	51	17	879	929	YES
ARF-GAG3	AACGATTCCAGTAAATCTGGCCCTTTTAG	1	30	10	914	943	NO
ARF-GAG4	TCCTGGCCCTTTAGAGACATCAGAAGGCTGTAG	2	33	11	930	962	YES
ARF-GAG5	ACAAACTGGGACAGCTACAAACATCCCTTCAGACAGC	2	54	18	963	1016	NO
ARF-GAG6	AGACATCAGAAGGCTGTAGACAATACTGGGACAGCTA	1	93	31	944	1036	NO
ARF-GAG7	AAGAGACCATCAATGAGGAAAGCTCAGAAATGGGATAG	1	81	27	1394	1474	YES
ARF-GAG8	ATTGCATCCAGTGTGATGAGGGCCCTATTGCCAGGCCA	2	63	21	1431	1493	NO
ARF-GAG9	CAGGAACACTAGTACCTTCAGGAACAATAAG	1	33	11	1499	1531	NO
ARF-GAG10	TACCTTCAGGAACAATAAGGATGGATGACACATAA	2	36	12	1512	1547	NO
ARF-GAG11	GAATGTATAGCCCTACAGCATCTTGACATAA	1	33	11	1613	1645	NO
ARF-GAG12	CCCTACAGCATTCTGGACATAAGCAAGGACCAAAAGG	2	48	16	1623	1670	NO
ARF-GAG13	GACAAGGACCAAGGAAACCCCTTAGAGACTATGTAG	1	36	12	1646	1681	NO
ARF-GAG14	AACTCTAAGAGCCGAGCAAGCTTCAACAGAGGTAAGAA	2	87	29	1695	1781	NO
ARF-GAG15	CAGAAACCTTGTGGTCAAAATGCGAACCCAGATTGT	1	51	17	1742	1792	NO
ARF-GAG16	GACTATTTAAAAAGCATTGGGACAGGAGCAGACTAG	2	84	28	1782	1865	NO
ARF-GAG17	TACAGAAAGGCAATTTAGGAACCAAGAAAGACTGTT	1	72	24	1922	1993	NO
ARF-GAG18	GTGTTTCAATTTGGGCAAGAGGGCCACATAGCCAAAA	2	54	18	1962	2015	NO
ARF-GAG19	CCAAAAATTCAGGGCCCTAGGAAAAAGGGCTGTTGG	1	66	22	1994	2059	NO
ARF-GAG20	GAAAAAGGCTGTGGAAATGTGGAAAGGAAAGGACAC	2	57	19	2016	2072	YES
ARF-GAG21	AAGATTGTACTGAGAGACAGGCTAATTTTTAG	1	33	11	2060	2092	NO
ARF-GAG/POL-1	GGAAAGTCTGGCCCTCCCAAGGGAAGGCCAGGGAA	1	135	45	2093	2227	YES
ARF-POL1	CAGTCTGGATGTGGGCGATGCATATTTTTAGTCCCT	0	42	14	2869	2910	YES
ARF-POL2	GAGAACCTCAAGATTTCTGGGAAGTCAATAGGAATAC	1	120	40	2795	2914	NO
ARF-POL3	ATGGACAGTACAGCCTATAGTGTGCGCAAAAAGGACA	1	54	18	3263	3316	NO
ARF-POL4	TGCTGCCAGAAAAGCAGCTGGAATGCAATGACATA	0	48	16	3283	3330	YES
ARF-POL5	CATACAGAAATAGTGGGAAAATGAATGGGCAAGTC	1	57	19	3317	3373	NO
ARF-POL6	ATTGGGCAAGTCAAGATTTATGACAGGATTAAGTAA	0	36	12	3343	3378	NO
ARF-POL7	AAATCTGAAAACAGGAAAGTATCAAGAATGAAGGGTC	1	48	16	3590	3637	NO
ARF-POL8	CGGACACAACTATCAGAAGACTGAGTTACAAGCAATT	0	45	15	3958	4002	YES
ARF-POL9	GTACAAGCAATTCATAGCTTGTGAGGATTCGGGATT	1	102	34	3983	4084	NO
ARF-POL10	ATAAATGGTCAAGTGGAAATCAGGAAAGTACTATTT	0	42	14	4195	4236	NO
ARF-POL11	ATGTACAGCTAAAAGGGGAAAGCCATGCATGGACAAGT	1	45	15	4355	4399	NO
ARF-POL12	AGAATTAAGAAAAATTAAGGACAGGTAAGGATCAGG	1	42	14	4697	4738	NO
ARF-POL13	AGAATTAACAAAAATAATTAACAAAAATTAACAAATTT	1	105	35	4862	4966	YES
ARF-POL14	CAACAGACATACAACTAAGAAATTAACAAAAATAAT	0	135	45	4843	4977	YES
ARF-POL15	CATAAAGTAGTCCCAAGAAAGAAAGCAAGATCATCA	1	66	22	4997	5062	NO
ARF-VIF1	CAACATATTGGGGTCTGCATACAGGAGAAAGAGACTGG	1	63	21	5240	5302	NO
ARF-VIF2	TTAGTCTAGGTTGTAATATCAAGCAGGACATAACAGG	1	42	14	5426	5467	NO
ARF-VIF3	CAAGGTAGGATCTCTCAGTACTGGCACTAGCAGCATT	2	75	25	5460	5534	NO
ARF-VIF4	AGCCACCTTTGGCTAGTGTAGGAAACTGA	1	30	10	5519	5548	NO
ARF-VIF5	CAGAGGACAGATGGAACAGCCCAAGACCAAGGG	1	60	20	5549	5608	NO
ARF-VPR1	ATGGGACACTAGAGCTTTTAGAGGAACTTAA	1	30	10	5609	5638	NO
ARF-VPR2	AGGAACTTAAGAGTGAAGCTGTAGACATTTCTAGGA	0	57	19	5629	5685	NO
ARF-VPR3	AACTTACGGGGTACTTGGGCAAGGAGTGGAAAGCATA	1	93	31	5702	5794	NO
ARF-VPR4	GAATTCGCAACAAGTCTGTTTATCCATTTAGAAATG	0	60	20	5743	5802	NO
ARF-VPR5	CAGAATAGGCGTACTCGACAGAGGAGAGCAAGAAAT	1	48	16	5795	5842	NO
ARF-TAT1	ACTAGAGCCCTGGGAAAGCTCCAGGAAAGTCAAGCTAA	2	36	12	5850	5885	NO
ARF-TAT/VP1	AAAGTGTGCTTTCATTCGCAAGTTTGTTCATGACAAA	2	237	79	5913	6149	NO
ARF-VP1	ACTAATAAGAAAGCAGAAAGGAGGCAAGTGA	2	33	11	6192	6224	NO
ARF-ENV1	AGGAGAAGTATCAGCATTGTGGAGATGGGGTGGGAA	2	66	22	6231	6296	NO
ARF-ENV2	TCTGTAGTGTACAGAAAAATTTGGGTCACAGTCTATT	2	174	58	6300	6473	YES
ARF-ENV3	AGAATGATACTAATACCAATAGTAGCGGGGAGAAATG	2	39	13	6624	6662	NO
ARF-ENV4	GTGTAACACTCAGTCAATACAGAGCCTGCCAAAGG	2	96	32	6798	6893	NO
ARF-ENV5	GCCAAATCCCATATTAATGTGCGCCGCTGTTTGTGC	0	54	18	6847	6900	NO
ARF-ENV6	AATGTAATAAAGACGTTCAATGGAAACAGGACCATGT	2	84	28	6894	6977	YES
ARF-ENV7	GACGTTCAATGGAAACAGGACCATGTACAAATGTGAG	0	126	42	6907	7032	NO
ARF-ENV8	TCITTAAGCAATCTCAGGAGGGGACCCAGAAATGTAA	2	39	13	7293	7331	NO
ARF-ENV9	GAAAAAGCAATGTATGCCCTCCCATCAGTGGACAAAT	2	72	24	7506	7577	YES
ARF-ENV10	ATGTTCACTAAATTAAGTGGGCTGTTAACAAGAGA	0	48	16	7546	7593	NO
ARF-ENV11	CAAGAGATGGTGGTAATAACAACATGGGTCAGAGATC	2	63	21	7578	7640	YES
ARF-ENV12	CAACAATGGTCCGAGATCTCAGACCTGGAGGAGGCC	0	63	21	7597	7659	NO
ARF-ENV13	CACCCCAAGGCAAGGAGAAAGTGTGACAGAGAGA	2	57	19	7704	7760	NO
ARF-ENV14	GAGCTTTGCTTCCGTTGGTCTTGGGAGCAGCAGGAAG	2	60	20	7761	7820	NO
ARF-ENV15	ACCATTAGGAGTAGCACCACCAAGGCAAGAGAGAGAC	0	165	55	7690	7854	NO
ARF-ENV16	CGGTACAGGCGAGCAATATTGTCTGATATAG	2	33	11	7827	7859	NO
ARF-ENV17	TATAGTGCAGCAGCAAAATTTGCTGAGGGCTATTG	0	39	13	7855	7893	NO
ARF-ENV18	GGGCTATTGAGGCGCAACAGCATCTGTGCAACTCACAT	2	93	31	7884	7976	NO
ARF-ENV19	GGGCGCAACAGCATCTGTGCAACTCAGCTGGGGCA	0	159	53	7894	8052	NO
ARF-ENV20	AGGATCAACAGCTCTGGGGATTGGGGTTGCTCTGGA	2	117	39	7977	8093	NO
ARF-ENV21	CATGACCTGGATGGAGTGGGACAGAGAAATTA	0	33	11	8089	8121	NO
ARF-ENV22	CTGGATGGAGTGGGACAGAGAAATTAACAATACACA	2	45	15	8094	8138	NO
ARF-ENV23	TTGAAAGATCGAAAAACGCAAGAAAAAGAAATGAACAA	2	54	18	8151	8204	NO
ARF-ENV24	ATAAATGGGCAAGTTTGGGAATTTGGTTAAACATAA	2	36	12	8205	8240	NO
ARF-ENV25	CATAACAATGGCTGGTATATAAAAATTTATCATAAT	0	99	33	8236	8334	NO
ARF-ENV26	AGAATAGTCTGTAACTGTCTCAATGCCACAGCCATAG	2	39	13	8637	8675	NO
ARF-ENV27	CTTGCTCAATGCCACAGCCATGACAGTACTGTA	0	33	11	8653	8685	NO
ARF-ENV/REV1	CACTTATCGGGACGATCTCGGAGCCTGTGCTCTTCA	2	54	18	8475	8528	NO
ARF-ENV/REV2	CGAGGATTTGGAACTTCTGGGACCGAGGGGTTGGGA	2	87	29	8550	8636	NO
ARF-ENV/NEF1	AAGAATAAGACGGCTTGGAAAGGATTTTGTATAAG	0	69	23	8749	8817	NO
ARF-NEF1	TGTGATGGATGGCTGCTGTAAGGAAAAAGATGAGAC	1	45	15	8813	8857	NO
ARF-NEF2	GACGAGCTAGCCAGCAGCAGATGGGTTGGGAGCAGT	0	51	17	8848	8898	NO
ARF-NEF3	GCCAGCAGGATGGGTTGGGAGCAGTATCTGAGAC	1	66	22	8858	8923	NO
ARF-NEF4	AAAAACATGGAGCAATCAACAAGTAGCAATACAGCAGCT	0	63	21	8899	8961	NO
ARF-NEF5	CAATGCTCTTGTGCTGCTGAGAAAGCACAAGAGGAGG	1	111	37	8939	9049	NO
ARF-NEF6	CCACTTTAAAAAGAAAGGGGGGACTGGAAGGGCTAA	1	69	23	9050	9118	NO
ARF-NEF7	TTCACTCCAAAGAACAGATATCTGTATCTGTGGA	0	111	37	9088	9198	NO
ARF-peptide leader TAT	ATAGAGTTAGGCAAGGATATCCACATTCGTTTCAGA	2	84	28	8331	8414	YES
ARF-peptide leader REV	GCAGGGATATCCACATTCGTTTCAGACCCACTCCC	0	303	101	8341	8643	YES

Table 1

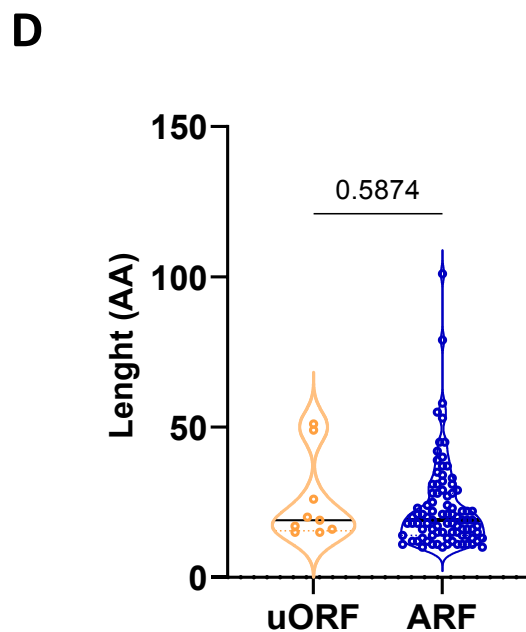
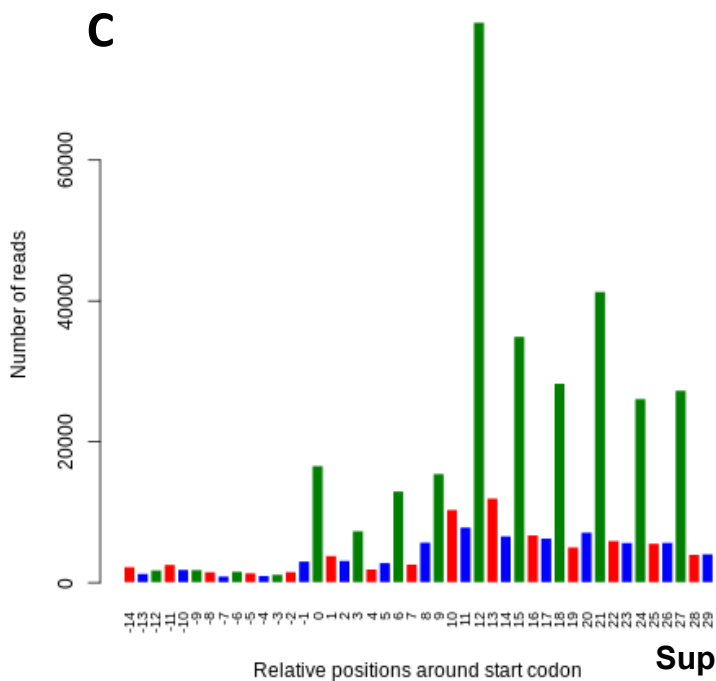
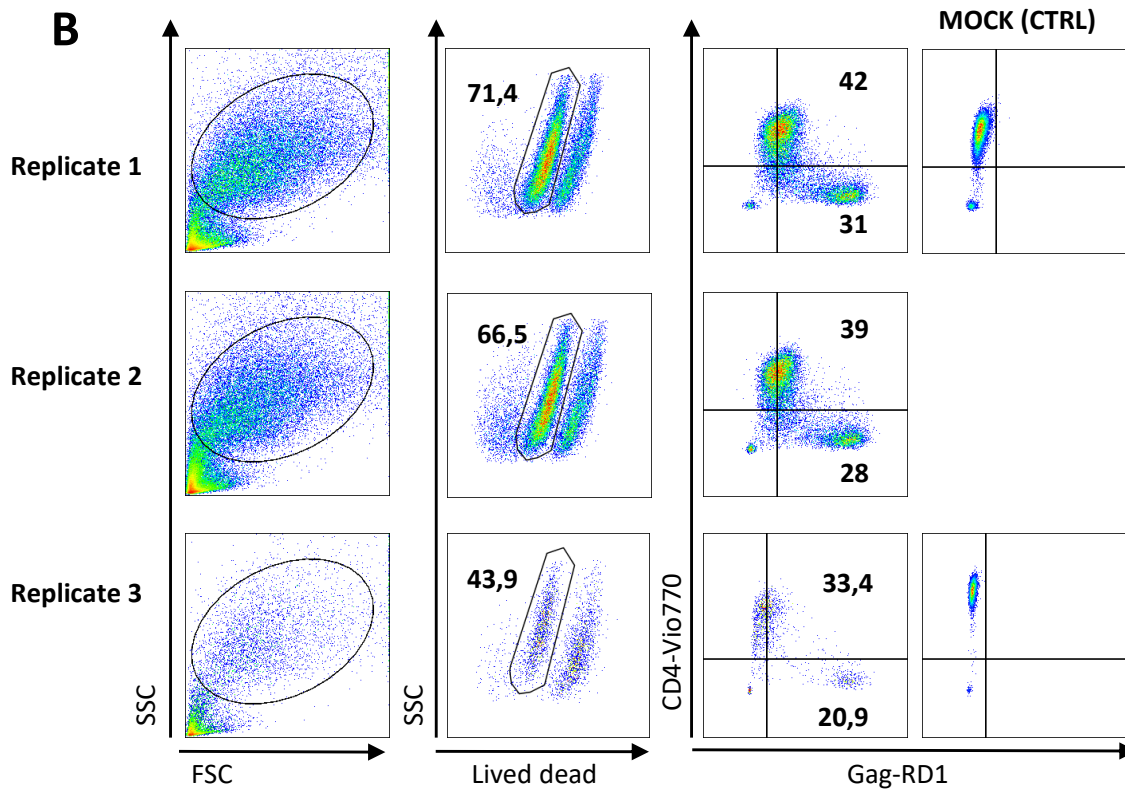
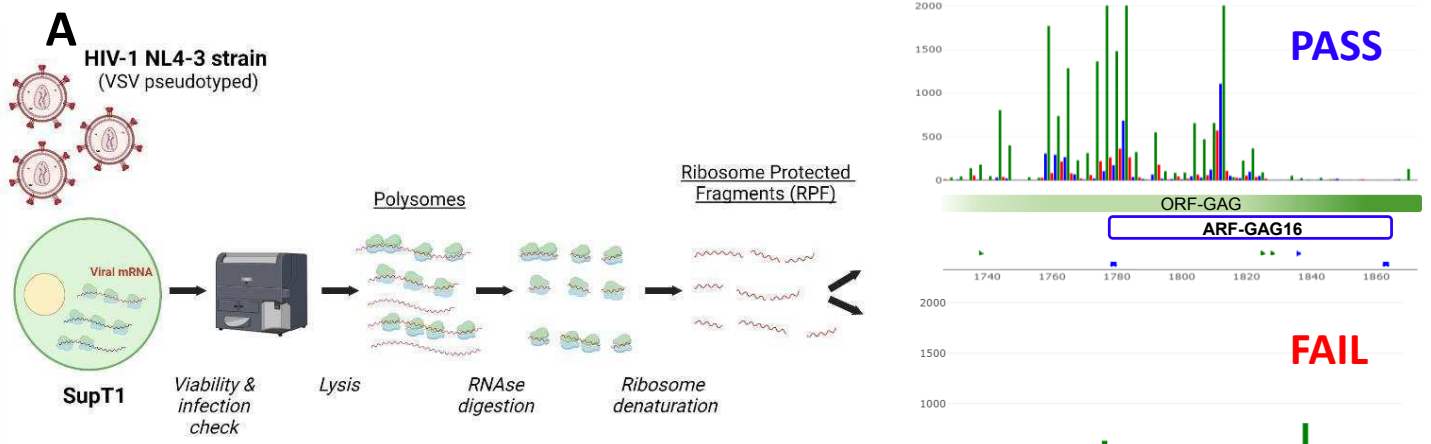
Peptide name	Short name of the peptide	Synthesized peptide sequence	length (AA)	pool
ARF-LTR2-A01	A01	LASIKLALSALVCPVSV	18	1
ARF-LTR4-A02	A02	LWVLEIPQTLVSVENL	17	1
ARF-LTR7-A03	A03	SQRRSLDAGLGLLKRARQEA	20	1
ARF-LTR8-1-A04	A04	TRERLSLDAGLGLLKSARQKRREL	24	1
ARF-LTR8-2-A05	A05	RQKRRELSRRRRLAEGLLKLC	23	X
ARF-LTR9-1-A06	A06	QWRPNRDLKAVKPE	15	1
ARF-LTR9-2-A07	A07	FEISRRRRLAEARTARGE	19	1
ARF-LTR9-3-A08	A08	GEGRRLVSTPKILTSGG	17	1
ARF-GAG/POL1-1-A09	A09	GRSGLPTREGQIGF	14	1
ARF-GAG/POL1-2-A10	A10	QPHQKRASGLGKROQLPLRSR	21	1
ARF-POL1-A11	A11	QYWMVVMHIFQFP	13	1
ARF-POL2-A12	A12	SRLLGSSIRNTTSRRVKEKI	21	9
ARF-POL3-B01	B01	MDSTAYSAARKQLDCQ	17	9
ARF-POL4-B02	B02	CCQKRTAGLSMTYRS	15	9
ARF-POL5-B03	B03	HTEVSGKIELGKSDLCRD	18	9
ARF-POL6-B04	B04	IGQVRFITGQGLK	11	1
ARF-POL8-B05	B05	LTQQIRRLSYKQFI	14	2
ARF-POL9-B06	B06	SSFAGFRIRSKHSNRLTI	18	2
ARF-POL10-B07	B07	INWSVLESQYFF	13	2
ARF-POL11-B08	B08	MSAKRRSHAWTSRL	14	2
ARF-POL12-B09	B09	RIKENYRTGKRSG	13	2
ARF-POL13-1-B10	B10	RITKTNKNSKFSG	14	2
ARF-POL13-2-B10	B11	LLQGQQRSTLERTSKASLER	20	2
ARF-POL14-1-B12	B12	QQTYYLKNYKLNKQKFKIF	19	2
ARF-POL14-2-C01	C01	NKLQKFKIFGITG	14	2
ARF-POL14-3-C02	C02	FKIFGITGTAIEHFGKQQSF	22	2
ARF-POL15-C03	C03	HKSSAKKSKDHQGLWKT DGR	21	3
ARF-GAG2-1-C04	C04	SVKARGKEKISIKTYSMGK	19	3
ARF-GAG2-2-C05	C05	TYSMGKQGARTI	12	3
ARF-GAG2-3-C06	C06	GARTIRSQSWPVRNIRRL	18	3
ARF-GAG5-C07	C07	TNTGTATTIPSDIRRT	17	3
ARF-GAG6-1-C08	C08	KHKAVDKYWDSYNHPF	17	3
ARF-GAG6-2-C09	C09	NHPFRQDQKNLNHYIQ	17	3
ARF-GAG7-C10	C10	RPSMRKLQNGIDCICMQGLLHQ	23	3
ARF-GAG8-C11	C11	IASSACRAYCTRPDERTKKG	20	3
ARF-GAG9-C12	C12	QELLVPPFRNK	10	3
ARF-GAG10-D01	D01	YPSGTNSMDDK	11	4
ARF-GAG11-D02	D02	GCIALPAFWT	10	9
ARF-GAG12-D03	D03	PYQHSGHKTRTKRTL	15	4
ARF-GAG13-D04	D04	DKDQRNPLETM	11	4
ARF-GAG14-D05	D05	SKSRASFTGGKLLDRNLVGP	22	4
ARF-GAG15-D06	D06	QKPCWSRMRTQIVRLF	16	4
ARF-GAG16-D07	D07	YFKSIRTYSYTRRNDSDM	18	4
ARF-GAG17-D08	D08	CRKAILGTERLRSVIVAKK	21	4
ARF-GAG18-D09	D09	VFQLWQRRRAHSQKLGQP	17	4
ARF-GAG19-D10	D10	PKIAGPLGKRAVGNVERKDTK	21	4
ARF-GAG20-D11	D11	EKGLEEMWKRRTPNERLY	18	4
ARF-ENV2-1-D12	D12	SVVLKKNCSQSIMGLYCGK	20	5
ARF-ENV2-2-E01	E01	YLCGKKQTPLYFVHQMLK	18	5
ARF-ENV2-3-E02	E02	KHMIQRYIMFGHPMPVYPQPT	22	5
ARF-ENV4-1-E03	E03	TPQSLHRPVQRYPL	14	5
ARF-ENV4-2-E04	E04	VQRYPLSQFPYIIVRPLVLF	21	5
ARF-ENV5-E05	E05	ANSHTLLYPGWFCDSKV	17	5
ARF-ENV6-1-E06	E06	SVMIRSSMEQDHVQMSAQY	19	5
ARF-ENV6-2-E07	E07	DHVQMSAQYVHVMELGQ	17	5
ARF-ENV7-E08	E08	ASSYNSTAVKWQSSRRRGSN	20	5
ARF-ENV8-E09	E09	SLINPQEGTQKL	12	5
ARF-ENV9-E10	E10	KQCMPLPSVEKLDVHQILQGY	22	5
ARF-ENV10-E11	E11	MFIKYYRAANKRWV	15	6
ARF-ENV11-E12	E12	QEMVITRPRSSDLEEEI	18	6
ARF-ENV12-F01	F01	HRDLQTRRRRYEGQLEK	17	6
ARF-ENV15-1-F02	F02	TIRSSTHQGKEKSG	14	6
ARF-ENV15-2-F03	F03	SSSGNRSYVPWVLGSSRKHVGR	22	6
ARF-ENV15-3-F04	F04	DGTGQTIIVWYSATAEQFAEGY	22	6
ARF-ENV18-1-F05	F05	ELLRNSICCN	11	6
ARF-ENV18-2-F06	F06	SQSGASSSRQESWLWKDT	19	9
ARF-ENV19-1-F07	F07	GATASVATHSLGHQITAPGKSP	21	6
ARF-ENV19-2-F08	F08	KIPKGSTAPRNLGLLWKTLLH	21	6
ARF-ENV19-3-F09	F09	NLGLLWKTLLHHCALY	18	9
ARF-ENV20-1-F10	F10	RINSSWFEFVALENSFAPL	19	6
ARF-ENV20-2-F11	F11	LENSFAPLMCLGILVGVILRF	22	9
ARF-ENV23-F12	F12	YTPYLKNRRTNKKRMKNYWNW	22	6
ARF-ENV25-G01	G01	NDSRRLDRFKNSFYCTFSK	20	7
ARF-peptide leader TAT-G02	G02	IELGRDTHYFRPASQPRGDP	22	7
ARF-ENV/REV1	G03	GLWNFWDAGGKPSNIGGISYSI	23	9
ARF-ENV26-G04	G04	RIVLACLMPQL	12	7
ARF-ENV/NEF1-G05	G05	KNKTGLRKFVWVAVSGQKVV	22	7
ARF-VIF1-G06	G06	QHIGVYQEKETGIWVRESP	20	7
ARF-VIF2-G07	G07	DIELVLGVNIKQDITR	16	7
ARF-VIF3-G08	G08	QGRISTVLGTSINSINKKDKAT	22	7
ARF-VIF5-G09	G09	QRIDGTSRRPRATEGAIQ	19	7
ARF-VPR2-G10	G10	RSLRMKLLIFLGHGFMA	18	7
ARF-VPR3-1-G11	G11	NLWGYLGRSGSLNK	14	7
ARF-VPR3-2-G12	G12	SLNKNSTAVYSFQNVVST	20	7
ARF-VPR4-H01	H01	EFCNNCLFISELGVNIAE	19	8
ARF-VPR5-H02	H02	QNRHYSTKESKWSQ	15	8
ARF-TAT/VPU1-1-H03	H03	KVLLSLPSLFHNKRLRHLLWQ	21	8
ARF-TAT/VPU1-2-H04	H04	KRLRHLWQEEAETATKSSSR	21	8
ARF-TAT/VPU1-3-H05	H05	SSRQSDSSFSIAVSSMCNATF	23	8
ARF-TAT/VPU1-4-H06	H06	NSSISSNNSNNSCVVHSIHR	22	8
ARF-ENV/REV2-H07	H07	HSSGSTCGACASSATTA	17	8
ARF-NEF2-1-H08	H08	DELSQSQQMGWEQYLETWKNM	22	8
ARF-NEF2-2-H09	H09	KNMEQSQVAIQTLMLIV	18	9
ARF-NEF5	H10	STRGGGGFSSQTSQTFK	18	8
ARF-NEF6	H11	PLFKRKGGTGRANLLPKKTRYP	22	8
ARF-NEF7	H12	FTRPKDKISLIGCSSTHK	18	8

Did not pass synthesis; not used

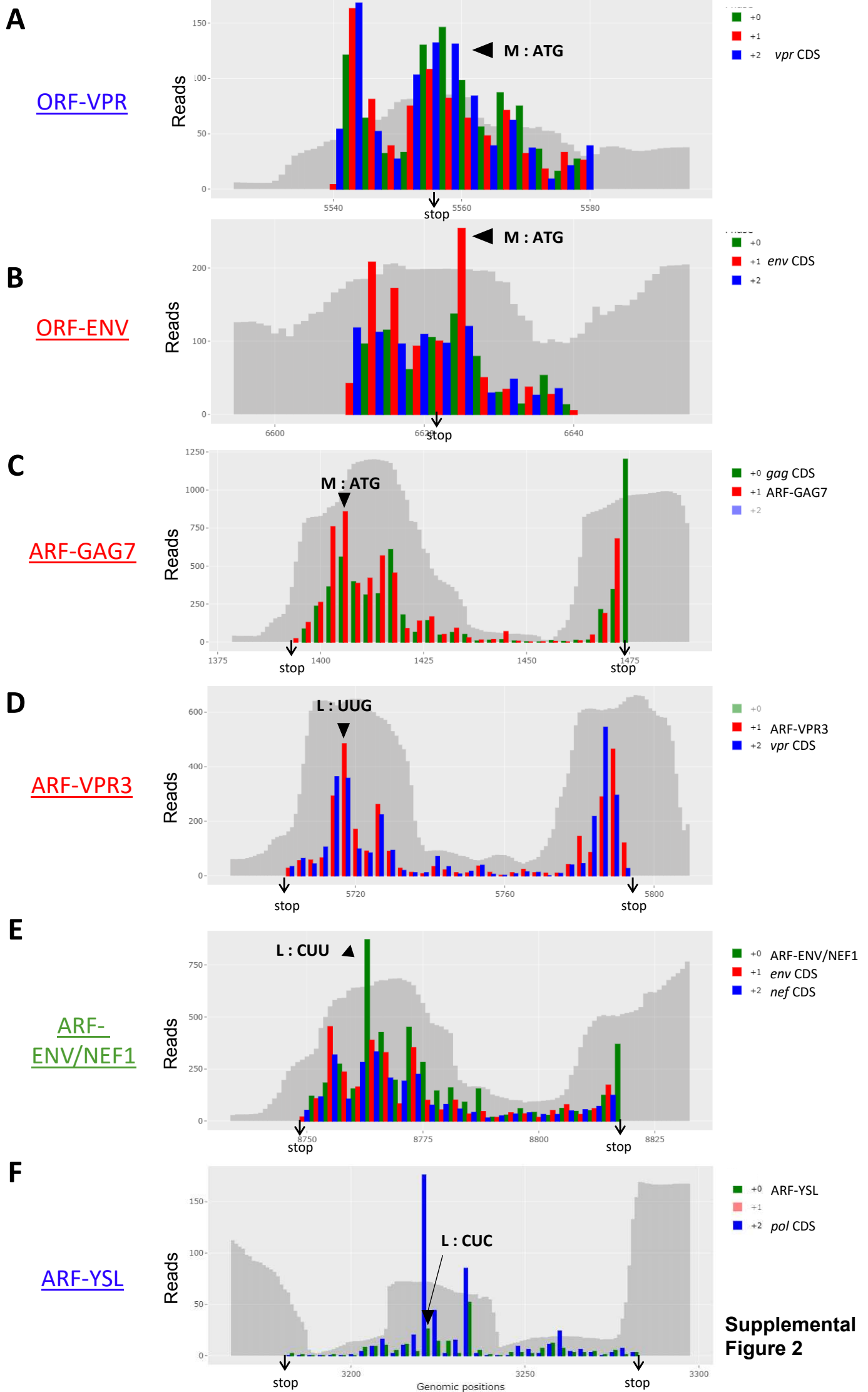
Table 2

Sample	Sample name used in the study	Age	Sex	Number of year of infection	Year of treatment	Treatment	CD4 T cell count	Viral load (copies/ml)	HLA-A1	HLA-A2	HLA-B1	HLA-B2	HLA-C1	HLA-C2	HLA-DR1	HLA-DR2	
34003	EC-1	59	F	31	22	did not receive treatment	1106	40	A3	A30	B57	B15:63	C14	C18	0102	1301	
63005	EC-4	59	H	30			647	61	A26	A33	B27	B51	C02	C12	ND	ND	ND
03002	EC-2	ND	ND	ND			ND	ND	A2	A11	B51:51	B55:56	C03	C15	ND	ND	ND
90001	EC-3	ND	ND	ND			ND	ND	A1:01	A68:01	B35:03	B37:01	C04:01	C06:02	ND	ND	ND
34,001	ART-2	61	M	22	22	edurant + kivexa	693	<40	A2	ND	B27	B57	C02	C06	ND	ND	
34,002	ART-3	63	M	32	21	atrilpa	488	<40	A2	A19	B27	B57	C02	C02	ND	ND	
DUB	ART-1	ND	ND	ND	ND	ND	ND	ND	A2	A3	B07	B56	C01	C07	01	15	

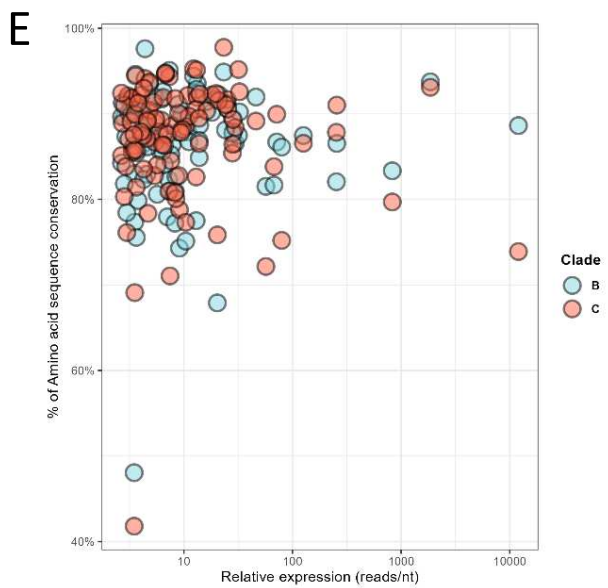
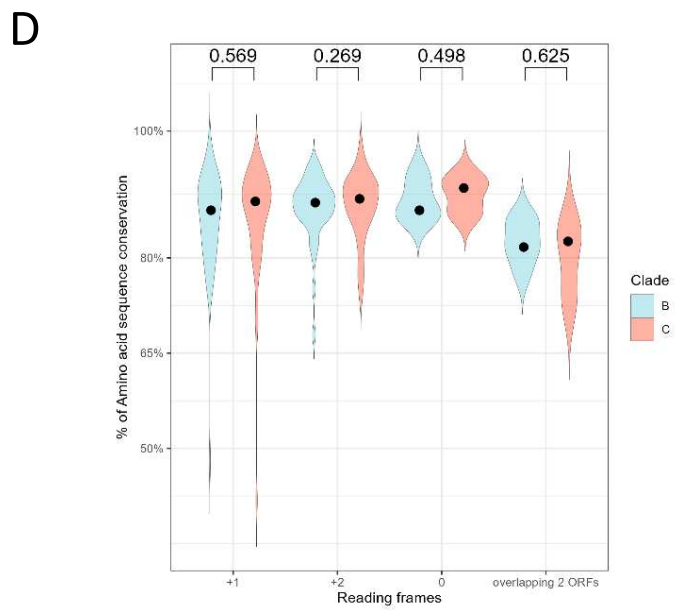
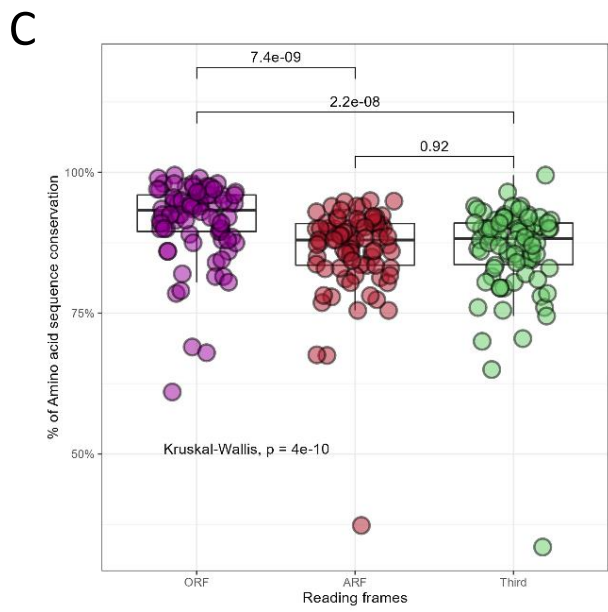
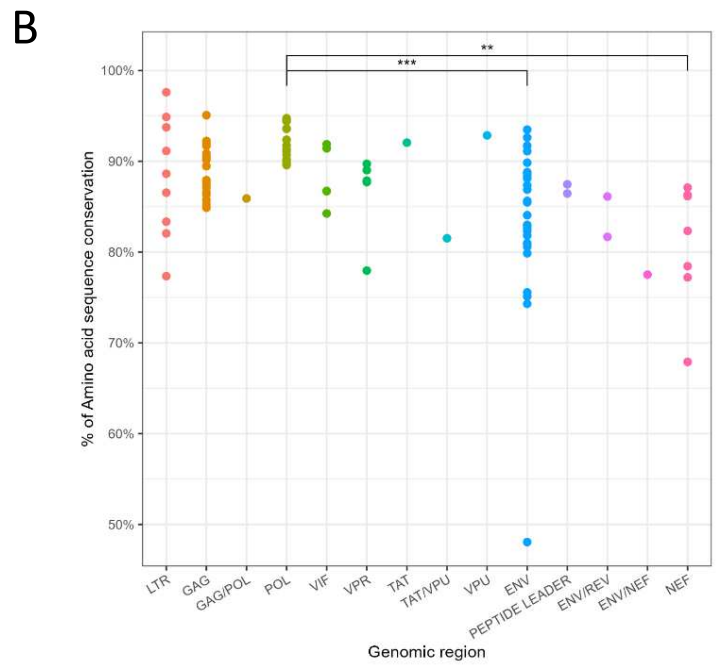
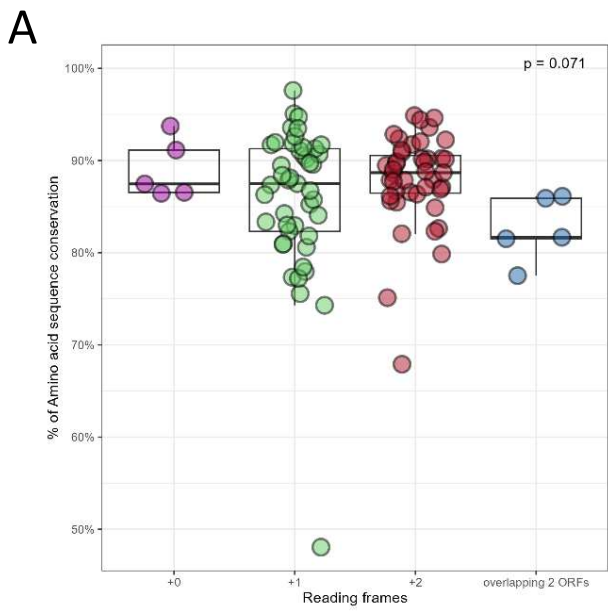
Table 3



Supplemental Figure 1

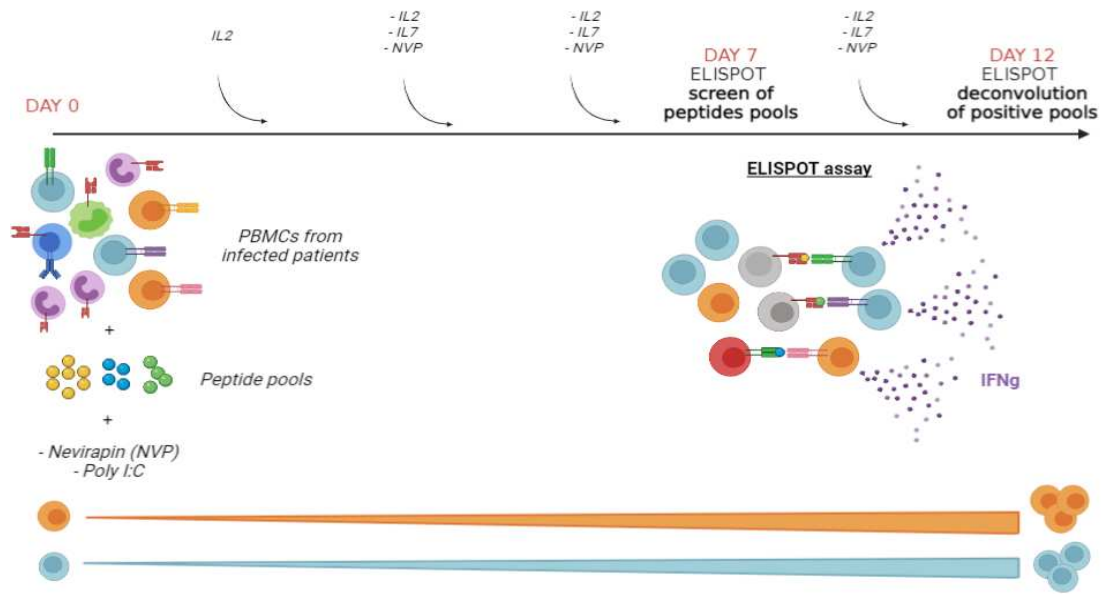


Supplemental Figure 2

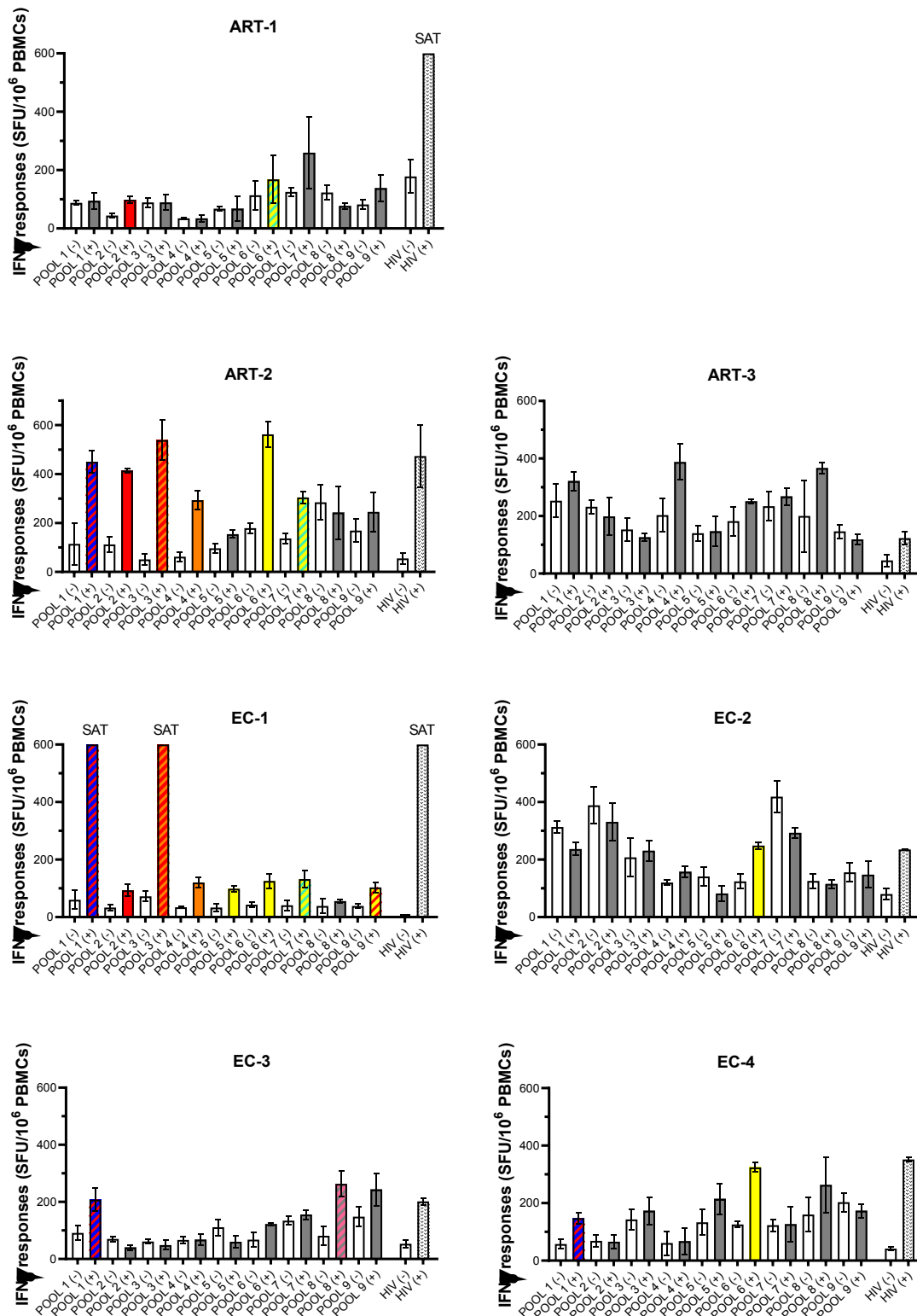


Supplemental Figure 3

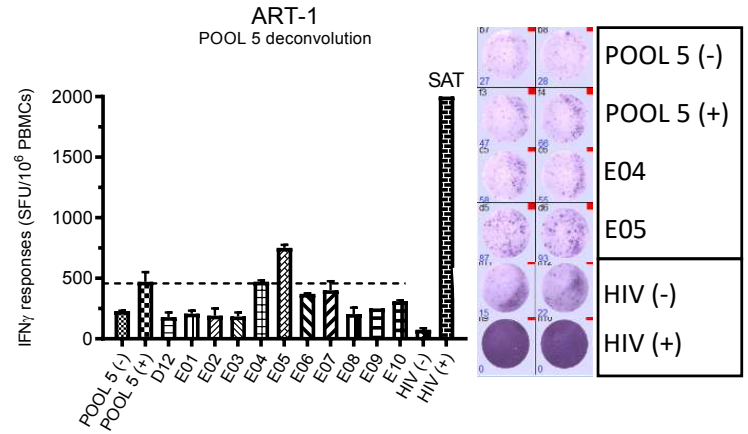
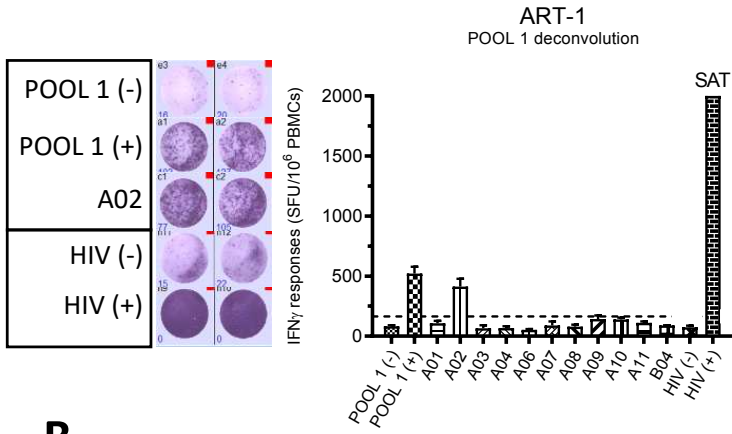
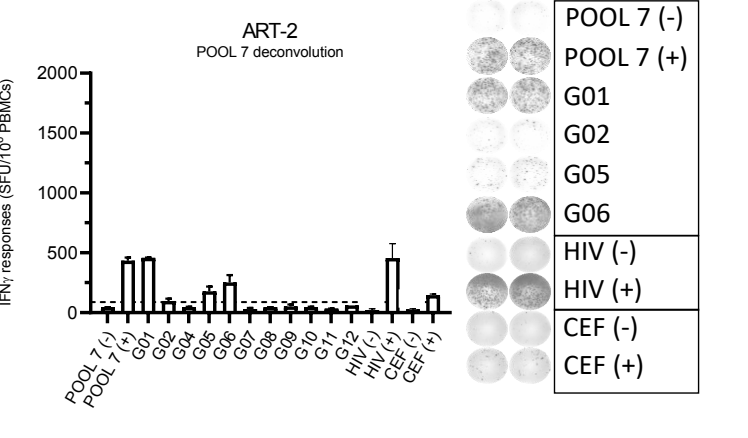
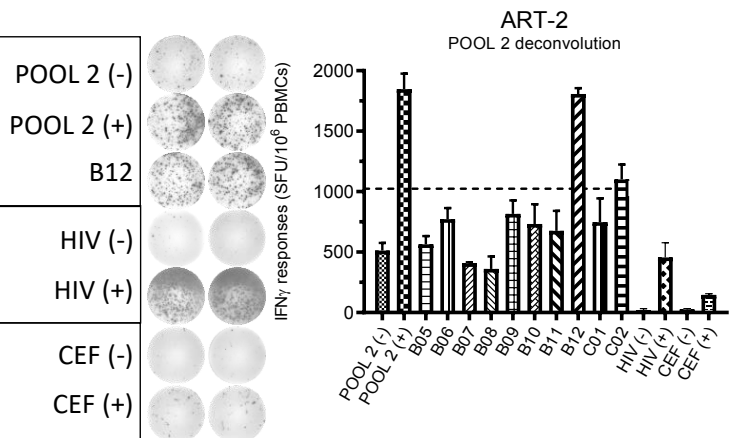
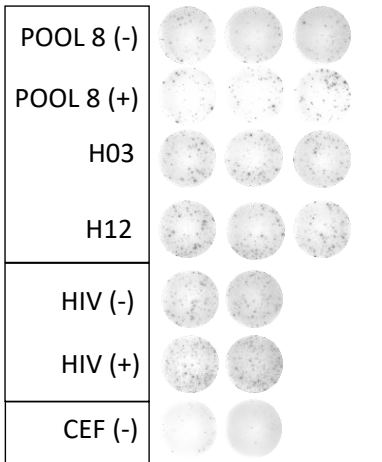
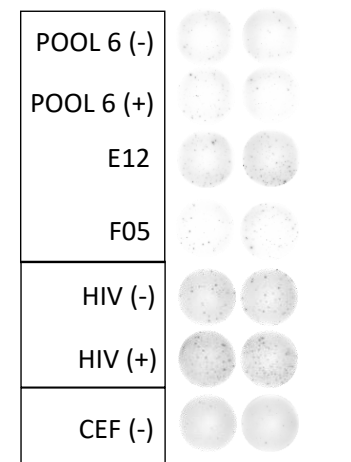
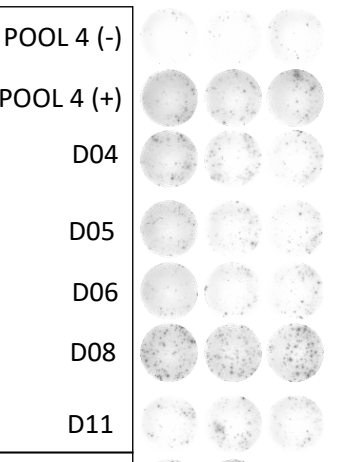
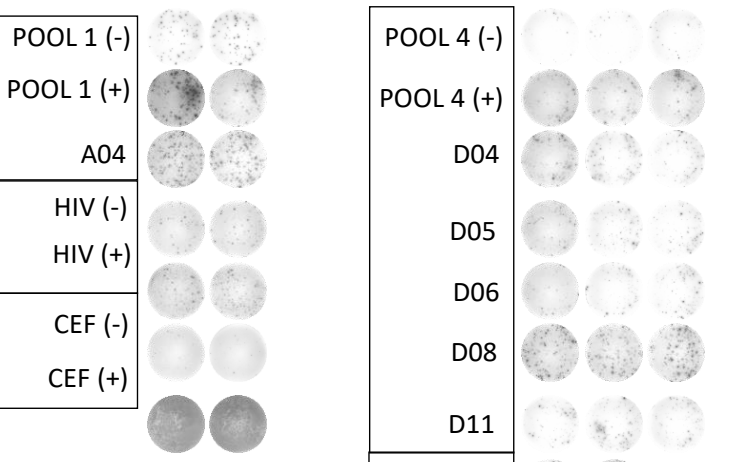
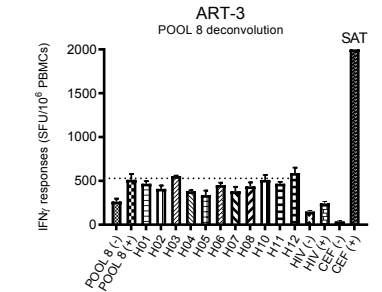
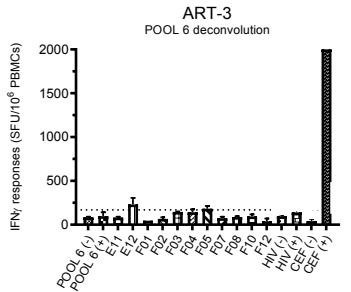
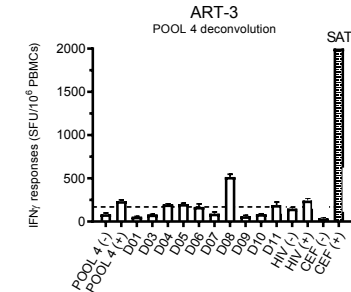
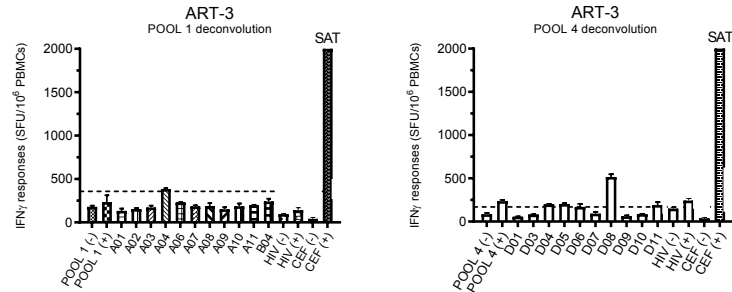
A

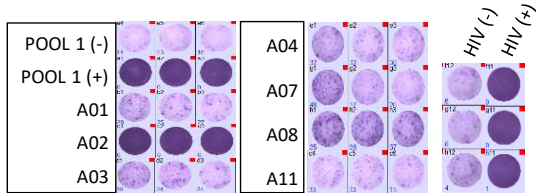
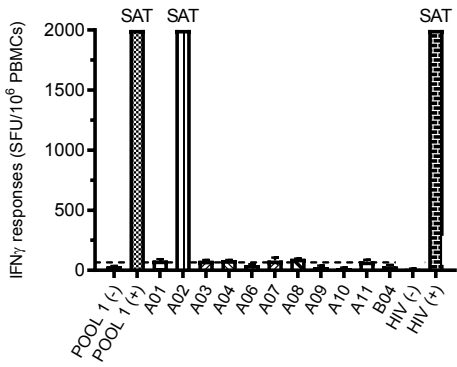
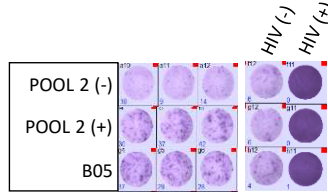
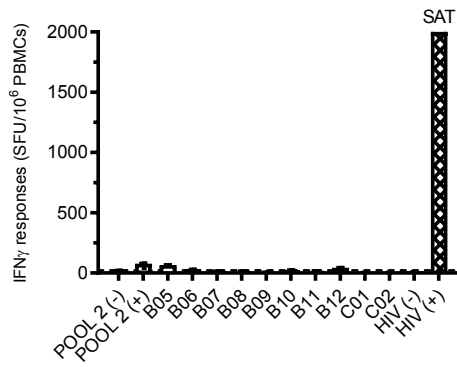
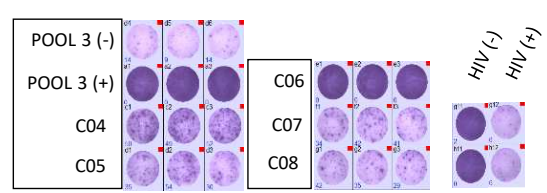
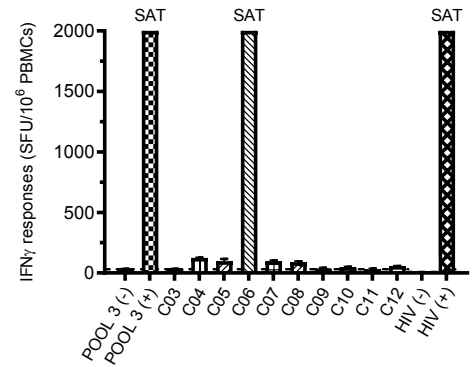
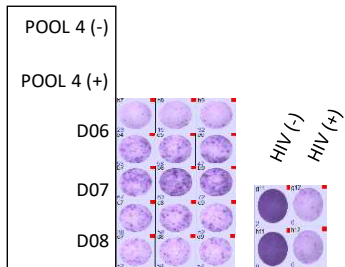
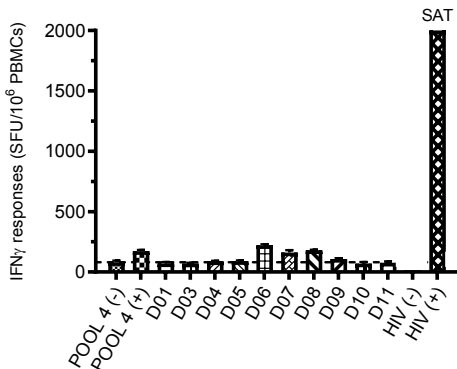
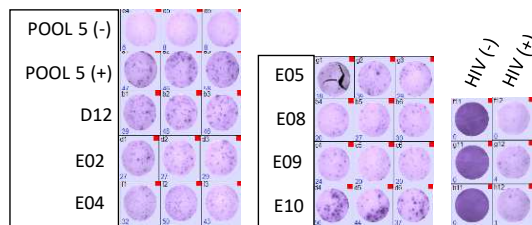
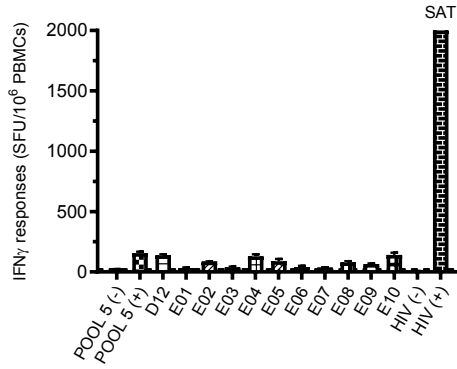
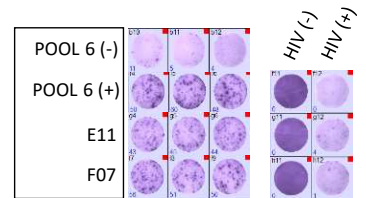
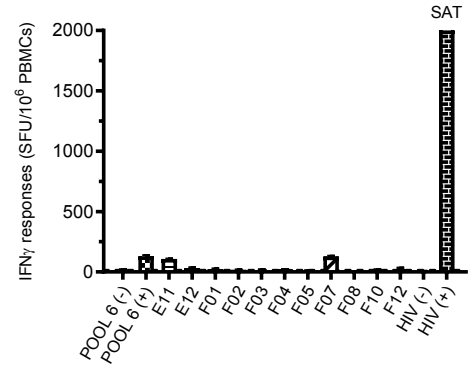
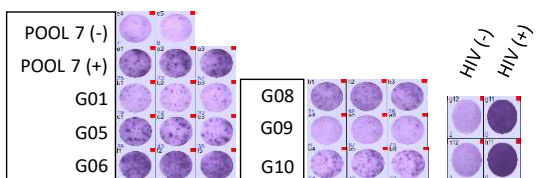
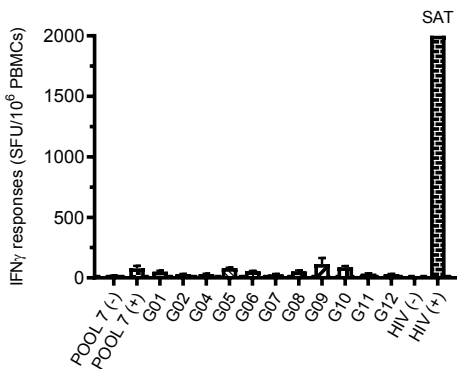
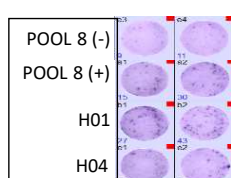
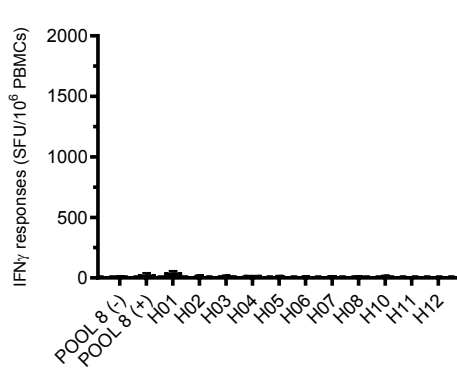
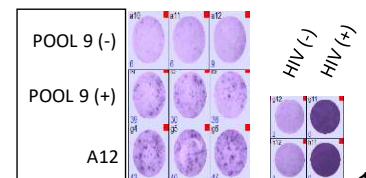
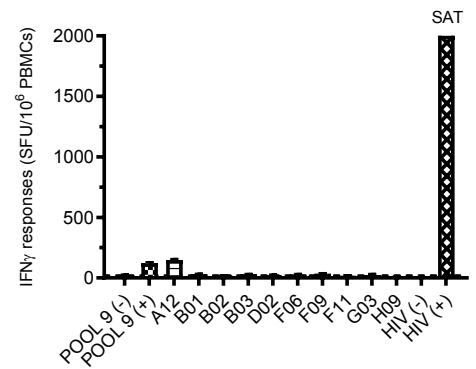


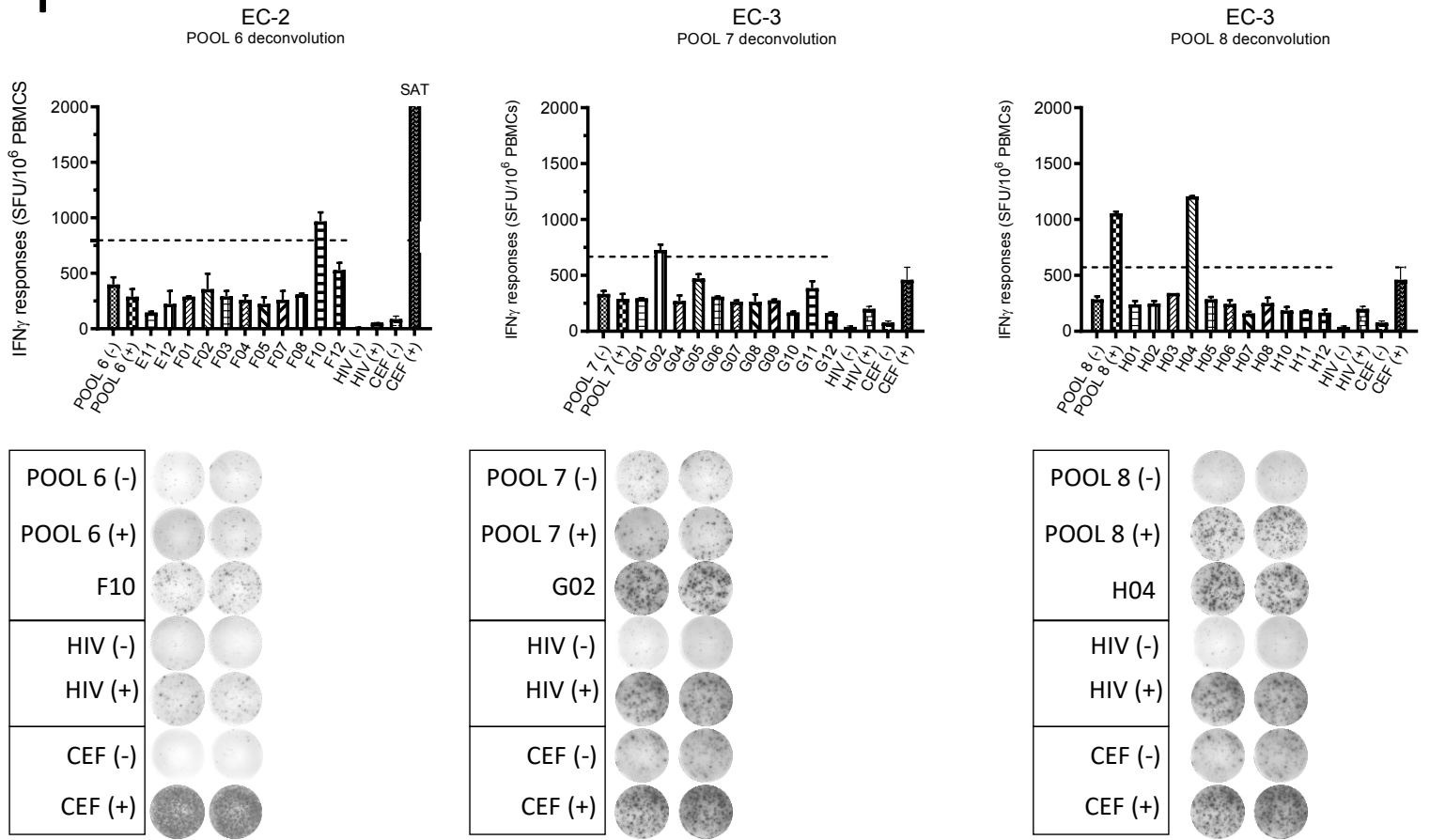
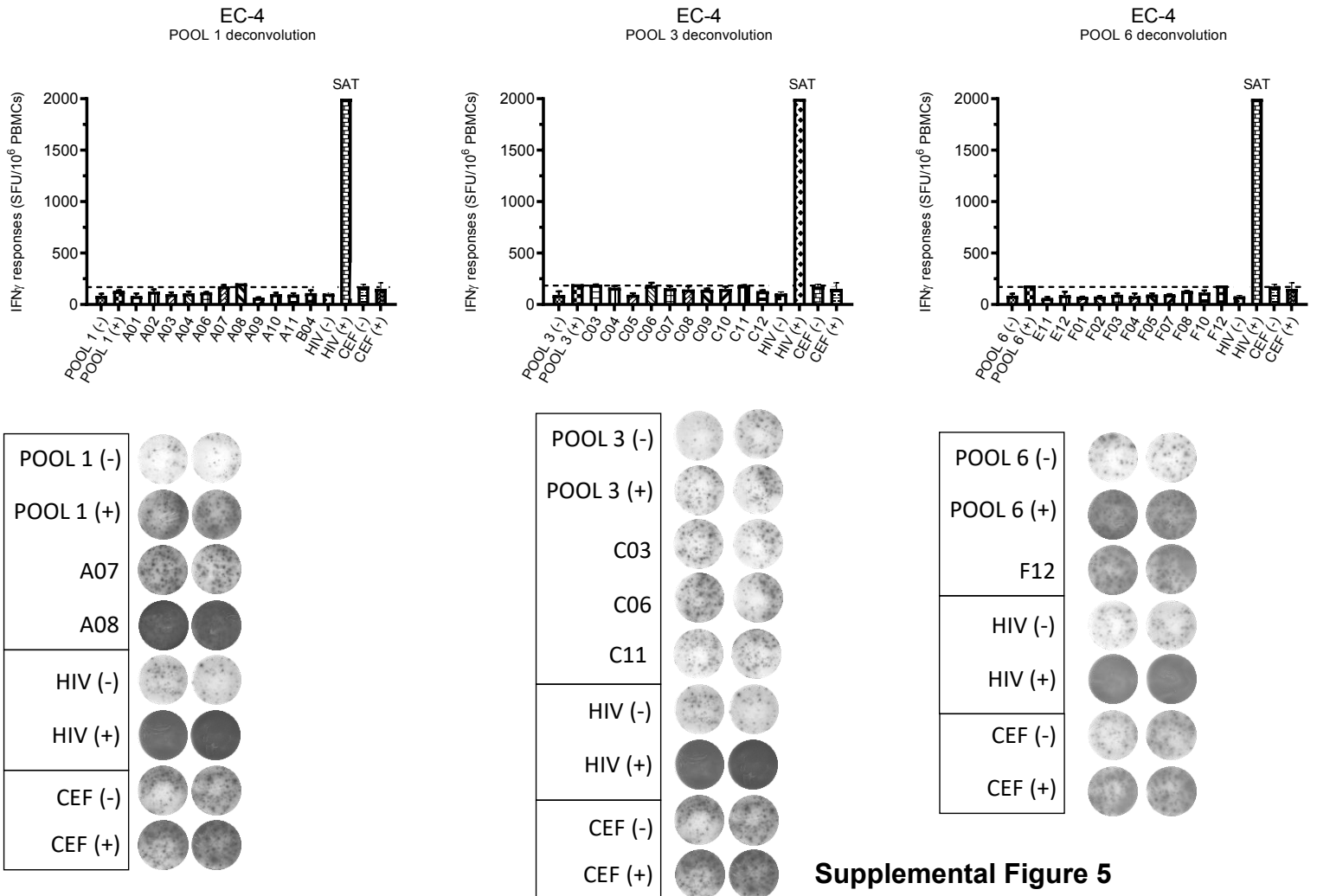
B



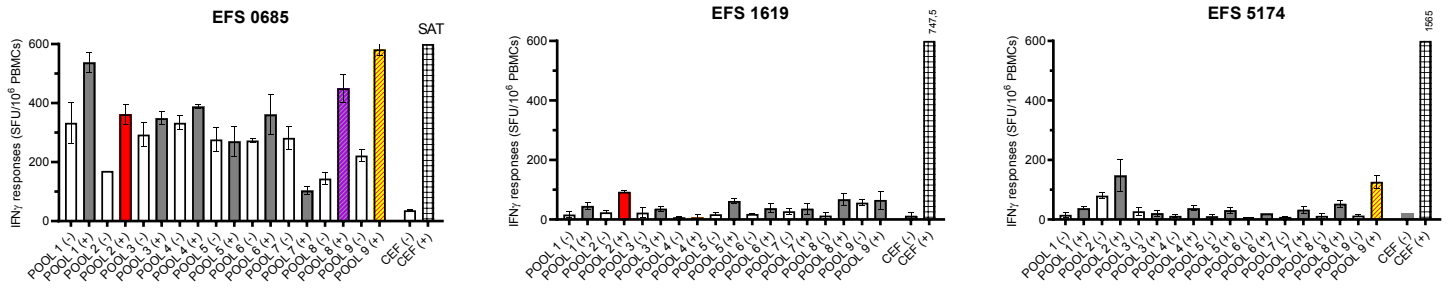
Supplemental
Figure 4

A**B****C****C****Supplemental Figure 5**

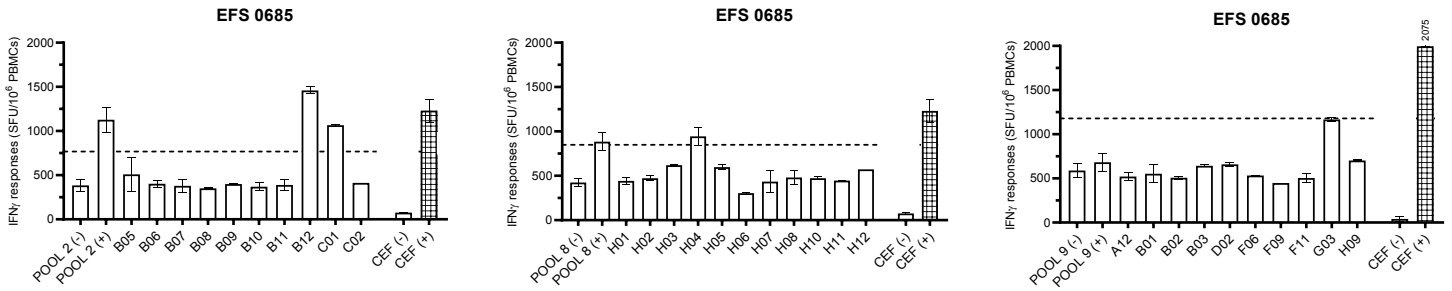
DEC-1
POOL 1 deconvolutionEC-1
POOL 2 deconvolutionEC-1
POOL 3 deconvolution**E**EC-1
POOL 4 deconvolutionEC-1
POOL 5 deconvolutionEC-1
POOL 6 deconvolutionEC-1
POOL 7 deconvolutionEC-1
POOL 8 deconvolutionEC-1
POOL 9 deconvolution**Supplemental
Figure 5**

F**G****Supplemental Figure 5**

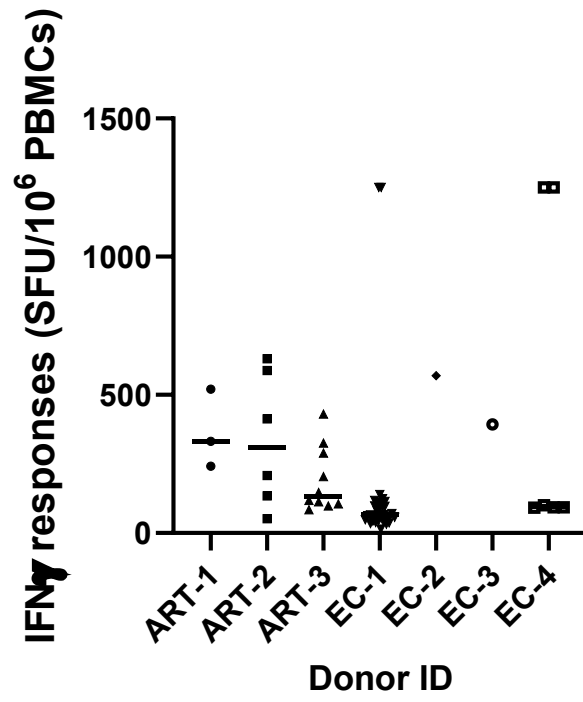
A



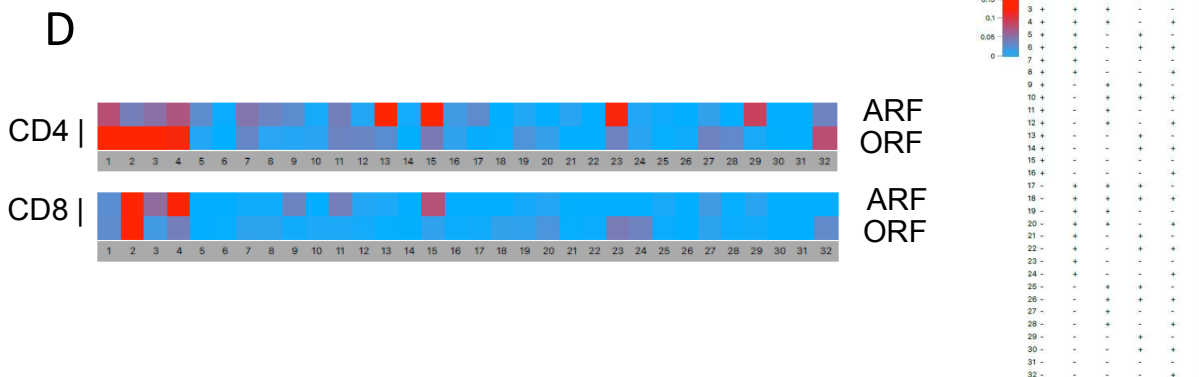
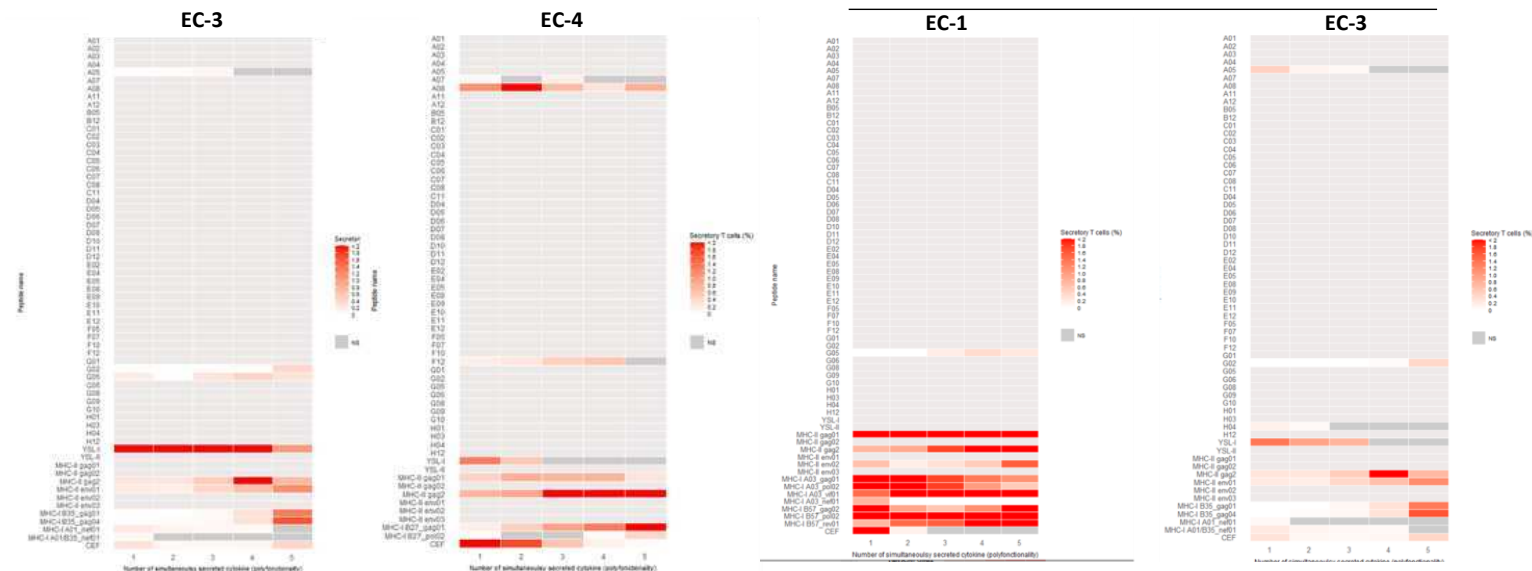
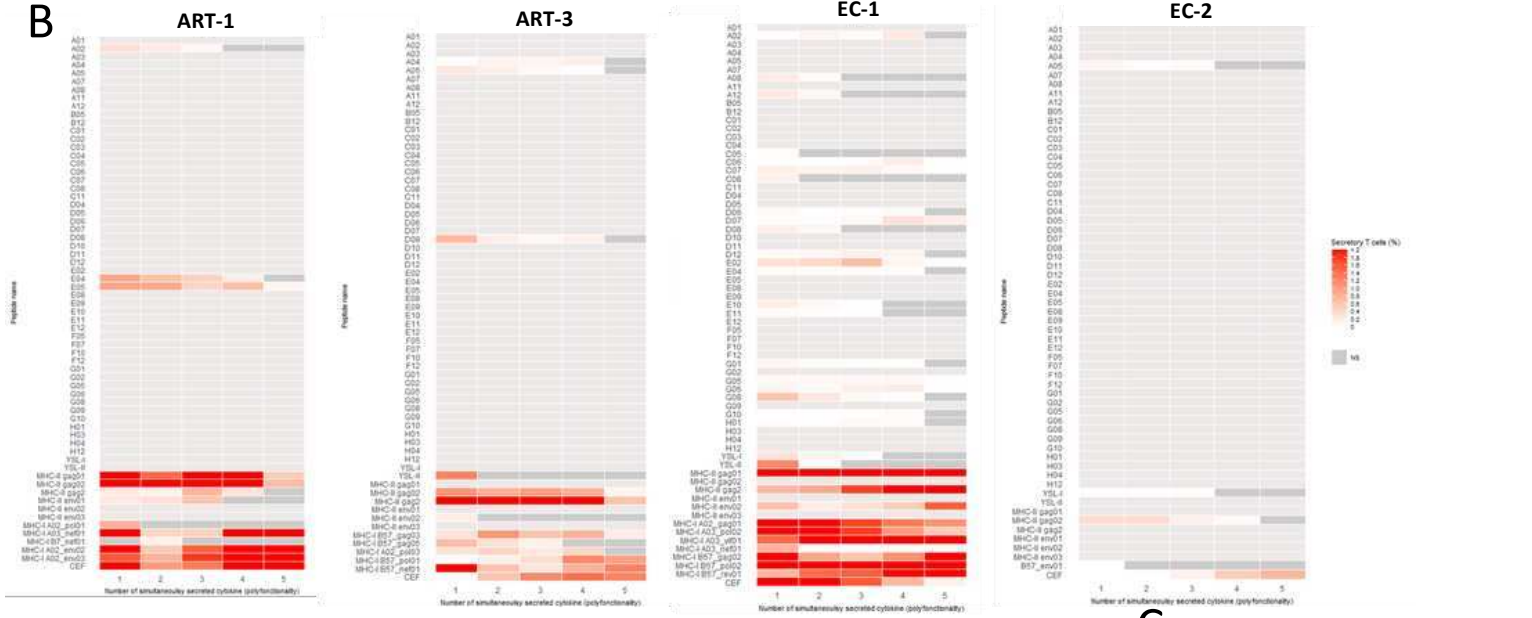
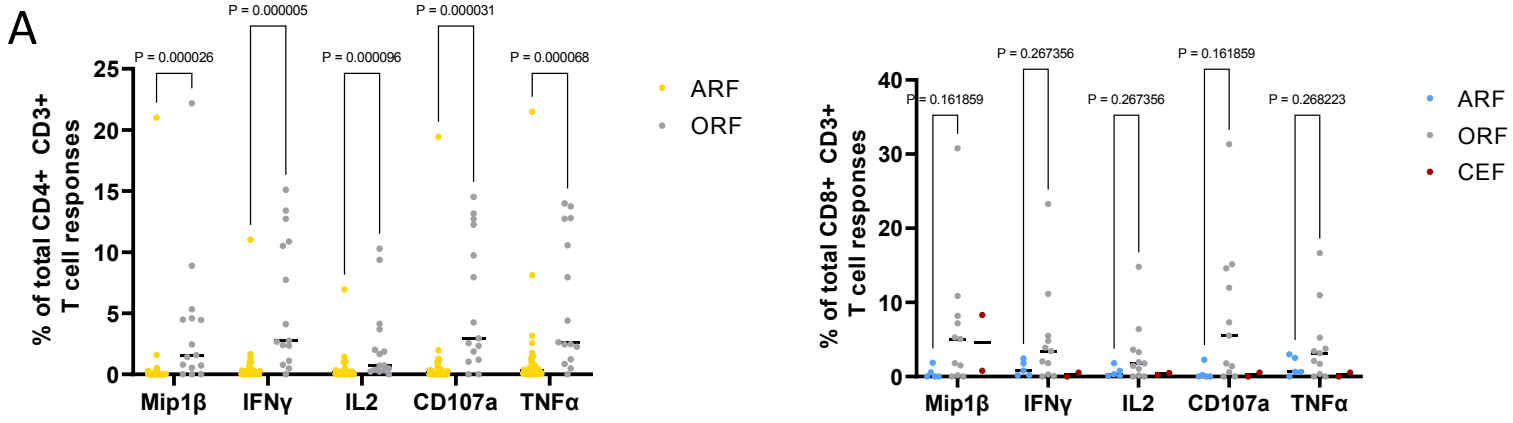
B



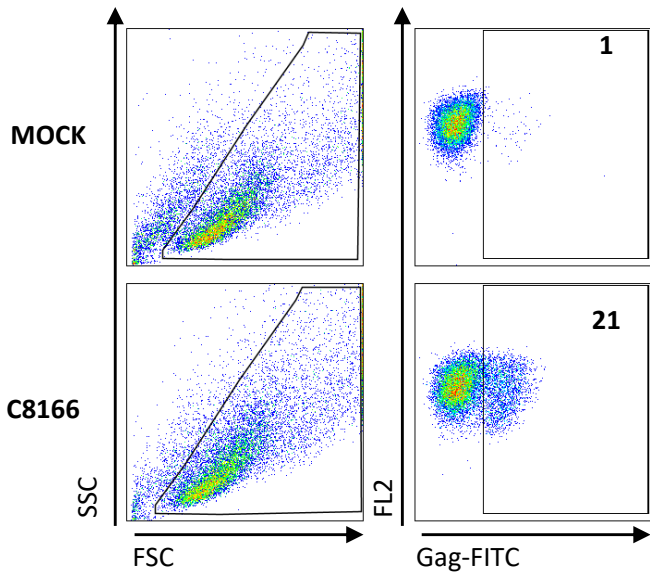
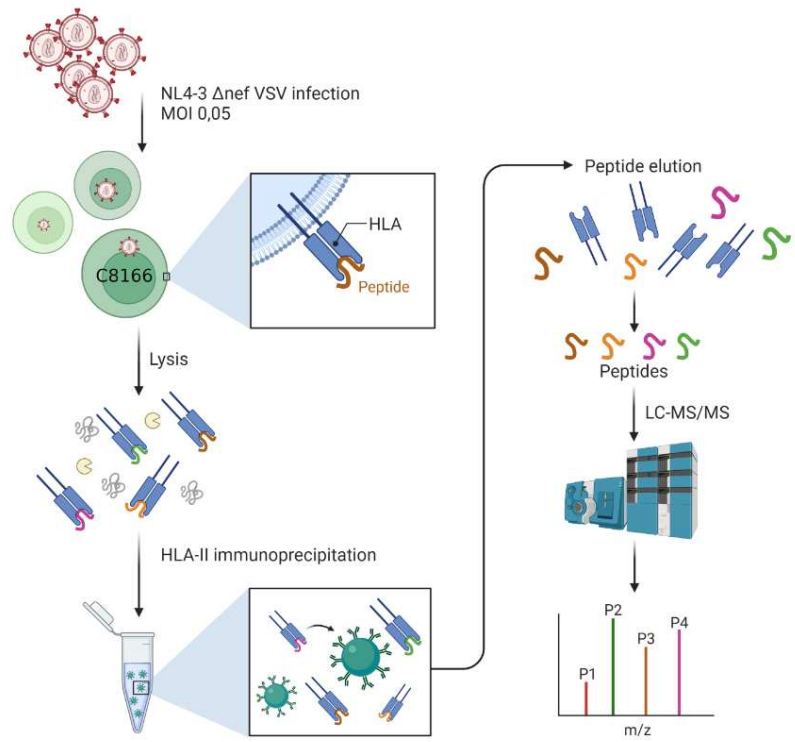
Supplemental Figure 6



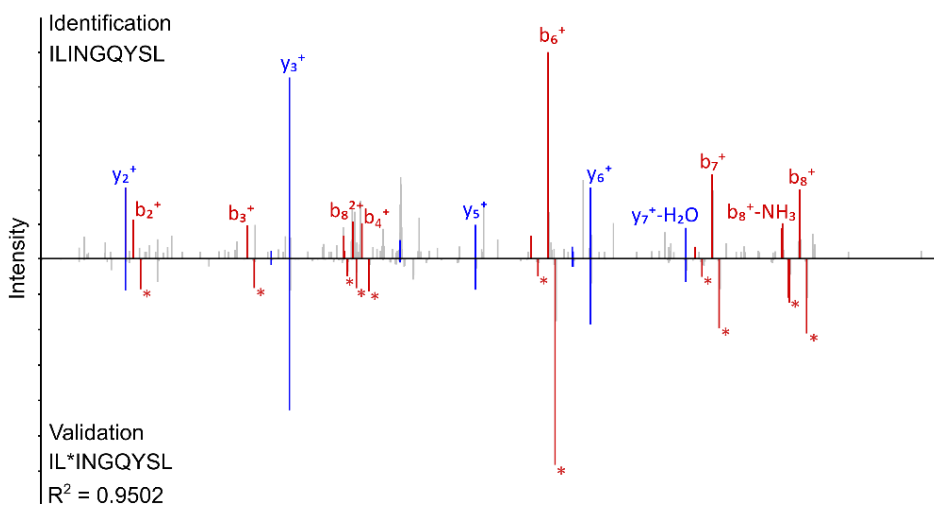
Supplemental Figure 7



Supplemental Figure 8

A**B****C**

Sample	HIV strain	% Infection	HLA type	HIV ligands	5' Position (nt)	CDS
C8166	NL4-3 Δ NEF	20%	DRB1*03:01; DRB1*04:01; DQB1*02;DQB1*03:01	RPEPTAPPEESFRFG EETTPSQK	2143	<i>gag</i>
				RPEPTAPPEESFRFG EETTPSQKQ	2143	<i>gag</i>
				SRPEPTAPPEESFRFG EETTPSQK	2140	<i>gag</i>
				ESFRFG EETTPSQK	2170	<i>gag</i>
				ESFRFG EETTPSQKQEP	2170	<i>gag</i>
				LQSRPEPTAPPEESFRFG EETTPS	2134	<i>gag</i>
				APPEESFRFG EETTPSQK	2158	<i>gag</i>
				APPEESFRFG EETTPSQKQ	2158	<i>gag</i>
				APPEESFRFG EETTPSQKQEP	2158	<i>gag</i>
				APPEESFRFG EETTPSQKQE	2158	<i>gag</i>
				SPEVIPMF SALSEGATP	1282	<i>gag</i>
				VDRFYKTLRAEQASQEVK	1678	<i>gag</i>
				DYVDRFYKTLRAEQASQEVK	1672	<i>gag</i>

D**Supplemental Figure 9**

Peptide name	Genomic region	Synthesized peptide sequence	pool
B06	<i>env</i>	KVSFEPIPIHYCAPAGFAILKC	HIV
B08	<i>env</i>	KVSFEPIPIHYCAPAGFAILK	HIV
B09	<i>env</i>	FEPIPIHYCAPAGFAILKC	HIV
B10	<i>env</i>	VSTVQCTHGIRPVVSTQLLL	HIV
C02	<i>env</i>	TGEIIGDIRQAHCNISktqW	HIV
D07	<i>gag</i>	EKAFSPEVIPMFALSEGAT	HIV
D09	<i>gag</i>	YVDRFYKTLRAEQASQEV	HIV
D10	<i>gag</i>	CKTILKALGPAATLEEMMTA	HIV

Supplemental Table 1

Peptide name	Short name of the peptide	Synthesized peptide sequence	length (AA)	Pool
ARF-LTR2-A01	A01	LASIKLALSALSVCPSVV	18	1
ARF-LTR4-A02	A02	LWVLEIPQTLVSVENL	17	1
ARF-LTR7-A03	A03	SQRRSLDAGLGLLKRARQEA	20	1
ARF-LTR8-1-A04	A04	TRERLSLDAGLGLLKSARQKRREL	24	1
ARF-LTR8-2-A05	A05	RQKRRELSRRRRLAEGLLKLC	23	9
ARF-LTR9-2-A07	A07	EEISRRRRLAEARTARGE	19	1
ARF-LTR9-3-A08	A08	GEGRRLVSTPKILTSGG	17	1
ARF-POL1-A11	A11	QYWMWVMHIFQFP	13	1
ARF-POL2-A12	A12	SRLGSSIRNTTSRRVKKEKI	21	9
ARF-POL8-B05	B05	LTQQIRRLSYKQFI	14	2
ARF-POL14-3-C02	C02	FKIFGITGTAEIHFGKDQQSF	22	2
ARF-POL15-C03	C03	HKSSAKKKSKDHQGLWKT DGR	21	3
ARF-GAG2-1-C04	C04	SVKARGKEKISIKTYSMGK	19	3
ARF-GAG2-2-C05	C05	TYSMGKQGARTI	12	3
ARF-GAG2-3-C06	C06	GARTIRSQSWPVRNIRRL	18	3
ARF-GAG5-C07	C07	TNTGTATTIPSDRIRRT	17	3
ARF-GAG6-1-C08	C08	KHQKAVDKYWDSYNHPF	17	3
ARF-GAG8-C11	C11	IASSACRAYCTRPDERTK GK	20	3
ARF-GAG13-D04	D04	DKDQRNPLETM	11	4
ARF-GAG14-D05	D05	SKSRASFTGGKKLDDRNLVGP K	22	4
ARF-GAG15-D06	D06	QKPCWSRMRTQIVRLF	16	4
ARF-GAG16-D07	D07	YFKSIRTSSYTRRNDDSM	18	4
ARF-GAG17-D08	D08	CRKAILGTKERLLSVSIVAK K	21	4
ARF-GAG20-D11	D11	EKGLLEMWKRRTPNERLY	18	4
ARF-ENV2-1-D12	D12	SVVLQKNCGSQSIMGYLCG K	20	5
ARF-ENV2-3-E02	E02	KHMIQRYIMFGPHMPVYPQT PT	22	5
ARF-ENV4-1-E03	E03	TPQSLHRPVQRYPL	14	5
ARF-ENV4-2-E04	E04	VQRYPLSQFYIIVPRLVLR F	21	5
ARF-ENV5-E05	E05	ANSHTLLYPGWFCDSKV	17	5
ARF-ENV7-E08	E08	ASSYNSTAVKWQSSRRRGS N	20	5
ARF-ENV8-E09	E09	SLINPQEGTQKL	12	5
ARF-ENV9-E10	E10	KQCMPLPSVEKLDVHQILQ GYY	22	5
ARF-ENV10-E11	E11	MFIKYYRAAINKRWW	15	6
ARF-ENV11-E12	E12	QEMVVITRPRSSDLEEEI	18	6
ARF-ENV18-1-F05	F05	ELLRRNSICCN	11	6
ARF-ENV19-1-F07	F07	GATASVATHSLGHQTAPGK SP	21	6
ARF-ENV20-1-F10	F10	RINSSWEFGVALENSFAPL	19	6
ARF-ENV23-F12	F12	YTPYLKNRRTNKKRMNKNY WNW	22	6
ARF-ENV25-G01	G01	NDSRRLDRFKNSFYCTFY SK	20	7
ARF-peptide leader TAT-G02	G02	IELGRDTHHYRFRPASQPRG DP	22	7
ARF-ENV/NEF1-G05	G05	KNKTGLRKGFAIRWVASGQ KV V	22	7
ARF-VIF1-G06	G06	QHIGVYIQEKETGIWVRESP	20	7
ARF-VIF3-G08	G08	QGRISTVLGTSSINSTKKDK AT	22	7
ARF-VIF5-G09	G09	QRIDGTSRRRPRATEGAIQ	19	7
ARF-VPR2-G10	G10	RSLRMKLLDIFLGHGFMA	18	7
ARF-VPR4-H01	H01	EFCNNCCLFISELGVNIAE	19	8
ARF-TAT/VPU1-1-H03	H03	KVLLSLPSLFHNKRLRHLLW Q	21	8
ARF-NEF7	H12	FTPRKDKISLICGSTTHK	18	8

Supplemental Table 2

HLA-restriction	ORF	sequence (AA)
A01/B35	Nef	YFPDWQNYT
A01	Env	RRGWEVLKY
A01	Rev	ISERILSTY
A01	Gag	GSEELRSLY
A01	Nef	WRFDSRLAFH
B35	Gag	PPIPVGDYI
B35	Pol	HPDIVIQY
B35	Nef	VPLRPMTY
B35	Env	DPNPQEVVL
B35/B51	Pol	IPLTEEAEL
B35	Gag	NSSKVSQNY
B35	Env	TAVPWNASW
B35	Pol	TVLDVGDAY
B35	Pol	VPLDEDFRKY
B35	Env	VPVWKEATITL
B35	Gag	WASRELERF
B35	Gag	HPVHAGPIA
B35	Gag	NPPIPVGDYI
A02	Gag	SLYNTVATL
A02	Pol	ILKEPVHGV
A02	Nef	VLEWRFD SRL
A02	Nef	PLTFGWCKYK
A02	Vpr	AIIRILOQL
A02	Pol	ALVEICTEM
A02	Gag	FLGKIWPSYK
A02	Nef	GAFDLSFFL
A02	Pol	LVGPTPVNI
A02	Env	RGPGRAFVTI
A02	Env	RIRQGLERA
A02	Env	SLLNATDIAV
A02	Pol	VIIQYMDDL
A02	Gag	YVDRFYKTL
A03	Gag	RLRPGGKKK
A03	Gag	KIRLRPGGK
A03	Pol	AIFQSSMTK
A03	Env	RLRDLLIVTR
A03	Nef	QVPLRPMTYK
A03	Nef	AVDLSHFLK
A03	Pol	AVFIHFKRK
A03	Vif	RIRTWKSLVK
A03	Gag	RLRPGGKKKY
A03	Pol	RMRGAHTNDVK
A24	Gag	KYKLGKHWV
A24	Nef	RYPLTFGW
B57	Vpr	AVRHFPRIW
B57	Gag	FSPEVPMF
B57	Nef	HTQGYFPDW
B57	Vif	ISKKAGWVF
B57	Gag	ISPRTLNAW
B57	Pol	IVLPEKDSW
B57	Gag	KAFSPEVI
B57	Gag	KAFSPEVPMF
B57	Gag	QASQEVKNW
B57	Pol	STTVKAACWW
B57	Gag	TSTLQEQIGW
B57	Rev	KAVRLIKFLY
B27	Pol	KTAVQMAVF
B27	Gag	KRWIILGLNK
B27	Gag	FRNQRKTVK
B27	Gag	IRLRPGGKK
B27	Pol	KRKGIGGY
B27	Nef	RRQDILDWI
B27	Vpr	VRHFPRIW
B27	Nef	LRPMTYKAA
B27	Gag	QRGNFRNQRK
B27	Gag	RRWIQLGLQK
B51	Gag	GRRGWALKY
B51	Pol	TAFTIPSI
B51	Env	RAIEAQQHL
B51	Vpr	EAVRHFPRI
B51	Pol	EKEGKISKI
B51	Vif	IPLGDAKLII
B51	Env	LPCRIKQJI
B51	Pol	RPLVTIKI
B51	Gag	RPGGKKKYKL
B44	Pol	EEMNLPGRW
B44	Env	AENLWVTVY
B44	Gag	AEQASQDVKNW
B44	Gag	LYNTVATLY
B44	Gag	RDYVDRFYKTL
B07	Gag	SPRTLNAWV
B07	Env	RPNNNTRKSI
B07	Env	IPRRIRQGL
B07	Nef	TPGPGVRYPL
B07	Gag	GGHKARVL
B07	Pol	SPAIFQSSM

Supplemental Table 3

DISCUSSION

1. Les cadres alternatifs de lecture du VIH

Le premier axe de mes travaux de thèse a porté sur la définition du translatome du VIH-1 (VIH). Pour cela, nous avons réalisé des expériences de ribosome profiling afin de définir l'ensemble des empreintes ribosomales sur les transcrits viraux et de dresser une liste exhaustive des cadres alternatifs de lecture (ARFs) du VIH activement traduits au sein des cellules infectées. Nous avons pu observer un alignement des fragments d'ARN protégés par les ribosomes sur les ORFs classiques du VIH démontrant une traduction active *in vivo* des gènes viraux. Par cette approche étendue aux régions génomiques non classiques, nous montrons l'existence de 98 ARFs répartis sur l'ensemble du génome viral, y compris la région 5'UTR.

Dans un premier temps, nous avons effectué ces expériences à l'aide d'une lignée de cellule T CD4⁺ SupT1 infectées avec l'isolat de laboratoire NL4-3. De manière remarquable, grâce à l'analyse des reads distribués sur les CDS annotés du génome humain, nous avons pu démontrer la périodicité des fragments d'ARN protégés par les ribosomes séquencés. En effet, la densité ribosomale est enrichie tous les trois nucléotides et dans la phase de lecture des CDS suggérant fortement que les ribosomes se déplaçaient bien codon par codon le long des ARNm lors de nos expériences. Bien que la périodicité ne soit pas toujours évaluée ou discutée dans certaines études, de nombreux papiers s'accordent sur le fait que c'est un contrôle qualité nécessaire aux expériences de ribosome profiling (Ingolia *et al.*, 2009b; Brar and Weissman, 2015; Lauria *et al.*, 2018; Xiao *et al.*, 2018). C'est une étape essentielle pour discriminer les ARN couverts de polysomes et en cours de traduction, des ARN non couverts correspondant à un signal de transcription. L'observation d'un signal périodique est la preuve la plus solide d'une traduction active, ce qui reflète le gage de qualité de nos données. Dans notre jeu de données, nous observons de plus que l'ensemble des ORFs du VIH sont couverts par les reads, ce qui signifie que les gènes classiques du VIH sont activement traduits. Nous observons également une forte diminution des reads alignés sur *pol* en comparaison à ceux qui s'alignent sur *gag*, expliquée par le mécanisme de décalage de cadre de lecture en -1 des ribosomes (FS) en cours de traduction de *gag*. L'expression relative, calculée en nombre de reads par nucléotide, nous permet de définir une efficacité du -1 FS à 6%, conformément à la littérature qui indique une efficacité du -1 FS entre 2,5 et 10% selon les modèles et techniques utilisées (Jacks *et al.*, 1988; Chen *et al.*, 1995, p. 19; PLANT and DINMAN, 2006). Ces contrôles internes de qualité démontrent la robustesse de notre ribosome profiling et nous ont permis d'étendre l'analyse aux évènements non canonique de traduction.

Nous avons identifié 98 ARFs codés dans le génome viral, allant de la région 5' UTR jusqu'au gène *nef*. Sur ces 98 ARFs, 9 sont localisés dans la région 5' UTR. N'étant pas une équipe dont l'axe de recherche est initialement orienté sur la génomique et la traduction des ARN viraux, nous étions dans un premier temps surpris par cette observation. En effet, la région 5' UTR a longtemps été considérée comme non codante mais nécessaire à diverses étapes du cycle viral, comme lors de la traduction où elle permet le recrutement et l'assemblage des différents acteurs de l'initiation (Araujo *et al.*, 2012; Leppke, Das and Barna, 2018; Zhang and Crumpacker, 2022). Mais finalement, les preuves d'une traduction active par les ribosomes de la région 5' UTR sont cohérentes avec les données de la littérature. En effet, plusieurs études utilisant le

ribosome profiling ont mis en évidence une traduction dans la région 5' UTR pour les génomes viraux et humains, avec une augmentation des reads sur la région 5'UTR par rapport au reste du génome (Ingolia *et al.*, 2009b; Ingolia, Lareau and Weissman, 2011; Lee *et al.*, 2012; Stern-Ginossar *et al.*, 2012; Arias *et al.*, 2014; Irigoyen *et al.*, 2018; Chong *et al.*, 2020; Finkel *et al.*, 2021; Ruiz Cuevas *et al.*, 2021; Ouspenskaia *et al.*, 2022). Le nombre importants de reads que nous observons également sur la région 5'UTR pourrait être dû à l'arrêt des ribosomes pendant l'initiation et au fait que les transcrit viraux (plus de 50 transcrits différents) portent tous la région 5'UTR (Calviello and Ohler, 2017; de Breyne and Ohlmann, 2018; Sertznig *et al.*, 2018). Les ARFs localisés en 5' UTR, connus sous le nom de uORFs, sont décrits comme des éléments régulateurs de l'expression des gènes, en inhibant la traduction des CDS situées en aval par le biais de divers mécanismes (Wethmar, 2014; Brar and Weissman, 2015; Orr *et al.*, 2020). Par conséquent, il est probable que les uORFs du VIH que nous avons identifié remplissent une fonction similaire. Toutefois, les données de la littérature indiquent que l'activité régulatrice des uORFs est modulée en réponse au stress (Starck *et al.*, 2016). Dans le modèle humain, les stimuli de stress provoquent la phosphorylation du complexe eIF2/GTP-ARnt-Meth, responsable d'une répression globale de la traduction. Toutefois, l'expression de certains gènes impliqués dans la survie cellulaire (ATF4, GADD4), la présentation des antigènes (HSP70 et les protéines UPR) et la sécrétion de cytokines est régulée positivement grâce aux uORFs (Barbosa, Peixeiro and Romão, 2013; Somers, Pöyry and Willis, 2013; Starck *et al.*, 2016; Young and Wek, 2016; Chen and Tarn, 2019). Dans ce cas de figure deux modèles sont proposés : la répression de la traduction médiée par les uORFs est détournée ou sert de réservoir à ribosomes pour favoriser la ré-initiation des CDSs en aval (Lu, Harding and Ron, 2004; Starck *et al.*, 2016; Bohlen *et al.*, 2020). Bien que le mécanisme d'action soit encore débattu, les auteurs de ces études s'entendent sur un effet bénéfique des uORFs sur la traduction de certains gènes lors de la réponse au stress cellulaire. Le même phénomène de ré-initiation à partir d'uORFs est également observé lors d'infections virales, telle que l'infection par le virus Ebola (Shabman *et al.*, 2013). Etant donné que l'infection à VIH induit également une réponse intégrée au stress (ISR), il est possible que le virus exploite la réponse au stress médiée par les uORFs pour favoriser la traduction de ses protéines et induire une persistance virale. Cependant, ce mécanisme ne serait effectif que sur les uORFs dont l'initiation dépend de eIF2 et d'un codon AUG. Afin d'étudier le rôle des uORF dans l'expression des gènes viraux, nous pourrions étudier l'impact de mutations des uORFs du VIH et analyser l'expression d'un gène rapporteur, cela dans différentes conditions de culture (avec et sans facteurs d'inducteurs de stress par exemple).

Les ARFs identifiés qui ne sont pas codés dans le 5' UTR se répartissent sur l'ensemble des gènes du VIH. Les ARFs localisés au sein de *gag*, *pol* et *env* sont surreprésentés dans notre jeu de donnée. Nous avons d'ailleurs constaté que le nombre d'ARFs identifiés au sein d'un ORF était corrélé à la taille de l'ORF. Dans les neuf études antérieures utilisant les approches HLA foot-print ou computationnelle, la majorité des ARFs du VIH ont été identifiés et localisés dans les régions *gag*, *pol*, et *nef* (Cardinaud *et al.*, 2004; Garrison *et al.*, 2009; Bansal *et al.*, 2010, 2015; Berger *et al.*, 2010, 2015; Champiat *et al.*, 2012; Ternette *et al.*, 2016; Peng *et al.*, 2018). Grâce au ribosome profiling, nous mettons en évidence la traduction *in vivo* d'ARFs au sein des

ORFs de *vif*, *vpr*, *vpu*, *tat*, *env* et *nef* ainsi que des ARFs chevauchant deux ORFs. Cette méthode nous a donc permis d'obtenir une vision globale et non biaisée des événements de traduction.

Parmi les 98 ARFs que nous identifions par ribosome profiling, 16 ARFs sont en commun avec les 103 ARFs du VIH décrits dans la littérature sur le brin *sens* de l'ARN (5'-3') (Cardinaud *et al.*, 2004; Garrison *et al.*, 2009; Bansal *et al.*, 2010, 2015; Berger *et al.*, 2010, 2015; Champiat *et al.*, 2012; Ternette *et al.*, 2016; Peng *et al.*, 2018). Nous pouvons nous poser la question suivante : pourquoi tous les ARFs de la littérature ne sont-ils pas retrouvés dans notre jeu de données de ribosome profiling ? Tout d'abord, le nombre d'ARFs décrit dans la littérature pourrait être surestimé en raison de l'utilisation d'approches computationnelles qui ne sont que des prédictions et ne reflètent pas fidèlement les événements de traduction qui se produisent *in vivo*. De plus, nous avons appliqué des critères rigoureux de sélection des ARFs, notamment une valeur minimale de 8 reads/codons, correspondant au troisième quartile de l'ensemble de notre jeu de données. Ceci est illustré par l'ARF chevauchant l'ORF de *gag* codant l'épitope Q9VF que le laboratoire a précédemment caractérisé dans des cellules infectées par le VIH (Cardinaud *et al.*, 2004, 2011) dont la couverture (5,9 reads/codons) est juste en dessous de notre critère de sélection. Nos données de ribosome profiling suggèrent donc un taux de traduction plus faible de cet ARF. Dans ce sens, il aurait été intéressant de réaliser un RNAseq en parallèle de notre RiboSeq pour évaluer l'efficacité de traduction (TE) des ORFs et des ARFs du VIH et de comparer la régulation de l'expression des gènes au niveau traductionnel entre les gènes humains et viraux. Outre des critères de sélection rigoureux, nous sommes capables d'identifier un ARF, ARF-ENV2, chevauchant l'ORF de *env* qui a été précédemment mis en évidence dans l'étude caractérisant Q9VF, codant pour les peptides Q10QM, G10QT, et M9PT (Cardinaud *et al.*, 2004).

Par ailleurs, dans notre jeu de données, nous n'identifions pas de reads correspondant à des gènes ou ARFs potentiellement traduits sur le brin *antisens* de l'ARN viral (3'-5'), tel que le gène ASP. ASP (« antisense protein-ASP ») est une protéine de 189 acides aminés codée par un cadre de lecture qui est localisé au sein de *env* et qui est traduit en *antisens* (Savoret *et al.*, 2021). A ce jour, il est difficile de classer ce cadre de lecture d'ORF ou d'ARF puisque son existence est encore débattue dans la communauté scientifique bien qu'une étude génétique semble suggérer que le gène codant pour ASP serait apparu dans le groupe M du VIH-1 lors de l'émergence de la pandémie (Cassan *et al.*, 2016). De plus, des anticorps spécifiques d'ASP ont été retrouvés dans le plasma d'individus infectés (C. Vanhée-Brossollet *et al.*, 1995; Savoret *et al.*, 2020a), et notre équipe a précédemment mis en évidence chez des individus infectés l'existence d'une réponse CTL dirigée contre des épitopes de ASP (Bet *et al.*, 2015). Ces études suggèrent fortement que la protéine ASP est exprimée *in vivo*. Les fonctions de ASP sont elles aussi débattues, mais ASP pourrait manipuler l'autophagie favorisant la réplication viral (Z. Liu *et al.*, 2019). Des études montrent que ASP semble être localisée à la membrane plasmique des cellules infectées et activées (Christine Vanhée-Brossollet *et al.*, 1995; Affram *et al.*, 2019; Savoret *et al.*, 2020b). Une deuxième explication pouvant justifier l'absence d'ASP dans notre ribosome profiling serait que ASP semble être majoritairement exprimée dans les cellules myéloïdes, alors que nous avons utilisé un modèle de cellules lymphoïdes pour nos expériences (Michael *et al.*, 1994; Laverdure *et al.*, 2012; Savoret *et al.*, 2021). Nos collaborateurs

n'identifient pas non plus d'empreintes ribosomales en *antisens* dans leur jeu de données de ribosome profiling obtenu sur le génome viral du VIH (Labaronne *et al.*, 2022). Le translatome du rétrovirus MLV n'a pas non plus mis en évidence l'existence de reads en *antisens* ou bien cela n'a pas été recherché (Irigoyen *et al.*, 2018).

Les résultats de notre étude suscitent une interrogation sur le mécanisme de traduction des ARFs. Bien que la traduction de certains ARFs puisse être dépendante de la présence d'un codon AUG, une grande partie des régions génomiques annotées comme non annotées eucaryotes et virales initient la traduction par des mécanismes non canoniques tels que le leaky scanning, le codon read-through, le ribosome shunting, la ré-initiation, les IRES, les FS, et les codons starts alternatifs (Ingolia, Lareau and Weissman, 2011; Malys and McCarthy, 2011; Firth and Brierley, 2012; Lee *et al.*, 2012; Stern-Ginossar *et al.*, 2012; Shatsky *et al.*, 2018; Yang and Wang, 2019; Andreev *et al.*, 2022). Dans notre modèle, nous pouvons exclure l'hypothèse du codon read-through car les ARFs ne sont pas codés dans la même phase de lecture que leurs ORFs respectives. Dans le but d'obtenir une vision globale des sites d'initiation des ARFs, nous avons réalisé une expérience de ribosome profiling en présence d'harringtonine (HAR). Les drogues inhibitrices de l'élongation telles que l'HAR ou la lactimidomycine sont couramment utilisées dans les expériences de ribosome profiling pour définir les sites d'initiation (Ingolia, Lareau and Weissman, 2011; Stern-Ginossar *et al.*, 2012; Yang *et al.*, 2015; Irigoyen *et al.*, 2018; Machkovech, Bloom and Subramaniam, 2019; Whisnant *et al.*, 2020; Finkel *et al.*, 2021). Bien que l'analyse bio-informatique des données soit encore en cours et délicate en raison de l'activité de l'HAR, qui ne semble pas complète aux vues de signaux de traduction en aval de l'AUG de certains ORFs, nous pouvons d'ores et déjà identifier les codons starts des CDS *env* et *vpr* ainsi que les potentiels sites d'initiation de 4 ARFs. Une alternative serait d'étudier les structures des ARFs afin de détecter des séquences glissantes, des pseudonoeuds ou des contextes Kozak favorables. Nous pourrions également envisager de reproduire ces expériences en essayant d'améliorer les conditions d'utilisation de l'HAR, même si les conditions optimales d'utilisation ont déjà été préalablement établies sur les cellules SupT1.

Nous avons également souhaité définir le translatome du VIH dans des lymphocytes T (LTs) CD4⁺ primaires. Le ribosome profiling de LTs CD4⁺ primaires infectés par l'isolat NL₄₋₃ en présence de cycloheximide est actuellement en cours d'analyse. A l'avenir, nous souhaiterions définir le translatome du VIH dans les cellules dendritiques (DCs) autologues afin d'étudier les événements de traduction dans deux cibles différentes de l'infection à VIH. Premièrement, les DCs étant des APCs professionnelles qui initient la réponse immune spécifique du VIH en présentant des antigènes viraux aux LTs, il serait intéressant d'identifier le panel d'ARFs du VIH activement traduits au sein des DCs infectées. Deuxièmement, la littérature décrit des différences de régulation de la traduction en fonctions des types cellulaires immuns (Vander Lugt *et al.*, 2017; Marchingo and Cantrell, 2022). Suite à la stimulation *in vitro* du TCR, Wolf et al ont démontré une augmentation de l'expression des protéines ribosomales et d'autres mécanismes de traduction dans les LTs activés, provoquant une augmentation de la synthèse des protéines en comparaison aux cellules T naïves (Wolf *et al.*, 2020). En plus de l'augmentation de l'expression des protéines ribosomales, la traduction des transcrits codant pour l'IFN γ , l'IL10 et l'IL4 est également régulée positivement (Marchingo and Cantrell, 2022).

Concernant les DCs, il a été montré une diminution de la transcription et de la traduction des protéines ribosomales lorsque les DCs sont activées (Amit *et al.*, 2009; Ceppi *et al.*, 2009). De plus, l'inhibition de la traduction dans les DCs a été corrélée avec des changements dans l'origine des peptides présentés par les molécules du CMH-I (Lelouard *et al.*, 2007). Aux vues de ces observations, il est donc possible que nous trouvions des différences dans l'expression et la nature des ARFs du VIH entre les T CD4⁺ et les DCs autologues. J'ai établi les conditions expérimentales permettant de générer un grand nombre de DCs à partir de monocytes dérivés du sang ainsi que les conditions d'infection avec des isolats du VIH de laboratoires et cliniques. Des premières expériences de ribosome profiling ont été effectuées mais nous sommes pour le moment limités par la difficulté à obtenir un signal périodique des reads sur les CDS du génome humain. Les conditions expérimentales sont peut-être à améliorer pour cette population cellulaire différente des LTs CD4⁺. Il faudrait par exemple évaluer la nécessité d'utiliser la cycloheximide pour arrêter la traduction, définir la cinétique d'incubation, ainsi que la nature et la concentration de la RNase utilisée pour digérer les polysomes et extraire les fragments d'ARNm protégés par les ribosomes.

En utilisant le ribosome profiling, nous avons pu définir le translatome du VIH et mettre en évidence que le génome non codant est loin d'être silencieux. Depuis le début du nouveau millénaire, une communauté scientifique s'est efforcée à prouver que l'« ADN poubelle » est en fait loin d'être inutile et code pour de petits ORFs (sORFs) in- ou out-of-frames, ayant pour certains un rôle clé dans la régulation de l'expression des gènes, la stabilité des chromosomes, et dans diverses maladies. Dans le cancer, une récente étude met en évidence que 85% des produits de traduction ne sont pas annotés dans la base de donnée du génome humain, et que la moitié d'entre eux codent des produits de traduction alternatifs (Ingolia, Hussmann and Weissman, 2019; Ruiz Cuevas *et al.*, 2021). Khurana et al montrent également que 99% des mutations associées aux tumeurs ont lieux dans les régions non codantes (Khurana *et al.*, 2016). Les mise en évidence de la traduction des régions génomiques non canoniques et les démonstrations de leurs implications dans divers processus cellulaires ou viraux, représentent une avancée significative dans notre compréhension de la génomique et des pathologies.

2. La conservation des ARFs

La seconde partie de mon travail de thèse a consisté à évaluer la conservation des ARFs dans les souches virales circulantes du clade B et C du VIH, dans l'optique d'une utilisation en tant que cible thérapeutique. Pour répondre à cette question, nous avons compilé des séquences du clade B et C du VIH disponibles sur la base de données publique du NIH afin de générer une banque de données interne au laboratoire.

Les ARFs identifiés par ribosome profiling ont une taille variant de 10 à 101 acides aminés (AA). Nos résultats montrent que les ARFs sont fortement conservés au sein des clades B et C du VIH avec conservation médiane en acides aminés de 87,7 et 88,7%, respectivement. Nous montrons également que la phase de l'ORF chevauchant est statistiquement plus conservée que la phase de l'ARF et la troisième phase. Il est intéressant de noter que 8 ARFs ont une conservation plus élevée que la séquence de l'ORF chevauchant.

Dans notre jeu de données, les ARFs sont définis, entre autres, de codon stop à codon stop. Nous gardons donc à l'esprit que la taille des ARFs est surestimée puisque nous n'avons pas encore défini leurs sites d'initiation.

Notre travail sur la conservation en acides aminés des ARFs a été effectué par l'alignement de chaque séquence à la banque de données interne en utilisant le logiciel UGENE. Les séquences incluses dans la banque de données interne ont été sélectionnées sur les critères suivants : des séquences complètes sans mutations aberrantes, une séquence par donneur et provenant de diverses études. Ces critères ont permis d'obtenir un échantillon représentatif en réduisant les biais d'analyse, en évitant les erreurs de séquençage et l'analyse répétitive de la même séquence. Cette démarche de sélection et d'analyse des séquences a également été utilisée par les D^{rs} Cassan et Gross dans l'analyse de la conservation de ASP (Cassan *et al.*, 2016). L'ARF-ENV3 est l'ARF le moins conservé de notre analyse, avec un pourcentage de conservation en AA de seulement 48,1 et 41,8% parmi les isolats du clade B et C, respectivement. Ceci pourrait être expliquée par la localisation d'ARF-ENV3 au sein de la séquence codante pour la boucle hypervariable V1 de Env (Palmer *et al.*, 1996; Curlin *et al.*, 2010; Hamoudi *et al.*, 2013; Pan *et al.*, 2015). De manière générale, la conservation en AA des ARFs ne semble pas être influencée par de la phase de lecture dans laquelle ils sont exprimés. En revanche, j'ai observé une différence très significative de la conservation des ARFs chevauchants les séquences codantes pour *pol* et *env*. Les ARFs situés dans la séquence codante pour Env ont une conservation en AA significativement plus faible, ce qui est en accord avec l'importante diversité génétique de cette région du génome viral (Stefic *et al.*, 2019).

La conservation des ARF pouvant être dictée par la séquence de l'ORF, j'ai comparé le pourcentage de conservation en AA des ARFs, des régions chevauchantes de l'ORF et de la troisième phase. Cette analyse a révélé une conservation significativement plus élevée des séquences des ORFs. De manière générale, pour garantir la fonctionnalité d'une protéine, les mutations au sein de l'ORF sont souvent silencieuses, impliquant ainsi une mutation sur le troisième nucléotide du codon pour pouvoir générer le même acide aminé. Toutefois, une telle mutation n'est pas silencieuse pour les deux autres phases, qui sont directement impactées par une mutation sur le premier ou le deuxième nucléotide du codon. Ces observations suggèrent que les ARFs que nous identifions par ribosome profiling ne sont pas soumis à une pression de sélection positive.

Les données obtenues sur la conservation en AA des ARFs peuvent être interprétées selon deux perspectives : la stratégie thérapeutique et l'évolution génomique.

En ce qui concerne la stratégie thérapeutique, la conservation des ARFs est-elle parmi les isolats circulants des clades B et C que l'on pourrait envisager de les utiliser comme cibles potentielles dans des stratégies de vaccination, particulièrement si les peptides qu'ils encodent sont reconnus par les lymphocytes T d'individus infectés. En raison de sa forte prévalence dans les pays du Sud, la conservation des ARFs dans le clade C offre un avantage thérapeutique supplémentaire. Bien que ces résultats soient encourageants, il serait intéressant d'effectuer une étude longitudinale des ARFs chez les individus infectés. En séquençant le génome viral à différents stades de la maladie, cette étude permettrait de déterminer si les ARFs sont soumis

à une pression de sélection par le système immunitaire. On pourrait également envisager d'évaluer le polymorphisme associé aux molécules HLA par HLA footprint.

La conservation des ARFs nous conduit à émettre l'hypothèse selon laquelle certains d'entre eux pourraient avoir une fonction virologique. Il s'agit d'une hypothèse plausible étant donné que l'analyse de la conservation des sORFs est couramment utilisée pour évaluer le potentiel fonctionnel des polypeptides codés à partir de ces séquences (Crowe, Wang and Rothnagel, 2006; Couso and Patraquim, 2017; Schlesinger and Elsässer, 2022; Wright *et al.*, 2022). Cependant, il est peu probable que la totalité des ARFs aient une fonction virologique (Ruiz-Orera *et al.*, 2018). Afin d'investiguer cette question et de faciliter la démarche, nous pourrions concentrer notre attention sur les 8 ARFs dont le pourcentage de conservation est supérieur à celui de l'ORF chevauchant. Utiliser des essais de perte ou de gain de fonction tels que le système CRISPR/Cas9, la mutagenèse dirigée, ou la surexpression, nous permettrait d'étudier le rôle des polypeptides codés par les ARFs dans la réplication virale. La capacité des mutants à se répliquer dans la cellule hôte pourrait être évaluée par titration virale, PCR quantitative ou par la détection des protéines virales par cytométrie en flux ou ELISA. L'altération de réplication indiquerait alors un rôle de l'ARF dans le cycle réplcatif. On pourrait également envisager d'ajouter un tag par knock-in ou expression ectopique dans les cellules cibles du VIH pour évaluer la localisation et les partenaires cellulaires des polypeptides codés par ARFs.

Une toute autre hypothèse à la fonctionnalité biologique, serait que les ARFs servent de réservoir génétique. La traduction pervasive est un phénomène selon lequel les régions génomiques non codantes sont traduites aléatoirement. Cette permissivité permet l'expression de nouveaux polypeptides qui peuvent être sélectionnés au cours de l'évolution et donner naissance à des protéines fonctionnelles. Dans ce modèle d'émergence de gènes *de novo*, les régions non codantes ont un rôle central en tant que réservoir de nouveauté génétique (Carvunis *et al.*, 2012; Schlötterer, 2015; Wilson *et al.*, 2017; Schmitz, Ullrich and Bornberg-Bauer, 2018; Papadopoulos *et al.*, 2021). La traduction d'une séquence non codante en protéine serait l'une des dernières étapes pour qu'un gène entre dans le monde du codant. Dans ce contexte, il serait pertinent d'examiner la distribution des reads sur l'ARF par rapport aux deux autres phases de lecture concurrentes. Dans une séquence codante, la traduction est en faveur de l'ORF, les deux autres phases de lecture en compétition sont donc traduites de manière pervasive. Une traduction en faveur de l'ARF serait alors représentée par une couverture de reads plus élevée que sur la phase de l'ORF, qui ne présenterait que peu ou pas de reads. Cette observation suggérerait une traduction sélective des ARFs et une évolution de la séquence non codante vers le monde du codant.

L'étude pourrait être complétée avec une analyse d'utilisation des codons. Parmi les codons, il existe les codons optimaux et suboptimaux. Le premier se définit, entre autres, par une sur-représentation d'ARNt appariés et par une vitesse de traduction plus rapide que pour les codons suboptimaux (Hanson and Collier, 2018). La conséquence est un temps d'arrêt du ribosome moins conséquent pour les codons optimaux que les codons suboptimaux pour un acide aminé donné. Des études sur la levure et les bactéries ont montré que l'utilisation de codons suboptimaux dans la région 5'UTR entraîne une baisse significative de l'expression des protéines. Cette diminution est provoquée par l'accumulation des ribosomes en aval de la

séquence suboptimale, perturbant ainsi l'initiation de la traduction (Rosenberg *et al.*, 1993; Chu *et al.*, 2014). Cependant, la présence de codons optimaux dans certaines régions génomiques peut entraîner un défaut de repliement co-translationnel du polypeptide, ce qui favorise la production de produits de traduction défectueux (Thanaraj and Argos, 1996; Zhang, Hubalewska and Ignatova, 2009; Pechmann and Frydman, 2013; Zhou *et al.*, 2015; Komar, 2016). Les données actuelles de la littérature montrent que le modèle d'utilisation des codons est fonctionnellement pertinent (Hanson and Coller, 2018). Par exemple, les protéines essentielles aux processus biologiques tels que la glycolyse et l'érythropoïèse sont (Hanson and Coller, 2018) codées principalement par des codons optimaux ; tandis que les protéines impliquées dans des réponses transitoires sont codées par des codons suboptimaux, de façon à être davantage régulées. En examinant le biais de codon de la séquence des ARFs, nous pourrions obtenir des indications sur la régulation de leurs expressions indiquant indirectement une fonctionnalité potentielle. Le mécanisme par lequel le biais de codon influe sur l'efficacité de traduction reste un sujet débattu, mais des études soutiennent un rôle du taux G/C. Dans cette ligne, nous pourrions envisager de comparer les taux G/C de la séquences codant les ORFs et les ARFs du VIH, ce qui pourrait nous aider à émettre des hypothèses sur le caractère codant des ARFs ou à observer un différentiel de comportement et de régulation de l'expression des ARFs (Newman *et al.*, 2016).

De plus, étudier la capacité des polypeptides codées par les ARFs à se replier et à former une structure en trois dimensions, appelée foldabilité, pourrait également apporter des éléments de réponse quant aux rôles potentiels des polypeptides codés par les ARFs dans le cycle de réplication du virus. En effet, le potentiel de foldabilité est un facteur important de l'émergence des gènes (Papadopoulos *et al.*, 2021) . Nous collaborons actuellement avec l'équipe de bio-informatique moléculaire de notre institut de recherche pour explorer ces approches.

3. Détection des polypeptides issus des ARFs à l'aide du système immunitaire

Nous avons ensuite cherché à déterminer la capacité des ARFs à coder des polypeptides viraux. Dans un premier temps, j'ai cherché à détecter des réponses T spécifiques de peptides dérivés d'ARFs dans les PBMCs de donneurs infectés par ELISPOT IFN γ . J'ai pu mettre en évidence que près de 50% des ARFs du génome du VIH, identifiés par ribosome profiling, codent des polypeptides immunogènes.

En effet, nous avons détecté un total de 60 réponses T spécifiques de peptides dérivés d'ARFs chez les 7 donneurs testés, correspondant à la détection de 46 peptides uniques. La médiane des réponses T spécifiques de peptides dérivés d'ARFs par individu était de 6. Nous avons observé une variation de la magnitude des réponses T allant de 13 à 1250 SFU/10⁶ PBMCs, en fonction du donneur, avec une médiane à 311 SFU/10⁶ PBMCs. De façon intéressante, chez certains individus, j'ai observé des niveaux d'activation des cellules T plus élevés contre les peptides dérivés d'ARFs que contre un pool de peptides contrôles issus de protéines canoniques du VIH ou de virus communs (virus de la grippe, EBV, et CMV). Deux arguments peuvent expliquer cette observation : une fréquence plus élevée de clones T spécifiques de

peptides dérivés d'ARFs chez les donneurs ART-2, ART-3, EC-2, EC-3 et EC-4, et/ou une capacité accrue à sécréter l'IFN γ .

Il est assez difficile de comparer nos résultats avec les cinq études qui ont cherché à détecter des réponses T spécifiques de peptides dérivés d'ARFs chez des individus infectés et pour lesquelles nous identifions des ARFs en commun (voir Annexe n°3) (Cardinaud *et al.*, 2004; Garrison *et al.*, 2009; Berger *et al.*, 2010; Champiat *et al.*, 2012; Peng *et al.*, 2018). Tout d'abord, les protocoles diffèrent : les précédentes études ont réalisé des ELISPOT IFN γ *ex vivo* en incubant les PBMCs décongelés sur la nuit, tandis que nous avons utilisé une méthode d'expansion *in vitro* sur 12 jours. Deuxièmement, ces études ne présentent pas leurs données à un niveau de détails aussi précis que le nôtre ; lorsqu'elles ont identifié un pool positif, elles n'ont pas toujours procédé à une déconvolution, et lorsque la déconvolution a été effectuée, les valeurs de SFU/10⁶ PBMCs ne sont pas toujours disponibles. Enfin, les critères de seuil de positivité ne sont pas identiques. Champiat et al, Garrison et al, ont établi le seuil d'une réponse positive comme étant au moins 50 SFU/10⁶ PBMCs et deux fois le bruit de fond, tandis que pour Berger et al, la réponse doit être d'au moins trois fois le bruit de fond (Garrison *et al.*, 2009; Berger *et al.*, 2010; Champiat *et al.*, 2012). Dans notre étude, le seuil de positivité correspond au minimum à deux fois le bruit de fond, avec une valeur minimale de 20 spots par puits. Il est à noter que l'expansion *in vitro* sur 12 jours des PBMCs représente le principal biais de notre analyse. Bien que cette méthode soit optimisée lorsque le nombre de PBMCs est limité ou lorsque le nombre de peptides à tester est important, l'expansion des cellules T *in vitro* n'est pas forcément représentative de la fréquence et de l'intensité *in vivo* des réponses T spécifiques de peptides dérivés des ARFs (Poschke *et al.*, 2020; Schuhmacher *et al.*, 2022). Il s'agit néanmoins d'une très bonne approche pour réaliser un premier criblage.

Nous sommes la première équipe à mettre en évidence des réponses T spécifiques de peptides dérivés de la région 5'UTR dans le génome du VIH. Cette découverte a d'ailleurs fait l'objet d'une collaboration menant à l'écriture du papier de Labaronne et al, où les auteurs s'intéressent à l'impact de l'infection sur l'hôte et au rôle de l'hélicase DDX3 sur les uORFs du VIH dans la régulation de l'expression des gènes viraux (voir Annexe n°4) (Labaronne *et al.*, 2022).

Nous n'observons pas de différences significatives dans la magnitude des réponses T spécifiques de peptides dérivés d'ARFs entre les individus sous traitement et les EC, à l'exception des réponses obtenues chez EC-1 qui présentent les plus faibles magnitudes. Pour confirmer cette tendance, il serait intéressant d'élargir la taille de notre échantillon afin d'augmenter la puissance statistique de nos analyses. Nous pourrions imaginer que la présence et l'intensité de réponses T spécifiques dérivés d'ARFs soient corrélées avec un profil EC comme l'ont montré Bansal et al en 2015 (Bansal *et al.*, 2015). Toutefois, Champiat et al ont mis en évidence que le traitement antirétroviral (TARV) diminue la magnitude des réponses T dirigées contre les peptides dérivés d'ARFs comparé aux individus virémiques mais élargit le répertoire des cellules T spécifiques (Champiat *et al.*, 2012). Les réponses T et les TARVs exercent tous deux une pression de sélection sur le VIH. Il est donc probable que la pression de sélection médiée par le système immunitaire ou le TARV favorise l'expression des ARFs. Afin de mener une analyse comparative complète, il serait intéressant d'étudier la capacité à monter une

réponse T spécifique de peptides dérivés d'ARFs chez 3 types de donneurs infectés : des donneurs virémiques, avirémiques sous traitement, et controllers.

Dans notre étude, des PBMCs provenant de 3 individus non infectés ont été utilisés comme contrôles négatifs. Sur les 96 peptides testés, seuls 3 peptides dérivés d'ARFs du VIH ont été reconnus par les PBMCs d'un seul donneur manifestant d'ailleurs une forte capacité à sécréter de l'IFN γ de manière non spécifique. Ces trois peptides dérivés d'ARFs ont été exclus de notre analyse mais il faut noter que des réponses T CD4⁺ mémoires préexistantes spécifiques du VIH ont déjà été identifiées chez des individus séronégatifs pour le VIH (Su *et al.*, 2013; Campion *et al.*, 2014), certainement dû à une forte réactivité croisée du récepteur de ces lymphocytes T.

4. Détection par immunopeptidomique des peptides dérivés d'ARFs présentés par les molécules du CMH à la surface des cellules infectées

L'immunopeptidomique est la technique ultime de omique pour confirmer l'expression et la traduction de sORFs, et démontrer de façon directe et non biaisée le rôle des peptides non canoniques en tant que source antigénique (Bassani-Sternberg *et al.*, 2015; Chong, Coukos and Bassani-Sternberg, 2022; Schlesinger and Elsässer, 2022; Wright *et al.*, 2022; Yewdell, 2022). A l'aide de cette approche, nous avons identifié un peptide dérivé d'ARF, nommé YSL, présenté par les molécules HLA-A*02:01 à la surface des cellules infectées.

La découverte du peptide YSL présenté par les molécules HLA-A*02 :01 résulte de la réanalyse des données d'immunopeptidomique publiées par nos collaborateurs (Ziegler *et al.*, 2020). Dans cette étude, Ziegler et al se sont intéressés à l'impact de l'infection à VIH sur le répertoire de peptides cellulaires et viraux présentés par les molécules du CMH-I et aux conséquences sur l'activation des cellules NK. Pour cela, des cellules primaires T CD4⁺ ont été infectées par l'isolat NL₄₋₃, lysées, les molécules HLA de classe I ont été ensuite immunoprécipitées, les épitopes élués, puis séparés par HPLC et analysés par spectrométrie de masse. Les auteurs se sont intéressés aux peptides canoniques du VIH et ont identifié 8 peptides dérivés d'ORFs du VIH sur 3 962 peptides identifiés. L'élargissement de l'analyse aux six cadres de lecture du génome viral, nous a permis de mettre en évidence un peptide dérivé d'ARF : YSL, présenté par les molécules HLA-A*02 :01. Suite à son identification, le spectre de masse du peptide YSL a été confirmé à l'aide de l'approche classique consistant à synthétiser le peptide et à le soumettre à une nouvelle analyse LC-MS/MS pour confirmer sa séquence peptidique. Afin de caractériser davantage cet épitope, nous pourrions envisager d'évaluer son affinité de liaison aux molécules HLA-A*02 :01, par le biais de tests de compétition classique utilisant une sonde radiomarquée, et de la comparer à celles de peptides immunodominants dérivés d'ORFs (Sidney *et al.*, 2013).

Le peptide YSL constitue le deuxième peptide dérivé d'ARFs du VIH mis en évidence par LC-MS/MS. Le premier a été mis en évidence par Ternette et al suite à l'analyse de l'immunopeptidome d'une lignée cellulaire T CD4⁺ infectée par un isolat de laboratoire (Ternette *et al.*, 2016). Sur les 24 individus infectés testés dans cette étude, le peptide alternatif

SR11 issu d'un ARF traduit en *antisens*, est reconnu par les PBMCs d'un seul donneur sur 24 testés avec une magnitude de 250 SFU/10⁶ PBMCs.

L'ARF codant YSL n'a pas passé le seuil de sélection de notre premier ribosome profiling avec une couverture de 2,5 reads/codons mais a été identifié dans le second jeu d'expériences effectuées en présence d'harringtonine, qui est actuellement en cours d'analyse pour l'identification des codons d'initiation. Le peptide YSL a été intégré à nos expériences de crible sur les PBMCs de donneurs infectés. Une séquence peptidique de YSL prédite pour se fixer sur les molécules du CMH de classe II a également été synthétisée et incorporée au crible. Le peptide YSL de classe I est reconnu par les PBMCs de deux individus infectés. La magnitude de la réponse T varie de 120 SFU/10⁶ PBMCs à une saturation complète des puits.

Le peptide YSL est une cible très intéressante en tant que candidat vaccin. En effet, l'ARF YSL est très conservé au sein des clades B et C du VIH avec un pourcentage à 90,8 et 92,8%, respectivement. De plus, la séquence codant pour YSL chevauche la séquence codant pour le site actif de la transcriptase inverse (RT), enzyme essentielle au cycle répliatif. C'est une région contrainte par la sélection positive, où les mutations d'échappement dans l'ORF sont limitées car elles affectent directement la capacité du virus à se répliquer. Au vu de la pression de sélection exercée par les drogues antirétrovirales, qui ciblent la RT, il est possible que, chez les individus infectés traités, des mutations d'échappement au sein de l'ARF YSL puissent apparaître. Parmi le clade B, nous observons plusieurs polymorphismes majoritaires au sein de la séquence de l'ARF codant YSL. Il est d'ailleurs intéressant de noter que nous observons un seul polymorphisme à la position 3 du peptide YSL (I->T), qui pourrait suggérer une pression de sélection médiée par le système immunitaire sur l'épitope. Cette substitution de l'isoleucine en thréonine dans l'épitope codé par l'ARF YSL n'affecte pas la séquence en acide aminé de l'ORF.

Parallèlement à ces travaux de réanalyse, nous avons réalisé des expériences d'immunopeptidomique sur des lignées cellulaires SupT1 et C8166 infectées par l'isolat de laboratoire NL₄₋₃, ainsi que des cellules TCD4⁺ primaires infectées par NL₄₋₃ et un isolat clinique (« transmitted founder virus », TF). Dans ces expériences, nous n'avons pas identifié de peptides dérivés d'ARFs présentés par les molécules du CMH. En résumé, nous avons détecté en moyenne 5 000 ligands des molécules HLA de classe I, enrichi en 9-mer. Cependant, nous avons identifié au maximum 11 peptides viraux présentés par les molécules HLA de classe I, ce qui représente seulement 0,16% des peptides identifiés. A l'aide la lignée C8166, nous avons également étudié l'immunopeptidome des molécules du CMH de Classe-II (HLA-DR et -DQ). Nous avons ainsi d'identifier 13 peptides viraux présentées à la surface de cellules T CD4⁺ infectées. Cette observation constitue la première démonstration par immunopeptidomique d'une présentation antigénique par les molécules du CMH de classe II de peptides endogènes au sein des cellules T infectées par le VIH.

Nous avons été d'abord surpris par cette faible proportion de peptides viraux détectés avant de s'apercevoir que c'est un ordre de grandeur couramment retrouvé dans les données de la littérature sur l'immunopeptidome du VIH (Ternette *et al.*, 2016; Ramarathinam *et al.*, 2018; Chikata *et al.*, 2019; Paes *et al.*, 2019; Ziegler *et al.*, 2020). Cette faible proportion ne peut pas

être expliquée par la modulation de l'expression des molécules du CMH à la surface cellulaire médiée par la protéine Nef, puisque les virus que nous avons utilisés dans nos expériences possèdent un codon stop en amont du gène *nef* qui abolit son expression. En revanche, l'approche LC-MS/MS présente quelques limitations qui entravent son utilisation dans une approche de découverte à large échelle des peptides viraux et des néoantigènes. Dans ce contexte, Chong et al rapportent une faible sensibilité et reproductivité de l'immunopeptidomique (Chong, Coukos and Bassani-Sternberg, 2022). Une des raisons qui peut expliquer ces observations est un taux faible de peptides présentés sur les molécules du CMH par rapport à l'abondance de leur source protéique. En effet, des études cinétiques par spectrométrie de masse avec marquages isotopiques stables des acides aminés en culture cellulaire mettent en évidence une faible corrélation entre le protéome et l'immunopeptidome (Milner *et al.*, 2006; Hoof *et al.*, 2012; Croft *et al.*, 2013). Cependant, Bassani-Sternberg et al apportent la nuance selon laquelle le renouvellement protéique influencerait davantage l'immunopeptidome plutôt que l'abondance protéique (Bassani-Sternberg *et al.*, 2015). Les protéines rapidement dégradées seraient davantage présentées à la surface des cellules. Ces observations rejoignent celles des D^{rs} Yewdell et Apcher qui décrivent les DRiPs et le cycle précoce de la traduction, comme les sources majeures de peptides présentés par les molécules du CMH de classe I (Apcher *et al.*, 2011; Antón and Yewdell, 2014). Une explication technique qui pourrait justifier la difficulté à détecter certains peptides est le faible rendement de l'immunoprécipitation de l'ordre de 0,5 à 3% (Hassan *et al.*, 2014).

Finalement, le prédicteur le plus robuste et sensible d'une présentation antigénique et d'une immunogénicité reste à ce jour la détection de clones T spécifiques de l'antigène (Regner *et al.*, 2001; Engelhard, Brickner and Zaring, 2002; Kotturi *et al.*, 2008; Bassani-Sternberg *et al.*, 2016).

5. Les lymphocytes T spécifiques des peptides dérivés des ARFs

Notre étude démontre que les régions présumées non codantes du VIH ont la capacité à produire des polypeptides viraux qui sont présentés par les molécules du CMH et reconnus par les LTs. Il est connu que les réponses T CD8⁺ mais aussi T CD4⁺ spécifiques du VIH jouent un rôle crucial dans le contrôle de la réplication virale (Kervevan and Chakrabarti, 2021; Rutishauser and Trautmann, 2022). En effet, la charge virale est en partie régulée par la présence de réponses T polyfonctionnelles spécifiques du VIH, qui se caractérisent par la production de cytokines et chimiokines telles que Mip1 β , IFN γ , IL2, CD107a et TNF α (Almeida *et al.*, 2007, 2009; Iglesias *et al.*, 2011; Vingert *et al.*, 2012). La polyfonctionnalité est également associée à l'avidité des TCR, qui est un corrélât majeur de la qualité des cellules T car elle leur permet de reconnaître des antigènes viraux à de faibles concentrations (Lissina *et al.*, 2016). A la lumière de ces observations, nous nous sommes posés les questions suivantes : les peptides dérivés d'ARFs sont-ils reconnus par les lymphocytes T CD8⁺ ou CD4⁺ ? Et quelle est la qualité fonctionnelle de ces réponses ? Pour caractériser les propriétés phénotypiques et fonctionnelles des réponses spécifiques des peptides dérivés d'ARFs nous avons effectué un marquage externe des molécules CD4 et CD8 et intracellulaire des cytokines (ICS) Mip1 β , IFN γ ,

IL2, CD107a et TNF α après 12 jours de culture *in vitro*. Nos résultats montrent que la majorité ces réponses est médiée par les LTs CD4 $^{+}$. Nous mettons également en évidence que les réponses T CD8 $^{+}$ et CD4 $^{+}$ spécifiques de peptides dérivés d'ARFs sont polyfonctionnelles, avec la sécrétion d'au moins 3 cytokines simultanément.

Chez l'ensemble des donneurs, nous avons observé que la majorité des réponses dirigées contre les peptides dérivés d'ARFs sont médiées par les LTs CD4 $^{+}$. Il est intéressant de noter que les réponses CD8 $^{+}$ spécifiques de peptides dérivés d'ARFs sont uniquement observées chez les EC. Nous avons été surpris de constater une prédominance de la réponse T CD4 $^{+}$ spécifique des peptides dérivés d'ARFs. Il n'est pas exclu que l'expansion *in vitro* introduise un biais technique en favorisant la prolifération et la différenciation de LTs CD4 $^{+}$. En effet, l'IL-2 pourrait jouer un rôle dans la génération *ex vivo* de LTs CD4 $^{+}$ cytotoxiques (CTL) à partir de PBMCs (Zheng *et al.*, 2007, 2008). Néanmoins, les peptides synthétiques utilisés dans nos expériences ont une taille moyenne de 18 acides aminés et peuvent donc fixer le sillon des molécules HLA de classe II. D'autre part, il a été récemment révélé dans le domaine du cancer que les cellules T CD4 $^{+}$ reconnaissent également des peptides dérivés de CDS non annotés et des néoantigènes, et que cette reconnaissance est associée à une régression de la maladie (Tran *et al.*, 2014; Malandro *et al.*, 2016; S. Liu *et al.*, 2019; Cachot *et al.*, 2021). Notamment, un certain nombre d'études ont montré une surreprésentation de la réponse T CD4 $^{+}$ spécifique, pouvant aller jusqu'à 67,5% des réponses totales, après vaccination avec des néoantigènes (Kreiter *et al.*, 2015; Ott *et al.*, 2017; Sahin *et al.*, 2017; Sonntag *et al.*, 2018; S. Liu *et al.*, 2019; Hu *et al.*, 2021). Il en est de même dans le modèle d'infection à HCMV où Ingolia et al ont mis en évidence une prévalence de réponses T CD4 $^{+}$ spécifiques de peptides dérivés de longs ARN (lncRNA) non codants (Ingolia *et al.*, 2014). Enfin, les études précédentes sur la capacité des LTs à reconnaître des peptides dérivés d'ARFs du VIH se sont exclusivement focalisées sur la réponse T CD8 $^{+}$, en étudiant les polymorphismes des séquences virales associées uniquement aux molécules HLA de classe I par exemple, ce qui pourrait expliquer pourquoi des réponses T CD4 $^{+}$ dirigées contre des peptides dérivés d'ARFs du VIH n'ont pas été précédemment rapportées (Cardinaud *et al.*, 2004; Maness *et al.*, 2007; Garrison *et al.*, 2009; Bansal *et al.*, 2010, 2015, 2015; Berger *et al.*, 2010, 2015; Champiat *et al.*, 2012; Ternette *et al.*, 2016; Peng *et al.*, 2018).

Nous montrons que 80% des réponses T CD4 $^{+}$ et CD8 $^{+}$ spécifiques des peptides dérivés d'ARFs présentent un profil trifonctionnel. En outre, certains peptides induisent même une réponse T ayant plus de fonctions que certaines réponses spécifiques de peptides dérivés d'ORFs. Chez l'un des trois donneurs pour lesquels une réponse T CD4 $^{+}$ spécifique de YSL-I est détectée, la fréquence des cellules T CD4 $^{+}$ sécrétant 4 cytokines est supérieure à celle observée pour la réponse spécifique au peptide dérivé de Env. Bien que la fréquence des cellules T CD4 $^{+}$ et CD8 $^{+}$ qui sécrètent simultanément 5 cytokines (ou pentafonctionnelles) soit supérieure suite à une stimulation antigénique par des peptides dérivés d'ORFs plutôt que d'ARFs, il n'y a pas de différence notable dans la capacité des cellules T CD4 $^{+}$ et CD8 $^{+}$ à sécréter au minimum 3 cytokines simultanément. De plus, nous observons une similitude des profils cytokiniques entre les cellules T CD4 $^{+}$ et CD8 $^{+}$ dirigées contre les peptides dérivés des ARFs et ORFs. Il est difficile de comparer la polyfonctionnalité observée chez les EC à celles des ART, au vu de notre faible

échantillonnage. Selon les données de la littérature disponibles sur les peptides dérivés d'ORFs, on s'attendrait à observer une réponse plus polyfonctionnelle chez les EC que chez les ARFs, majoritairement caractérisés par une réponse monofonctionnelle. De manière générale, nos résultats sont encourageants sachant que la polyfonctionnalité corrèle mieux à la diminution de la charge virale que la fréquence des réponses T.

Notre étude met en évidence une réponse polyfonctionnelle contre les peptides dérivés d'ARFs, qui partage des caractéristiques fonctionnelles similaires avec les réponses spécifiques des peptides dérivés d'ORFs. Cependant, il convient de nuancer notre propos : après 12 jours de culture, la notion de polyfonctionnalité est relative. C'est pourquoi, nous sommes en train d'analyser la polyfonctionnalité *ex vivo* de certaines réponses spécifiques de peptides dérivés d'ARFs, afin de consolider nos affirmations et d'évaluer une fréquence des cellules T spécifiques qui se rapproche au plus près des phénomènes *in vivo*. Nous envisageons également d'évaluer la capacité à lysier les cellules infectées de certaines réponses T spécifiques de peptides dérivés d'ARFs.

Puisque l'avidité du TCR est corrélée à la polyfonctionnalité de la réponse, nos observations semblent indiquées une haute avidité des TCR des lymphocytes T spécifiques des peptides dérivés d'ARFs (Almeida *et al.*, 2007, 2009; Iglesias *et al.*, 2011; Lissina *et al.*, 2016). La mesure de l'EC50 ainsi qu'un séquençage single-cell peuvent être envisagés sur les cellules T spécifiques triées, pour déterminer l'avidité et le clonotype des TCR de cellules spécifiques de peptides dérivés d'ARF.

6. Perspectives sur les ARFs du VIH : un modèle exploratoire

Les résultats de notre étude mettent en lumière la grande diversité de codage du VIH. Nous démontrons que des ARFs sont activement traduits au sein des lymphocytes T CD4⁺ infectés, et que ces séquences codent des polypeptides viraux reconnus par les lymphocytes T. Ces résultats soulèvent beaucoup de questions fondamentales et conceptuelles sur les mécanismes d'expression et les intérêts biologiques à traduire des régions génomiques alternatives. Nos données confirment le précepte selon lequel les sORFs non annotés et présumés non codants ont un rôle en tant que source antigénique (Apcher *et al.*, 2011, 2022; Apcher, Daskalogianni and Fåhraeus, 2015; Laumont *et al.*, 2018; Smith *et al.*, 2018; Zhang *et al.*, 2019; Erhard *et al.*, 2020; Yewdell and Hollý, 2020; Ruiz Cuevas *et al.*, 2021; Weingarten-Gabbay *et al.*, 2021; Chong, Coukos and Bassani-Sternberg, 2022; Ouspenskaia *et al.*, 2022; Schlesinger and Elsässer, 2022; Wright *et al.*, 2022).

6.1 Réservoir antigénique : cause ou conséquence de la traduction des régions non codantes ?

Le réservoir antigénique que représente les peptides dérivés d'ARFs est-il une cause ou une conséquence de la traduction des régions non codantes ? Il est important d'explorer cette question afin de mieux comprendre l'impact de la modulation de l'expression des gènes sur la

réponse immunitaire, quelle fasse suite à une infection virale ou à une rupture de l'homéostasie.

En me basant que les concepts décrits dans la littérature, je propose les modèles suivants qui sont schématisés en Figure 53 :

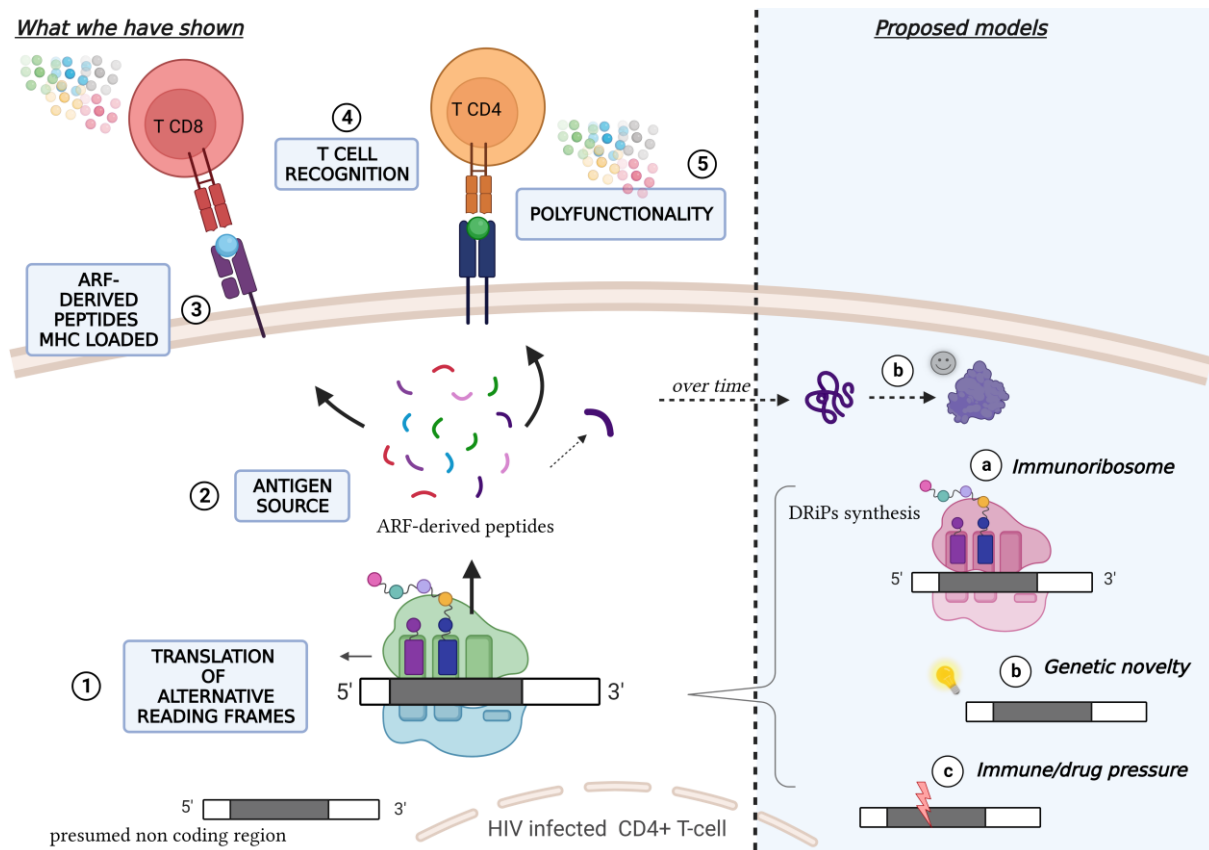


Figure 53. Les ARFs du VIH : résultats du projet et modèles proposés

Les conséquences

La traduction pervasive des régions non codantes permettrait au virus la création d'une réserve de polypeptides non fonctionnels et déstructurés. Au cours du temps, la sélection de certains polypeptides donnerait naissance à de nouveaux gènes ou permettrait leur intégration dans les protéines virales préexistantes. La traduction pervasive assurerait l'évolution et la diversité du VIH en permettant l'émergence de nouvelles protéines virales fonctionnelles et structurées. Cependant, lors de la traduction pervasive, un grand nombre de ces polypeptides ne seraient pas sélectionnés mais dégradés et présentés par les molécules du CMH à la surface de la cellule infectée. L'avantage génomique serait donc contre balancé par une exposition et une reconnaissance par le système immunitaire. Selon ce modèle, la protéine ASP serait l'exemple illustrant l'émergence d'un gène viral *de novo*. Des études doivent encore être menées pour déterminer ses fonctions dans le cycle répliatif, mais ASP est foldable ce qui lui permet un ancrage à la membrane plasmique. La présence de groupements hydrophobes est d'ailleurs l'une des caractéristiques observées dans les protéines virales fonctionnelles issues de sORFs, telles que les viroporines Vpu et M2, issue des virus VIH et IAV, respectivement (DiMaio, 2014; Sze and Tan, 2015; Finkel, Stern-Ginossar and Schwartz, 2018).

Dans un contexte où le stress induit par l'infection modulerait la traduction des uORFs pour favoriser l'expression des gènes viraux, là encore, l'effet bénéfique serait diminué par la reconnaissance des peptides issus de uORFs par les cellules T.

Les causes

Le D^r Yewdell avance que le réservoir immunologique est l'une des causes de la traduction non canonique. Selon sa théorie, une sous-population de ribosome spécialisée dans la synthèse des DRiPs, appelée immunoribosome, assurerait l'immunosurveillance en fournissant des peptides non canoniques pour présentation par les molécules du CMH-I (Yewdell, 2002; Antón and Yewdell, 2014; Yewdell and Hollý, 2020; Dersh, Hollý and Yewdell, 2021). Cette hypothèse est plausible quand on sait que l'hétérogénéité de la composition des ribosomes permet une sélection différentielle des transcrits pour la traduction, et que la déplétion d'une ou plusieurs protéines ribosomales confère aux ribosomes des fonctions distinctes (Dinman, 2016; Shi *et al.*, 2017; Wei *et al.*, 2019). Ce postulat peut donc être généralisé au contexte infectieux. Les immunoribosomes, en traduisant des régions génomiques qui ne correspondent pas à des CDS, permettraient une stimulation antigénique du système immunitaire avant même la traduction des protéines virales canoniques. Les immunoribosomes seraient un mécanisme à part entière de défense en alertant rapidement le système immunitaire du danger en traduisant des DRiPs, ici définis en tant que produits issus de la traduction de régions non codantes. Une étude sur le virus de l'EBV a mis en évidence l'existence d'une protéine virale nommée EBNA1 capable d'inhiber sa traduction et sa dégradation par le protéasome, bloquant ainsi la production des DRiPs de EBNA1 (Apcher, Fahraeus and Manoury, 2004). Ces travaux suggèrent que le VIH pourrait également développer des mécanismes pour inhiber la traduction des DRiPs viraux.

La question que l'on peut se poser est si la génération de DRiPs ne pourrait pas être la conséquence d'un biais de codon ? En effet, comme nous l'avons vu précédemment, le profil des codons impact la biosynthèse des protéines. La présence de codons suboptimaux aux sites importants de structures permet un ralentissement de la traduction qui facilite le repliement correct des protéines (Thanaraj and Argos, 1996). L'optimisation des codons dans ces régions entraîne la production de protéines non structurées et non fonctionnelles qui sont dégradées rapidement (Hanson and Coller, 2018). Il est important de noter que l'optimisation des codons effectuée dans la stratégie du vaccin trivalent de Merck entraîne une augmentation des événements non canoniques de traduction (Bansal *et al.*, 2010). Finalement, le biais de codon pourrait être un des mécanismes utilisés par l'immunoribosome qui sélectionnerait des régions génomiques riches en codons optimaux.

Une autre cause générant la traduction de régions génomiques non canoniques pourrait être la pression de sélection médiée par le système immunitaire, les traitements antirétroviraux (TARVs), et les ARN non codants de l'hôte tels que les miRNA. A cet égard, on pourrait s'attendre à ce que les mutations d'échappement générées par la pression immunologique ou médicamenteuse *in vivo*, en perturbant les trois phases de lecture, facilite l'expression d'ARFs ou d'autres produits de traduction non canoniques.

En attendant avec impatience de nouvelles études qui répondront à ces différentes interrogations, j'aime l'idée que ces mécanismes puissent coexister, et qu'il serait davantage question d'une continuelle balance bénéfices-risques entre les mécanismes viraux et ceux de défenses de l'hôte.

6.2 Réponses lymphocytaires T spécifiques des peptides dérivés d'ARFs et présentation antigénique

Notre étude démontre que les peptides dérivés d'ARFs sont majoritairement reconnus par les LTs CD4⁺. Il serait intéressant de caractériser davantage leurs phénotypes : à quelle sous-population appartiennent-ils ? Quels sont leur état d'activation, d'épuisement et de sénescence ?

Il est stimulant d'envisager une prévalence d'une sous-population de LTs CD4⁺ CTLs dans la reconnaissance des peptides dérivés d'ARFs. Ce postulat se justifie par la présence d'un profil cytokinique dominé par la sécrétion de CD107a, IFN γ , et TNF α (Appay *et al.*, 2002; Zaunders *et al.*, 2004, 2005; Couturier *et al.*, 2014). Nos données sont néanmoins trop préliminaires pour affirmer une sécrétion d'IL2 restreinte aux T CD4⁺ CD107a⁻, qui serait un phénotype compatible avec un profil cytokinique (Appay *et al.*, 2002). Dans l'étude de Johnson *et al.*, les LTs CD4⁺ spécifiques du VIH CD107a⁺ et IFN γ ⁺, ont présenté un profil transcriptionnel et une activité suppressive similaire aux LTs CD8⁺ (Johnson *et al.*, 2015). L'induction d'une réponse T CD4⁺ CTL suite à une stimulation antigénique par les peptides dérivés d'ARFs représenterait une perspective thérapeutique encourageante puisque : 1) les CD4⁺ CTLs sont détectés en phase aigüe et chronique d'infection, 2) les EC montrent une fréquence des CD4⁺ CTLs circulants plus élevée, 3) les individus ayant une fréquence plus élevée de CD4⁺ CD57⁺ ou granzyme A⁺ ont une progression vers la maladie plus lente.

Outre les sous-populations lymphocytaires T impliquées, les voies de présentation antigéniques des peptides dérivés d'ARFs restent entièrement à définir. Il y a de fortes probabilités que les diverses voies alternatives de présentation des antigènes jouent un rôle dans l'initiation de la réponse adaptative par les DCs. L'activation d'une réponse CD4⁺ et CD8⁺ spécifiques de peptides dérivés d'ARFs permet de proposer les modèles suivants :

- A) Présentation des peptides dérivés d'ARFs par les molécules du CMH-I :
 - a. Lorsque les DCs sont infectées de manière productive, les peptides endogènes sont chargés sur les molécules du CMH-I par la voie canonique (i.e. dégradation des protéines par le système UPS, import dans le RE via les protéines TAP, digestion et chargement du peptide par les protéines du PLC, export à la membrane plasmique par la voie sécrétoire).
 - b. Les peptides dérivés d'ARFs sont présentés par les molécules du CMH-I par le biais de la présentation croisée suite à la capture de virions ou de cellules infectées par les DCs.
- B) Présentation des peptides dérivés d'ARFs par les molécules du CMH-II :
 - a. Les peptides dérivés d'ARFs sont internalisés par les DCs puis pris en charge par les différents acteurs de la voie de présentation des antigènes de classe II (i.e.

endocytose des peptides, fusion des endosomes précoces avec des lysosomes, acheminement au MIIC, digestion du peptide et chargement du peptide par l'activité des cathepsines et HLA-DM, export à la membrane plasmique par des vésicules de transports).

- b. Dans les DCs infectées, les peptides issus de la néosynthèse de protéines virales peuvent être chargés sur les molécules du CMH-II selon diverses voies alternatives de chargement : le chargement dans les endosomes tardifs ou le RE indépendant ou dépendant de TAP et HLA-DM, l'autophagie ou via l'implication du récepteur autophagique TAX1BP1 (Coulon *et al.*, 2016; Purnima, Ganesan and Eisenlohr, 2017; Sarango *et al.*, 2022).

Etant donné que seulement 1 à 3% des DCs sont infectées de manière productive par le VIH (Smed-Sørensen *et al.*, 2005), je privilégierais l'implication de la voie de présentation croisée ainsi que la voie canonique dans la présentation des peptides dérivés d'ARFs par les molécules du CMH-I et -II, respectivement. En effet, la présentation croisée du VIH par les DCs est largement documentée dans la littérature (Moris *et al.*, 2006b; Rodriguez-Plata *et al.*, 2012; Dinter *et al.*, 2015; Boucau and Le Gall, 2019; Martín-Moreno and Muñoz-Fernández, 2019). Dans l'optique où les peptides dérivés d'ARFs seraient internalisés dans la particule virale, la fusion des lysosomes avec les corps multivésiculaires formés suite à l'endocytose du virus par les DCs, permettrait également la liaison des peptides aux molécules du CMH-II et la présentation croisée par les molécules du CMH-I.

6.3 Les applications thérapeutiques

L'utilisation des peptides dérivés d'ARFs en tant que cibles thérapeutiques est une perspective encourageante, qui a prouvé et mérité sa place « autour de la table » des candidats vaccins. Dans le domaine du cancer, une récente étude a mis en évidence que 85% des produits de traduction ne sont pas annotés dans la base de donnée du génome humain, et que la moitié d'entre eux codent des produits de traduction alternatifs (Ingolia, Hussmann and Weissman, 2019; Ruiz Cuevas *et al.*, 2021). De plus, Khurana et al ont également montré que 99% des mutations associées aux tumeurs ont lieux dans les régions non codantes (Khurana *et al.*, 2016). Toutes les études discutées dans ce manuscrit démontrent l'implication des régions génomiques non codantes dans les divers processus immunologiques et virologiques. Ces travaux représentent une avancée significative dans notre compréhension de la génomique et des pathologies. Il semble évident que l'entière potentiel des régions non codantes en tant que cibles thérapeutiques n'est pas suffisamment exploité dans le domaine des infections virales, contrairement à la cancérologie où l'utilisation des néoantigènes en tant qu'immunothérapie a fait ses preuves (Bassani-Sternberg *et al.*, 2016; Brennick *et al.*, 2017; Verdegaal and van der Burg, 2017; Bobisse *et al.*, 2018; Arnaud *et al.*, 2021; Chen *et al.*, 2021; Dersh, Hollý and Yewdell, 2021; Jürgens and Wethmar, 2022). L'ensemble de nos travaux élargit la liste des polypeptides viraux conservés qui pourraient être des cibles potentielles pour les stratégies HIV cure. La démonstration d'une initiation de réponses T CD4⁺ et CD8⁺ par les peptides dérivés d'ARFs est d'autant plus avantageuse. Bien que l'on connaisse l'importance d'initier une réponse T CD8⁺ notamment dans les stratégies HIV cure, les preuves que la réponse CD4⁺ CTL est un composant majeur de l'immunité protectrice contre les infections

virales s'additionnent (Juno *et al.*, 2017). Dans l'essai RV144, bien que l'efficacité protectrice partielle semble associée à l'induction d'anticorps spécifiques, des cellules T CD4⁺ mémoires polyfonctionnelle spécifiques de Env ont été détectées, et caractérisées par une forte expression de CD107a et une activité cytotoxique testée *ex vivo* (Rerks-Ngarm *et al.*, 2009; de Souza *et al.*, 2012; Haynes *et al.*, 2012; Lin *et al.*, 2015). S'il s'avère que les réponses T CD4 dirigées contre les peptides dérivés d'ARFs soient cytotoxiques, cela renforcerait d'autant plus l'intérêt à les intégrer dans des essais vaccinaux.

A l'image de l'utilisation des néoantigènes dans les immunothérapies du cancer, les peptides dérivés d'ARFs du VIH pourrait être utiliser en tant que vaccin peptidique, à ARN, ou en ciblant les DCs. Dans cette dernière approche, le candidat vaccin prophylactique « CD40.HIVRI.En », qui est un anticorps monoclonal spécifique du récepteur CD40 couplé à un épitope de l'enveloppe du VIH montre des résultats prometteurs avec des réponses immunes précoces, fortes, et durables (Cheng *et al.*, 2018; Godot *et al.*, 2020; Ceglia *et al.*, 2021; Pastor *et al.*, 2022). La conception d'un tel vaccin permet d'élargir la gamme des réponses induites en changeant ou combinant la nature de(s) l'épitope(s) couplé(s).

L'épitope YSL-I paraît être un candidat vaccin prometteur, de par sa forte conservation, sa localisation au sein du site actif de la transcriptase inverse soumis à la pression de sélection positive, et de la fréquence et polyfonctionnalité des réponses T spécifiques *in vitro*. Des expériences de réponse immunitaire thérapeutique du peptide YSL-I en modèle souris humanisées mériteraient d'être entreprises.

CONCLUSION GÉNÉRALE

L'étude du génome non codant connaît un réel essor depuis ces dix dernières années grâce à l'utilisation de méthodes à grande échelle de omiques, telles que la transcriptomique, le ribosome profiling, la protéomique, et l'immunopeptidomique. Ces méthodes ont permis de mettre en évidence le rôle des ARN dits non codants dans la régulation de l'expression des gènes, l'évolution génomique, le développement embryonnaire, la régulation de processus biologique et dans l'immunité en tant que source antigénique. Dans le cas du cancer et de certaines infections virales, des peptides issus de la traduction non canonique de ces séquences présumées non codantes sont présentés par les molécules du CMH et reconnus par les lymphocytes T.

Dans ce travail de thèse, nous avons donc adopté une approche similaire à grande échelle appliquée à l'infection à VIH. Contrairement à certains virus pour lesquels le ribosome profiling a permis d'établir une liste exhaustive des ARN activement traduits au sein des cellules infectées, aucune démonstration n'avait été faite pour le génome du VIH.

Le ribosome profiling réalisé sur des cellules T CD4⁺ infectées par le VIH a permis la mise en évidence de la traduction d'une centaine de cadre alternatifs de lecture (ARFs). Les ARFs du VIH sont répartis sur l'ensemble du génome viral, et une signature traductionnelle est également identifiée dans la région 5' UTR. La séquence en acide aminé des ARFs est hautement conservée au sein des souches circulantes du clade B et C du VIH, certaines présentant même une conservation supérieure à celle de l'ORF chevauchante. A l'aide deux approches complémentaires, biochimique et cellulaire, j'ai pu montrer que des polypeptides viraux sont exprimés à partir ces ARF, dont près de 50% reconnus par les PBMCs d'individus infectés. Les peptides reconnus par les cellules T dérivent d'ARFs de toutes les régions génomiques, y compris la région 5' UTR. Les réponses T spécifiques de peptides dérivés d'ARFs sont observées aussi bien chez les individus sous traitement que chez les Elite Controllers, avec la majorité des réponses T spécifiques médiées par des lymphocytes T CD4⁺ polyfonctionnels sécrétant au moins 3 cytokines simultanément.

Notre étude établit un lien direct entre la traduction de l'ARNm non canonique du VIH et la présentation antigénique. Les peptides dérivés d'ARFs découverts étendent la liste des polypeptides viraux conservés qui pourraient être ciblés par des stratégies de vaccination efficaces et durables contre le VIH. Ces résultats remettent en question l'idée préconçue de « l'ADN poubelle », et soulignent les rôles biologiques et thérapeutiques des événements de traduction non canoniques et des régions non codantes. En conséquence, il est possible que les ARFs du VIH soient tout aussi polyvalents que les séquences canoniques, et qu'un répertoire immunologique plus vaste reste à explorer. Ces découvertes offrent une lueur d'espoir dans la lutte contre le VIH.

ANNEXES

Table 2 | Recent MS-based discovery studies where immunogenic noncanonical peptides were found

Publication	Tumor type	Type of noncanonical antigen	Strategy/healthy reference	Peptide immunogenicity response/tested	Immunogenic peptide sequences	Test method	Immunogenic sequence detected in other samples
Laumont et al. ⁷¹	Epstein-Barr virus-transformed B cell line (non-tumoral)	<ul style="list-style-type: none"> ● Novel antisense ● Outside exonic regions 	Six-frame translation of B cell transcriptome	2/4	<ul style="list-style-type: none"> ● MKQIKGGSL ● QPNLRVSTV 	<ul style="list-style-type: none"> ● PBMCs cocultured with autologous monocyte-derived DCs pulsed with the peptide. ● ELISpot on isolated CD8⁺ T cells with peptide-pulsed CD8⁻ APCs 	Not reported
Zhang et al. ⁶⁴	Melanoma	RNA-edited peptides from CCNI R75G	Search against large immunopeptidomics tumor and healthy collection	2/2	<ul style="list-style-type: none"> ● LLDGFLATV ● SLLDGFLATV 	<ul style="list-style-type: none"> ● Melanoma TILs and IFNG ELISpot ● Tumor cell killing via pulsing peptides on T2 cells and coculture with TILs 	CCNI-edited peptides are more abundant in cancer immunopeptidomics samples (ovarian cancer, melanoma and breast cancer)
Laumont et al. ⁷⁷	Mouse tumor EL4	<ul style="list-style-type: none"> ● EREs ● Noncoding exons ● UTR ● Exon junctions 	TECs as normal control	5/5	<ul style="list-style-type: none"> ● VNYLHRNV ● VNYIHRNV ● TVPLNHNTL ● VTPVYQHL ● IILEFHSL 	Immunization of C57BL/6 mice with peptide-pulsed DCs before challenge with live tumor cells	Not reported
Chong et al. ⁷⁶	Melanoma	Novel ORF of ABCB5	Ribo-seq-inferred database of matched melanoma sample	1/135 (for 1 specific melanoma sample)	<ul style="list-style-type: none"> ● KYKDRTNILF 	Incubation of peptide with either autologous TILs or peripheral CD8 ⁺ T cells and IFNG ELISpot	Shared in three other melanoma samples at the peptide level
Faridi et al. ⁹⁶	Melanoma	Cis spliced peptides	De novo sequencing and library (re)-searches	6/26 selected	<ul style="list-style-type: none"> ● KLLILELHV ● LILGLLTKV ● LLSLLIPAL ● LLSLLLPAI ● ILSLIPAL ● LLEALEQL 	PBMCs from healthy donors and patients with melanoma stimulated with pooled and individual peptides, with TNF measured by ICS	Shared across multiple patients

PBMCs, peripheral blood mononuclear cells; DCs, dendritic cells; APCs, antigen-presenting cells; IFNG, interferon gamma; ELISpot, enzyme-linked immunospot.

Chong et al., 2022

ANNEXE 2

LITERATURE											
Material & Methods					ARF identified						
Author	Date	Clade	Strain	Technology	Total ARF	Forward	Overlapping ORFs	Reverse ARF	Data set available	Total ARF-derived peptides used for screen (forward)	Unique ARF screened (forward) - excluded overlapping peptides
Champliat	2012	B	HXB2	Bioinformatical	83	13	<u>Forward</u> gag (2), pol (5), vif(1), env(5)	70	Yes	13	13
Bansal	2010	C	Donors	HLA foot-print	391		<u>Forward</u> gag (1), pol (4)		No	5	5
Berger	2010	B	Donors	HLA foot-print	70	70	<u>Forward</u> gag, pol and nef		No	22	18
Garrison	2009	B	HXB2	Bioinformatical	>2 ?	2	Tat and rev	ND	No	2	2
Maness	2007	SIV	mac239	MacVector 8 trial version	1	1	env			1	1
Ternette	2016	B	IIIB	Immunopeptidomic	1		gag	1	Yes		
Berger	2015	B	Donors	Bioinformatical/Los alamos	5		pol and env	5	Yes		
Bansal	2015	B	Donors	Bioinformatical	20	2	<u>Forward</u> pol (2)	18	Yes	2	2
Maness	2010	SIV	mac239	Bioinformatical	1	1	<u>Forward</u> env (1)		Yes	1	1
Peng	2018	B	Donors	HLA foot-print	66	13	<u>Forward</u> gag (4), pol (9)	53	Yes	0	0
Cardinaud	2004	B	LAI	Bioinformatical	3	3	<u>Forward</u> gag (1), pol (1), env (1)		Yes	6	3

Mise en contexte et contribution aux travaux publiés en collaboration : Labaronne et al. (Annexe 4)

L'équipe du D^r Ricci s'est intéressée à caractériser l'impact de l'infection à VIH sur le transcriptome et translatome de la cellule hôte. Ils ont ensuite identifié les signatures traductionnelles dans la région 5' UTR des transcrits viraux dans le but d'identifier des petits cadres ouverts de lectures appelés uORFs (upstream open reading frames). Enfin, l'équipe a évalué l'effet de la dead-box hélicase DDX3, connue pour lever l'inhibition de la traduction médiée par les uORFs et favoriser la traduction de la CDS en aval, sur la régulation des transcrits du VIH.

Par des expériences de omiques et de biochimies, Labaronne et al montrent que l'infection à VIH impacte fortement la transcription des gènes cellulaires mais la modulation de la traduction n'est restreinte qu'à un groupe de transcrits. Les transcrits impliqués dans la biosynthèse des protéines ribosomales ont une expression diminuée, ce qui suggérait une réponse cellulaire au stress induit par l'infection, tandis que ceux intervenant dans la réponse immune et le métabolisme sont augmentés. L'analyse du ribosome profiling a différents points de suivi révèle une traduction de la région 5'UTR tout au long de l'infection. En combinant des expériences de ribosome profiling en présence d'un inhibiteur de la traduction et en utilisant des gènes rapporteurs où les putatifs codons d'initiation non-AUG ont été mutés, les auteurs démontrent que de nombreux uORFs initient leur traduction par le biais de codon non-AUG. Pour finir, les auteurs montrent que les uORFs du VIH exercent une activité de répression de la traduction sur la CDS en aval, et que cette inhibition est modulée par l'activité de DDX3.

Ma contribution à ces travaux a été d'apporter la preuve que les uORFs ont la capacité à coder des polypeptides viraux. Cette partie de l'étude s'est focalisée sur les uORFs communs aux deux jeux de données de ribosome profiling, c'est-à-dire celui obtenu par notre laboratoire et celui obtenu par Labaronne et al. Les peptides dérivés de uORFs ont été sélectionnés sur la base différents paramètres, notamment la conservation des séquences parmi les isolats du VIH, et plus important encore, sur la prédiction de leur capacité à fixer des allèles fréquents des molécules du CMH. Parmi les 7 peptides synthétisés, 5 ont été reconnus par les PBMCs d'individus infectés. De manière remarquable, trois individus sur les sept testés ont montré des réponses T spécifiques dirigées contre les peptides dérivés de uORFs. Certaines réponses montraient des magnitudes équivalentes aux réponses induites par les peptides dérivés d'ORFs.

Extensive uORF translation from HIV-1 transcripts conditions DDX3 dependency for expression of main ORFs and elicits specific T cell immune responses in infected individuals

Emmanuel Labaronne¹, Didier Décimo^{2*}, Lisa Bertrand^{3*}, Laura Guiguettaz¹, Thibault J.M. Sohier¹, David Cluet¹, Valérie Vivet-Boubou⁴, Clara Dahoui², Pauline François³, Isabelle Hatin³, Olivier Lambotte^{5,6,7,8}, Assia Samri⁹, Brigitte Autran⁹, Lucie Etienne², Caroline Goujon¹⁰, Jean-Christophe Paillart⁴, Olivier Namy³, Berta Cecilia Ramirez³, Théophile Ohlmann², Arnaud Moris³, Emiliano P. Ricci^{1#}

¹Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, 46 allée d'Italie F-69364 Lyon, France.

²CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007, Lyon, France.

³Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91190, Gif-sur-Yvette, France.

⁴CNRS, Architecture et Réactivité de l'ARN, UPR 9002, IBMC-2 Allée Konrad Roentgen, Université de Strasbourg, F-67000 Strasbourg, France.

⁵Université Paris-Saclay, Le Kremlin Bicêtre, France.

⁶INSERM UMR-1184, Center for Immunology of Viral Infections and Autoimmune Diseases, Le Kremlin Bicêtre, France.

⁷CEA, DSV/iMETI, Division of Immuno-Virology, IDMIT, Fontenay-aux-Roses, France.

⁸AP-HP, Service de Médecine Interne-Immunologie Clinique, Hôpitaux Universitaires Paris-Saclay, Le Kremlin Bicêtre, France.

⁹Sorbonne Université, INSERM, CNRS, Center for Immunology and Microbial Infections (CIMI-Paris), 75013, Paris, France.

* Equal contribution.

Correspondence: emiliano.ricci@ens-lyon.org

Abstract

Human immunodeficiency virus type-1 (HIV-1) is a complex retrovirus which relies on alternative splicing, translational and post-translational mechanisms to produce more than 15 functional proteins from its single ~10kb transcriptional unit. Here, we have applied ribosome profiling and nascent protein labeling at different time points during infection of CD4⁺ T lymphocytes to characterize the translational landscape of cellular and viral transcripts during the course of infection. Our results indicate a strong impact of viral infection on host cellular transcript levels but a modest impact on global translation rates. Analysis of ribosome profiling reads from viral transcripts reveals extensive and productive non-AUG translation of small peptides from multiple upstream open reading-frames (uORFs) located in the 5' long terminal repeat. Remarkably, these uORFs derived peptides elicit specific T cell responses in HIV-infected individuals. uORFs are conserved among other retroviruses and, together with the TAR sequence, condition the dependency on DDX3 for efficient translation of the main viral open-reading frames.

Introduction

HIV-1 is a single-stranded, positive-sense, diploid and enveloped RNA virus that belongs to the Retroviridae family. Upon infection, the genomic RNA is reverse-transcribed into a double-stranded DNA molecule that is then integrated into the host cell genome as a provirus or in some instances remains unintegrated in a circular form. The provirus DNA (integrated or circular) is transcribed by host RNA polymerase II (Pol II) into a single capped and polyadenylated RNA. Like other complex retroviruses, this single RNA species undergoes extensive alternative splicing to produce a multitude of matured transcripts coding for regulatory proteins in addition to the canonical Gag, Pol and Env proteins (Emery et al., 2017; Nguyen Quang et al., 2020; Ocwieja et al., 2012). Alternative splicing of the viral transcript and expression of the different viral proteins is temporally regulated during the course of infection. Fully spliced transcripts coding for regulatory proteins are preferentially exported to the cytoplasm during early stages of the replication cycle. Partially spliced mRNAs and the unspliced genomic RNA (the latter also coding for Gag and Gag-Pol) are preferentially exported at late stages of the replication cycle through a process that requires the viral encoded protein Rev (Felber et al., 1989).

Translation of viral transcripts is also tightly regulated during the replication cycle through multiple mechanisms of translation initiation for both the unspliced genomic RNA and several spliced variants (de Breyne and Ohlmann, 2018; Guerrero et al., 2015; Jaafar and Kieft, 2019). Indeed, the 5' untranslated region (UTR) of viral transcripts is relatively long (>300 nucleotides) and contains several extensive RNA structures (such as TAR, Poly(A), PBS and the packaging signal stem-loops) that are essential at different steps of the replication cycle but may represent an obstacle for 43S

ribosomes to bind viral mRNAs and reach the translation start codon (Ricci et al., 2008a). As such, internal ribosome entry sites (IRESes) have been described in the 5'UTR of unspliced and some spliced viral transcripts as well as in the coding region of Gag for unspliced RNAs (Brasey et al., 2003; de Breyne and Ohlmann, 2018; Daudé et al., 2016; Deforges et al., 2017; Herbreteau et al., 2005; Locker et al., 2011; Plank et al., 2013; Ricci et al., 2008b; Vallejos et al., 2012). Furthermore, the viral protease encoded from the full-length RNA is capable of mediating proteolytic cleavage of the translation initiation factors eIF4G, eIF4G2 and PABP, which inhibits cap-dependent translation and 43S ribosomal scanning (Alvarez et al., 2003; Castelló et al., 2009; Ohlmann et al., 2002; Prévôt et al., 2003; Ricci et al., 2008b; Ventoso et al., 2001). Translation initiation on viral RNAs can also occur through a cap-dependent process that is facilitated by many host proteins including RNA helicases and RNA binding proteins that facilitate recruitment and scanning of 40S on viral transcripts (Bolinger et al., 2010; de Breyne and Ohlmann, 2018; Chatel-Chaix et al., 2004; Dugré-Brisson et al., 2004; García-de-Gracia et al., 2021; Hartman et al., 2006; Ramos et al., 2021; Singh et al., 2022; Soto-Rifo et al., 2012, 2013). Translation of the full-length RNA is also regulated at the elongation step, through a frameshifting signal which allows the expression of the structural Gag polyprotein and the fusion Gag-Pol polyprotein that contains viral proteins with enzymatic activity such as the protease, reverse-transcriptase and integrase (Jacks et al., 1988).

The regulation of viral protein translation has a strong influence on the initiation of immune responses to the virus, specifically on T cell mediated immunity. T cells recognize, on infected cells, virus-derived peptides presented by major histocompatibility complex (MHC) molecules. In particular, CD8⁺ T cells recognize viral peptide, presented by MHC class-I molecules, that are derived from the processing, by proteasomes, of the native viral proteins and/or, misfolded or truncated viral polypeptides (Wei et al., 2019; Yewdell, 2020). There is a direct link between translation, protein degradation and the loading of MHC-I molecules (Erhard et al., 2020; Pierre, 2009; Wei et al., 2019), in particular in HIV-infected cells (Casartelli et al., 2010; Schubert et al., 2000). Interestingly, CD4⁺ T cells can also recognize HIV peptides, presented by MHC class-II molecules, that are derived from newly synthesized viral proteins (Coulon et al., 2016).

Translation of HIV-1 and closely related lentiviruses has been mainly studied using *in vitro* translation extracts or reporter constructs that have been instrumental to uncover the many mechanisms governing HIV-1 expression. However, a global view of the expression and translation of viral transcripts during a productive replication cycle is still missing. Here, we have performed ribosome profiling and RNA sequencing from cytoplasmic extracts obtained from infected cells and cells with a latent HIV-1 infection in which the replication cycle is coordinately induced. Our results highlight the temporal regulation of expression and translation of viral and cellular transcripts and indicate a modest impact of infection on overall cellular translation rates. Ribosome

profiling data reveals extensive non-AUG translation events from upstream open reading frames (uORFs) located in the 5'UTR of spliced and unspliced viral mRNAs. We further demonstrate that uORFs negatively modulate translation from the main viral ORFs. We then searched for cellular factors that might alleviate the negative effect exerted by uORFs on translation of viral proteins from canonical ORFs and identified the host DEAD-box protein DDX3, which is implicated in regulating uORF translation (Calviello et al., 2021; Chen et al., 2018a; Guenther et al., 2018; Linsalata et al., 2019) as well as translation of HIV-1 viral transcripts (Soto-Rifo et al., 2012, 2013). We demonstrate that DDX3 requirement for efficient expression of HIV-1 viral transcripts alleviates the negative effect that uORFs exert on translation of viral proteins from canonical ORFs. uORF translation from viral transcripts also occurs in the closely related lentivirus HIV-2 and in the more phylogenetically distant retrovirus HTLV-1 (Human T Lymphotropic Virus 1), suggesting that it may be a conserved feature among retroviruses. Finally, using IFN γ ELISPOT assays in PBMCs from HIV-1 infected individuals, we can detect specific T cell responses directed against uORF-derived peptides thus indicating that uORFs encode MHC-ligands with potential relevance in mediating an immune response against infected cells *in vivo*.

Results

HIV-1 infection induces changes in abundance and translation of specific host mRNAs

Infection with HIV-1 has been previously proposed to down-regulate cap-dependent translation through the cleavage of translation initiation factors and the arrest of infected cells in G2-M phase (Alvarez et al., 2003; Castelló et al., 2009; Ohlmann et al., 2002; Sharma et al., 2012; Ventoso et al., 2001). However, this has mainly been tested under over-expression of the viral protease or transfection of plasmids coding for the infectious provirus. A global assessment of cellular and viral translation at the transcriptome-wide scale under productive infection is still missing. To monitor the impact of HIV-1 infection on gene expression, we infected SupT1 CD4⁺ T cells with HIV-1 (NL4-3 strain) to obtain more than 90% of cells expressing p24, 24 hours after infection, as measured by flow cytometry analysis (Supplementary Figure 1A). Mock-infected cells were grown in parallel as a control. Cytoplasmic lysates from HIV-1 and mock-infected cells were recovered at 0, 1, 12, 24 and 36 hours post infection (hpi) to monitor changes in cytoplasmic transcript abundance and ribosome loading by RNA-seq and Ribosome Profiling (Ribo-seq), respectively (Figure 1A). For time course experiments, micrococcal nuclease was used to obtain ribosome footprints instead of RNase I, which has been shown to lead to ribosome degradation (Cenik et al., 2015; Gerashchenko and Gladyshev, 2017; Ricci et al., 2014). However, because tri-nucleotide

periodicity is not as clear when working with micrococcal nuclease, samples at 24hpi were also prepared with RNase I in order to track the reading frame of ribosomes more accurately (Supplementary figure 1B).

Comparison of RNA-seq and Ribo-seq reads at 1hpi indicates that infection induces a mild (less than two-fold) but significant change in the abundance of a handful of cellular transcripts (Figure 1B, Top-left panel). However, their ribosome-loading is not affected, indicating that they have probably not engaged efficiently with the translation machinery or that the cell is buffering their expression at the translational level. At 12hpi, the overall distribution of RNA-seq and Ribo-seq reads correlates well and indicates global but mild (mostly not-significant) changes in transcript abundance (Figure 1B, Top-right panel). At 24 and 36hpi, we observed a strong and global effect on gene expression, with thousands of cellular transcripts being differentially expressed (Figure 1B, Bottom panels). Surprisingly, most changes are driven at the level of cytoplasmic transcript abundance (which can result from changes in transcription, nuclear export or transcript stability), with only a minor subset of genes (423 in total) being regulated at the translational level 36hpi.

To perform differential gene expression and gene-ontology analysis, we further used ribosome profiling reads. Indeed, these correspond to the combination of transcript abundance and ribosome loading and should therefore better correlate with protein output than RNA-seq reads. Supporting this hypothesis, we found that quantifying gene expression with Ribosome profiling data correlates better with proteomics data of HIV-infected SupT1 cells (Golumbeanu et al., 2019) than RNAseq datasets (Supplementary Figure 1C). Gene ontology analysis on up-regulated genes at 36hpi shows an enrichment for proteins involved in cell communication, signaling, immune system processes, and cholesterol pathway, among others (Figure 1C). Down-regulated genes are enriched in RNA splicing, rRNA processing, ribosomal large and small subunit biogenesis, interleukin-7-mediated pathway, cell cycle, regulation of translation, IRES-dependent viral translation initiation (Figure 1C). Gene ontology analysis performed on differentially translated transcripts (423 genes in total 36hpi, 301 up-regulated and 122 down-regulated) did not show any enrichment of specific functional category. Nevertheless, down-translated genes have significantly lower GC content in their CDS and longer CDS than control genes and vice versa for up-translated genes (Supplementary Figure 1D). This suggests a possible role of the endogenous host restriction factor Schlafen 11 (SLFN11), which inhibits HIV-1 replication through GC content and codon-usage (Li et al., 2012). Further analysis of the abundance of differentially expressed transcripts across all time-points indicate that most genes display a gradual trajectory, with mild but coherent changes in cytoplasmic abundance at 12hpi, which become more pronounced as infection progresses (Supplementary Figure 1E).

Global translation rates are modestly affected by HIV-1 infection

In the absence of synthetic exogenous RNA spike-in controls, high-throughput sequencing approaches, such as RNA-seq and ribosome profiling, can inform about relative changes in gene expression but are not able to accurately monitor bulk changes in intracellular transcript abundance or ribosome occupancy (Jiang et al., 2011). To test whether HIV-1 infection is accompanied by changes in overall translation efficiency in infected cells, we probed absolute levels of protein synthesis in infected cells using O-propargyl-puromycin (OPP), a clickable puromycin analog that incorporates into elongating polypeptide chains (Liu et al., 2012) (Figure 1D). For this, SupT1 cells were mock-infected or infected with HIV-1 and OPP incorporation measured at 12, 24 and 36hpi (Figure 1E). As a positive control, we infected HEK293T cells with Sindbis virus (SINV), an alphavirus that has been reported to induce a strong translation shutoff in host infected cells (Garcia-Moreno et al., 2019). As expected, SINV infection induced a strong decrease in OPP incorporation as soon as 12hpi (Supplementary Figure 2A, right panel). However, no significant differences in OPP incorporation were detected between mock-infected and HIV-infected cells at any of the tested time points (Figure 1E). A slight decrease of translation was nevertheless detected for both mock-infected and HIV-1 infected cells at 36hpi, probably due to the higher density of cells in the culture well.

To further validate these results at later time points of infection, SupT1 cells were infected at MOI 0.5 with VSVg pseudotyped, genetically modified single-round HIV-1 virions in which there is a frameshift mutation in the envelope coding sequence and the GFP coding sequence is placed upstream of Nef. This strategy allows to track infected cells and avoid the important cell mortality that is associated with the replication and spreading of wild-type HIV-1 virions at longer time-points of infection. To monitor translation and infection, the level of OPP incorporation and GFP expression was measured at 24, 36, 48, 72 and 96hpi. As observed (Figure 1F), translation rates in GFP positive cells (indicative of HIV-1 replication) increased slightly at 24h compared to the mock-infected control and to the GFP negative cells from the infected condition (corresponding either to non-infected cells or cells in which expression of viral proteins was still not detected). At 48 and 72h after infection, translation efficiency increased both for GFP positive and negative cells from the infected condition compared to mock-infected cells, while at 96h post-infection translation efficiency was similar in all tested conditions.

Finally, we infected activated primary CD4⁺ T-cells obtained from two different donors with wild-type HIV-1 virions (NL4.3 strain) and monitored translation rates at 48 and 72hpi among p24-positive and -negative cells (Figure 1G). In this set-up, we were able to detect a modest decrease in OPP incorporation (close to 20%) in the former compared to the latter, at both time points and for both donors.

Taken together, while no global translational shut-off was detected upon infection, we observed important changes in transcript abundance of host mRNAs and a milder effect on translation of specific host mRNAs in a GC-content dependent manner.

Dynamics of HIV-1 cytoplasmic transcript expression

HIV-1 proviral DNA is transcribed by RNA polymerase II as a single-transcription unit that can be spliced to produce close to 100 different transcripts coding for viral proteins (Nguyen Quang et al., 2020; Ocwieja et al., 2012). These transcripts are expressed in a temporally coordinated manner to produce all accessory and core viral proteins required to efficiently assemble new viral particles. To quantify cytoplasmic levels of HIV-1 transcripts, we collected data from long read sequencing (Nanopore) of NL4-3 infected primary CD4 cells to identify all HIV-1 transcripts expressed in infected cells (Nguyen Quang et al., 2020) and used this information as a template to reconstruct isoform expression from our RNA sequencing reads. A simulated *in silico* dataset of RNA sequencing reads generated from the list of known HIV-1 transcripts, with a similar length distribution to our RNA sequencing libraries, validated our deconvolution approach (Supplementary Figure 3A). At 1hpi, we could only detect unspliced viral RNAs corresponding to incoming genomic RNAs (Figure 2A and Supplementary Figure 3B). At 12hpi, upon reverse-transcription and integration of most of the proviral DNA, viral transcripts were dominated by the fully spliced transcript coding for Nef, although we could still detect important levels of unspliced RNAs in the cytoplasm of infected cells (Figure 2A). The relative amount of fully spliced transcripts (coding for Nef, Tat, Rev) strongly decreased at 24 and 36hpi (Figure 2A). On the contrary, relative abundance of single-spliced and unspliced transcripts, which require Rev for their nuclear export (Cullen, 2003; Harris and Hope, 2000; Yedavalli et al., 2004), increased at 24 and 36hpi (Figure 2A). These results are in agreement with the known expression kinetics of viral transcripts during the replication cycle of HIV-1 (Davis et al., 1997; Kim et al., 1989).

To bypass the heterogeneity introduced by variable timing in reverse-transcription of incoming genomic RNAs, integration and transcription of proviral DNAs, we performed a time-course analysis of cytoplasmic RNA expression and ribosome profiling from U937 cells in which a stably integrated HIV-1 provirus lies in a latent form (Folks et al., 1987). In this system, expression of viral transcripts is triggered in a synchronous manner through incubation of cells with phorbol 12-myristate 13-acetate (PMA) and ionomycin. U937 with a latent integrated HIV-1 provirus were induced and cytoplasmic extracts recovered at 0, 3, 6, 9, 12, 15, 18 and 24 h to monitor RNA levels and mRNA translation. Differential gene expression analysis indicated a strong impact of PMA/ionomycin treatment on cellular gene expression and translation (Figure 2B and 2C). The overlap between differentially expressed transcripts obtained from PMA/ionomycin activated U937

cells and SupT1 cells infected with HIV-1 virions is nevertheless moderate (Supplementary figure 3C), suggesting that changes observed in U937 are mainly driven by PMA/ionomycin and not by expression of viral transcripts. Similarly to results obtained in SupT1 cells, viral transcripts were dominated by the fully spliced transcript coding for Nef at early time points (Figure 2D). Expression of unspliced genomic RNA was also abundant at early time-points and constantly increased throughout time in parallel with the decrease in Nef expression. Surprisingly, relative expression of all other viral transcripts was much lower than that of Nef and the genomic RNA and appeared comparatively constant throughout time.

These results indicate that cytoplasmic levels of fully and partially spliced viral transcripts, as well as that of the unspliced genomic RNA are regulated in a temporal manner, as extensively described in the literature (Davis et al., 1997; Kim et al., 1989). However, expression and translation of unspliced genomic RNAs is already abundant at early time points of infection.

Translational landscape of HIV-1 transcripts

Having characterized the dynamics of the cytoplasmic viral transcriptome, we set out to monitor its translational landscape. As described above, we could detect RNA-seq reads corresponding to cytoplasmic incoming unspliced HIV-1 genomic RNAs as soon as 1hpi (Figure 2A). Surprisingly, we could also detect ribosome footprints on viral transcripts at 1h post infection, suggesting that a fraction of incoming viral RNAs undergo translation (Figure 3A and 3B). As expected, ribosome footprints on incoming genomic RNAs were mainly located on the Gag coding sequence and very few on Pol, which depends on ribosome frameshifting for translation (Figure 3B). However, we could also detect a significant amount of ribosome profiling signal downstream of Pol, which was unexpected because unspliced mRNAs should only code for Gag and Gag-Pol. Ribosome profiling reads downstream of the Gag-Pol coding sequence could correspond to bona-fide 80S ribosomes footprints originating from alternative translation initiation mechanisms or to footprints of RNA-binding proteins co-sedimenting with 80S ribosomes. To quantify the extent of incoming genomic RNAs undergoing translation, we calculated ribosome density on the Gag and Gag-Pol coding sequences at 1, 12, 24 and 36hpi (Figure 3C). As expected, ribosome density (calculated by normalizing ribosome footprinting reads in the Gag-Pol coding region by RNA-seq reads from the same region) was significantly lower at 1h post infection than at all other time-points (2.46, 4.42 and 5.75 times lower at 1hpi than at 12, 24 and 36 hpi, respectively). This suggests that only a small fraction of incoming genomic RNAs are capable of uncoating from viral cores and undergo translation instead of reverse-transcription (Figure 3C).

At later time points, translation of viral transcripts was detected on all canonical coding sequences (Figure 3B and 3D). To quantify translation efficiency, we calculated ribosome occupancy from the

non-overlapping part of each canonical open reading frame, with exception of Rev which overlaps with Tat and Nef. As observed, ribosome occupancy on Pol is much lower than in all other viral ORFs (Figure 3D), owing to its specific translation mechanism which depends on the frameshifting of ribosomes translating Gag (Jacks et al., 1988; Namy et al., 2006). Using the ratio of ribosome footprints in the Pol and Gag coding sequences, we obtained the percentage of frameshifting, which is close to 5%, in agreement with previous observations (Mouzakis et al., 2013). The percentage of frameshifting is robust in the 4~5% range at all time points, being slightly higher at 12hpi than 24 and 36hpi (Figure 3E), suggesting that it is not subjected to strong dynamic regulation during the replication cycle.

Ribosome occupancy on viral coding sequences is however dynamically regulated during the course of infection. Ribosome occupancy on Gag increases from 12 to 24hpi and remains relatively steady at 36hpi. Ribosome occupancy on Env, Vpr and Vpu increases at 36hpi, while on the contrary that of Nef decreases from 12 to 36hpi (Figure 3D). Taken together, similarly to transcription and splicing of viral transcripts, translation of viral ORFs is tightly regulated throughout the replication cycle.

Mapping translation initiation sites on viral transcripts

To characterize ORFs translated from HIV-1 genome and detect new ones, we use Ribocode software (Xiao et al., 2018) that allows the detection of translated ORF analyzing the periodicity of ribosome profiling signals. We performed this analysis on samples prepared using RNaseI or RNaseA+T1, and concatenate results in table 2.

To go further, we mapped translation initiation sites on viral transcripts. For that, U937 cells in which productive replication was induced by addition of PMA/ionomycin (to obtain synchronised expression of viral transcripts) for 0, 3, 6, 9, 12, 15, 18 and 24h were incubated for 15 minutes with the translation initiation inhibitor harringtonine to allow for elongating ribosomes to run-off from translating mRNAs while blocking initiating ribosomes at their translation start sites (Figure 4 and Supplementary Figure 4A) (Ingolia et al., 2011). As shown in figure 4A, this strategy led to a clear enrichment of ribosomes at the predicted start codon on host transcripts. Enrichment of ribosome profiling fragments (RPFs) at canonical translation initiation sites was also detected in viral transcripts (Figure 4B and supplementary figure 4A). The HIV-1 genomic RNA has been shown to contain an IRES that is located within the Gag coding region (downstream of the canonical Gag start codon) which allows translation of an N-terminal truncated form of Gag (also known as p40) from an in-frame AUG-start codon (Daudé et al., 2016; Deforges et al., 2017; Locker et al., 2011; Ricci et al., 2008a; Weill et al., 2010). We were able to detect a low albeit specific signal corresponding to initiating RPFs at the p40 AUG codon (Figure 4C), suggesting that the IRES is

functional *in cellula* although it is orders of magnitude less efficient than cap-dependent translation within the context of infection.

Multiple translation events from non-AUG start codons in the 5'UTR of viral transcripts

Although we could detect initiating RPFs on canonical AUG initiation codons, the majority of initiating ribosomes within viral transcripts originated from the LTR (Figure 4B and Supplementary Figure 4A). This signal was also observed in regular ribosome profiling samples prepared with cycloheximide (in the absence of Harringtonine) in SupT1 and in primary human CD4⁺ T lymphocytes (Puray-Chavez et al., 2020) infected with HIV-1 virions, as well as in U937 cells upon induction of HIV-1 transcription (Figure 5A). Careful observation of the ribosome profiling signal indicated that it overlapped with small ORFs starting with a near-cognate AUG codon. However, knowing the highly structured nature of the LTR and the fact that multiple RNA-binding proteins (including Gag) could be tightly bound to it and co-sediment with 80S ribosomes upon RNase treatment to obtain ribosome footprints, we could not exclude that these signals may originate from footprints of RNA-binding proteins other than ribosomes. Nevertheless, Fragment Length Organization Similarity Score (FLOSS) analysis of 5'UTR viral reads (Ingolia et al., 2014) from infected primary human T CD4⁺ lymphocytes and U937 cells, indicated that these reads had a similar length pattern than ribosome footprints mapped to the CDS of cellular coding mRNAs (Figure 5B). This, strongly suggests that reads originating from the 5'UTR of viral transcripts are bona-fide ribosome footprints.

To further confirm these results, we prepared ribosome profiling libraries from HIV-1 infected cells using the ribolace kit (Clamer et al., 2018), which relies on a biotinylated puromycin analog to specifically pull-down 80S ribosomes competent for elongation. Ribosome profiling libraries were also prepared from cytoplasmic lysates incubated with 1 M of KCl prior to RNase treatment, in order to dissociate RNA-binding proteins from mRNAs as well as 80S ribosomes that are not engaged in active translation and lack a nascent peptide (Mills et al., 2016). Finally, we transfected HEK293T cells, expressing Flag-tagged ribosomes from the endogenous eL8 (RPL7a) locus, with the pNL4.3 plasmid and prepared ribosome profiling libraries by ribosome immunoprecipitation. Ribosome immunoprecipitation bypasses the use of sucrose sedimentation to recover 80S ribosomes and avoids co-sedimenting RNP complexes that could contaminate samples. As observed in Figure 6A, we could observe RPFs in the 5'UTR of viral transcripts in all tested protocols, strongly arguing for genuine uORF translation events. As a final confirmation of these results, we created reporter constructs consisting of the 5'UTR of the HIV-1 genomic RNA, a mutant version where most non-AUG initiation codons were replaced by UAG stop codons (Figure 6B, see No-uORFs 5'UTR) and a second mutant where 4 of the putative non-AUG initiation codons were

replaced by AUG codons (Figure 6B, see AUG-uORFs 5'UTR), followed by the renilla luciferase coding sequence. Compensatory mutations were introduced in mutant 5'UTRs in order to maintain their overall secondary structure, which was probed by SHAPE (Gilmer et al., 2022; Merino et al., 2005; Wilkinson et al., 2006) (Supplementary Figure 5A and B). These reporter plasmids were *in vitro* transcribed to generate capped and polyadenylated RNAs that were transfected into HEK293T cells. 2h30 post-transfection, cells were collected to prepare ribosome profiling samples. As shown in Figure 6B RPFs were detected in the renilla luciferase coding sequence for all constructs tested. However, RPFs in the 5'UTR were only detected in the wild-type and AUG-uORFs mutant transcripts while almost no signal could be detected in the No-uORFs 5'UTR.

Taken together, these results strongly suggest the presence of multiple non-AUG translation initiation sites occurring in the 5'UTR of viral transcripts that might drive productive translation of small ORF upstream of the canonical viral ORF.

uORF translation is conserved among retroviruses

Having shown translation from uORFs originating from HIV-1 unspliced and spliced transcripts, we tested whether this was a conserved feature among retroviruses. For this, we prepared ribosome profiling from HIV-2 (Lentivirus) and HTLV-1 (Deltaretrovirus) infected cells. As shown in Supplementary figure 6C and 6D, non-AUG translation was detected at uORFs located in the 5'UTR of viral transcripts for both viruses. Interestingly, RPFs originating from non-AUG putative uORFs were also detected in the 5'UTR of Murine Leukemia Virus (Gammaretrovirus) as recently shown (Irigoyen et al., 2018)

These results indicate that non-AUG translation occurring at uORFs within the 5'UTR of viral transcripts is a conserved feature of many retroviruses.

DDX3 bypasses the negative effects of uORF on HIV-1 translation

Upstream open-reading frames generally inhibit translation of the main ORF by restricting the fraction of initiating/scanning 40S ribosomal subunits that can reach the canonical start codon (Zhang et al., 2019). To test whether uORFs located in the 5'UTR of viral transcripts restrict translation from the main ORF, we used the HIV-1 WT 5'UTR, AUG-uORFs 5'UTR and No-uORFs 5'UTR renilla luciferase reporter mRNAs described in figure 6B in addition to a control reporter mRNA bearing the 5'UTR of the human beta-globin transcript. To compare translation efficiencies from the different reporter RNAs, we first performed *in vitro* translation reactions using rabbit reticulocyte lysate (RRL). As expected, translation driven by the wild-type HIV-1 5'UTR is less efficient than that driven by the beta-globin 5'UTR in RRL (Figure 6C). Translation of the No-uORFs mutant RNA is also more efficient than that of the wild-type RNA, therefore suggesting that uORFs are able to repress translation of the main ORF by preventing scanning 40S to efficiently

reach the AUG canonical codon (Figure 6C). Translation of the AUG-uORFs mutant RNA is however less efficient than that of the wild-type version, therefore suggesting that initiation on non-canonical codons is not as efficient as that occurring when they are replaced by AUG codons. The DEAD box helicase DDX3 has been shown, in yeast and humans, to mediate bypass of near-cognate start codon uORF initiation and stimulate translation from the main ORF in a large number of cellular transcripts with complex secondary structures similar to those present in the HIV-1 5'UTR (Calviello et al., 2021; Chen et al., 2018a; Guenther et al., 2018). Furthermore, efficient translation of the HIV-1 genomic RNA has been previously shown to rely on DDX3 (Soto-Rifo et al., 2012, 2013). To test whether DDX3 is involved in bypassing uORF translation from the HIV-1 5'UTR, we down-regulated DDX3 in HEK293 using CRISPR-Cas9 and transfected the reporter RNAs to test their translation. As expected, translation driven by the short beta-globin 5'UTR is not significantly sensitive to DDX3 inhibition, while that driven by the wild-type HIV-1 is strongly reduced upon DDX3 down-regulation (Figure 6D and 6E). Translation driven by the UAG-mutant was significantly more resistant to DDX3 depletion than that of the wild-type RNA thus indicating that removal of upstream non-AUG start codons from the HIV-1 5'UTR ("No-uORFs" reporter) partially removes DDX3-dependency for translation of the main ORF (Figure 6D and E). Interestingly, replacing the non-AUG start codons with AUG ("AUG-uORFs" reporter) almost completely removes the dependency on DDX3 for translation of the main ORF (Figure 6D and E). Overall these results strongly suggest that canonical HIV-1 ORF expression is regulated by viral 5' LTR uORFs, in concert with DDX3. Finally, requirement of DDX3 for the translation of HIV-1 viral transcripts was shown to depend on the trans-activator response region (TAR) located immediately downstream the cap-structure (Soto-Rifo et al., 2012). As expected, removing TAR from our wild-type reporter RNA led to a loss of DDX3 dependency for its translation (Figure 6F). This was also the case when TAR was removed from the "No uORFs" reporter RNA therefore suggesting that TAR is required for DDX3 to modulate translation of the main ORF in a uORF-dependent manner.

Peptides derived from 5' LTR uORFs elicit HIV-specific T cell responses in HIV-infected individuals.

To further demonstrate the existence of translation events from non-AUG start codons in the 5'UTR of HIV transcripts and test their potential role in the context of infected patients, we made use of the capacity of T cells to recognize peptide-derived from viral proteins and/or polypeptides presented by MHC molecules. HIV-specific T cells can indeed recognize peptide encoded by classical ORF but also by alternative reading frames (ARF) from HIV-1 genome (Bansal et al., 2015; Berger et al., 2010; Bet et al., 2015; Cardinaud et al., 2004, 2011). To this end, we selected 5

peptides derived from 3 different uORFs based on different parameters including ribosomal coverage, conservation among HIV isolates and more importantly on the prediction to bind a high number of HLA alleles with a high affinity. The peptides were used to screen for T cell responses in PBMCs of HIV-infected individuals. PBMCs from treated (ART) and untreated but aviremic (so called elite controllers, EC) individuals were stimulated with a pool of peptide containing the 5 uORF-derived peptides (POOL) and cultured in the presence of cytokines (IL2 and IL7) in order to expand peptide-specific memory T cells. On day 7 (not shown) and day 12 (Figure 7A-D and Supplementary Figure 6A-D), T cell responses against the POOL were assessed using IFN γ -Elispot. In addition, on day 12, except for donor EC-3, T cell responses to individual peptides of the pool (A02; A03; A06; A07; A08) were also evaluated (Figure 7A-D and Supplementary Figure 6A-D). As positive controls for T cell expansion and activation, PBMCs were also expanded with a pool of HIV Env- and Gag-derived peptides (HIV (+)) or a commercial pool of common virus-derived peptides (CEF (+)). Out of 7 donors tested, 6 reacted to HIV protein-derived peptides (HIV (+)). Note that the ART-3 donor, who did not show a significant T cell response to HIV protein-derived peptides, did however react to the CEF pool (Figure 7A-D and/or Supplementary Figure 6A-D). Remarkably, 3 individuals showed uORF-derived peptide-specific T cells responses, with the individual EC-1 reacting to peptides derived from the 3 uORF tested (Figure 7A). Depending on the peptide, the magnitude of uORF-derived peptide-specific T cells responses (IFN γ ⁺ spots / 10⁶ PBMCs) varied and could reach plateau responses equivalent to responses induced by the pool of HIV protein-derived peptides (see peptide A08 and A02 Figure 7C and D, respectively). These results show that uORF-derived peptides elicit potent T cell responses in HIV-infected individuals. This further implies that uORFs encode polypeptides in the course of natural infection *in vivo*.

Discussion

HIV-1 is a complex retrovirus capable of expressing a multitude of viral proteins from a single transcriptional unit, through a combination of alternative splicing as well as translational and post-translational mechanisms. Expression of viral proteins is tightly coordinated across time during the replication cycle (Lefebvre et al.; Nguyen Quang et al., 2020; Ocwieja et al., 2012) and it has been proposed that HIV-1 infection is associated with a translational shutoff of cellular transcripts that could favor translation of viral transcripts (Alvarez et al., 2003; Castelló et al., 2009; Ohlmann et al., 2002; Prévôt et al., 2003; Sharma et al., 2012; Ventoso et al., 2001).

Our results show that HIV-1 infection induces a strong perturbation of the cellular cytoplasmic transcriptome and a significant change in the translational status of a restricted group of transcripts, with no detectable changes in the global translation rate of the host cell. This contrasts with other RNA viruses, such as picornaviruses or alphaviruses, that induce a rapid and strong translational

shut-off of cellular transcripts, which favors translation of viral transcripts. However, a lack of translational shut-off is not surprising because HIV-1 directly relies on the host transcription machinery for synthesis of viral mRNAs which are both capped and polyadenylated and can be translated in a cap-dependent manner (Berkhout et al., 2011; de Breyne and Ohlmann, 2018; Miele et al., 1996; Ricci et al., 2008a).

Cellular transcripts which expression is down-regulated upon HIV-1 infection are mainly related to the regulation of gene expression and include factors involved in ribosome biogenesis, translation, tRNA biogenesis and aminoacylation, mRNA splicing and processing, which could result from a cellular response to the stress induced by infection. However, these changes are not associated with a significant decrease in global translation rates, but rather with transcript specific changes in translation efficiency, at least within the first 36 hours of infection. In addition to the above mentioned factors, our data revealed that transcripts coding for several host proteins involved in promoting HIV-1 expression are down-regulated upon infection, these include the DEAD-Box protein DDX3 (Rao et al., 2021; Soto-Rifo et al., 2013; Yedavalli et al., 2004), FTSJ3 (Ringard et al., 2019), RNA helicase A (Bolinger et al., 2010; Hartman et al., 2006) among others (See Table1). Cellular transcripts which expression is up-regulated upon infection are mostly associated with the plasma membrane and related to the immune response (including interferon signaling, cell migration and adhesion molecules), cellular stress response pathways and cholesterol metabolism. Interestingly, cholesterol is an important factor to recruit Gag at the plasma membrane (Dick et al., 2012; Favard et al., 2019; Nieto-Garai et al., 2021) and cholesterol depletion has been shown to reduce virion release from infected cells (Ono et al., 2007). Thus, over-expression of factors involved in cholesterol metabolism upon infection could favor Gag assembly and particle release at late time points of the replication cycle.

Cytoplasmic abundance and translation of viral transcripts as measured by RNA-seq and ribosome profiling at different times post-infection indicates a relative sequential expression of fully spliced, partially spliced and unspliced viral transcripts in infected cells (Figure 2). Nevertheless, unspliced mRNA expression and translation is still detected at early replication times. This could be partially explained by the fact that experiments were performed in a bulk population of cells with unsynchronized infection events. However, early expression of Gag/Pol could be important for the replication cycle in addition to their canonical roles at late replication steps. For instance, Integrase is known to interact with TAR RNA and Tat, and has been recently shown to regulate proviral DNA transcription at early time points of infection (Elliott et al., 2020; Liu et al., 2021; Rocchi et al., 2021; Winans and Goff, 2020).

Monitoring translation on viral transcripts revealed the presence of multiple uORFs in the 5'UTR (Figure 5A). Raw ribosome profiling signal originating from these non-AUG uORFs is as important

as that from canonical viral ORFs (Figure 5A), especially in primary CD4⁺ T cells. This is however due to the fact that the section of the 5'UTR, which contains the uORFs is located upstream of the first splice-donor site and therefore is common to all viral transcripts. When probing translation from a single viral mRNA species (Figure 6B), ribosome profiling signal from uORFs was lower than that from the main ORF. Nonetheless, ribosome profiling signals from uORFs could be readily detected under high-salt conditions, ribosome immunoprecipitation or through the Ribolace protocol (Figure 6A). Also mutating all non-AUG start codons into UAA stop-codons led to a complete loss of ribosome profiling signal (Figure 6B), indicating that translation from uORFs might be productive. To further demonstrate that uORFs encoded polypeptides, we reveal in HIV-infected individuals the existence of T cells specific to uORF-derived peptides. Note that, using the same protocol, no uORF-derived peptide-specific T cell responses could be detected in the PBMCs of HIV-negative individuals. In addition, in HIV⁺ individual tested, uORF-derived peptide-specific T cells responses could reach the magnitude induced by a pool of peptide from classical HIV proteins. Overall, these results strongly argue that uORFs encode polypeptides in the course of natural infection *in vivo*.

Remarkably, removal of uORF start codons lead to a mild but significant increase in translation from the main ORF, suggesting that they exert a negative effect on viral protein expression. Nevertheless, this negative effect is partially suppressed by the DEAD-box protein DDX3 (Figure 7D). Interestingly, translation from uORFs is conserved in other retroviruses, such as HIV-2, HTLV and MLV (Supplementary figure 6C, 6D and (Irigoyen et al., 2018)). This raises a question regarding the potential roles of these uORFs. Is it a consequence of the structural and sequence constraints imposed by the multiple functional elements located in the 5'UTR (TAR, poly(A), PBS, dimerization and encapsidation signals)? Do they play a *cis*-regulatory role by restraining the number of ribosomes reaching the main ORF? Cellular mRNAs with uORFs are known to be efficiently translated under stress conditions, similar to those induced by infection (Meijer and Thomas, 2002; Morris and Geballe, 2000). Dynamic regulation of uORF translation through cellular factors could also modulate translation from canonical ORFs in a dynamic fashion during the replication cycle. HIV-1 has been described to contain IRES sequences both in its 5'UTR and in the coding region of Gag (Brasey et al., 2003; de Breyne et al., 2012; Deforges et al., 2017; Locker et al., 2011; Plank et al., 2013; Vallejos et al., 2012; Weill et al., 2010). These IRESes could bypass uORFs to maintain efficient expression of viral proteins. However, our results indicate that translation from the IRES located in the Gag coding region is inefficient in the context of infection and that addition of AUG codons in the 5'UTR strongly represses translation from the main ORF, as previously described (Berkhout et al., 2011), arguing for a limited contribution of IRES translation compared to cap-dependent translation, at least in the context of reporter transcripts.

Peptides originating from uORFs could have a biological role as described for other small peptides produced from cellular uORFs (Jayaram et al., 2021). They might also be presented by MHC class I molecules and used by the immune system to mount potent antiviral T cell responses (Antón and Yewdell, 2014; Ouspenskaia et al., 2021). We highlight here that HIV-1 uORFs can regulate the expression of canonical HIV proteins and that this can be further modulated by host factors such as DDX3. In addition, we show that uORFs induce T cell responses in treated and untreated HIV-infected individuals. In the future it will be interesting to characterize the contribution of HIV uORF-specific T cell responses in the control of viral replication and disease progression and to ask whether the initiation of uORF-specific T cell might be counteracted by cellular factors such as DDX3 or cellular stress conditions that are known to regulate uORF translation.

Material and Methods

Plasmids

LentiCRISPRv2 was a gift from Feng Zhang (Addgene plasmid #52961 (Sanjana et al., 2014)). pMD2.G was a gift from Didier Trono (Addgene plasmid # 12259; <http://n2t.net/addgene:12259>; RRID:Addgene_12259). pBlades plasmids coding for single-guide RNAs targeting the genes coding for eL8 (RPL7a), DDX3 and YFP and were constructed as described in (Mangeot et al., 2019, 2021).

Primers and DNA sequences

Sequence of the sgRNA targeting eL8 used to introduce the Flag-tag is 5' GTACAGCTCACTCACCATCTT 3'. Sequence of the single-stranded DNA oligo used for insertion of a Flag-tag sequence in the eL8 locus by homology directed repair is 5' TTACCCACAATTCCCTTTCTTTCTCTCTCCTCCCGCCGCCCAAGATGGACTACAAAGAC GATGACGACAAGGTGAGTGAGCTGTAGTTCCGTGGCACTATAGCCAGGTTCCGGCTGTA T 3'.

Cell lines and cell culture

HIV-1 latently infected U937 Cells (U1) were obtained from the AIDS reagent program (Reagent number 165) and cultured in RPMI medium supplemented with 10% fetal calf serum (FCS). Induction of productive HIV-1 replication in latently infected U937 cells was obtained by complementing the medium with 100 ng/mL of PMA and 500 ng/mL of Ionomycin. SupT1 cells were obtained from ATCC and cultured in RPMI medium supplemented with 10% fetal calf serum (FCS).

HEK293T Flag-eL8 (Rpl7a) cells cultured in DMEM medium supplemented with 10% fetal calf serum (FCS). They were produced by transfection of lentiCRISPRv1 together with pBLADES-eL8, that expresses a sgRNA targeting the eL8 locus near the translation start codon, and a single-stranded DNA donor template for homology-directed repair containing homology arms to eL8 and a Flag-tag downstream the AUG start-codon. Upon transfection, a clonal population expressing Flag-eL8 was isolated by limit dilution and screening of isolated clones.

Primary cells

Blood from healthy donors was obtained from the Etablissement Français du Sang, under agreement n°21PLER2019-0106. Peripheral blood mononuclear cells (PBMCs) were isolated by centrifugation through a Ficoll® Paque Plus cushion (Sigma-Aldrich). Primary human T cells were purified by positive selection using CD3 MicroBeads (Miltenyi Biotec), as previously described (Goujon et al., 2013). Primary T cells were cultured in RPMI supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin, and stimulated for 48h with 10 µg/ml phytohemagglutinin (PHA) (Fisher Scientific) and 50 U/mL interleukin-2 (IL-2, Miltenyi Biotec) prior to infection.

Viral production and infection

Wild-type HIV-1 NL4-3 was produced by standard polyethylenimine transfection of HEK293T. The culture medium was changed 6h later, and virus-containing supernatants were harvested 42h later. Viral particles were filtered, purified by ultracentrifugation through a sucrose cushion (25% weight/volume in Tris-NaCl-EDTA buffer) for 75 min at 4°C and 28,000 rpm using a SW 32 TI rotor (Beckman Coulter), resuspended in serum-free DMEM medium and stored in small aliquots at -80°C. Viral particles were titrated using an HIV-1 p24Gag Alpha-Lisa kit and an Envision plate reader (Perkin Elmer). VSV-G pseudotyped HIV-1 (NL4.3) and non replicating VSV-G pseudotyped HIV-1-GFP virions were produced by co-transfecting the wild-type HIV-1 NL4-3 or HIV-1 NL4-3 Δ-Env GFP plasmids, with the pMD2.G plasmid coding for the VSV-G envelope glycoprotein.

Infection of SupT1 cells was carried out as follows. Cells were concentrated at 10 million cells per mL and incubated for 1 h at 37°C in the presence of VSV-G pseudotyped HIV-1 virions (NL4.3 strain) at MOI 5. Following this incubation, cells were diluted in RPMI medium (supplemented with 10 % FCS) to a final concentration of 1 million cells per mL and collected at the indicated time points. Cells were then analyzed with the MACSQuant VYB cytometer.

Infection of primary T cells was carried out as follows. Activated cells (see above in the “**primary cells**” section) were plated at 250,000 cell per well in a 96-well plate and infected with 10 or 100 ng

p24Gag for 4 h prior to media replacement. At 48h and 72h post-infection, supernatants were harvested for viral production quantification by Alpha-Lisa.

Flow cytometry analysis of p24 expression in infected cells

Cells infected with HIV-1 virions (NL4.3 strain) were fixed in PBS-paraformaldehyde 4% for 20 min at room temperature and permeabilized with PBS-0.5% Triton X-100 for 20 min. Cells were then incubated with the fluorescently labeled anti-p24 antibody (Clone: KC57, Beckman Coulter, 6604667; 1/200 dilution).

Measurement of protein synthesis using O-propargyl-puromycin (OPP)

After infection with VSVg-pseudotyped HIV-1-GFP or wild-type HIV1 virions (NL4.3 strain), cells were treated with 10 μ M OPP (Immagina Biotechnology, OP001-26) for 30 min at 37°C. Then, cells were fixed in PBS-paraformaldehyde 4% for 20 min at room temperature and permeabilized with PBS-0.5% Triton X-100 for 20 min. For cells infected with HIV1 virus NL4.3 strain, a fluorescent labeling was first made with anti-p24 antibody (Clone: KC57, Beckman Coulter, 6604667; 1/200 dilution). Then, fluorescent labeling of the OPP was made with the Click-IT™ Plus Alexa Fluor™ 488 Picolyl Azide (Thermo Fisher Scientific, C10641), according to the manufacturer's instructions.

For cells infected with VSVg-pseudotyped HIV-1-GFP, fluorescent labeling of the OPP was made with the Click-IT™ Plus Alexa Fluor™ 555 Picolyl Azide (Thermo Fisher Scientific, C10642), according to the manufacturer's instructions. Finally, the cells were analyzed with the MACSQuant VYB cytometer.

Ribosome profiling

Ribosome profiling samples were prepared as described in (Heyer et al., 2015) with the following modifications. At each time point of infection, 10 mL of the cell culture suspension were collected and immediately mixed with 40 mL of ice-cold 1X PBS supplemented with 120 μ g/mL of cycloheximide. Cells were then pelleted at 500 g for 5 min at 4°C, the supernatant was discarded and the cell pellet resuspended in 10 mL of ice-cold 1X PBS + cycloheximide (100 μ g/mL). Cells were pelleted again at 500 g for 5 min, the supernatant was discarded and the cell pellet lysed in 1 ml of lysis buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 μ g/ml cycloheximide and 1 \times Protease-Inhibitor Cocktail EDTA-free (Roche)). Lysate was homogenized with a P1000 pipettor by gentle pipetting up and down for a total of eight strokes and incubated at 4°C for 10 min. The lysate was centrifuged at 1,300 g for 10 min at 4°C, the supernatant recovered and the absorbance at 260 nm measured. For ribosome footprinting, 5 A260 units of the cleared cell lysates were incubated either for 30 min at RT either with 3 μ g of

micrococcal nuclease (Sigma), or 300 units of RNase T1 (Fermentas) and 500 ng of RNase A (Ambion) or for 1 h at 4°C with 350 units of RNase I (Ambion). After nuclease treatment, samples were loaded on top of a 10–50% (w/v) linear sucrose gradient (20 mM HEPES-KOH, pH 7.4, 5 mM MgCl₂, 100 mM KCl, 2 mM DTT and 100 µg/mL of cycloheximide) and centrifuged in a SW-40Ti rotor at 35,000 rpm for 2 h 40 min at 4°C. The collected 80S fraction was complemented with SDS to 1% final and proteinase K (200 µg/mL) and then incubated at 42°C for 45 min. After proteinase K treatment, RNA was extracted with one volume of phenol (pH 4.5)/chloroform/isoamyl alcohol (25:24:1). The recovered aqueous phase was supplemented with 20 µg of glycogen, 300 mM sodium acetate, pH 5.2, and 10 mM MgCl₂. RNA was precipitated with three volumes of 100% ethanol at –20°C overnight. After a wash with 70% ethanol, RNA was resuspended in 5 µl of water and the 3' ends dephosphorylated with PNK (New England BioLabs) in MES buffer (100 mM MES-NaOH, pH 5.5, 10 mM MgCl₂, 10 mM β-mercaptoethanol and 300 mM NaCl) at 37°C for 3 h. Dephosphorylated RNA footprints were then resolved on a 15% acrylamide (19:1), 8 M urea denaturing gel for 1 h 30 min at 35 W and fragments ranging from 26 nt to 32 nt size-selected from the gel. Size-selected RNAs were extracted from the gel slice by overnight nutation at RT in RNA elution buffer (300 mM NaCl, and 10 mM EDTA). The recovered aqueous phase was supplemented with 20 µg of glycogen, 300 mM sodium acetate, pH 5.2, and 10 mM MgCl₂. RNA was precipitated with three volumes of 100% ethanol at –20°C overnight. After a wash with 70% ethanol, RNA was resuspended in 5 µl of water and subjected to cDNA library construction as described below.

For ribosome profiling samples exposed to high-salt in order to induce dissociation of non-elongating 80S ribosomes, the protocol was similar to that described above with the following modification. 1 M KCl (final concentration) was added to 5 A260 units of the cleared lysates and incubated for 30 min at 4°C. After this, samples were passed through Zeba Spin desalting columns 7K MWCO (Thermo Scientific) pre-equilibrated with cell lysis buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl₂, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 µg/mL cycloheximide and 1× Protease-Inhibitor Cocktail EDTA-free (Roche)). Then, samples were treated with nucleases as described above to obtain ribosome footprints.

Ribolace ribosome footprinting experiments were performed following the manufacturer's protocol (Immagina BioTechnology) with the following change, RNase I was replaced with 7 µg of micrococcal nuclease (Sigma) as described above.

Ribosome immunoprecipitation-based profiling was performed as follows. Two 10cm plates of HEK293T cells expressing FLAG-eL8 (RPL7a) from the endogenous locus were transfected with

the pNL4.3-HIV-1 plasmid. 24 h after transfection, medium was removed, cells were washed with ice-cold 1X PBS supplemented with cycloheximide (final concentration of 100 µg/mL) and scraped from the plates. Cells were pelleted at 500g for 5 min and lysed in 1 mL of ice-cold lysis buffer (25mM Tris pH 7.4, 150mM NaCl, 15 mM MgCl₂, 1mM DTT, 1X cOmplete EDTA-free protease inhibitor (Roche), 100µg/mL CHX and 1% Triton X-100) for 10 min on ice. Cell debris and nuclei were removed by centrifugation at 1,300 g for 10 min at 4°C. Cytoplasmic lysate 254nm UV absorption was measured and adjusted to 25 A.U./mL and lysate was treated with 60U RNase T1 and 10 ng RNase A per A.U. for 30 min at 25°C to generate ribosome footprints. Lysate was then diluted-up to 10 mL with ice-cold lysis buffer. 150 µL of anti-FLAG agarose beads suspension (Sigma-Aldrich, A2220) were washed 3 times with 10mL lysis buffer and added to the diluted sample. Ribosome binding was allowed to proceed for 1h at 4°C on rotating wheel. Beads were then washed for a total of 5 times in 15mL tubes using 10mL of lysis buffer for 5 min at 4°C, and tubes were changed after the 1st and 4th wash to minimize contaminants recovery. Beads were washed one final time in a 2 mL low-protein binding tube with 1 mL and ribosomes were eluted using 150 µL of lysis buffer supplemented with 500 µg/mL FLAG peptide (DYKDDDDK peptide - GenScript) for 1 h at 4°C. Eluates were finally digested with proteinase K (200 µg/mL final) for 45 min at 42 °C and RNAs were purified using acid Phenol:Chloroform purification. cDNA libraries were generated as described below.

Ribosomal RNA depletion

For RNA-seq samples, cytoplasmic RNAs were depleted from ribosomal RNAs using antisense DNA oligos complementary to rRNA and RNaseH as previously described (Adiconis et al., 2013). Following DNase treatment to remove DNA oligos, rRNA-depleted RNAs were fragmented using RNA fragmentation reagent (Ambion, Cat: AM8740) for 3 min at 70°C followed by inactivation with the provided “Stop” buffer.

cDNA library preparation for Ribosome profiling and RNA-sequencing

High-throughput sequencing libraries were prepared as described in (Heyer et al., 2015). Briefly, RNAs (either from ribosome footprints or fragmented RNAs from total cytoplasmic RNAs) were dephosphorylated at their 3' end using PNK (New England Biolabs, Cat: M0201) in the following buffer: 100 mM Tris-HCl pH 6.5, 100 mM magnesium acetate, 5 mM β-mercaptoethanol and incubate at 37 °C during 3 h. RNA fragments with a 3'-OH were ligated to a preadenylated DNA adaptor. Then, ligated RNAs were reverse transcribed with Superscript III (Invitrogen) with a barcoded reverse-transcription primer that anneals to the preadenylated adaptor. After reverse transcription, cDNAs were resolved in a denaturing gel (10% acrylamide and 8 M urea) for 1 h and

45 min at 35 W. Gel-purified cDNAs were then circularized with CircLigase II (Lucigen, Cat: CL4115K) and PCR-amplified with Illumina's paired-end primers 1.0 and 2.0. PCR amplicons (12-14 cycles for RNA-seq and 4-6 cycles for ribosome profiling) were gel-purified and submitted for sequencing on the Illumina HiSeq 2000 platform.

Bioinformatic analysis

First, fastq files were demultiplexed using Flexi-splitter software. Next, adaptor sequences were trimmed from RNAseq and Riboseq reads and bad quality reads were filtered with fastp software (v0.20.1) (Chen et al., 2018b). Filtered reads were then mapped on rRNA and tRNA sequenced with bowtie2 v2.3.4.1 (Langmead and Salzberg, 2012) to remove contaminants. Non-aligned reads are then mapped against the human genome with hisat2 v2.1.0 (Kim et al., 2019). Reads that are not mapped against the human genome are mapped against HIV-1 genome with Hisat2 v2.1.0. Duplicated reads were removed using a home-made script. Deduplicated reads were used to quantify gene expression with htseq-count v0.11.2 (Anders et al., 2015). Finally, bam files were used to generate normalized bigwig files with deeptools v3.5.1 (Ramírez et al., 2016).

The periodicity of Ribo-seq experiments was checked with plastid v0.5.1 (Dunn and Weissman, 2016). To identify new ORFs from our datasets, we first mapped filtered reads against hybrid human/HIV-1 genome with STAR v2.7.3a (Dobin et al., 2013) to generate bam files with transcript coordinates. These files are then used by RiboCode v1.2.11 (Xiao et al., 2018), RiboTish v0.2.5 (Zhang et al., 2017) and RiboTricer v1.3.2 (Choudhary et al., 2020) to identify new ORFs in the HIV-1 genome. FLOSS scores were calculated with scripts from (Ingolia et al., 2014).

Patients and Samples

EC (n=4) were recruited from the CO21 CODEX cohort implemented by the ANRS (Agence nationale de recherches sur le SIDA et les hépatites virales). PBMCs were cryopreserved at enrolment. ECs were defined as HIV-infected individuals maintaining viral loads (VL) under 400 copies of HIV RNA/mL without cART for more than 5 years. HIV-infected efficiently treated patients (ART) (n=3) were recruited at Kremlin Bicêtre Hospital. They were treated for at least 1 year (mean of 10 years) and had an undetectable viral load using standard assays. HIV-negative individuals (n=3) were anonymous blood donors (Établissement Français du Sang).

Ethic statement.

All the subjects provided their written informed consent to participate in the study. The CO21 CODEX cohort and this sub-study were funded and sponsored by ANRS and approved by the Ile de France VII Ethics Committee. The study was conducted according to the principles expressed in the Declaration of Helsinki.

12-day in vitro T cell amplification prior Elispot-assay

PBMCs were thawed according to standard procedure and rested 2-3 h in IMDM (Gibco/ThermoFisher Scientific) containing 5% human AB serum (SAB, Institut Jacques Boy), recombinant human IL-2 (rhIL2, 10 U/ml, Miltenyi) and DNase (1 U/ml, New England Biolabs), washed and enumerated. $5-9 \times 10^6$ PBMCs were then seeded/well of a 24-well plate in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids and sodium pyruvate (all from Gibco/ThermoFisher Scientific), Nevirapine (NVP, 1,2 μ M, AIDS reagent program) to inhibit potential viral replication, and poly I:C (2 μ g/ml, InvivoGen) to facilitate the presentation of long peptides by antigen presenting cells (Schuhmacher et al., 2020). PBMCs were immediately loaded with peptides (Vivitide) all at 10 μ g/ml except for the CEF pool (CEF extended peptide pool for human CD8 T cells, Mabtech) used at 5 μ g/ml and cultured overnight at 37°C, 5% CO₂. HIV and/or CEF peptide pools were used as positive control for the expansion of HIV Env- and Gag-specific and common anti-virus-specific T cells, respectively. On day 1 and 3, T cell media were then complemented with rhIL2 (10 U/ml) and rhIL-7 (20 ng/ml, Miltenyi), respectively, and maintained through out the culture. On days 3, 5, 7, and 9, when the cell layer was > 70% confluent, cells were splitted 1:2 before the addition of rhIL-2, rhIL-7 and NVP. On day 7, cells were harvested, counted and a fraction submitted to IFN- γ Elispot assay using the uORF-derived peptide pool and positive controls: HIV and/or CEF peptide pools. The remaining cells were maintained in culture and submitted, on day 12, to IFN- γ Elispot assay using individual uORF peptides and the controls.

Peptide for T cell experiments

Peptide Name	uORF number based on the position of the start codon	AA-sequence (black: NL4-3 / red: common extension among HIV clade B strains)	Peptide Synthetized by
A02	uORF#7	LWVLEIPQTLVSVENL	Vivitide
A03	uORF#12	SQRRSLDAGLGLLKRRARQEA	Vivitide
A06	uORF#11	QWRPNRDLKAKVKPE	Vivitide
A07	uORF#11	EEISRRRTRLAEARTARGE	Vivitide
A08	uORF#11	GEGRRLVSTPKILTSGG	Vivitide
Pool	Mixture of the above-mentioned peptides	See A02, A03, A06, A07, A08	Homemade pool
HIV	Homemade mix of HIV Env- and Gag-derived peptides	gp120 (207-228)KVSFEPIPIHYCAPAGFAILKC gp120 (207-227) KVSFEPIPIHYCAPAGFAILK gp120 (210-228) FEPIPIHYCAPAGFAILKC gp120 (242-261) VSTVQCTHGIRPVVSTQLLL gp120 (320-338) TGEIIGDIRQAHCNISktqW gagp24 (161-180) EKAFSPEVPMFSALSEGAT gagp24 (296-313) YVDRFYKTLRAEQASQEV gagp24 (330-349) CKTILKALGPAATLEEMMTA	Vivitide Homemade pool
CEF	Commercial mix of peptide derived	Peptide from 8-12 aa length. See datasheet for peptide sequences	Mabtech

	from cytomegalovirus Epstein-Barr virus and influenza virus		
--	---	--	--

Enzyme-linked immunospot (IFN- γ Elispot) assay

From 10^5 cells/well were seeded Elispot plates (MSIPN4550, Millipore) in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids and sodium pyruvate, loaded with either 50, 10 or 2 $\mu\text{g/ml}$ of uORF-derived peptides, HIV and CEF peptides, respectively and incubated at 37°C for 16h. Elispot plates were pre-coated with anti-IFN γ primary antibody and after cell-incubation IFN γ revealed using anti-IFN γ secondary antibody conjugated with biotin (both from Mabtech) as described (Cardinaud et al., 2017). The Elispot analysis was performed in technical triplicates or duplicates. Spots were counted with the AID ELISPOT Reader according to standard protocols. Responses were considered positive when IFN γ production was superior to 20 spots/ 10^6 PBMCs and at least twofold higher than background (cells loaded with the DMSO containing medium used to solubilized the peptides).

hSHAPE analysis

Benzoyl cyanide (BzCN) was used to acylate the 2'-hydroxyl group of the unconstrained nucleotides in the RNA structure, followed by interrogation of each nucleotide using two sets of identical but differentially labeled primers: one set (AS1 and AS2) corresponds to the sequence 5'-TCGCTTTCAAGTCCCTGTTCG-3' (complementary to HIV-1 nt 189-209) and the second set (AS3 and AS4) correspond to 5'-TTCTTTCCCCCTGGCCTTAAC -3' (complementary to luciferase sequence nt 395-415). One primer within each set was labeled with either VIC (AS1 & AS3) or NED (AS2 & AS4). The NED-labeled primers from each set were used to prepare a ddG sequencing ladder from the untreated RNA samples. The VIC-labeled primers were used for reverse transcription of the modified or DMSO-treated control RNAs.

Briefly, 1 pmol of *in vitro* transcribed RNA was denatured at 90°C for 2 min and then cooled on ice for 2 min, followed by the addition of excess yeast tRNA (2 μg) and RNasin (5U) in 10 μl HEPES Buffer (30 mM HEPES pH 8, 300 mM KCl, 5 mM MgCl_2). The RNA was then folded at 37°C for 20 min and modified by 3 μl of 300 mM BzCN in DMSO one min at room temperature. After adding 82 μl of water, the chemically modified RNA was extracted (Roti-Phenol/Chloroform/Isoamyl alcohol), ethanol precipitated and resuspended in 7 μl of water. Similarly, for the control (unmodified RNA sample), 3 μl of anhydrous DMSO was added instead of BzCN and treated in the same manner.

For elongation of both the modified and control samples, 1 μl of AS1 or AS3 (1 μM) were added to the resuspended RNA and incubated at 90°C for 2 min, then cooled on ice for 2 min. Two μl of 5X

RT buffer was added to each of the samples and incubated at room temperature for 10 min. Reverse transcription was initiated by addition of 10 μ l of the elongation mix (2 μ l of 5X RT Buffer, 0.6 μ l of 25 mM dNTP and 2 U of AMV RT (Life Science)) and incubation at 42°C for 20 min, 50°C for 30 min and 60°C for 10 min. For the ddG sequencing ladder, 2 pmol of untreated RNA and 1 μ l of the A2 or AS4 (2 μ M) were used and treated as above except for the composition of the sequencing mix (2 μ l of 5X RT Buffer, 2 μ l of 100 μ M ddGTP, 6 μ l of G10 (0.25 mM dGTP, 1mM dATP, 1mM dCTP, 1mM dTTP) and 2U of AMV RT (Life Science)). For each experiment, 80 μ l of water were added and cDNA was extracted using Roti Aqua-Phenol/Chloroform/Isoamylalcohol (Carl Roth). The aqueous phase of modified or unmodified samples were pooled with the aqueous phase of the ddG sequencing ladder. The samples were then ethanol precipitated and resuspended in 10 μ l of HiDi Formamide (ABI). The samples were then denatured 5 min at 90°C, cooled on ice for 5 min, centrifuged, and loaded onto a 96-well plate for sequencing (Applied Biosystems 3130xl genetic analyser).

The electropherograms obtained were analyzed with QuShape algorithm (Karabiber et al., 2013) to extract reactivity data for each sample. The mean reactivity data from three independent experiments were applied as constraints to the RNA sequence in RNAstructure (version 6.1; (Reuter and Mathews, 2010)). The dot bracket file obtained was then used to draw the RNA 2D structure into Structure Editor graphical tool, a module of the RNAstructure software.

Acknowledgments

We gratefully acknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for all computing resources. We thank all participants of the ANRS CODEX cohort and the NIH AIDS Research and Reference Reagent Program for providing drugs. We thank, Eléonore Pérès, Madeleine Duc-Dodon, Louis Gazzolo, Pierre Jalinot and Vincent Mocquet for kindly sharing immortalized T cells infected with HTLV-1. We thank all members of the RMI2 team and Antoine Corbin for manuscript proofreading. We thank Andrea Cimarelli, head of the LP2L

team at CIRI. We also thank the Dormeur Foundation, Vaduz, for providing AID ELISPOT Reader at CIMI and Bernard Maillere (CEA) and his team for granting access to their ELISPOT reader.

Funding sources

This work was funded by Agence Nationale des Recherches sur le SIDA et les Hépatites Virales (ANRS – ECTZ3306), Fondation FINOVI and Sidaction. Emmanuel Labaronne was supported by a postdoctoral fellowship from Agence Nationale des Recherches sur le SIDA et les Hépatites Virales (ANRS). Work in Emiliano Ricci's laboratory is further supported by the European Research Council (ERC-StG-LS6-805500) under the European Union's Horizon 2020 research and innovation programs as well as the ATIP-Avenir program. Lisa Bertrand is a formal Sidaction fellow (Aide aux Equipes) and her current PhD-salary is supported by ANRS. Lucie Etienne is supported by the CNRS and funding from ANRS/MI ECTZ118944. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Contributions

EPR and TO conceived the initial study with help from DD. EPR, EL and DD designed all experiments with the exception of the Elispot tests presented in Figure 6B-F and Supplementary Figure 5C-F, which were conceived, designed and performed by AM, BCR and LB with technical assistance from ON, PF and IH. EL, DD and LG performed most experiments, with technical assistance from DC. EL performed most bioinformatics analyses. OL, AS, BA provided samples from HIV-1 infected individuals and technical support for Elispot experiments. TS created the transgenic cell-lines expressing flag-tagged ribosomes and performed ribosome immunoprecipitation experiments from infected cells. JCP and VVB performed SHAPE analysis on mutated reporter RNAs. CG, CD, LG and LE performed infection experiments in primary human CD4⁺ T cells. EPR wrote the manuscript (with the exception of the Elispot section which was written by AM) with contributions from all authors.

References

- Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D.S., Busby, M.A., Berlin, A.M., Sivachenko, A., Thompson, D.A., Wysocker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* 10, 623–629. <https://doi.org/10.1038/nmeth.2483>.
- Alvarez, E., Menéndez-Arias, L., and Carrasco, L. (2003). The eukaryotic translation initiation factor 4GI is cleaved by different retroviral proteases. *J Virol* 77, 12392–12400. <https://doi.org/10.1128/jvi.77.23.12392-12400.2003>.

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
- Antón, L.C., and Yewdell, J.W. (2014). Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors. *J Leukoc Biol* 95, 551–562. <https://doi.org/10.1189/jlb.1113599>.
- Bansal, A., Mann, T., Sterrett, S., Peng, B.J., Bet, A., Carlson, J.M., and Goepfert, P.A. (2015). Enhanced Recognition of HIV-1 Cryptic Epitopes Restricted by HLA Class I Alleles Associated With a Favorable Clinical Outcome. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 70, 1–8. <https://doi.org/10.1097/QAI.0000000000000700>.
- Berger, C.T., Carlson, J.M., Brumme, C.J., Hartman, K.L., Brumme, Z.L., Henry, L.M., Rosato, P.C., Piechocka-Trocha, A., Brockman, M.A., Harrigan, P.R., et al. (2010). Viral adaptation to immune selection pressure by HLA class I–restricted CTL responses targeting epitopes in HIV frameshift sequences. *Journal of Experimental Medicine* 207, 61–75. <https://doi.org/10.1084/jem.20091808>.
- Berkhout, B., Arts, K., and Abbink, T.E.M. (2011). Ribosomal scanning on the 5′-untranslated region of the human immunodeficiency virus RNA genome. *Nucleic Acids Research* 39, 5232–5244. <https://doi.org/10.1093/nar/gkr113>.
- Bet, A., Maze, E.A., Bansal, A., Sterrett, S., Gross, A., Graff-Dubois, S., Samri, A., Guihot, A., Katlama, C., Theodorou, I., et al. (2015). The HIV-1 Antisense Protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* 12, 15. <https://doi.org/10.1186/s12977-015-0135-y>.
- Bolinger, C., Sharma, A., Singh, D., Yu, L., and Boris-Lawrie, K. (2010). RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res* 38, 1686–1696. <https://doi.org/10.1093/nar/gkp1075>.
- Brasey, A., Lopez-Lastra, M., Ohlmann, T., Beerens, N., Berkhout, B., Darlix, J.-L., and Sonenberg, N. (2003). The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J Virol* 77, 3939–3949. <https://doi.org/10.1128/jvi.77.7.3939-3949.2003>.
- de Breyne, S., and Ohlmann, T. (2018). Focus on Translation Initiation of the HIV-1 mRNAs. *IJMS* 20, 101. <https://doi.org/10.3390/ijms20010101>.
- de Breyne, S., Chamond, N., Décimo, D., Trabaud, M.-A., André, P., Sargueil, B., and Ohlmann, T. (2012). *In vitro* studies reveal that different modes of initiation on HIV-1 mRNA have different levels of requirement for eukaryotic initiation factor 4F: eIF4F requirement for HIV-1 mRNA translation. *FEBS Journal* 279, 3098–3111. <https://doi.org/10.1111/j.1742-4658.2012.08689.x>.
- Calviello, L., Venkataramanan, S., Rogowski, K.J., Wyler, E., Wilkins, K., Tejura, M., Thai, B., Krol, J., Filipowicz, W., Landthaler, M., et al. (2021). DDX3 depletion represses translation of mRNAs with complex 5′ UTRs. *Nucleic Acids Res* 49, 5336–5350. <https://doi.org/10.1093/nar/gkab287>.
- Cardinaud, S., Moris, A., Février, M., Rohrlich, P.-S., Weiss, L., Langlade-Demoyen, P., Lemonnier, F.A., Schwartz, O., and Habel, A. (2004). Identification of Cryptic MHC I–restricted Epitopes Encoded by HIV-1 Alternative Reading Frames. *Journal of Experimental Medicine* 199, 1053–1063. <https://doi.org/10.1084/jem.20031869>.

Cardinaud, S., Consiglieri, G., Bouziat, R., Urrutia, A., Fourati, S., Malet, I., Guernon, J., van Endert, P., Lemonnier, F.A., Appay, V., et al. (2011). CTL Escape Mediated by Proteasomal Destruction of an HIV-1 Cryptic Epitope. *PLoS Pathogens* 7, 15. .

Cardinaud, S., Urrutia, A., Rouers, A., Coulon, P.-G., Kervevan, J., Richetta, C., Bet, A., Maze, E.A., Larsen, M., Iglesias, M.-C., et al. (2017). Triggering of TLR-3, -4, NOD2, and DC-SIGN reduces viral replication and increases T-cell activation capacity of HIV-infected human dendritic cells. *Eur. J. Immunol.* 47, 818–829. <https://doi.org/10.1002/eji.201646603>.

Casartelli, N., Guivel-Benhassine, F., Bouziat, R., Brandler, S., Schwartz, O., and Moris, A. (2010). The antiviral factor APOBEC3G improves CTL recognition of cultured HIV-infected T cells. *Journal of Experimental Medicine* 207, 39–49. <https://doi.org/10.1084/jem.20091933>.

Castelló, A., Franco, D., Moral-López, P., Berlanga, J.J., Alvarez, E., Wimmer, E., and Carrasco, L. (2009). HIV-1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4GI and PABP cleavage. *PLoS One* 4, e7997. <https://doi.org/10.1371/journal.pone.0007997>.

Cenik, C., Cenik, E.S., Byeon, G.W., Grubert, F., Candille, S.I., Spacek, D., Alsallakh, B., Tilgner, H., Araya, C.L., Tang, H., et al. (2015). Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res* 25, 1610–1621. <https://doi.org/10.1101/gr.193342.115>.

Chatel-Chaix, L., Clément, J.-F., Martel, C., Bériault, V., Gatignol, A., DesGroseillers, L., and Moulard, A.J. (2004). Identification of Staufin in the Human Immunodeficiency Virus Type 1 Gag Ribonucleoprotein Complex and a Role in Generating Infectious Viral Particles. *Mol Cell Biol* 24, 2637–2648. <https://doi.org/10.1128/MCB.24.7.2637-2648.2004>.

Chen, H.-H., Yu, H.-I., Yang, M.-H., and Tarn, W.-Y. (2018a). DDX3 Activates CBC-eIF3-Mediated Translation of uORF-Containing Oncogenic mRNAs to Promote Metastasis in HNSCC. *Cancer Res* 78, 4512–4523. <https://doi.org/10.1158/0008-5472.CAN-18-0282>.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018b). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.

Choudhary, S., Li, W., and D Smith, A. (2020). Accurate detection of short and long active ORFs using Ribo-seq data. *Bioinformatics* 36, 2053–2059. <https://doi.org/10.1093/bioinformatics/btz878>.

Clamer, M., Tebaldi, T., Lauria, F., Bernabò, P., Gómez-Biagi, R.F., Marchioretto, M., Kandala, D.T., Minati, L., Perenthaler, E., Gubert, D., et al. (2018). Active Ribosome Profiling with RiboLace. *Cell Rep* 25, 1097-1108.e5. <https://doi.org/10.1016/j.celrep.2018.09.084>.

Coulon, P.-G., Richetta, C., Rouers, A., Blanchet, F.P., Urrutia, A., Guerbois, M., Piguet, V., Theodorou, I., Bet, A., Schwartz, O., et al. (2016). HIV-Infected Dendritic Cells Present Endogenous MHC Class II-Restricted Antigens to HIV-Specific CD4+ T Cells. *J Immunol* 197, 517–532. <https://doi.org/10.4049/jimmunol.1600286>.

Cullen, B.R. (2003). Nuclear RNA export. *J Cell Sci* 116, 587–597. <https://doi.org/10.1242/jcs.00268>.

Daudé, C., Décimo, D., Trabaud, M.-A., André, P., Ohlmann, T., and de Breyne, S. (2016). HIV-1 sequences isolated from patients promote expression of shorter isoforms of the Gag polyprotein. *Arch Virol* 161, 3495–3507. <https://doi.org/10.1007/s00705-016-3073-7>.

- Davis, A.J., Li, P., and Burrell, C.J. (1997). Kinetics of viral RNA synthesis following cell-to-cell transmission of human immunodeficiency virus type 1. *J Gen Virol* 78 (Pt 8), 1897–1906. <https://doi.org/10.1099/0022-1317-78-8-1897>.
- Deforges, J., de Breyne, S., Ameer, M., Ulryck, N., Chamond, N., Saaidi, A., Ponty, Y., Ohlmann, T., and Sargueil, B. (2017). Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. *Nucleic Acids Res* 45, 7382–7400. <https://doi.org/10.1093/nar/gkx303>.
- Dick, R.A., Goh, S.L., Feigenson, G.W., and Vogt, V.M. (2012). HIV-1 Gag protein can sense the cholesterol and acyl chain environment in model membranes. *Proc Natl Acad Sci U S A* 109, 18761–18766. <https://doi.org/10.1073/pnas.1209408109>.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Dugré-Brisson, S., Elvira, G., Boulay, K., Chatel-Chaix, L., Mouland, A.J., and Desgroseillers, L. (2004). Interaction of Staufen1 with the 5' end of mRNA facilitates translation of these RNAs. *33*, 4797–4812. <https://doi.org/10.1093/nar/gki794>.
- Dunn, J.G., and Weissman, J.S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics* 17, 958. <https://doi.org/10.1186/s12864-016-3278-x>.
- Elliott, J.L., Eschbach, J.E., Koneru, P.C., Li, W., Puray-Chavez, M., Townsend, D., Lawson, D.Q., Engelman, A.N., Kvaratskhelia, M., and Kutluay, S.B. (2020). Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. *ELife* 9, e54311. <https://doi.org/10.7554/eLife.54311>.
- Emery, A., Zhou, S., Pollom, E., and Swanstrom, R. (2017). Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J Virol* 91, e02515-16. <https://doi.org/10.1128/JVI.02515-16>.
- Erhard, F., Dölken, L., Schilling, B., and Schlosser, A. (2020). Identification of the Cryptic HLA-I Immunopeptidome. *Cancer Immunol Res* 8, 1018–1026. <https://doi.org/10.1158/2326-6066.CIR-19-0886>.
- Favard, C., Chojnacki, J., Merida, P., Yandrapalli, N., Mak, J., Eggeling, C., and Muriaux, D. (2019). HIV-1 Gag specifically restricts PI(4,5)P2 and cholesterol mobility in living cells creating a nanodomain platform for virus assembly. *Science Advances* <https://doi.org/10.1126/sciadv.aaw8651>.
- Felber, B.K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T., and Pavlakis, G.N. (1989). rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 86, 1495–1499. <https://doi.org/10.1073/pnas.86.5.1495>.
- Folks, T.M., Justement, J., Kinter, A., Dinarello, C.A., and Fauci, A.S. (1987). Cytokine-induced expression of HIV-1 in a chronically infected promonocyte cell line. *Science* 238, 800–802. <https://doi.org/10.1126/science.3313729>.
- García-de-Gracia, F., Gaete-Argel, A., Riquelme-Barrios, S., Pereira-Montecinos, C., Rojas-Araya, B., Aguilera, P., Oyarzún-Arrau, A., Rojas-Fuentes, C., Acevedo, M.L., Chnaiderman, J., et al. (2021). CBP80/20-dependent translation initiation factor (CTIF) inhibits HIV-1 Gag synthesis by

targeting the function of the viral protein Rev. *RNA Biology* 18, 745–758.
<https://doi.org/10.1080/15476286.2020.1832375>.

Garcia-Moreno, M., Noerenberg, M., Ni, S., Järvelin, A.I., González-Almela, E., Lenz, C.E., Bach-Pages, M., Cox, V., Avolio, R., Davis, T., et al. (2019). System-wide Profiling of RNA-Binding Proteins Uncovers Key Regulators of Virus Infection. *Molecular Cell* 74, 196-211.e11.
<https://doi.org/10.1016/j.molcel.2019.01.017>.

Gerashchenko, M.V., and Gladyshev, V.N. (2017). Ribonuclease selection for ribosome profiling. *Nucleic Acids Res* 45, e6–e6. <https://doi.org/10.1093/nar/gkw822>.

Gilmer, O., Mailler, E., Paillart, J.-C., Mouhand, A., Tisné, C., Mak, J., Smyth, R.P., Marquet, R., and Vivet-Boudou, V. (2022). Structural maturation of the HIV-1 RNA 5' untranslated region by Pr55^{Gag} and its maturation products. *RNA Biology* 19, 191–205.
<https://doi.org/10.1080/15476286.2021.2021677>.

Golumbeanu, M., Desfarges, S., Hernandez, C., Quadroni, M., Rato, S., Mohammadi, P., Telenti, A., Beerenwinkel, N., and Ciuffi, A. (2019). Proteo-Transcriptomic Dynamics of Cellular Response to HIV-1 Infection. *Sci Rep* 9. <https://doi.org/10.1038/s41598-018-36135-3>.

Goujon, C., Moncorgé, O., Bauby, H., Doyle, T., Ward, C.C., Schaller, T., Hué, S., Barclay, W.S., Schulz, R., and Malim, M.H. (2013). Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* 502, 559–562. <https://doi.org/10.1038/nature12542>.

Guenther, U.-P., Weinberg, D.E., Zubradt, M.M., Tedeschi, F.A., Stawicki, B.N., Zagore, L.L., Brar, G.A., Licatalosi, D.D., Bartel, D.P., Weissman, J.S., et al. (2018). The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature* 559, 130–134.
<https://doi.org/10.1038/s41586-018-0258-0>.

Guerrero, S., Batisse, J., Libre, C., Bernacchi, S., Marquet, R., and Paillart, J.-C. (2015). HIV-1 replication and the cellular eukaryotic translation apparatus. *Viruses* 7, 199–218.
<https://doi.org/10.3390/v7010199>.

Harris, M.E., and Hope, T.J. (2000). RNA export: insights from viral models. *Essays Biochem* 36, 115–127. <https://doi.org/10.1042/bse0360115>.

Hartman, T.R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D.R., and Boris-Lawrie, K. (2006). RNA helicase A is necessary for translation of selected messenger RNAs. *Nature Structural & Molecular Biology* 13, 509–516. <https://doi.org/10.1038/nsmb1092>.

Herbreteau, C.H., Weill, L., Décimo, D., Prévôt, D., Darlix, J.-L., Sargueil, B., and Ohlmann, T. (2005). HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat Struct Mol Biol* 12, 1001–1007. <https://doi.org/10.1038/nsmb1011>.

Heyer, E.E., Ozadam, H., Ricci, E.P., Cenik, C., and Moore, M.J. (2015). An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Research* 43, e2–e2. <https://doi.org/10.1093/nar/gku1235>.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* 147, 789–802.
<https://doi.org/10.1016/j.cell.2011.10.002>.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome Profiling Reveals Pervasive Translation Outside of

Annotated Protein-Coding Genes. *Cell Rep* 8, 1365–1379.
<https://doi.org/10.1016/j.celrep.2014.07.045>.

Irigoyen, N., Dinan, A.M., Brierley, I., and Firth, A.E. (2018). Ribosome profiling of the retrovirus murine leukemia virus. *Retrovirology* 15, 10. <https://doi.org/10.1186/s12977-018-0394-5>.

Jaafar, Z.A., and Kieft, J.S. (2019). Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology* 17, 110. <https://doi.org/10.1038/s41579-018-0117-x>.

Jacks, T., Powert, M.D., Masiarzt, F.R., Luciw, P.A., Barr, P.J., and Varmus, H.E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. 4. .

Jayaram, D.R., Frost, S., Argov, C., Liju, V.B., Anto, N.P., Muraleedharan, A., Ben-Ari, A., Sinay, R., Smoly, I., Novoplansky, O., et al. (2021). Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2018899118.
<https://doi.org/10.1073/pnas.2018899118>.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 21, 1543–1551.
<https://doi.org/10.1101/gr.121095.111>.

Karabiber, F., McGinnis, J.L., Favorov, O.V., and Weeks, K.M. (2013). QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* 19, 63–73. <https://doi.org/10.1261/rna.036327.112>.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37, 907–915.
<https://doi.org/10.1038/s41587-019-0201-4>.

Kim, S.Y., Byrn, R., Gropman, J., and Baltimore, D. (1989). Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J Virol* 63, 3708–3713. <https://doi.org/10.1128/JVI.63.9.3708-3713.1989>.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.

Lefebvre, G., Desfarges, S., Uyttebroeck, F., Muñoz, M., Beerenwinkel, N., Rougemont, J., Telenti, A., and Ciuffi, A. Analysis of HIV-1 Expression Level and Sense of Transcription by High-Throughput Sequencing of the Infected Cell. *Jvi.Asm.Org*.

Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., Jones, T.E., Landry, S., Pan, T., Weitzman, M.D., et al. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* 491, 125–128. <https://doi.org/10.1038/nature11433>.

Linsalata, A.E., He, F., Malik, A.M., Glineburg, M.R., Green, K.M., Natla, S., Flores, B.N., Krans, A., Archbold, H.C., Fedak, S.J., et al. (2019). DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation. *EMBO Reports* 20, e47498.
<https://doi.org/10.15252/embr.201847498>.

Liu, J., Xu, Y., Stoleru, D., and Salic, A. (2012). Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *PNAS* 109, 413–418. <https://doi.org/10.1073/pnas.1111561108>.

- Liu, S., Koneru, P.C., Li, W., Pathirage, C., Engelman, A.N., Kvaratskhelia, M., and Musier-Forsyth, K. (2021). HIV-1 integrase binding to genomic RNA 5'-UTR induces local structural changes in vitro and in virio. *Retrovirology* 18, 37. <https://doi.org/10.1186/s12977-021-00582-0>.
- Locker, N., Chamond, N., and Sargueil, B. (2011). A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3. *Nucleic Acids Research* 39, 2367–2377. <https://doi.org/10.1093/nar/gkq1118>.
- Mangeot, P.E., Risson, V., Fusil, F., Marnef, A., Laurent, E., Blin, J., Mournetas, V., Massouridès, E., Sohier, T.J.M., Corbin, A., et al. (2019). Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nature Communications* 10. <https://doi.org/10.1038/s41467-018-07845-z>.
- Mangeot, P.E., Guiguettaz, L., Sohier, T.J.M., and Ricci, E.P. (2021). Delivery of the Cas9/sgrRNA Ribonucleoprotein Complex in Immortalized and Primary Cells via Virus-like Particles (“Nanoblades”). *J Vis Exp* <https://doi.org/10.3791/62245>.
- Meijer, H.A., and Thomas, A.A.M. (2002). Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *The Biochemical Journal* 367, 1–11. <https://doi.org/10.1042/BJ20011706>.
- Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005). RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* 127, 4223–4231. <https://doi.org/10.1021/ja043822v>.
- Miele, G., Mouland, A., Harrison, G.P., Cohen, E., and Lever, A.M. (1996). The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J Virol* 70, 944–951. <https://doi.org/10.1128/jvi.70.2.944-951.1996>.
- Mills, E.W., Wangen, J., Green, R., and Ingolia, N.T. (2016). Dynamic Regulation of a Ribosome Rescue Pathway in Erythroid Cells and Platelets. *Cell Reports* 17, 1–10. <https://doi.org/10.1016/j.celrep.2016.08.088>.
- Morris, D.R., and Geballe, A.P. (2000). Upstream open reading frames as regulators of mRNA translation. *Molecular and Cellular Biology* 20, 8635–8642. .
- Mouzakis, K.D., Lang, A.L., Vander Meulen, K.A., Easterday, P.D., and Butcher, S.E. (2013). HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res* 41, 1901–1913. <https://doi.org/10.1093/nar/gks1254>.
- Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J.C., and Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *441*, 4. .
- Nguyen Quang, N., Goudey, S., Ségéral, E., Mohammad, A., Lemoine, S., Blugeon, C., Versapuech, M., Paillart, J.-C., Berlioz-Torrent, C., Emiliani, S., et al. (2020). Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection. *Retrovirology* 17, 25. <https://doi.org/10.1186/s12977-020-00533-1>.
- Nieto-Garai, J.A., Arboleya, A., Otaegi, S., Chojnacki, J., Casas, J., Fabriàs, G., Contreras, F.-X., Kräusslich, H.-G., and Lorizate, M. (2021). Cholesterol in the Viral Membrane is a Molecular Switch Governing HIV-1 Env Clustering. *Adv Sci (Weinh)* 8, 2003468. <https://doi.org/10.1002/advs.202003468>.

- Ocwieja, K.E., Sherrill-Mix, S., Mukherjee, R., Custers-Allen, R., David, P., Brown, M., Wang, S., Link, D.R., Olson, J., Travers, K., et al. (2012). Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res* *40*, 10345–10355. <https://doi.org/10.1093/nar/gks753>.
- Ohlmann, T., Prévôt, D., Décimo, D., Roux, F., Garin, J., Morley, S.J., and Darlix, J.-L. (2002). In Vitro Cleavage of eIF4GI but not eIF4GII by HIV-1 Protease and its Effects on Translation in the Rabbit Reticulocyte Lysate System. *Journal of Molecular Biology* *318*, 9–20. [https://doi.org/10.1016/S0022-2836\(02\)00070-0](https://doi.org/10.1016/S0022-2836(02)00070-0).
- Ono, A., Waheed, A.A., and Freed, E.O. (2007). Depletion of cellular cholesterol inhibits membrane binding and higher-order multimerization of human immunodeficiency virus type 1 Gag. *Virology* *360*, 27–35. <https://doi.org/10.1016/j.virol.2006.10.011>.
- Ouspenskaia, T., Law, T., Clauser, K.R., Klaeger, S., Sarkizova, S., Aguet, F., Li, B., Christian, E., Knisbacher, B.A., Le, P.M., et al. (2021). Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat Biotechnol* 1–9. <https://doi.org/10.1038/s41587-021-01021-3>.
- Pierre, P. (2009). Immunity and the regulation of protein synthesis: surprising connections. *Current Opinion in Immunology* *8*. .
- Plank, T.-D.M., Whitehurst, J.T., and Kieft, J.S. (2013). Cell type specificity and structural determinants of IRES activity from the 5' leaders of different HIV-1 transcripts. *Nucleic Acids Res* *41*, 6698–6714. <https://doi.org/10.1093/nar/gkt358>.
- Prévôt, D., Décimo, D., Herbreteau, C.H., Roux, F., Garin, J., Darlix, J.-L., and Ohlmann, T. (2003). Characterization of a novel RNA-binding region of eIF4GI critical for ribosomal scanning. *The EMBO Journal* *22*, 1909–1921. <https://doi.org/10.1093/emboj/cdg175>.
- Puray-Chavez, M., Lee, N., Tanneti, K., Wang, Y., Vuong, H.R., Liu, Y., Horani, A., Huang, T., Gunsten, S.P., Case, J.B., et al. (2020). The translational landscape of SARS-CoV-2 and infected cells (Microbiology).
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* *44*, W160–W165. <https://doi.org/10.1093/nar/gkw257>.
- Ramos, H., Monette, A., Niu, M., Barrera, A., López-Ulloa, B., Fuentes, Y., Guizar, P., Pino, K., DesGroseillers, L., Mouland, A.J., et al. (2021). The double-stranded RNA-binding protein, Stauf1, is an IRES-transacting factor regulating HIV-1 cap-independent translation initiation. *Nucleic Acids Research* gkab1188. <https://doi.org/10.1093/nar/gkab1188>.
- Rao, S., Lungu, C., Crespo, R., Steijaert, T.H., Gorska, A., Palstra, R.-J., Prins, H.A.B., van Ijcken, W., Mueller, Y.M., van Kampen, J.J.A., et al. (2021). Selective cell death in HIV-1-infected cells by DDX3 inhibitors leads to depletion of the inducible reservoir. *Nat Commun* *12*, 2475. <https://doi.org/10.1038/s41467-021-22608-z>.
- Reuter, J.S., and Mathews, D.H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* *11*, 129. <https://doi.org/10.1186/1471-2105-11-129>.
- Ricci, E.P., Soto Rifo, R., Herbreteau, C.H., Decimo, D., and Ohlmann, T. (2008a). Lentiviral RNAs can use different mechanisms for translation initiation. *Biochem Soc Trans* *36*, 690–693. <https://doi.org/10.1042/BST0360690>.

- Ricci, E.P., Herbreteau, C.H., Decimo, D., Schaupp, A., Datta, S.A.K., Rein, A., Darlix, J.-L., and Ohlmann, T. (2008b). In vitro expression of the HIV-2 genomic RNA is controlled by three distinct internal ribosome entry segments that are regulated by the HIV protease and the Gag polyprotein. *RNA* 14, 1443–1455. <https://doi.org/10.1261/rna.813608>.
- Ricci, E.P., Kucukural, A., Cenik, C., Mercier, B.C., Singh, G., Heyer, E.E., Ashar-Patel, A., Peng, L., and Moore, M.J. (2014). Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol* 21, 26–35. <https://doi.org/10.1038/nsmb.2739>.
- Ringear, M., Marchand, V., Decroly, E., Motorin, Y., and Bennasser, Y. (2019). FTSJ3 is an RNA 2'-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature* 565, 500–504. <https://doi.org/10.1038/s41586-018-0841-4>.
- Rocchi, C., Louvat, C., Miele, A., Batisse, J., Guillon, C., Ballut, L., Lener, D., Negroni, M., Ruff, M., Gouet, P., et al. (2021). The HIV-1 Integrase C-Terminal domain induces TAR RNA structural changes promoting Tat binding (Biochemistry).
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 11, 783–784. <https://doi.org/10.1038/nmeth.3047>.
- Schubert, U., Norbury, C.C., Yewdell, J.W., and Bennink, J.R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *404*, 5. .
- Schuhmacher, J., Heidt, S., Balchen, T., Richardson, J.R., Schmeltz, C., Sonne, J., Schweiker, J., Rammensee, H.-G., Thor Straten, P., Røder, M.A., et al. (2020). Vaccination against RhoC induces long-lasting immune responses in patients with prostate cancer: results from a phase I/II clinical trial. *J Immunother Cancer* 8, e001157. <https://doi.org/10.1136/jitc-2020-001157>.
- Sharma, A., Yilmaz, A., Marsh, K., Cochrane, A., and Boris-Lawrie, K. (2012). Thriving under Stress: Selective Translation of HIV-1 Structural Protein mRNA during Vpr-Mediated Impairment of eIF4E Translation Activity. *PLoS Pathog* 8, e1002612. <https://doi.org/10.1371/journal.ppat.1002612>.
- Singh, G., Seufzer, B., Song, Z., Zucko, D., Heng, X., and Boris-Lawrie, K. (2022). HIV-1 hypermethylated guanosine cap licenses specialized translation unaffected by mTOR. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2105153118. <https://doi.org/10.1073/pnas.2105153118>.
- Soto-Rifo, R., Rubilar, P.S., Limousin, T., de Breyne, S., Décimo, D., and Ohlmann, T. (2012). DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs: Translation initiation mediated by DDX3. *The EMBO Journal* 31, 3745–3756. <https://doi.org/10.1038/emboj.2012.220>.
- Soto-Rifo, R., Rubilar, P.S., and Ohlmann, T. (2013). The DEAD-box helicase DDX3 substitutes for the cap-binding protein eIF4E to promote compartmentalized translation initiation of the HIV-1 genomic RNA. *Nucleic Acids Research* 41, 6286–6299. <https://doi.org/10.1093/nar/gkt306>.
- Vallejos, M., Carvajal, F., Pino, K., Navarrete, C., Ferres, M., Huidobro-Toro, J.P., Sargueil, B., and López-Lastra, M. (2012). Functional and Structural Analysis of the Internal Ribosome Entry Site Present in the mRNA of Natural Variants of the HIV-1. *PLoS ONE* 7, e35031. <https://doi.org/10.1371/journal.pone.0035031>.
- Ventoso, I., Blanco, R., Perales, C., and Carrasco, L. (2001). HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation. *PNAS* 98, 12966–12971. <https://doi.org/10.1073/pnas.231343498>.

Wei, J., Kishton J, R., Angel, M., Conn S, C., Dalla-Venezia, N., Marcel, V., Vincent, A., Catez, F., Ferré, S., Ayadi, L., et al. (2019). Ribosomal Proteins Regulate MHC Class I Peptide Generation for Immunosurveillance. *Mol Cell* 73, 1162–1173. <https://doi.org/doi:10.1016/j.molcel.2018.12.020>.

Weill, L., James, L., Ulryck, N., Chamond, N., Herbreteau, C.H., Ohlmann, T., and Sargueil, B. (2010). A new type of IRES within gag coding region recruits three initiation complexes on HIV-2 genomic RNA. *Nucleic Acids Research* 38, 1367–1381. <https://doi.org/10.1093/nar/gkp1109>.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* 1, 1610–1616. <https://doi.org/10.1038/nprot.2006.249>.

Winans, S., and Goff, S.P. (2020). Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. *PLOS Pathogens* 16, e1009147. <https://doi.org/10.1371/journal.ppat.1009147>.

Xiao, Z., Huang, R., Xing, X., Chen, Y., Deng, H., and Yang, X. (2018). De novo annotation and characterization of the translome with ribosome profiling data. *Nucleic Acids Res* 46, e61. <https://doi.org/10.1093/nar/gky179>.

Yedavalli, V.S.R.K., Neuveut, C., Chi, Y.-H., Kleiman, L., and Jeang, K.-T. (2004). Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* 119, 381–392. <https://doi.org/10.1016/j.cell.2004.09.029>.

Yewdell, J.W. (2020). DRiPs get molecular. *Current Opinion in Immunology* 7. .

Zhang, H., Wang, Y., and Lu, J. (2019). Function and Evolution of Upstream ORFs in Eukaryotes. *Trends in Biochemical Sciences* <https://doi.org/10.1016/j.tibs.2019.03.002>.

Zhang, P., He, D., Xu, Y., Hou, J., Pan, B.-F., Wang, Y., Liu, T., Davis, C.M., Ehli, E.A., Tan, L., et al. (2017). Genome-wide identification and differential analysis of translational initiation. *Nat Commun* 8, 1749. <https://doi.org/10.1038/s41467-017-01981-8>.

Figure legends

Figure 1. Transcriptional and translational changes in HIV-1 infected cells. **A.** Schematic representation of the procedure to monitor transcript abundance and translation in HIV-1 infected cells. Briefly, SupT1 cells were infected or not (Mock) with HIV-1 (NL4.3 strain) at MOI 5. At 0, 1, 12, 24 and 36 hours post infection (hpi), cells were lysed to recover the cytoplasmic fraction and prepare ribosome profiling and RNA-seq libraries subjected to high-throughput sequencing on the Illumina HiSeq platform. **B.** Scatter-plot of the fold-change (log₂) in cytoplasmic RNA-seq and Ribo-Seq of the Mock-infected and HIV-1 infected cells at each time point of infection. Orange dots (“Transcript-level only”) corresponds to genes exclusively regulated at the transcript abundance level. Blue dots (“Translation only”) correspond to genes which display differences in ribosome

occupancy while transcript abundance remains unchanged. Green dots (“Translationally intensified”) correspond to genes with significant changes in transcript abundance and significantly further changes, in the same direction, in ribosome occupancy upon infection. Red dots (“Translationally buffered”) correspond to transcripts displaying significant changes in transcript abundance but for which there is compensation at the translational level to maintain unchanged ribosome occupancy levels upon infection. **C.** Gene ontology analysis of differentially expressed genes at each time point. **D.** (Left panel) Scheme describing nascent protein labeling using O-propargyl-puromycin (OPP). Briefly, cells are incubated with OPP and fixed in paraformaldehyde before a fluorophore is covalently linked through a click-reaction. Cells are then analyzed by flow cytometry to monitor signal intensity at a single-cell level. **E.** Flow cytometry analysis of OPP signal in SupT1 cells infected with HIV-1, 12, 24 and 36 hpi. **F.** OPP signal in SupT1 cells infected with single-round recombinant HIV-1 virions coding for GFP at 24, 48, 72 and 96 hpi. **G.** OPP signal in primary Human CD4⁺ T cells infected with HIV-1, 48 and 72 hpi and either positive or negative for p24 expression (p24⁺ or p24⁻). Results from two independent donors are displayed separately for each time point tested.

Figure 2. Relative and absolute expression of viral transcripts during infection. **A.** (Left panel) Relative cytoplasmic amount of viral transcripts (the sum of all transcripts at any given time point corresponding to 100%) in SupT1 cells, 1, 12, 24 and 36 hpi. (Right panel) Overall abundance of viral transcripts within RNA-seq libraries displayed in transcripts per million (TPM). **B.** MA-plot (log₂ fold change Vs mean average) of cytoplasmic RNA-seq reads from U937 cells with a latent HIV-1 integrated proviral DNA, 24 hours after induction of proviral DNA expression. Blue dots correspond to differentially expressed genes with an adjusted p-value < 0.05. **C.** Same as **B.** but using ribosome density values. **D.** Relative (left panel) and overall (right panel) abundance of viral transcripts in U937 cells, 3, 6, 9, 12, 15, 18 and 24 hours after induction of proviral DNA expression.

Figure 3. Translational landscape of viral transcripts. **A.** HIV-1 genomic structure. **B.** Distribution of RNA-seq (red) and ribosome profiling (blue) reads across the HIV-1 genome in SupT1 cells at 1, 12, 24 and 36 hpi. **C.** Translation efficiency of the genomic RNA (across both Gag and Pol coding sequences) at each time point of infection. **D.** Translation efficiency of canonical viral mRNAs, 12, 24 and 36 hpi. **E.** Percentage of Gag-Pol ribosome frameshifting at each time point of infection.

Figure 4. Translation initiation sites in viral transcripts. **A.** Distribution of ribosome P-sites around annotated start and stop codons in all cellular transcripts in harringtonine and cycloheximide libraries. **B.** Distribution of ribosome profiling reads across the HIV-1 genome obtained from harringtonine and cycloheximide treated cells. **C.** Distribution of ribosome profiling reads in the first 500 nucleotides of the Gag CDS obtained from harringtonine and cycloheximide treated cells. The canonical AUG start codon of Gag (p55 isoform) at position +336, as well as the position of other out-of-frame start codons predicted by Ribocode and lastly the position of the Gag (p40 isoform) start codon at position +759 are annotated in the figure.

Figure 5. Non-AUG translation initiation sites in the 5'UTR of viral transcripts. **A.** (Top panel) Distribution of ribosome profiling reads across the HIV-1 genome from infected U937, SupT1 and primary CD4⁺ T lymphocytes. (Bottom panel) Close-up view of the unspliced HIV-1 5'UTR showing the distribution of ribosome profiling reads and the position of putative non-AUG start-codons. **B.** Distribution of FLOSS (Fragment Length Distribution Score) values for cellular and viral transcripts computed from ribosome profiling reads from SupT1 (left panel), U937 (middle panel) or primary CD4 T cells (right panel).

Figure 6. Productive translation from uORFs in viral transcripts and role of DDX3 in alleviating the negative impact of uORFs on translation from downstream main ORFs **A.** Distribution of ribosome profiling reads across the 5'UTR of unspliced viral mRNAs in U937 cytoplasmic lysates incubated with 1M KCl (high salt wash), or from Ribolace experiments in SupT1 cells (Actively translating ribosomes), or from ribosome immunoprecipitation in transfected HEK293T cells (HEK RPL7a IP). **B.** Distribution of ribosome profiling reads on luciferase coding reporter transcripts bearing the wild-type HIV-1 5'UTR (WT 5'UTR), a mutant version in which all non-AUG uORF start codons were mutated to UAA stop codons (no uORFs 5'UTR), a mutant in which all non-AUG uORF start codons were mutated into AUG codons (AUG uORFs 5'UTR). **C.** Relative renilla luciferase activity (normalized against the Globin 5'UTR reporter mRNA) of the HIV-1 “wild-type”, “No uORFs” 5'UTR and “AUG uORFs” 5'UTR reporter mRNAs upon *in vitro* translation in the rabbit reticulocyte lysate. **D.** Western-blot analysis of β -actin (left panel) and DDX3 (right panel) expression in HEK293T cells transfected with plasmids coding for the Cas9 and a sgRNA targeting the YFP sequence (sgYFP) or the DDX3 coding sequence (sgDDX3). **E.** (Left panel) Relative renilla luciferase activity (normalized against the Globin 5'UTR reporter mRNA) from reporter mRNAs transfected in HEK293T cells in which DDX3 was knocked-down (sgDDX3) or not (sgYFP) using CRISPR-Cas9. (right panel) Fold-change of luciferase activity for each reporter mRNA between the control and DDX3 depleted condition computed from the data

presented in the left panel. **F.** Relative renilla luciferase activity (normalized against the Globin 5'UTR reporter mRNA) from viral reporter mRNAs in which the TAR loop is present or absent, transfected into HEK293T cells in which DDX3 expression was knocked-down (sgDDX3) or not (sgYFP) using CRISPR-Cas9. For panels C, E and F, a Student t-test was performed to assess the differences between the mean values of compared conditions (* corresponding to a p-value <0.05).

Figure 7. Identification of polypeptides that initiate specific T cell responses in treated and untreated HIV infected individuals. A to D. Based on the nucleotide sequence of NL4-3, peptides potentially encoded by uORFs (see peptides “A02; A03; A06; A07; A08” in panel **A** of this figure) of the 5'LTR were synthesized and used to screen for T cell responses in PBMCs of HIV-infected individuals. PBMCs from treated (ART) and untreated but aviremic (so called elite controllers, EC) individuals were stimulated with a pool of uORF-derived peptides (POOL) and cultured in the presence of IL-2 and IL-7 cytokines in order to expand peptide-specific memory T cells. On day 7 (not shown) and day 12, T cell responses against the POOL were assessed using IFN γ -Elispot. In addition, on day 12, except for donor EC-3, individual peptides of the pool (A02; A03; A06; A07; A08) were assessed using IFN γ -Elispot. As positive control for T cell expansion and activation, a pool of immunogenic peptides from HIV Env and Gag proteins was also used. **A.** Number of uORF-derived peptides recognized by each HIV-infected individuals. The color code indicates from which uORF the peptides are derived. **B, C and D.** left panels, detailed IFN γ -Elispot data from the 3 individuals presenting T cell responses, expressed as spot forming units (SFU) per million PBMCs; right panels pictures of the raw data from the Elispot plates in technical triplicate (**B and D**) or duplicate (**C**). POOL (-) and POOL (+): PBMCs expanded with the pool of uORF-derived peptides but restimulated on the day of the Elispot with medium or the POOL, respectively. A02; A03; A06; A07 and A08 name of individual uORF-derived peptides used for restimulation. HIV (-) and HIV (+): PBMC expanded with the pool of Env- and Gag-derived peptides but restimulated for the Elispot assay with medium or the same pool of HIV peptides, respectively. Responses were considered positive when IFN γ production was superior to 20 spots/10⁶ PBMCs and at least twofold higher than background production from cells restimulated with medium (dotted lines). SAT: saturated signal, where counts cannot be estimated due to overwhelm IFN γ secretion by activated T cells.

Supplementary Figure 1. A. Flow cytometry analysis of p24 expression in mock (blue) or HIV-1 infected (pink) SupT1 cells at 12, 24 or 36h pi. **B.** Distribution of ribosome P-sites in the -40 to +100 nucleotide region surrounding the annotated AUG start codons of all cellular transcripts obtained from ribosome profiling libraries prepared from MNase or RNaseI-treated SupT1

cytoplasmic cell lysates. **C.** Scatter-plot of protein abundance from SupT1 cells infected with obtained by mass spectrometry (*y*-axis) from (Golumbeanu et al., 2019) against Ribo-seq (left-panel, *x*-axis) or RNA-seq (right-panel, *x*-axis) per gene read counts obtained from SupT1 cells 24 hpi from our study. **D.** Random forest analysis of features associated with down-translated (top panel), up-translated (middle panel) transcripts. (Bottom panel) Violin plots of CDS GC content of down-translated (down), non-regulated (N.S.) and up-translated (up) genes at 36 hpi. **E.** Gene trajectory clustering across infection time points in SupT1 cells and its corresponding gene-ontology analysis from differentially expressed-transcripts.

Supplementary Figure 2. A. Density plot of flow cytometry signal for OPP incorporation in HEK293T cells infected with Sindbis virus (at MOI 5) from 0 to 18 hpi. **B.** Flow cytometry analysis of OPP signal in SupT1 cells mock infected or infected with HIV-1 (NL4.3 strain at MOI 5) at 12, 24 and 36 hpi (results presented correspond to gating on live cells).

Supplementary Figure 3. A. Scatter plot of a simulated HIV-1 RNA-seq dataset to test the Kallisto transcript-specific quantification pipeline based on annotated full-length transcripts obtained by Nanopore cDNA sequencing described in (Nguyen Quang et al., 2020). True HIV-1 transcript counts are presented in the *x*-axis, while Kallisto deconvoluted transcript counts are presented in the *y*-axis. **B.** Sashimi plot indicating all exon-exon junction mapping reads from our RNA-seq and Ribo-seq datasets at 1 and 24 hpi.

Supplementary Figure 4. A. Distribution of ribosome profiling reads across the HIV-1 genome obtained from Harringtonine-treated U937 cells in which HIV-1 transcription was induced for 3, 6, 9, 12, 15, 18 and 24 h. The position of each viral gene and its corresponding canonical AUG start codon are depicted.

Supplementary Figure 5. A. Secondary structure of the HIV-1 wild-type 5'UTR and **B.** No-uORFs 5'UTR mutant followed by the renilla luciferase coding region as obtained by SHAPE-analysis. **C.** Distribution of ribosome profiling reads across the 5'UTR of unspliced viral mRNAs in HIV-2 infected U937 cells incubated with cycloheximide or harringtonine before lysis. **D.** Distribution of ribosome profiling reads across the 5'UTR of unspliced viral mRNA of HTLV-1 in immortalized T cells obtained from HTLV-1 infected humanized mice.

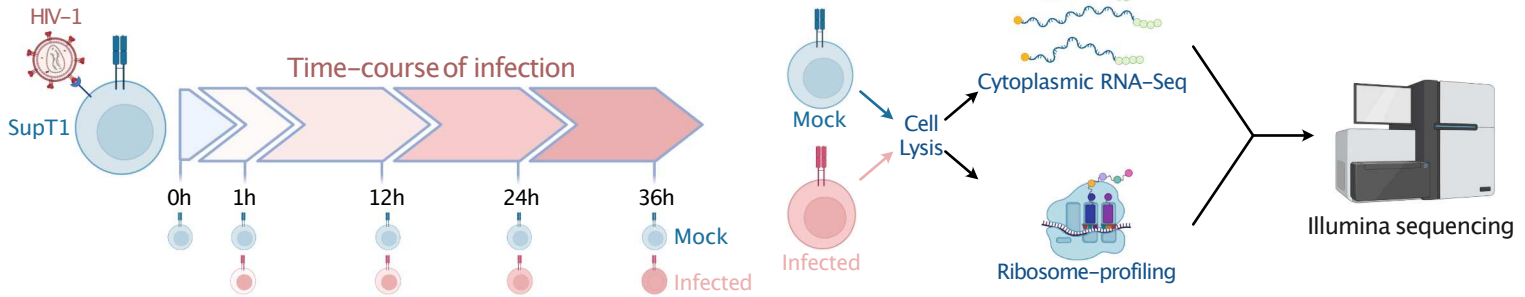
Supplementary Figure 6. A to D. Four HIV infected individuals with no 5'LTR uORF-specific T cell responses. IFN γ -Elispot data from the 4 individuals, from Figure 6 (**C**), presenting no T cell

responses against 5’LTR uORF-derived peptides, expressed as spot forming units (SFU) per million PBMCs. As in Figure 6C to 6F, memory T cells from ART and EC were stimulated/expanded with the pool of uORF derived peptides (POOL), the HIV pool (HIV) or an additional control pool (CEF) composed of peptides derived from CMV, EBV and Flu viruses. **(C, D, E)**, on day 12 T cell responses against the POOL, HIV and CEF pools as well as individual 5’LTR uORF-derived peptides (A02; A03; A06; A07; A08) were assessed using IFN γ -Elispot. **(F)**, T cell responses of donor EC-3 were assessed on day 7 with the indicated pools. Since no reactivity to the pool of 5’LTR uORF-derived peptides (POOL) was detected, T cell responses to the individual peptides were not assessed on day 12 for donor EC-3. POOL (-) and POOL (+): PBMCs expanded with the pool of uORF peptides but restimulated on the day of the Elispot with medium or the POOL, respectively. A02; A03; A06; A07 and A08 name of individual uORF-derived peptides used for restimulation. HIV (-) and HIV (+): PBMCs expanded with the pool of Env- and Gag-derived peptides but restimulated for the Elispot assay with medium or the same pool of HIV peptides, respectively. CEF (-) and CEF (+) PBMCs expanded with the pool of virus-derived peptides but restimulated for the Elispot assay with medium or the same CEF pool, respectively. Responses were considered positive when IFN γ production was superior to 20 spots/10⁶ PBMCs and at least twofold higher than background production from cells restimulated with medium (dotted lines). SAT: saturated signal, where counts cannot be estimated due to overwhelm IFN γ secretion by activated T cells.

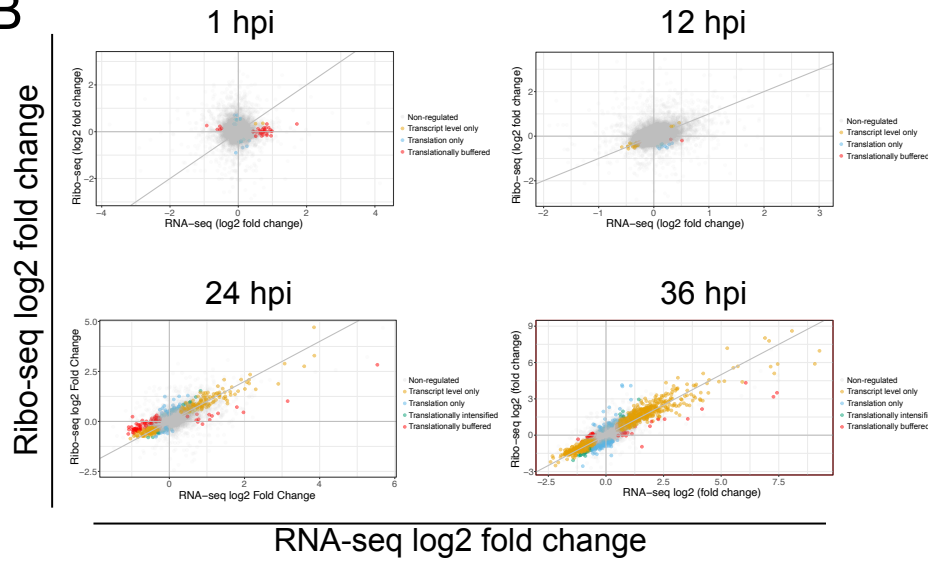
Supplementary table 1. Differential gene expression and translation analysis of SupT1 infected cells and U937 cells in which latent HIV-1 is induced.

Supplementary table 2. ORF identification from viral transcripts based on ribosome profiling reads as obtained with Ribocode.

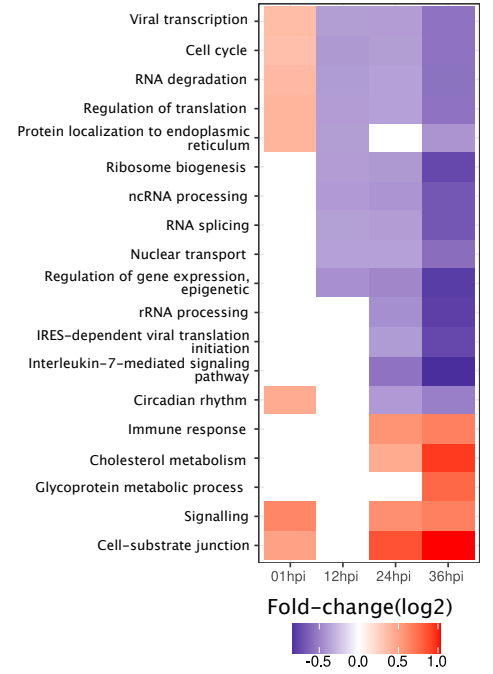
A



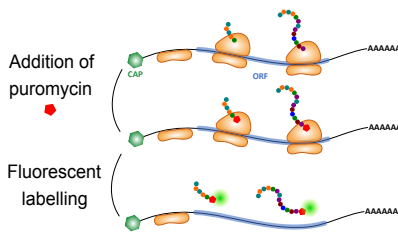
B



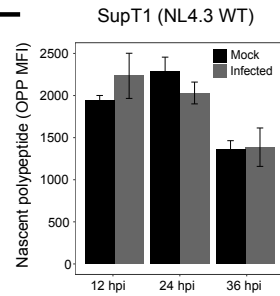
C



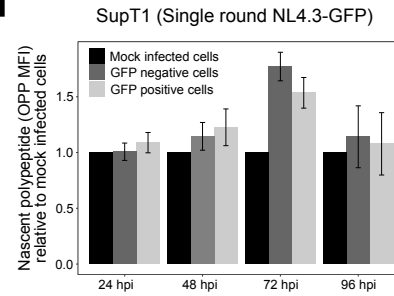
D



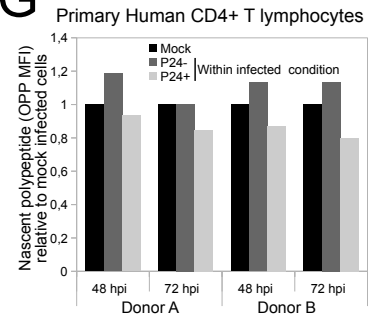
E

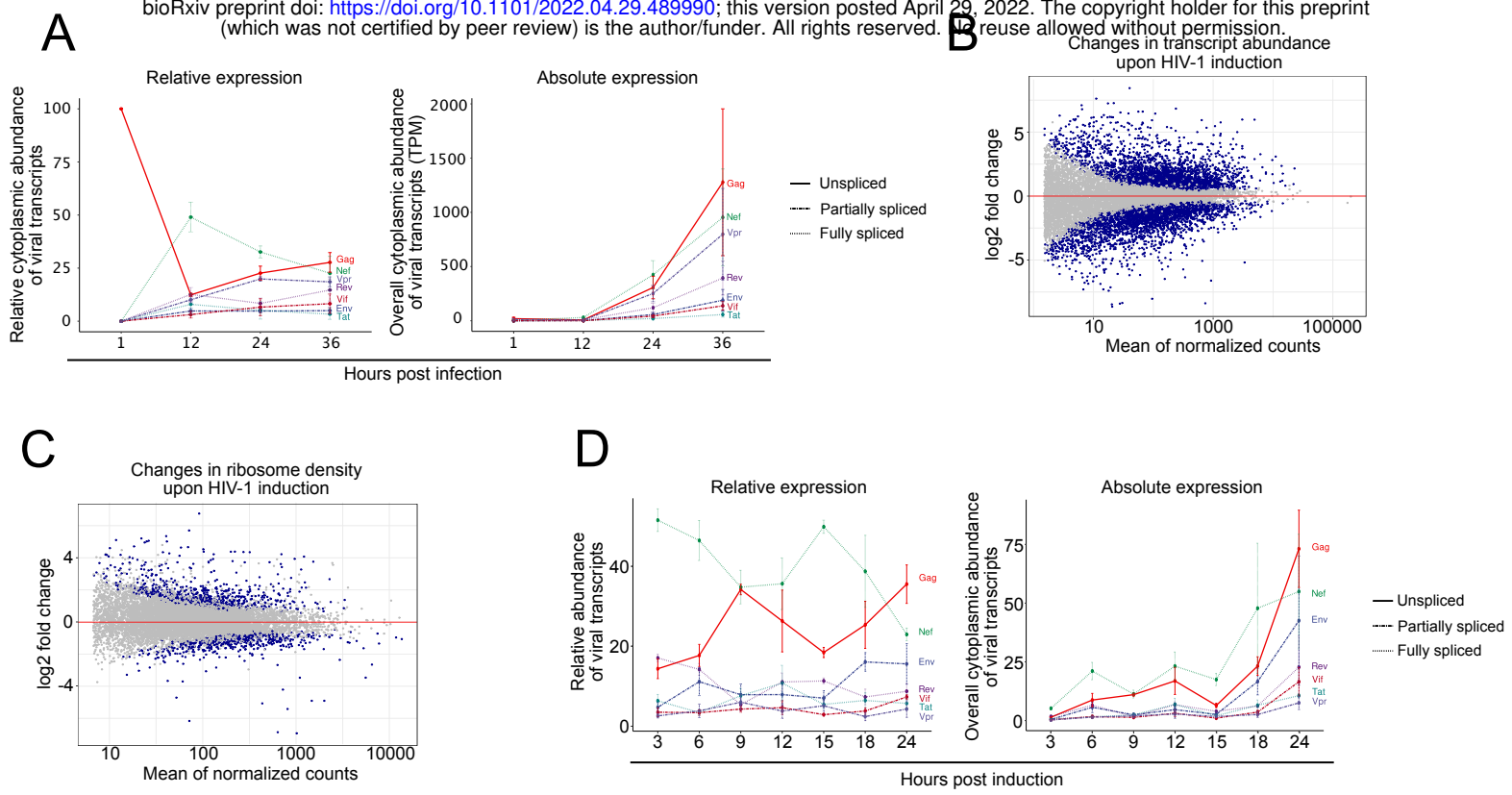


F

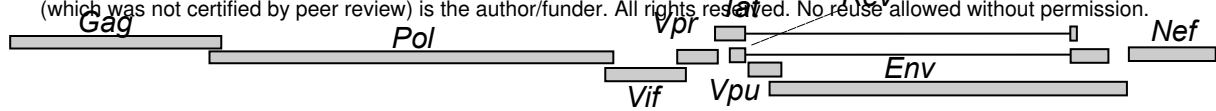


G

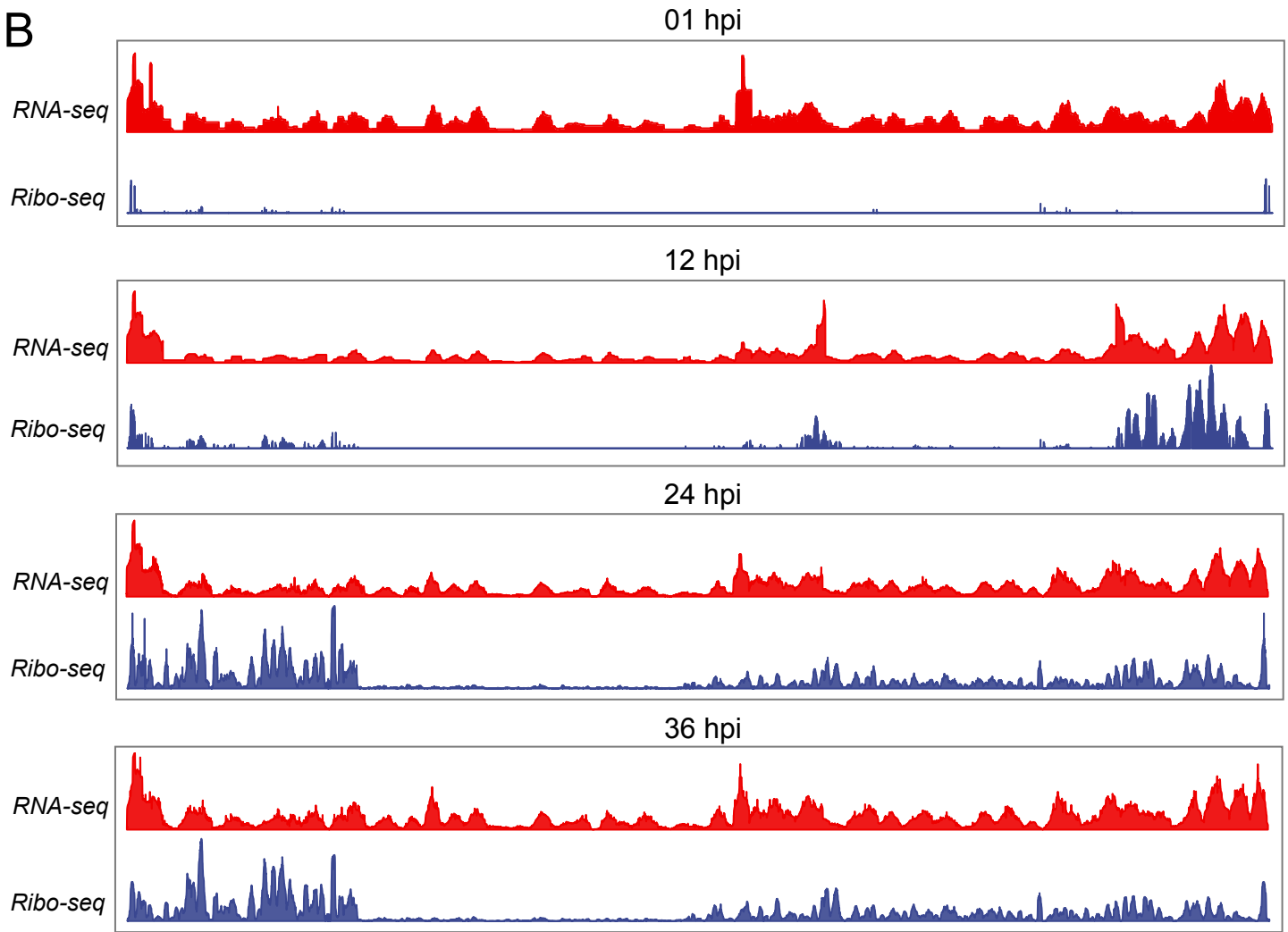




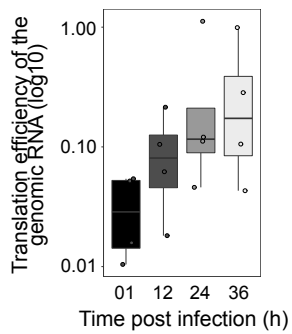
A



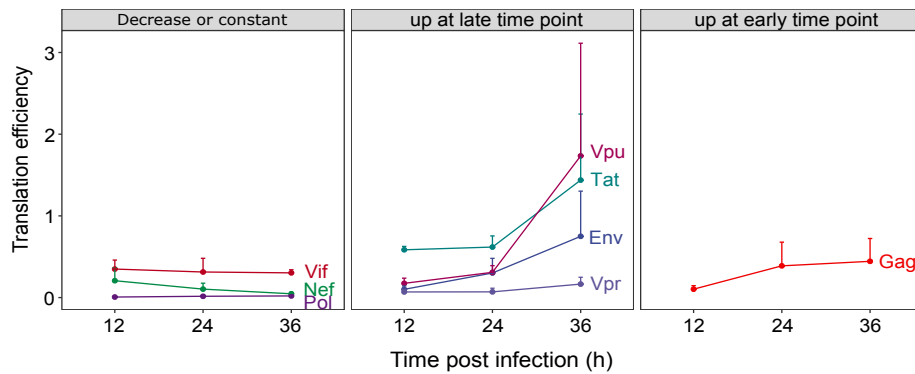
B



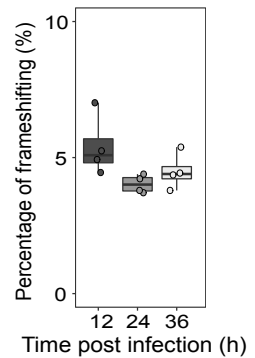
C



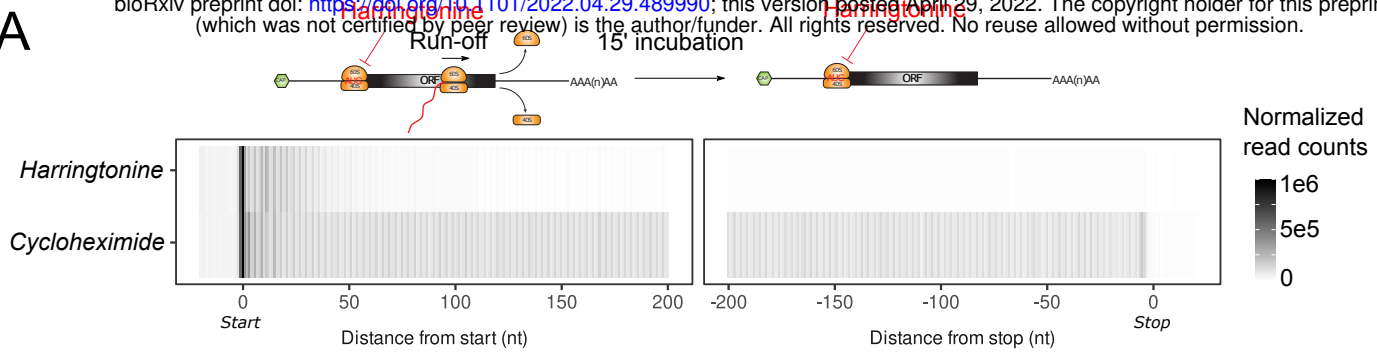
D



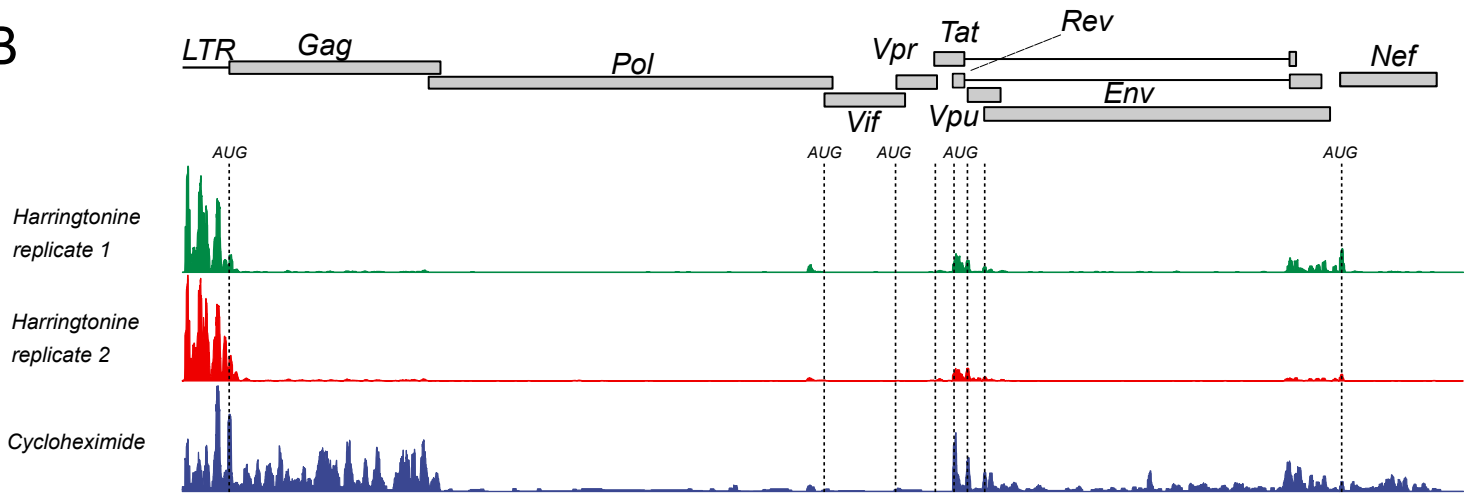
E



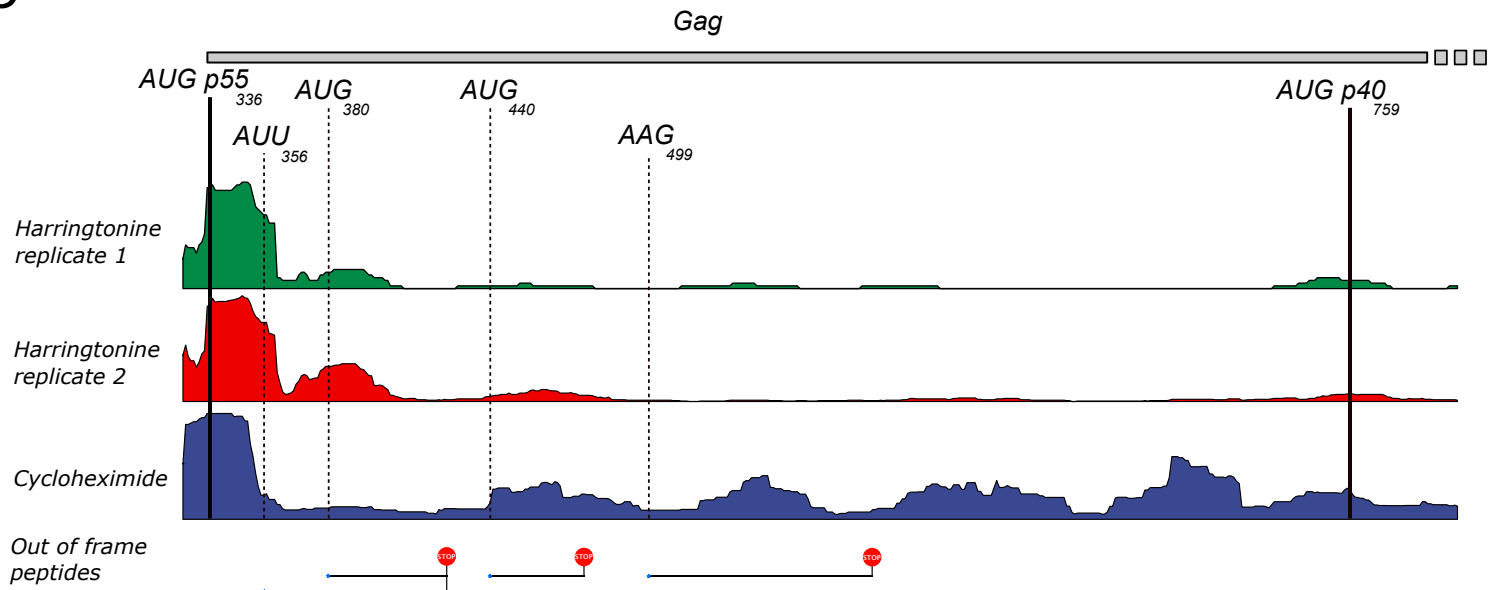
A



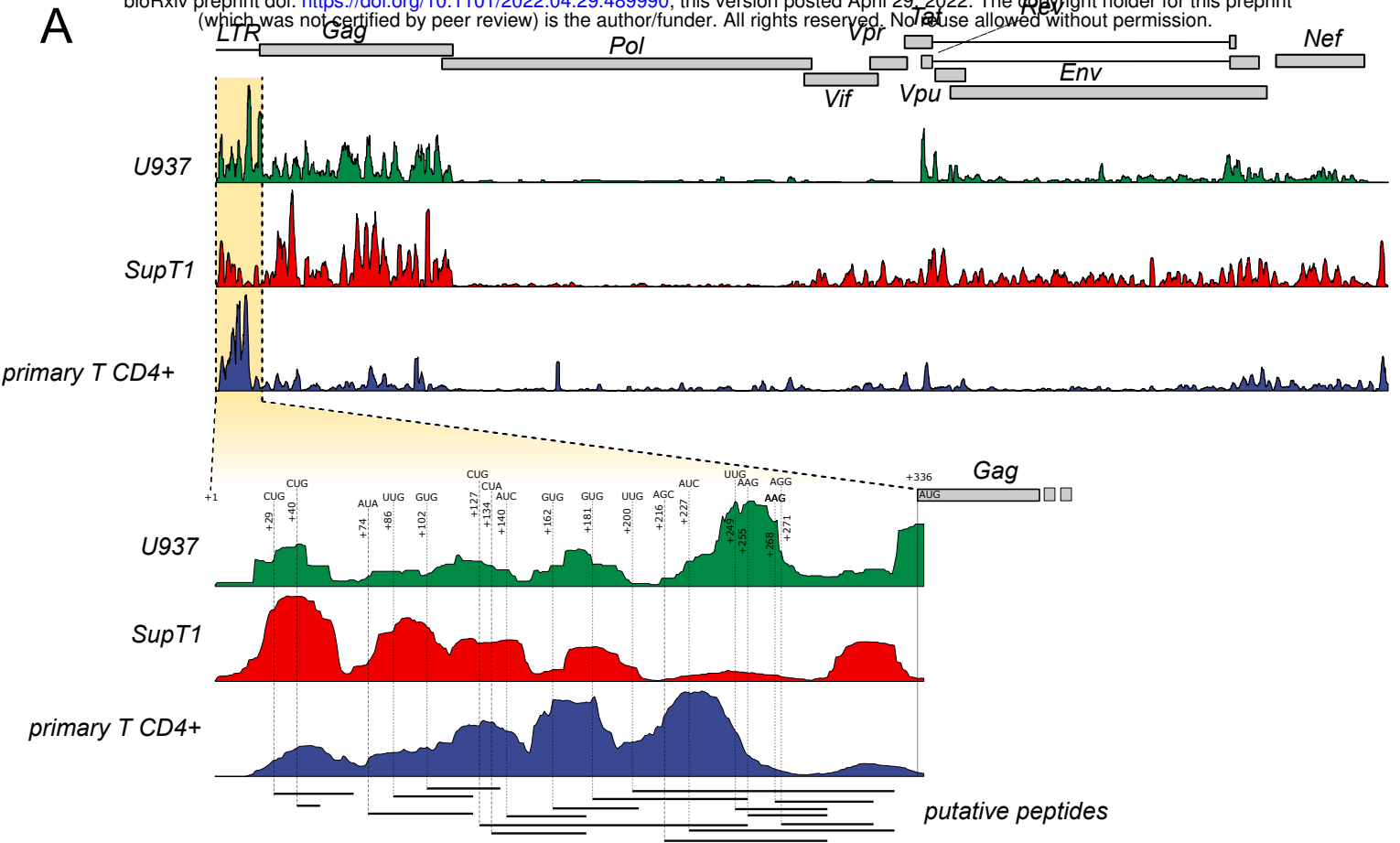
B



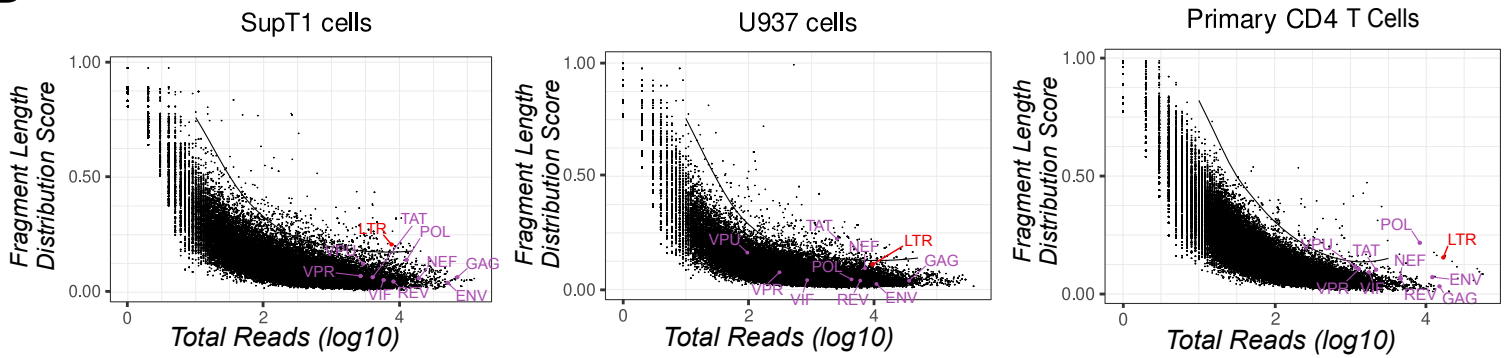
C

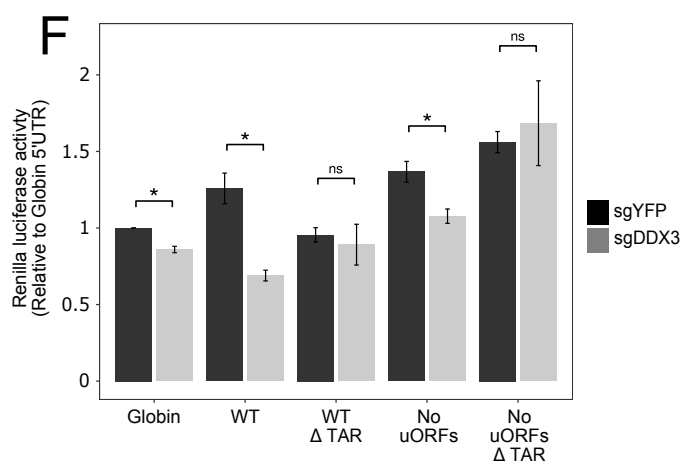
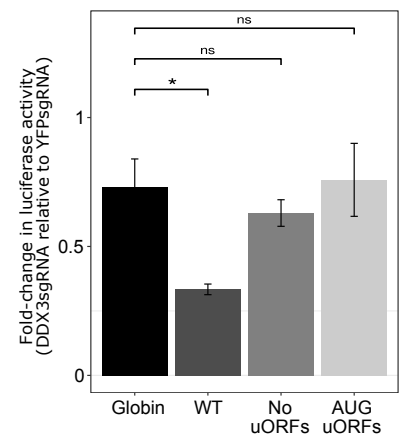
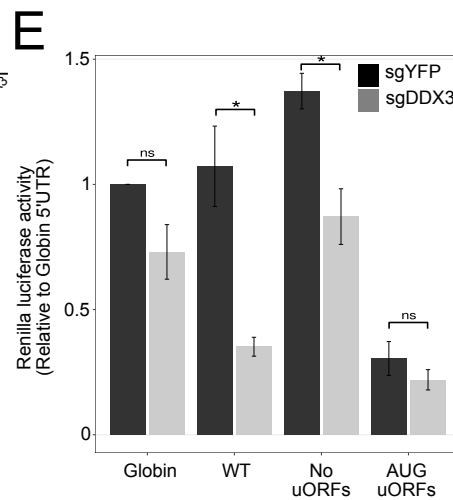
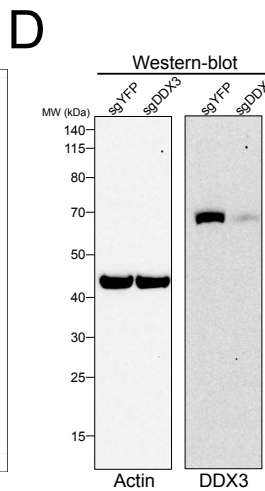
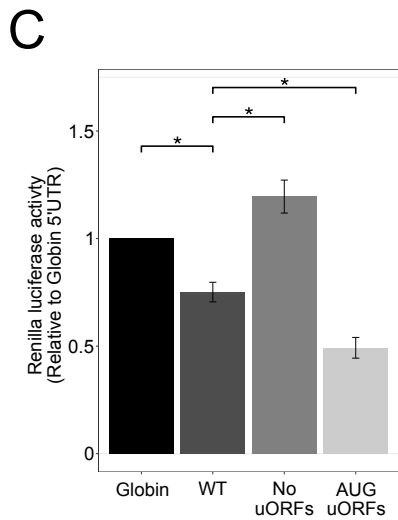
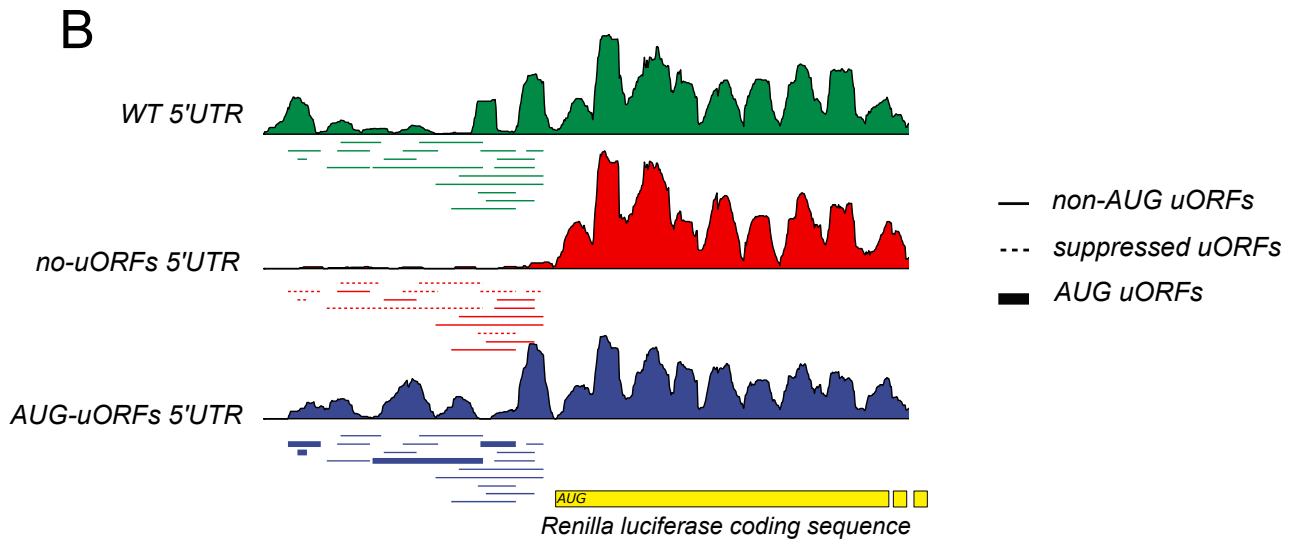
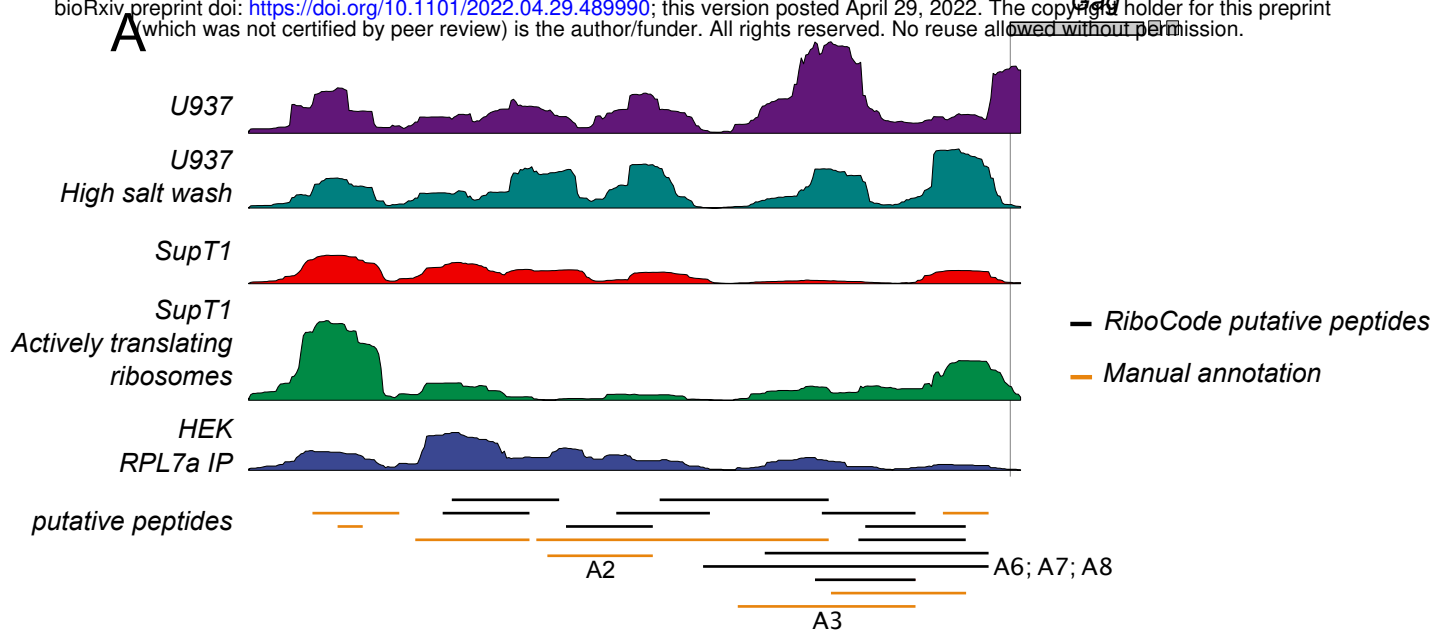


A

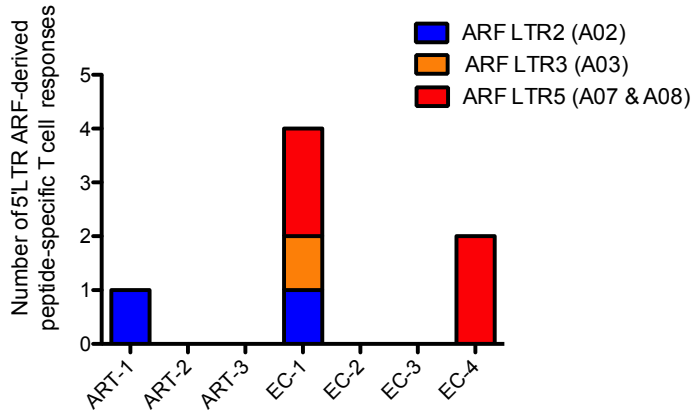


B

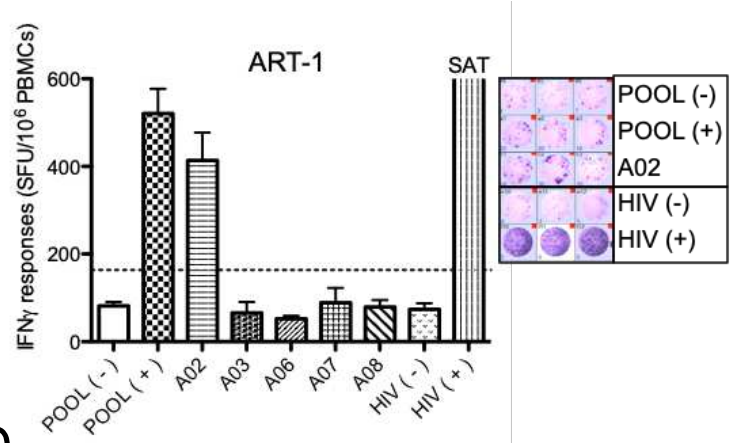




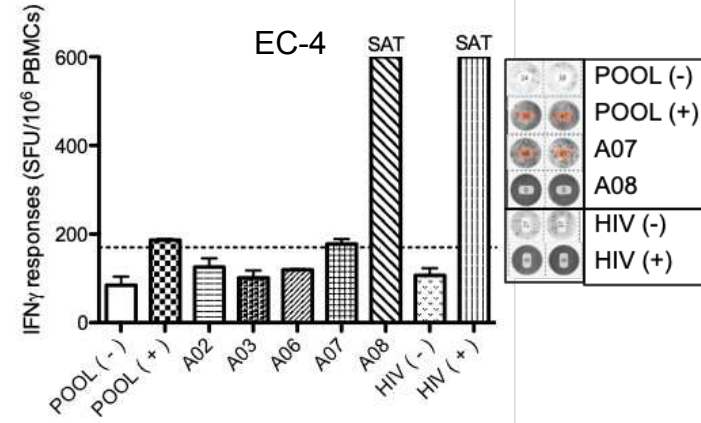
A



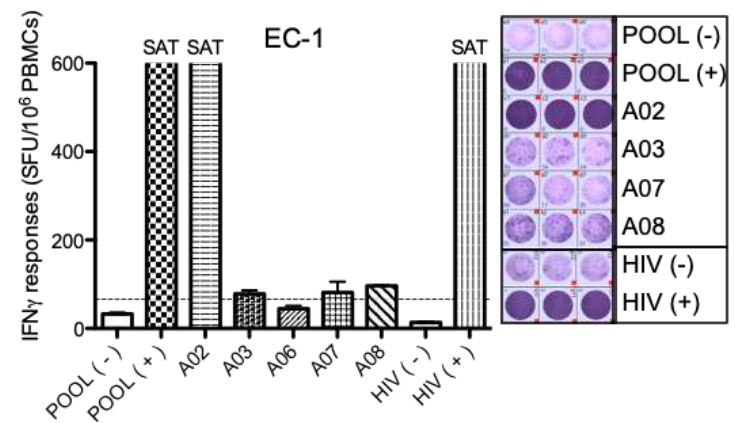
B



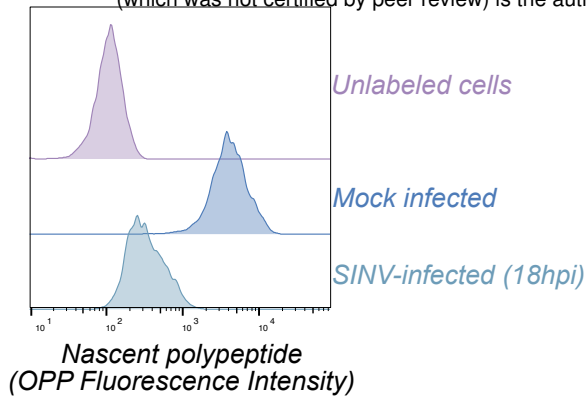
C



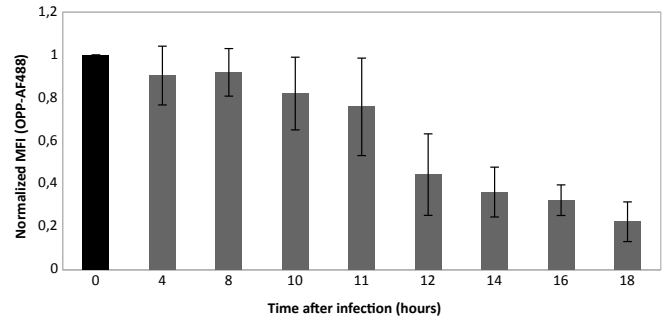
D



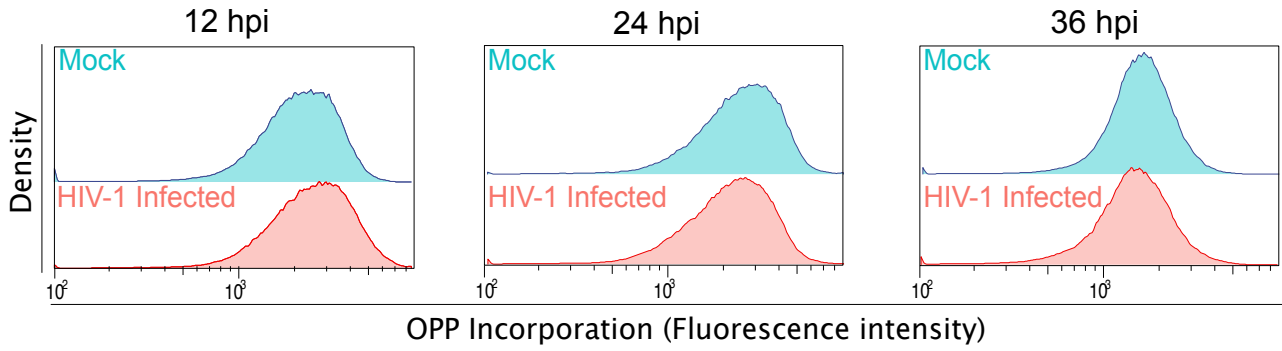
A



Translation efficiency upon Sindbis virus infection - HEK293T cells

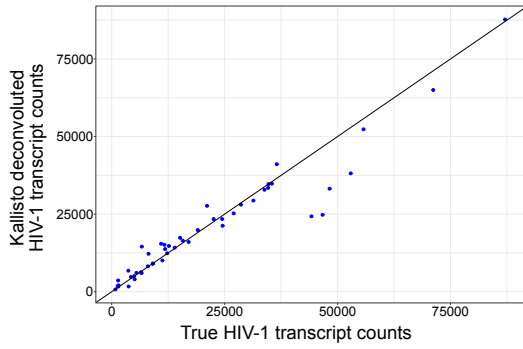


B



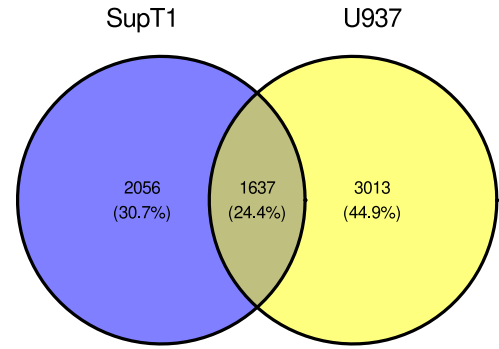
A

Deconvolution from simulated HIV-1 RNA-Seq dataset

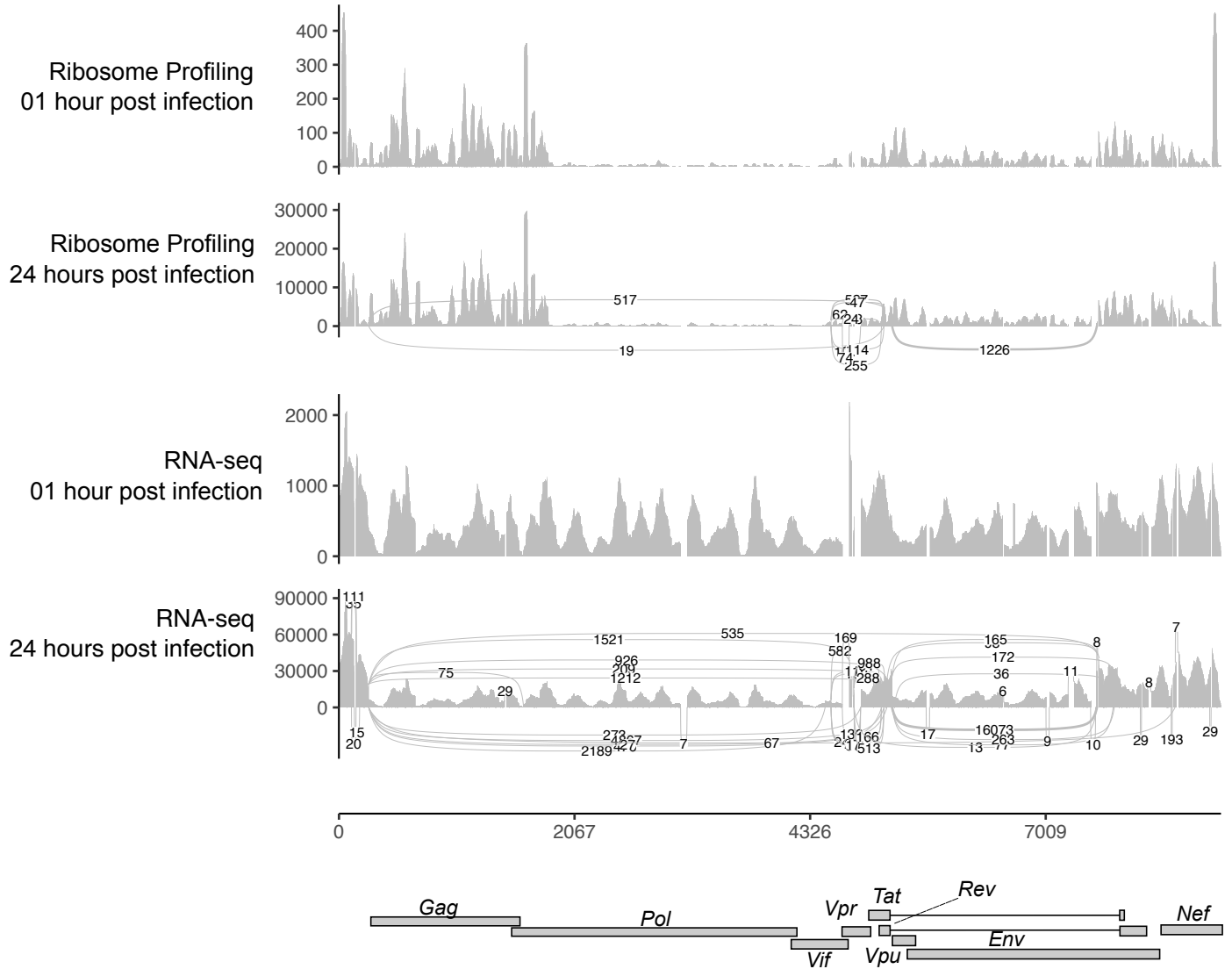


C

Differentially expressed genes from ribosome profiling data

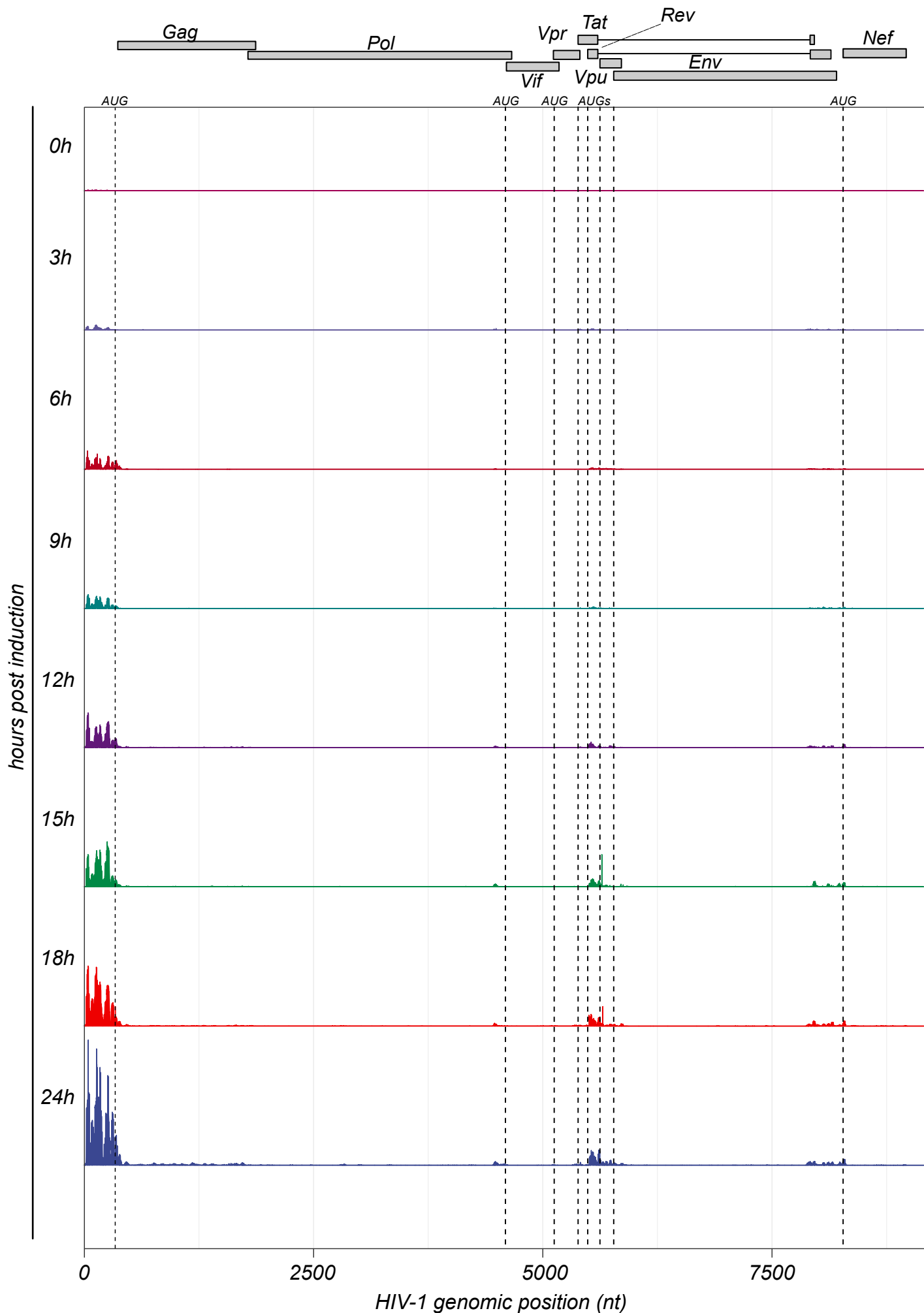


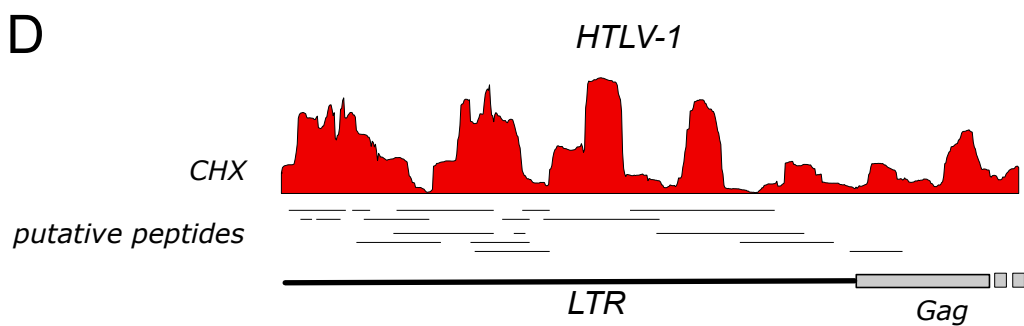
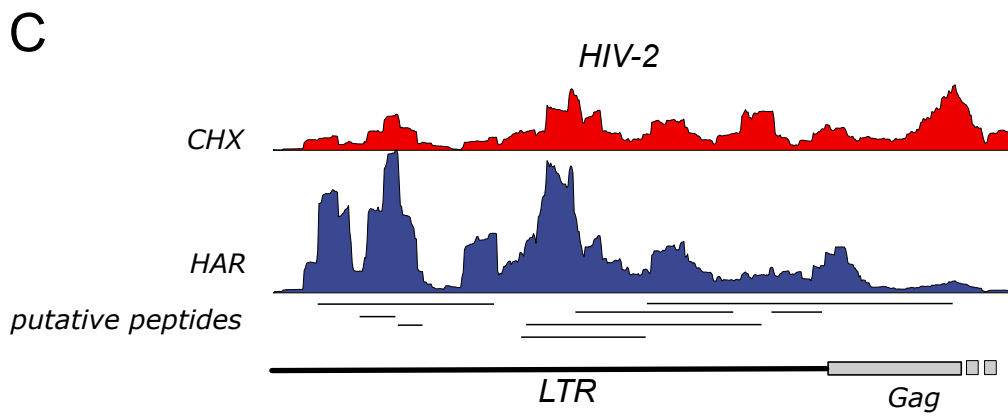
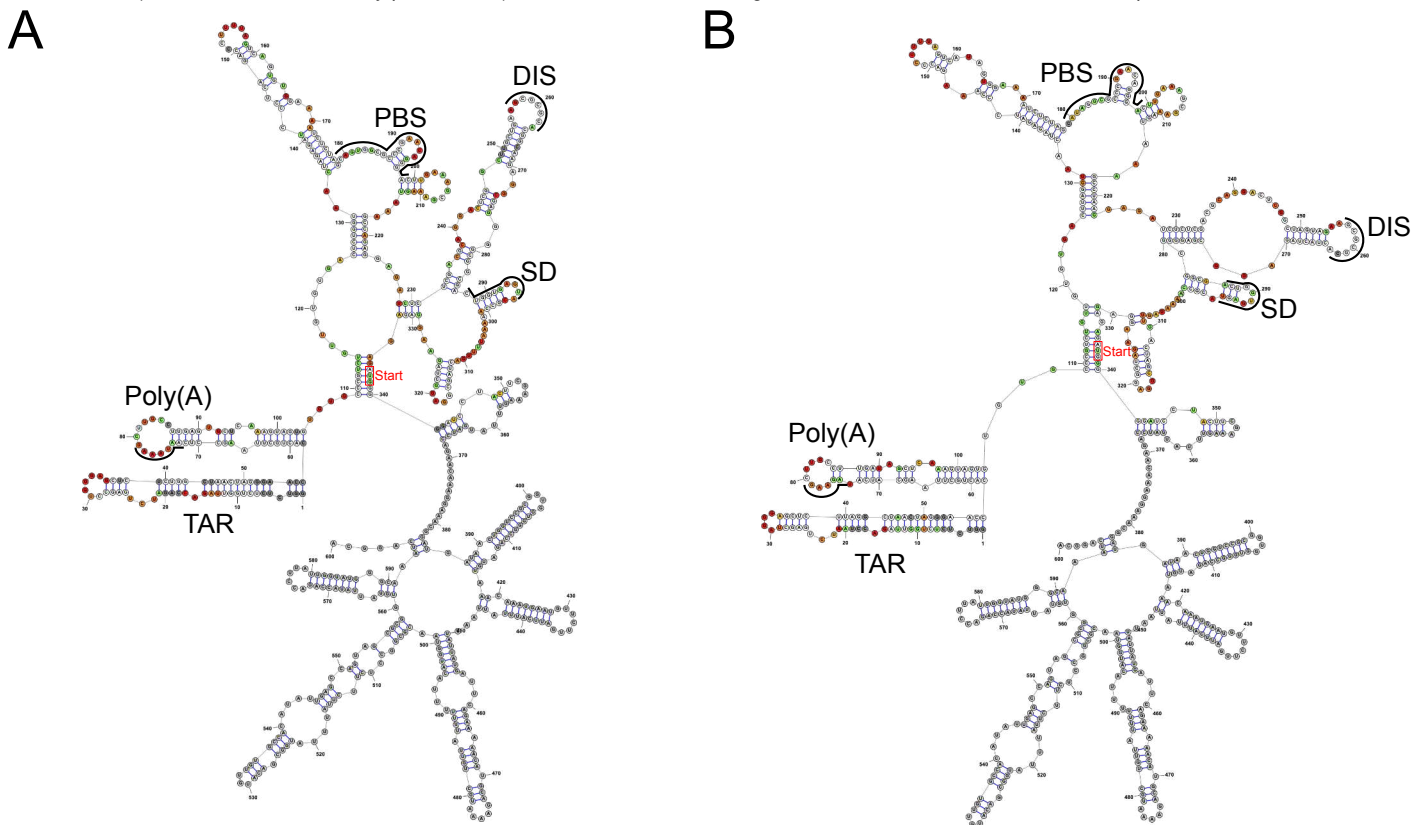
B

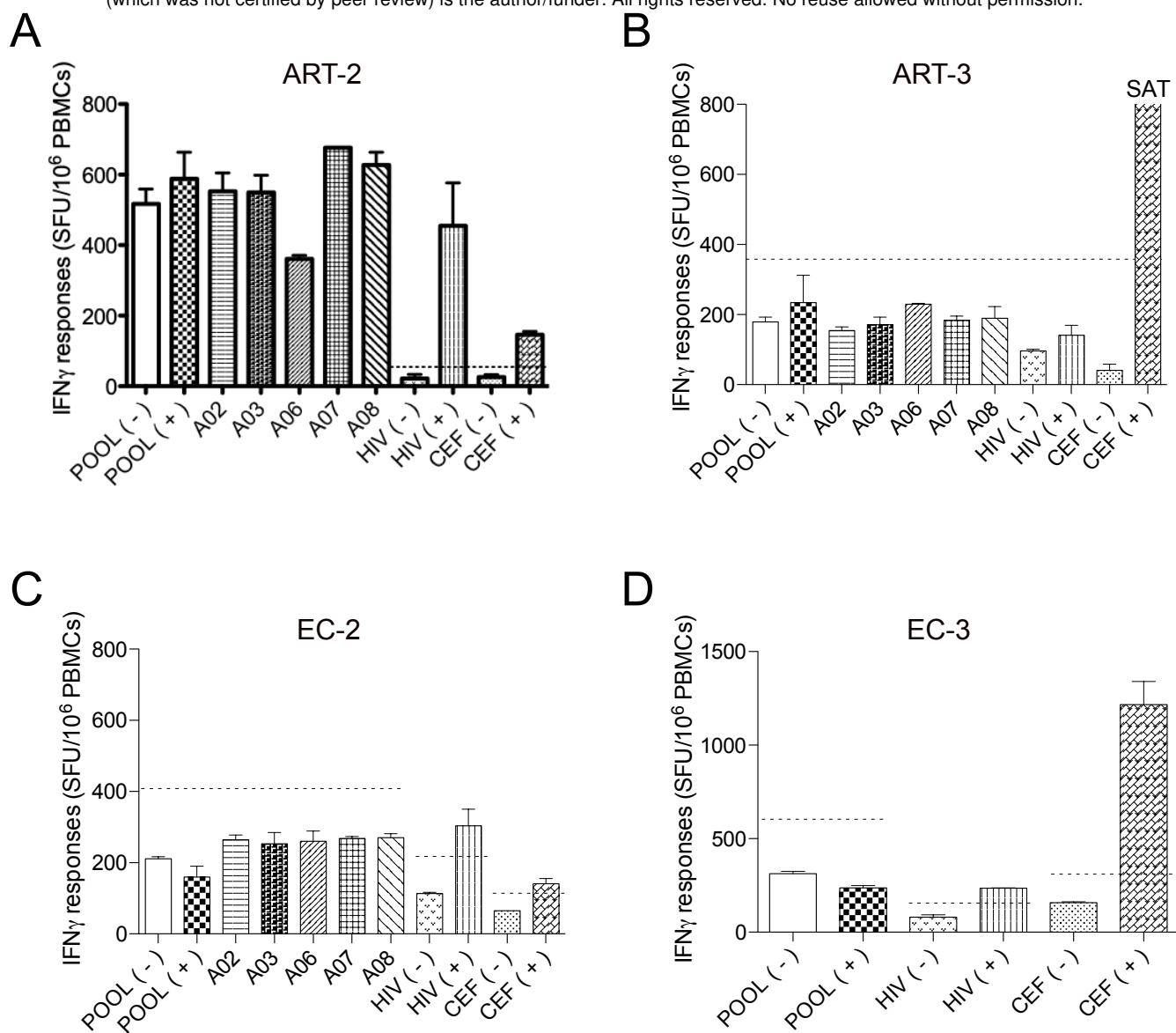


A

Ribosome profiling in presence of Harringtonine from U937 cells induced with PMA and iomycin







BIBLIOGRAPHIE

- Abi Habib, J. *et al.* (2022) 'Functional Differences between Proteasome Subtypes', *Cells*, 11(3), p. 421. Available at: <https://doi.org/10.3390/cells11030421>.
- Ackerman, M.E. *et al.* (2016) 'Polyfunctional HIV-Specific Antibody Responses Are Associated with Spontaneous HIV Control', *PLoS pathogens*, 12(1), p. e1005315. Available at: <https://doi.org/10.1371/journal.ppat.1005315>.
- Addo, M.M. *et al.* (2003) 'Comprehensive epitope analysis of human immunodeficiency virus type 1 (HIV-1)-specific T-cell responses directed against the entire expressed HIV-1 genome demonstrate broadly directed responses, but no correlation to viral load', *Journal of Virology*, 77(3), pp. 2081–2092. Available at: <https://doi.org/10.1128/jvi.77.3.2081-2092.2003>.
- Aderem, A. (2002) 'How to Eat Something Bigger than Your Head', *Cell*, 110(1), pp. 5–8. Available at: [https://doi.org/10.1016/S0092-8674\(02\)00819-X](https://doi.org/10.1016/S0092-8674(02)00819-X).
- Adiko, A.C. *et al.* (2015) 'Intracellular Transport Routes for MHC I and Their Relevance for Antigen Cross-Presentation', *Frontiers in Immunology*, 6. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2015.00335> (Accessed: 6 March 2023).
- Admon, A. (2021) 'Are There Indeed Spliced Peptides in the Immunopeptidome?', *Molecular & cellular proteomics: MCP*, 20, p. 100099. Available at: <https://doi.org/10.1016/j.mcpro.2021.100099>.
- Affram, Y. *et al.* (2019) 'The HIV-1 Antisense Protein ASP Is a Transmembrane Protein of the Cell Surface and an Integral Protein of the Viral Envelope', *Journal of Virology*, 93(21), pp. e00574-19. Available at: <https://doi.org/10.1128/JVI.00574-19>.
- Ahlenstiel, C.L. *et al.* (2020) 'Block and Lock HIV Cure Strategies to Control the Latent Reservoir', *Frontiers in Cellular and Infection Microbiology*, 10, p. 424. Available at: <https://doi.org/10.3389/fcimb.2020.00424>.
- Alizon, M. *et al.* (1984) 'Molecular cloning of lymphadenopathy-associated virus', *Nature*, 312(5996), pp. 757–760. Available at: <https://doi.org/10.1038/312757a0>.
- Allfrey, V.G., Mirsky, A.E. and Osawa, S. (1957) 'Protein synthesis in isolated cell nuclei', *The Journal of General Physiology*, 40(3), pp. 451–490. Available at: <https://doi.org/10.1085/jgp.40.3.451>.
- Almeida, J.R. *et al.* (2007) 'Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover', *The Journal of Experimental Medicine*, 204(10), pp. 2473–2485. Available at: <https://doi.org/10.1084/jem.20070784>.
- Almeida, J.R. *et al.* (2009) 'Antigen sensitivity is a major determinant of CD8+ T-cell polyfunctionality and HIV-suppressive activity', *Blood*, 113(25), pp. 6351–6360. Available at: <https://doi.org/10.1182/blood-2009-02-206557>.
- Alvarez, E. *et al.* (2006) 'HIV protease cleaves poly(A)-binding protein', *The Biochemical Journal*, 396(2), pp. 219–226. Available at: <https://doi.org/10.1042/BJ20060108>.
- Amit, I. *et al.* (2009) 'Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses', *Science (New York, N.Y.)*, 326(5950), pp. 257–263. Available at: <https://doi.org/10.1126/science.1179050>.
- Anderson, J.L. *et al.* (2007) 'Both linear and discontinuous ribosome scanning are used for translation initiation from bicistronic human immunodeficiency virus type 1 env mRNAs', *Journal of Virology*, 81(9), pp. 4664–4676. Available at: <https://doi.org/10.1128/JVI.01028-06>.
- Anderson, P.L. *et al.* (2012) 'Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men', *Science Translational Medicine*, 4(151), p. 151ra125. Available at: <https://doi.org/10.1126/scitranslmed.3004006>.
- Andreev, D. *et al.* (2008) 'The bacterial toxin RelE induces specific mRNA cleavage in the A site of the eukaryote ribosome', *RNA (New York, N.Y.)*, 14(2), pp. 233–239. Available at: <https://doi.org/10.1261/rna.693208>.

- Andreev, D.E. *et al.* (2022) 'Non-AUG translation initiation in mammals', *Genome Biology*, 23(1), p. 111. Available at: <https://doi.org/10.1186/s13059-022-02674-2>.
- Andrews, S.J. and Rothnagel, J.A. (2014) 'Emerging evidence for functional peptides encoded by short open reading frames'. Available at: <https://doi.org/10.1038/nrg3520>.
- ANRS | Résultats finaux de l'essai ANRS IPERGAY (2016). Available at: <https://www.anrs.fr/fr/presse/communiqués-de-presse/326/resultats-finaux-de-lessai-anrs-ipergay-la-tres-haute-efficacite> (Accessed: 28 March 2023).
- Antón, L.C. and Yewdell, J.W. (2014) 'Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors', *Journal of Leukocyte Biology*, 95(4), pp. 551–562. Available at: <https://doi.org/10.1189/jlb.1113599>.
- Apcher, S. *et al.* (2011) 'Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation', *Proceedings of the National Academy of Sciences*, 108(28), pp. 11572–11577. Available at: <https://doi.org/10.1073/pnas.1104104108>.
- Apcher, S. *et al.* (2022) 'mRNA translation from an antigen presentation perspective: A tribute to the works of Nilabh Shastri', *Molecular Immunology*, 141, pp. 305–308. Available at: <https://doi.org/10.1016/j.molimm.2021.12.010>.
- Apcher, S., Daskalogianni, C. and Fåhræus, R. (2015) 'Pioneer translation products as an alternative source for MHC-I antigenic peptides', *Molecular Immunology*, 68(2, Part A), pp. 68–71. Available at: <https://doi.org/10.1016/j.molimm.2015.04.019>.
- Apcher, S., Fahraeus, R. and Manoury, B. (2004) 'Epstein–Barr virus: exploiting the immune system by interfering with defective ribosomal products', *Microbes and Infection*, 6(13), pp. 1212–1218. Available at: <https://doi.org/10.1016/j.micinf.2004.08.003>.
- Appay, V. *et al.* (2002) 'Characterization of CD4+ CTLs Ex Vivo', *The Journal of Immunology*, 168(11), pp. 5954–5958. Available at: <https://doi.org/10.4049/jimmunol.168.11.5954>.
- Appay, V. and Sauce, D. (2008) 'Immune activation and inflammation in HIV-1 infection: causes and consequences', *The Journal of Pathology*, 214(2), pp. 231–241. Available at: <https://doi.org/10.1002/path.2276>.
- Araujo, P.R. *et al.* (2012) 'Before It Gets Started: Regulating Translation at the 5' UTR', *Comparative and Functional Genomics*, 2012, p. 475731. Available at: <https://doi.org/10.1155/2012/475731>.
- Arias, C. *et al.* (2014) 'KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features', *PLoS pathogens*, 10(1), p. e1003847. Available at: <https://doi.org/10.1371/journal.ppat.1003847>.
- Ariumi, Y. (2014) 'Multiple functions of DDX3 RNA helicase in gene regulation, tumorigenesis, and viral infection', *Frontiers in Genetics*, 5. Available at: <https://www.frontiersin.org/articles/10.3389/fgene.2014.00423> (Accessed: 13 February 2023).
- Arnaud, M. *et al.* (2021) 'The Promise of Personalized TCR-Based Cellular Immunotherapy for Cancer Patients', *Frontiers in Immunology*, 12, p. 701636. Available at: <https://doi.org/10.3389/fimmu.2021.701636>.
- Arnaud, M. *et al.* (2022) 'Sensitive identification of neoantigens and cognate TCRs in human solid tumors', *Nature Biotechnology*, 40(5), pp. 656–660. Available at: <https://doi.org/10.1038/s41587-021-01072-6>.
- Autran, B. *et al.* (1997) 'Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease', *Science (New York, N.Y.)*, 277(5322), pp. 112–116. Available at: <https://doi.org/10.1126/science.277.5322.112>.
- Baeten, J.M. *et al.* (2012) 'Antiretroviral prophylaxis for HIV prevention in heterosexual men and women', *The New England Journal of Medicine*, 367(5), pp. 399–410. Available at: <https://doi.org/10.1056/NEJMoa1108524>.

- Bailey, J.R. *et al.* (2006) 'Neutralizing antibodies do not mediate suppression of human immunodeficiency virus type 1 in elite suppressors or selection of plasma virus variants in patients on highly active antiretroviral therapy', *Journal of Virology*, 80(10), pp. 4758–4770. Available at: <https://doi.org/10.1128/JVI.80.10.4758-4770.2006>.
- Bansal, A. *et al.* (2010) 'CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription', *The Journal of Experimental Medicine*, 207(1), pp. 51–59. Available at: <https://doi.org/10.1084/jem.20092060>.
- Bansal, A. *et al.* (2015) 'Enhanced Recognition of HIV-1 Cryptic Epitopes Restricted by HLA Class I Alleles Associated With a Favorable Clinical Outcome', *Journal of Acquired Immune Deficiency Syndromes (1999)*, 70(1), pp. 1–8. Available at: <https://doi.org/10.1097/QAI.0000000000000700>.
- Barbosa, C., Peixeiro, I. and Romão, L. (2013) 'Gene Expression Regulation by Upstream Open Reading Frames and Human Disease', *PLoS Genetics*, 9(8), p. e1003529. Available at: <https://doi.org/10.1371/journal.pgen.1003529>.
- Baril, M. *et al.* (2003) 'Efficiency of a programmed -1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M', *RNA (New York, N.Y.)*, 9(10), pp. 1246–1253. Available at: <https://doi.org/10.1261/rna.5113603>.
- Bar-On, Y. *et al.* (2018) 'Safety and antiviral activity of combination HIV-1 broadly neutralizing antibodies in viremic individuals', *Nature Medicine*, 24(11), pp. 1701–1707. Available at: <https://doi.org/10.1038/s41591-018-0186-4>.
- Barré-Sinoussi, F. *et al.* (1983) 'Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS)', *Science*, 220(4599), pp. 868–871. Available at: <https://doi.org/10.1126/science.6189183>.
- Basrai, M.A., Hieter, P. and Boeke, J.D. (1997) 'Small Open Reading Frames: Beautiful Needles in the Haystack', *Genome Research*, 7(8), pp. 768–771. Available at: <https://doi.org/10.1101/gr.7.8.768>.
- Bassani-Sternberg, M. *et al.* (2015) 'Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation', *Molecular & cellular proteomics: MCP*, 14(3), pp. 658–673. Available at: <https://doi.org/10.1074/mcp.M114.042812>.
- Bassani-Sternberg, M. *et al.* (2016) 'Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry', *Nature Communications*, 7, p. 13404. Available at: <https://doi.org/10.1038/ncomms13404>.
- Bassani-Sternberg, M. and Coukos, G. (2016) 'Mass spectrometry-based antigen discovery for cancer immunotherapy', *Current Opinion in Immunology*, 41, pp. 9–17. Available at: <https://doi.org/10.1016/j.coi.2016.04.005>.
- Bastide, A., Yewdell, J.W. and David, A. (2018) 'The RiboPuromycylation Method (RPM): an Immunofluorescence Technique to Map Translation Sites at the Sub-cellular Level', *Bio-protocol*, 8(1), p. e2669. Available at: <https://doi.org/10.21769/BioProtoc.2669>.
- Bazzini, A.A. *et al.* (2014) 'Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation', *The EMBO Journal*, 33(9), pp. 981–993. Available at: <https://doi.org/10.1002/embj.201488411>.
- Bazzini, A.A., Lee, M.T. and Giraldez, A.J. (2012) 'Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish', *Science (New York, N.Y.)*, 336(6078), pp. 233–237. Available at: <https://doi.org/10.1126/science.1215704>.
- Belyakov, I.M. and Berzofsky, J.A. (2004) 'Immunobiology of Mucosal HIV Infection and the Basis for Development of a New Generation of Mucosal AIDS Vaccines', *Immunity*, 20(3), pp. 247–253. Available at: [https://doi.org/10.1016/S1074-7613\(04\)00053-6](https://doi.org/10.1016/S1074-7613(04)00053-6).

- Benati, D. *et al.* (2016) 'Public T cell receptors confer high-avidity CD4 responses to HIV controllers', *The Journal of Clinical Investigation*, 126(6), pp. 2093–2108. Available at: <https://doi.org/10.1172/JCI83792>.
- Benitez-Cantos, M.S. *et al.* (2020) 'Translation initiation downstream from annotated start codons in human mRNAs coevolves with the Kozak context', *Genome Research*, 30(7), pp. 974–984. Available at: <https://doi.org/10.1101/gr.257352.119>.
- Ben-Shem, A. *et al.* (2011) 'The Structure of the Eukaryotic Ribosome at 3.0 Å Resolution', *Science*, 334(6062), pp. 1524–1529. Available at: <https://doi.org/10.1126/science.1212642>.
- Berger, C.T. *et al.* (2010) 'Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences', *The Journal of Experimental Medicine*, 207(1), pp. 61–75. Available at: <https://doi.org/10.1084/jem.20091808>.
- Berger, C.T. *et al.* (2015) 'Immune screening identifies novel T cell targets encoded by antisense reading frames of HIV-1', *Journal of Virology*, 89(7), pp. 4015–4019. Available at: <https://doi.org/10.1128/JVI.03435-14>.
- Berget, S.M., Moore, C. and Sharp, P.A. (1977) 'Spliced segments at the 5' terminus of adenovirus 2 late mRNA', *Proceedings of the National Academy of Sciences of the United States of America*, 74(8), pp. 3171–3175. Available at: <https://doi.org/10.1073/pnas.74.8.3171>.
- Berry, F. *et al.* (2018) 'Les facteurs du pore nucléaire : un collectif gagnant pour la translocation et l'intégration du VIH-1', *médecine/sciences*, 34(6–7), pp. 512–515. Available at: <https://doi.org/10.1051/medsci/20183406006>.
- Berthelot, K. *et al.* (2004) 'Dynamics and processivity of 40S ribosome scanning on mRNA in yeast', *Molecular Microbiology*, 51(4), pp. 987–1001. Available at: <https://doi.org/10.1046/j.1365-2958.2003.03898.x>.
- Berzofsky, J.A. *et al.* (1988) 'Antigenic peptides recognized by T lymphocytes from AIDS viral envelope-immune humans', *Nature*, 334(6184), pp. 706–708. Available at: <https://doi.org/10.1038/334706a0>.
- Bet, A. *et al.* (2015) 'The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection', *Retrovirology*, 12, p. 15. Available at: <https://doi.org/10.1186/s12977-015-0135-y>.
- Betts, M.R. *et al.* (2001) 'Analysis of total human immunodeficiency virus (HIV)-specific CD4(+) and CD8(+) T-cell responses: relationship to viral load in untreated HIV infection', *Journal of Virology*, 75(24), pp. 11983–11991. Available at: <https://doi.org/10.1128/JVI.75.24.11983-11991.2001>.
- Betts, M.R. *et al.* (2006) 'HIV nonprogressors preferentially maintain highly functional HIV-specific CD8+ T cells', *Blood*, 107(12), pp. 4781–4789. Available at: <https://doi.org/10.1182/blood-2005-12-4818>.
- Betts, M.R. and Koup, R.A. (2004) 'Detection of T-Cell Degranulation: CD107a and b', in *Methods in Cell Biology*. Academic Press (Cytometry, 4th Edition: New Developments), pp. 497–512. Available at: [https://doi.org/10.1016/S0091-679X\(04\)75020-7](https://doi.org/10.1016/S0091-679X(04)75020-7).
- Bevan, M.J. (1976) 'Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay', *The Journal of Experimental Medicine*, 143(5), p. 1283. Available at: <https://doi.org/10.1084/jem.143.5.1283>.
- Birnstiel, M.L. and Hyde, B.B. (1963) 'Protein synthesis by isolated pea nucleoli', *The Journal of Cell Biology*, 18(1), pp. 41–50. Available at: <https://doi.org/10.1083/jcb.18.1.41>.
- Biswas, P. *et al.* (2004) 'The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy', *Journal of Virology*, 78(4), pp. 2082–2087. Available at: <https://doi.org/10.1128/jvi.78.4.2082-2087.2004>.
- Blackburn, S.D. *et al.* (2009) 'Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection', *Nature Immunology*, 10(1), pp. 29–37. Available at: <https://doi.org/10.1038/ni.1679>.

- Blander, J.M. (2018) 'Regulation of the Cell Biology of Antigen Cross-Presentation', *Annual Review of Immunology*, 36, pp. 717–753. Available at: <https://doi.org/10.1146/annurev-immunol-041015-055523>.
- Blees, A. *et al.* (2017) 'Structure of the human MHC-I peptide-loading complex', *Nature*, 551(7681), pp. 525–528. Available at: <https://doi.org/10.1038/nature24627>.
- Blum, J.S., Wearsch, P.A. and Cresswell, P. (2013) 'Pathways of Antigen Processing', *Annual review of immunology*, 31, pp. 443–473. Available at: <https://doi.org/10.1146/annurev-immunol-032712-095910>.
- Bobisse, S. *et al.* (2018) 'Sensitive and frequent identification of high avidity neo-epitope specific CD8 + T cells in immunotherapy-naive ovarian cancer', *Nature Communications*, 9(1), p. 1092. Available at: <https://doi.org/10.1038/s41467-018-03301-0>.
- Boes, M. *et al.* (2002) 'T-cell engagement of dendritic cells rapidly rearranges MHC class II transport', *Nature*, 418(6901), pp. 983–988. Available at: <https://doi.org/10.1038/nature01004>.
- Bohlen, J. *et al.* (2020) 'DENR promotes translation reinitiation via ribosome recycling to drive expression of oncogenes including ATF4', *Nature Communications*, 11(1), p. 4676. Available at: <https://doi.org/10.1038/s41467-020-18452-2>.
- Boon, T. and Van Pel, A. (1989) 'T cell-recognized antigenic peptides derived from the cellular genome are not protein degradation products but can be generated directly by transcription and translation of short subgenic regions. A hypothesis', *Immunogenetics*, 29(2), pp. 75–79. Available at: <https://doi.org/10.1007/BF00395854>.
- Boritz, E., Palmer, B.E. and Wilson, C.C. (2004) 'Human immunodeficiency virus type 1 (HIV-1)-specific CD4+ T cells that proliferate in vitro detected in samples from most viremic subjects and inversely associated with plasma HIV-1 levels', *Journal of Virology*, 78(22), pp. 12638–12646. Available at: <https://doi.org/10.1128/JVI.78.22.12638-12646.2004>.
- Borrow, P. *et al.* (1994) 'Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection', *Journal of Virology*, 68(9), pp. 6103–6110. Available at: <https://doi.org/10.1128/JVI.68.9.6103-6110.1994>.
- Boucau, J. and Le Gall, S. (2019) 'Antigen processing and presentation in HIV infection', *Molecular Immunology*, 113, pp. 67–74. Available at: <https://doi.org/10.1016/j.molimm.2018.03.027>.
- Boufassa, F. *et al.* (2011) 'CD4 dynamics over a 15 year-period among HIV controllers enrolled in the ANRS French observatory', *PloS One*, 6(4), p. e18726. Available at: <https://doi.org/10.1371/journal.pone.0018726>.
- Boyer, Z. and Palmer, S. (2018) 'Targeting Immune Checkpoint Molecules to Eliminate Latent HIV', *Frontiers in Immunology*, 9. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02339> (Accessed: 29 March 2023).
- Brar, G.A. and Weissman, J.S. (2015) 'Ribosome profiling reveals the what, when, where and how of protein synthesis', *Nature Reviews. Molecular Cell Biology*, 16(11), pp. 651–664. Available at: <https://doi.org/10.1038/nrm4069>.
- Brenchley, J.M. *et al.* (2004) 'T-cell subsets that harbor human immunodeficiency virus (HIV) in vivo: implications for HIV pathogenesis', *Journal of Virology*, 78(3), pp. 1160–1168. Available at: <https://doi.org/10.1128/jvi.78.3.1160-1168.2004>.
- Brenchley, J.M. *et al.* (2008) 'Differential Th17 CD4 T-cell depletion in pathogenic and nonpathogenic lentiviral infections', *Blood*, 112(7), pp. 2826–2835. Available at: <https://doi.org/10.1182/blood-2008-05-159301>.
- Brennick, C.A. *et al.* (2017) 'Neoepitopes as cancer immunotherapy targets: key challenges and opportunities', *Immunotherapy*, 9(4), pp. 361–371. Available at: <https://doi.org/10.2217/imt-2016-0146>.
- de Breyne, S. *et al.* (2012) 'In vitro studies reveal that different modes of initiation on HIV-1 mRNA have different levels of requirement for eukaryotic initiation factor 4F', *The FEBS journal*, 279(17), pp. 3098–3111. Available at: <https://doi.org/10.1111/j.1742-4658.2012.08689.x>.

- de Breyne, S. and Ohlmann, T. (2018) 'Focus on Translation Initiation of the HIV-1 mRNAs', *International Journal of Molecular Sciences*, 20(1), p. 101. Available at: <https://doi.org/10.3390/ijms20010101>.
- Brierley, I. (1995) 'Ribosomal frameshifting on viral RNAs', *Journal of General Virology*, 76(8), pp. 1885–1892. Available at: <https://doi.org/10.1099/0022-1317-76-8-1885>.
- Brierley, I. and Dos Ramos, F.J. (2006) 'Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV', *Virus Research*, 119(1), pp. 29–42. Available at: <https://doi.org/10.1016/j.virusres.2005.10.008>.
- Briggs, J.A.G. *et al.* (2003) 'Structural organization of authentic, mature HIV-1 virions and cores', *The EMBO journal*, 22(7), pp. 1707–1715. Available at: <https://doi.org/10.1093/emboj/cdg143>.
- Brumme, Z.L. *et al.* (2008) 'Marked Epitope- and Allele-Specific Differences in Rates of Mutation in Human Immunodeficiency Type 1 (HIV-1) Gag, Pol, and Nef Cytotoxic T-Lymphocyte Epitopes in Acute/Early HIV-1 Infection', *Journal of Virology*, 82(18), pp. 9216–9227. Available at: <https://doi.org/10.1128/JVI.01041-08>.
- Brunet, M.A. *et al.* (2021) 'The FUS gene is dual-coding with both proteins contributing to FUS-mediated toxicity', *EMBO reports*, 22(1), p. e50640. Available at: <https://doi.org/10.15252/embr.202050640>.
- Buck, C.B. *et al.* (2001) 'The Human Immunodeficiency Virus Type 1 gag Gene Encodes an Internal Ribosome Entry Site', *Journal of Virology*, 75(1), pp. 181–191. Available at: <https://doi.org/10.1128/JVI.75.1.181-191.2001>.
- Buckner, C.M. *et al.* (2016) 'Maintenance of HIV-Specific Memory B-Cell Responses in Elite Controllers Despite Low Viral Burdens', *The Journal of Infectious Diseases*, 214(3), pp. 390–398. Available at: <https://doi.org/10.1093/infdis/jiw163>.
- Budkevich, T.V. *et al.* (2014) 'Regulation of the Mammalian Elongation Cycle by Subunit Rolling: A Eukaryotic-Specific Ribosome Rearrangement', *Cell*, 158(1), pp. 121–131. Available at: <https://doi.org/10.1016/j.cell.2014.04.044>.
- Bullock, T.N. and Eisenlohr, L.C. (1996) 'Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames', *The Journal of Experimental Medicine*, 184(4), pp. 1319–1329. Available at: <https://doi.org/10.1084/jem.184.4.1319>.
- Buranapraditkun, S. *et al.* (2017) 'Preservation of Peripheral T Follicular Helper Cell Function in HIV Controllers', *Journal of Virology*, 91(14), pp. e00497-17. Available at: <https://doi.org/10.1128/JVI.00497-17>.
- Burgevin, A. *et al.* (2008) 'A Detailed Analysis of the Murine TAP Transporter Substrate Specificity', *PLOS ONE*, 3(6), p. e2402. Available at: <https://doi.org/10.1371/journal.pone.0002402>.
- Busch, R. *et al.* (1996) 'Invariant chain protects class II histocompatibility antigens from binding intact polypeptides in the endoplasmic reticulum', *The EMBO journal*, 15(2), pp. 418–428.
- Bushman, F.D. *et al.* (2009) 'Host cell factors in HIV replication: meta-analysis of genome-wide studies', *PLoS pathogens*, 5(5), p. e1000437. Available at: <https://doi.org/10.1371/journal.ppat.1000437>.
- Cachot, A. *et al.* (2021) 'Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer', *Science Advances*, 7(9), p. eabe3348. Available at: <https://doi.org/10.1126/sciadv.abe3348>.
- Caillat, C. *et al.* (2020) 'Neutralizing Antibodies Targeting HIV-1 gp41', *Viruses*, 12(11), p. 1210. Available at: <https://doi.org/10.3390/v12111210>.
- Calviello, L. and Ohler, U. (2017) 'Beyond Read-Counts: Ribo-seq Data Analysis to Understand the Functions of the Transcriptome', *Trends in Genetics*, 33(10), pp. 728–744. Available at: <https://doi.org/10.1016/j.tig.2017.08.003>.
- Calvo, S.E., Pagliarini, D.J. and Mootha, V.K. (2009) 'Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans', *Proceedings of the National Academy of Sciences*, 106(18), pp. 7507–7512. Available at: <https://doi.org/10.1073/pnas.0810916106>.

- Campion, S.L. *et al.* (2014) 'Proteome-wide analysis of HIV-specific naive and memory CD4(+) T cells in unexposed blood donors', *The Journal of Experimental Medicine*, 211(7), pp. 1273–1280. Available at: <https://doi.org/10.1084/jem.20130555>.
- Cao, X. *et al.* (2021) 'Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24', *Nature Communications*, 12(1), p. 508. Available at: <https://doi.org/10.1038/s41467-020-20841-6>.
- Cao, X. and Slavoff, S.A. (2020) 'Non-AUG start codons: Expanding and regulating the small and alternative ORFeome', *Experimental Cell Research*, 391(1), p. 111973. Available at: <https://doi.org/10.1016/j.yexcr.2020.111973>.
- Carbannelle, D. *et al.* (2013) 'The Melanoma Antigens MELOE-1 and MELOE-2 Are Translated from a Bona Fide Polycistronic mRNA Containing Functional IRES Sequences', *PLOS ONE*, 8(9), p. e75233. Available at: <https://doi.org/10.1371/journal.pone.0075233>.
- Cardinaud, S. *et al.* (2004) 'Identification of Cryptic MHC I-restricted Epitopes Encoded by HIV-1 Alternative Reading Frames', *Journal of Experimental Medicine*, 199(8), pp. 1053–1063. Available at: <https://doi.org/10.1084/jem.20031869>.
- Cardinaud, S. *et al.* (2011) 'CTL Escape Mediated by Proteasomal Destruction of an HIV-1 Cryptic Epitope', *PLOS Pathogens*, 7(5), p. e1002049. Available at: <https://doi.org/10.1371/journal.ppat.1002049>.
- Carlson, E. (1966) 'The gene: a critical history', in. Available at: <https://www.semanticscholar.org/paper/The-gene%3A-a-critical-history-Carlson/08410ff808a3ff6a789c52dbe7eda7b1dc677b8a> (Accessed: 13 February 2023).
- Carlson, J.M. *et al.* (2008) 'Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag', *PLoS computational biology*, 4(11), p. e1000225. Available at: <https://doi.org/10.1371/journal.pcbi.1000225>.
- Carneiro, J. and Leblond, C.P. (1959) 'Continuous protein synthesis in nuclei, shown by radioautography with H³-labeled amino acids', *Science (New York, N.Y.)*, 129(3346), pp. 391–392. Available at: <https://doi.org/10.1126/science.129.3346.391>.
- Caron, E. *et al.* (2015) 'Analysis of Major Histocompatibility Complex (MHC) Immunoepitomes Using Mass Spectrometry', *Molecular & cellular proteomics: MCP*, 14(12), pp. 3105–3117. Available at: <https://doi.org/10.1074/mcp.O115.052431>.
- Carvunis, A.-R. *et al.* (2012) 'Proto-genes and de novo gene birth', *Nature*, 487(7407), pp. 370–374. Available at: <https://doi.org/10.1038/nature11184>.
- Casazza, J.P. *et al.* (2001) 'Decay kinetics of human immunodeficiency virus-specific CD8+ T cells in peripheral blood after initiation of highly active antiretroviral therapy', *Journal of Virology*, 75(14), pp. 6508–6516. Available at: <https://doi.org/10.1128/JVI.75.14.6508-6516.2001>.
- Cassan, E. *et al.* (2016) 'Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic', *Proceedings of the National Academy of Sciences of the United States of America*, 113(41), pp. 11537–11542. Available at: <https://doi.org/10.1073/pnas.1605739113>.
- Castelló, A. *et al.* (2009) 'HIV-1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4G1 and PABP cleavage', *PloS One*, 4(11), p. e7997. Available at: <https://doi.org/10.1371/journal.pone.0007997>.
- Castelo-Szekely, V. *et al.* (2019) 'Charting DENR-dependent translation reinitiation uncovers predictive uORF features and links to circadian timekeeping via Clock', *Nucleic Acids Research*, 47(10), pp. 5193–5209. Available at: <https://doi.org/10.1093/nar/gkz261>.
- Castle, J.C. *et al.* (2012) 'Exploiting the mutanome for tumor vaccination', *Cancer Research*, 72(5), pp. 1081–1091. Available at: <https://doi.org/10.1158/0008-5472.CAN-11-3722>.

- Ceglia, V. *et al.* (2021) 'Anti-CD40 Antibody Fused to CD40 Ligand Is a Superagonist Platform for Adjuvant Intrinsic DC-Targeting Vaccines', *Frontiers in Immunology*, 12, p. 786144. Available at: <https://doi.org/10.3389/fimmu.2021.786144>.
- Cen, S. *et al.* (2004) 'Incorporation of pol into human immunodeficiency virus type 1 Gag virus-like particles occurs independently of the upstream Gag domain in Gag-pol', *Journal of Virology*, 78(2), pp. 1042–1049. Available at: <https://doi.org/10.1128/jvi.78.2.1042-1049.2004>.
- Ceppi, M. *et al.* (2009) 'Ribosomal protein mRNAs are translationally-regulated during human dendritic cells activation by LPS', *Immunome Research*, 5, p. 5. Available at: <https://doi.org/10.1186/1745-7580-5-5>.
- Chaillon, A. *et al.* (2020) 'HIV persists throughout deep tissues with repopulation from multiple anatomical sources', *The Journal of Clinical Investigation*, 130(4), pp. 1699–1712. Available at: <https://doi.org/10.1172/JCI134815>.
- Champiat, S. *et al.* (2012) 'Influence of HAART on Alternative Reading Frame Immune Responses over the Course of HIV-1 Infection', *PLOS ONE*, 7(6), p. e39311. Available at: <https://doi.org/10.1371/journal.pone.0039311>.
- Chang, C.H. *et al.* (1994) 'Class II transactivator (CIITA) is sufficient for the inducible expression of major histocompatibility complex class II genes', *The Journal of Experimental Medicine*, 180(4), pp. 1367–1374. Available at: <https://doi.org/10.1084/jem.180.4.1367>.
- Chang, Y.N. *et al.* (1995) 'Direct interactions between autoantigen La and human immunodeficiency virus leader RNA', *Journal of Virology*, 69(1), pp. 618–619. Available at: <https://doi.org/10.1128/JVI.69.1.618-619.1995>.
- Chen, B. (2019) 'Molecular Mechanism of HIV-1 Entry', *Trends in Microbiology*, 27(10), pp. 878–891. Available at: <https://doi.org/10.1016/j.tim.2019.06.002>.
- Chen, H. *et al.* (2009) 'Differential neutralization of human immunodeficiency virus (HIV) replication in autologous CD4 T cells by HIV-specific cytotoxic T lymphocytes', *Journal of Virology*, 83(7), pp. 3138–3149. Available at: <https://doi.org/10.1128/JVI.02073-08>.
- Chen, H. *et al.* (2012) 'TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection', *Nature Immunology*, 13(7), pp. 691–700. Available at: <https://doi.org/10.1038/ni.2342>.
- Chen, H.-H. *et al.* (2018) 'DDX3 Activates CBC-eIF3-Mediated Translation of uORF-Containing Oncogenic mRNAs to Promote Metastasis in HNSCC', *Cancer Research*, 78(16), pp. 4512–4523. Available at: <https://doi.org/10.1158/0008-5472.CAN-18-0282>.
- Chen, H.-H. and Tarn, W.-Y. (2019) 'uORF-mediated translational control: recently elucidated mechanisms and implications in cancer', *RNA biology*, 16(10), pp. 1327–1338. Available at: <https://doi.org/10.1080/15476286.2019.1632634>.
- Chen, J. *et al.* (2020) 'Pervasive functional translation of noncanonical human open reading frames', *Science (New York, N.Y.)*, 367(6482), pp. 1140–1146. Available at: <https://doi.org/10.1126/science.aay0262>.
- Chen, R. *et al.* (2021) 'Identification of Mhc Peptides Using Mass Spectrometry for Neoantigen Discovery and Cancer Vaccine Development', *Mass Spectrometry Reviews*, 40(2), pp. 110–125. Available at: <https://doi.org/10.1002/mas.21616>.
- Chen, W. *et al.* (2001) 'Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens', *The Journal of Experimental Medicine*, 193(11), pp. 1319–1326. Available at: <https://doi.org/10.1084/jem.193.11.1319>.
- Chen, X. *et al.* (1995) 'Structural and functional studies of retroviral RNA pseudoknots involved in ribosomal frameshifting: nucleotides at the junction of the two stems are important for efficient ribosomal frameshifting', *The EMBO Journal*, 14(4), pp. 842–852.

- Cheng, L. *et al.* (2018) 'TLR3 agonist and CD40-targeting vaccination induces immune responses and reduces HIV-1 reservoirs', *The Journal of Clinical Investigation*, 128(10), pp. 4387–4396. Available at: <https://doi.org/10.1172/JCI99005>.
- Cheroutre, H. and Husain, M.M. (2013) 'CD4 CTL: living up to the challenge', *Seminars in Immunology*, 25(4), pp. 273–281. Available at: <https://doi.org/10.1016/j.smim.2013.10.022>.
- Chevalier, M.F. *et al.* (2011) 'HIV-1-specific interleukin-21+ CD4+ T cell responses contribute to durable viral control through the modulation of HIV-specific CD8+ T cell function', *Journal of Virology*, 85(2), pp. 733–741. Available at: <https://doi.org/10.1128/JVI.02030-10>.
- Chicz, R.M. *et al.* (1993) 'Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles', *The Journal of Experimental Medicine*, 178(1), pp. 27–47. Available at: <https://doi.org/10.1084/jem.178.1.27>.
- Chikata, T. *et al.* (2019) 'Identification of Immunodominant HIV-1 Epitopes Presented by HLA-C*12:02, a Protective Allele, Using an Immunopeptidomics Approach', *Journal of Virology*, 93(17), pp. e00634–19. Available at: <https://doi.org/10.1128/JVI.00634-19>.
- Chong, C. *et al.* (2020) 'Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes', *Nature Communications*, 11(1), p. 1293. Available at: <https://doi.org/10.1038/s41467-020-14968-9>.
- Chong, C., Coukos, G. and Bassani-Sternberg, M. (2022) 'Identification of tumor antigens with immunopeptidomics', *Nature Biotechnology*, 40(2), pp. 175–188. Available at: <https://doi.org/10.1038/s41587-021-01038-8>.
- Chow, L.T. *et al.* (1977) 'An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA', *Cell*, 12(1), pp. 1–8. Available at: [https://doi.org/10.1016/0092-8674\(77\)90180-5](https://doi.org/10.1016/0092-8674(77)90180-5).
- Chu, D. *et al.* (2014) 'Translation elongation can control translation initiation on eukaryotic mRNAs', *The EMBO journal*, 33(1), pp. 21–34. Available at: <https://doi.org/10.1002/embj.201385651>.
- Chun, T.W. *et al.* (1998) 'Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection', *Proceedings of the National Academy of Sciences of the United States of America*, 95(15), pp. 8869–8873. Available at: <https://doi.org/10.1073/pnas.95.15.8869>.
- Cicala, C. *et al.* (2009) 'The integrin alpha4beta7 forms a complex with cell-surface CD4 and defines a T-cell subset that is highly susceptible to infection by HIV-1', *Proceedings of the National Academy of Sciences of the United States of America*, 106(49), pp. 20877–20882. Available at: <https://doi.org/10.1073/pnas.0911796106>.
- Ciechanover, A. *et al.* (1980) 'ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation', *Proceedings of the National Academy of Sciences of the United States of America*, 77(3), pp. 1365–1368. Available at: <https://doi.org/10.1073/pnas.77.3.1365>.
- Clavel, F. *et al.* (1986) 'Isolation of a new human retrovirus from West African patients with AIDS', *Science (New York, N.Y.)*, 233(4761), pp. 343–346. Available at: <https://doi.org/10.1126/science.2425430>.
- Cochrane, A.W., Chen, C.H. and Rosen, C.A. (1990) 'Specific interaction of the human immunodeficiency virus Rev protein with a structured region in the env mRNA', *Proceedings of the National Academy of Sciences of the United States of America*, 87(3), pp. 1198–1202. Available at: <https://doi.org/10.1073/pnas.87.3.1198>.
- Coffin, J. and Swanstrom, R. (2013) 'HIV pathogenesis: dynamics and genetics of viral populations and infected cells', *Cold Spring Harbor Perspectives in Medicine*, 3(1), p. a012526. Available at: <https://doi.org/10.1101/cshperspect.a012526>.
- Collora, J.A. *et al.* (2022) 'Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones', *Immunity*, 55(6), pp. 1013–1031.e7. Available at: <https://doi.org/10.1016/j.immuni.2022.03.004>.

- Colomer-Lluch, M. *et al.* (2018) 'Restriction Factors: From Intrinsic Viral Restriction to Shaping Cellular Immunity Against HIV-1', *Frontiers in Immunology*, 9. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02876> (Accessed: 6 April 2023).
- Coulie, P.G. *et al.* (1995) 'A mutated intron sequence codes for an antigenic peptide recognized by cytolytic T lymphocytes on a human melanoma', *Proceedings of the National Academy of Sciences of the United States of America*, 92(17), pp. 7976–7980. Available at: <https://doi.org/10.1073/pnas.92.17.7976>.
- Coulon, P.-G. *et al.* (2016) 'HIV-Infected Dendritic Cells Present Endogenous MHC Class II–Restricted Antigens to HIV-Specific CD4+ T Cells', *The Journal of Immunology*, 197(2), pp. 517–532. Available at: <https://doi.org/10.4049/jimmunol.1600286>.
- Couso, J.-P. and Patraquim, P. (2017) 'Classification and function of small open reading frames', *Nature Reviews. Molecular Cell Biology*, 18(9), pp. 575–589. Available at: <https://doi.org/10.1038/nrm.2017.58>.
- Couturier, J. *et al.* (2014) 'HIV replication in conjunction with granzyme B production by CCR5+ memory CD4 T cells: Implications for bystander cell and tissue pathologies', *Virology*, 462–463, pp. 175–188. Available at: <https://doi.org/10.1016/j.virol.2014.06.008>.
- Craigie, R. and Bushman, F.D. (2012) 'HIV DNA Integration', *Cold Spring Harbor Perspectives in Medicine*, 2(7), p. a006890. Available at: <https://doi.org/10.1101/cshperspect.a006890>.
- Cranage, M. *et al.* (2008) 'Prevention of SIV rectal transmission and priming of T cell responses in macaques after local pre-exposure application of tenofovir gel', *PLoS medicine*, 5(8), pp. e157; discussion e157. Available at: <https://doi.org/10.1371/journal.pmed.0050157>.
- Crick, F.H.C. (1966) 'Codon—anticodon pairing: The wobble hypothesis', *Journal of Molecular Biology*, 19(2), pp. 548–555. Available at: [https://doi.org/10.1016/S0022-2836\(66\)80022-0](https://doi.org/10.1016/S0022-2836(66)80022-0).
- Croft, N.P. *et al.* (2013) 'Kinetics of antigen expression and epitope presentation during virus infection', *PLoS pathogens*, 9(1), p. e1003129. Available at: <https://doi.org/10.1371/journal.ppat.1003129>.
- Crowe, M.L., Wang, X.-Q. and Rothnagel, J.A. (2006) 'Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides', *BMC genomics*, 7, p. 16. Available at: <https://doi.org/10.1186/1471-2164-7-16>.
- Curlin, M.E. *et al.* (2010) 'HIV-1 Envelope Subregion Length Variation during Disease Progression', *PLOS Pathogens*, 6(12), p. e1001228. Available at: <https://doi.org/10.1371/journal.ppat.1001228>.
- Dani, A. *et al.* (2004) 'The pathway for MHCII-mediated presentation of endogenous proteins involves peptide transport to the endo-lysosomal compartment', *Journal of Cell Science*, 117(Pt 18), pp. 4219–4230. Available at: <https://doi.org/10.1242/jcs.01288>.
- Day, C.L. *et al.* (2006) 'PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression', *Nature*, 443(7109), pp. 350–354. Available at: <https://doi.org/10.1038/nature05115>.
- De Plaen, E. *et al.* (1988) 'Immunogenic (tum-) variants of mouse tumor P815: cloning of the gene of tum- antigen P91A and identification of the tum- mutation', *Proceedings of the National Academy of Sciences of the United States of America*, 85(7), pp. 2274–2278. Available at: <https://doi.org/10.1073/pnas.85.7.2274>.
- Deeks, S.G. *et al.* (2012) 'Towards an HIV cure: a global scientific strategy', *Nature Reviews Immunology*, 12(8), pp. 607–614. Available at: <https://doi.org/10.1038/nri3262>.
- Deeks, S.G. and Walker, B.D. (2007) 'Human Immunodeficiency Virus Controllers: Mechanisms of Durable Virus Control in the Absence of Antiretroviral Therapy', *Immunity*, 27(3), pp. 406–416. Available at: <https://doi.org/10.1016/j.immuni.2007.08.010>.
- Deforges, J. *et al.* (2017) 'Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame', *Nucleic Acids Research*, 45(12), pp. 7382–7400. Available at: <https://doi.org/10.1093/nar/gkx303>.

- Delbos, F., Malard, S. and Congy, N. (2018) 'Human leukocyte antigen: genetics, structure and functions'.
- DeLong, T. *et al.* (2016) 'Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion', *Science (New York, N.Y.)*, 351(6274), pp. 711–714. Available at: <https://doi.org/10.1126/science.aad2791>.
- Deng, H., Gao, K. and Jankovic, J. (2014) 'The role of FUS gene variants in neurodegenerative diseases', *Nature Reviews. Neurology*, 10(6), pp. 337–348. Available at: <https://doi.org/10.1038/nrneurol.2014.78>.
- Dengjel, J. *et al.* (2005) 'Autophagy promotes MHC class II presentation of peptides from intracellular source proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 102(22), pp. 7922–7927. Available at: <https://doi.org/10.1073/pnas.0501190102>.
- Denton, P.W. *et al.* (2008) 'Antiretroviral pre-exposure prophylaxis prevents vaginal transmission of HIV-1 in humanized BLT mice', *PLoS medicine*, 5(1), p. e16. Available at: <https://doi.org/10.1371/journal.pmed.0050016>.
- Denzin, L.K. and Cresswell, P. (1995) 'HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading', *Cell*, 82(1), pp. 155–165. Available at: [https://doi.org/10.1016/0092-8674\(95\)90061-6](https://doi.org/10.1016/0092-8674(95)90061-6).
- Dersh, D., Hollý, J. and Yewdell, J.W. (2021) 'A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion', *Nature Reviews Immunology*, 21(2), pp. 116–128. Available at: <https://doi.org/10.1038/s41577-020-0390-6>.
- Dever, T.E., Dinman, J.D. and Green, R. (2018) 'Translation Elongation and Recoding in Eukaryotes', *Cold Spring Harbor Perspectives in Biology*, 10(8), p. a032649. Available at: <https://doi.org/10.1101/cshperspect.a032649>.
- Dever, T.E., Ivanov, I.P. and Sachs, M.S. (2020) 'Conserved uORF Nascent Peptides that Control Translation', *Annual review of genetics*, 54, pp. 237–264. Available at: <https://doi.org/10.1146/annurev-genet-112618-043822>.
- DiMaio, D. (2014) 'Viral miniproteins', *Annual Review of Microbiology*, 68, pp. 21–43. Available at: <https://doi.org/10.1146/annurev-micro-091313-103727>.
- Dinman, J.D. (2016) 'Pathways to Specialized Ribosomes: The Brussels Lecture', *Journal of Molecular Biology*, 428(10 Pt B), pp. 2186–2194. Available at: <https://doi.org/10.1016/j.jmb.2015.12.021>.
- Dinter, J. *et al.* (2015) 'Variable processing and cross-presentation of HIV by dendritic cells and macrophages shapes CTL immunodominance and immune escape', *PLoS pathogens*, 11(3), p. e1004725. Available at: <https://doi.org/10.1371/journal.ppat.1004725>.
- Dmitriev, S.E. *et al.* (2010) 'GTP-independent tRNA delivery to the ribosomal P-site by a novel eukaryotic translation factor', *The Journal of Biological Chemistry*, 285(35), pp. 26779–26787. Available at: <https://doi.org/10.1074/jbc.M110.119693>.
- Doitsh, G. *et al.* (2014) 'Pyroptosis drives CD4 T-cell depletion in HIV-1 infection', *Nature*, 505(7484), pp. 509–514. Available at: <https://doi.org/10.1038/nature12940>.
- Dolan, B.P. *et al.* (2010) 'RNA Polymerase II Inhibitors Dissociate Antigenic Peptide Generation from Normal Viral Protein Synthesis: A Role for Nuclear Translation in Defective Ribosomal Product Synthesis?', *The Journal of Immunology*, 185(11), pp. 6728–6733. Available at: <https://doi.org/10.4049/jimmunol.1002543>.
- Dolstra, H. *et al.* (1999) 'A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia', *The Journal of Experimental Medicine*, 189(2), pp. 301–308. Available at: <https://doi.org/10.1084/jem.189.2.301>.
- Dongre, A.R. *et al.* (2001) 'In vivo MHC class II presentation of cytosolic proteins revealed by rapid automated tandem mass spectrometry and functional analyses', *European Journal of Immunology*, 31(5), pp. 1485–1494. Available at: [https://doi.org/10.1002/1521-4141\(200105\)31:5<1485::AID-IMMU1485>3.0.CO;2-A](https://doi.org/10.1002/1521-4141(200105)31:5<1485::AID-IMMU1485>3.0.CO;2-A).

Du, K. *et al.* (2018) 'Recent advances in the discovery and development of TLR ligands as novel therapeutics for chronic HBV and HIV infections', *Expert Opinion on Drug Discovery*, 13(7), pp. 661–670. Available at: <https://doi.org/10.1080/17460441.2018.1473372>.

Dugré-Brisson, S. *et al.* (2005) 'Interaction of Staufen1 with the 5' end of mRNA facilitates translation of these RNAs', *Nucleic Acids Research*, 33(15), pp. 4797–4812. Available at: <https://doi.org/10.1093/nar/gki794>.

Dulude, D., Baril, M. and Brakier-Gingras, L. (2002) 'Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1', *Nucleic Acids Research*, 30(23), pp. 5094–5102. Available at: <https://doi.org/10.1093/nar/gkf657>.

Dunn, P. (2003) 'Gregor Mendel, OSA (1822–1884), founder of scientific genetics', *Archives of Disease in Childhood Fetal and Neonatal Edition*, 88(6), pp. F537–F539. Available at: <https://doi.org/10.1136/fn.88.6.F537>.

Dyer, W.B. *et al.* (2008) 'Mechanisms of HIV non-progression; robust and sustained CD4+ T-cell proliferative responses to p24 antigen correlate with control of viraemia and lack of disease progression after long-term transfusion-acquired HIV-1 infection', *Retrovirology*, 5, p. 112. Available at: <https://doi.org/10.1186/1742-4690-5-112>.

Eiz-Vesper, B. and Schmetzer, H.M. (2020) 'Antigen-Presenting Cells: Potential of Proven und New Players in Immune Therapies', *Transfusion Medicine and Hemotherapy*, 47(6), pp. 429–431. Available at: <https://doi.org/10.1159/000512729>.

Elliott, T., Bodmer, H. and Townsend, A. (1996) 'Recognition of out-of-frame major histocompatibility complex class I-restricted epitopes in vivo', *European Journal of Immunology*, 26(5), pp. 1175–1179. Available at: <https://doi.org/10.1002/eji.1830260532>.

van Els, C.A. *et al.* (2000) 'A single naturally processed measles virus peptide fully dominates the HLA-A*0201-associated peptide display and is mutated at its anchor position in persistent viral strains', *European Journal of Immunology*, 30(4), pp. 1172–1181. Available at: [https://doi.org/10.1002/\(SICI\)1521-4141\(200004\)30:4<1172::AID-IMMU1172>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1521-4141(200004)30:4<1172::AID-IMMU1172>3.0.CO;2-J).

ENCODE Project Consortium (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. Available at: <https://doi.org/10.1038/nature11247>.

van Endert, P.M. *et al.* (1994) 'A sequential model for peptide binding and transport by the transporters associated with antigen processing', *Immunity*, 1(6), pp. 491–500. Available at: [https://doi.org/10.1016/1074-7613\(94\)90091-4](https://doi.org/10.1016/1074-7613(94)90091-4).

Engelhard, V.H. *et al.* (1993) 'Mass spectrometric analysis of peptides associated with the human class I MHC molecules HLA-A2.1 and HLA-B7 and identification of structural features that determine binding', *Chemical Immunology*, 57, pp. 39–62.

Engelhard, V.H., Brickner, A.G. and Zarlign, A.L. (2002) 'Insights into antigen processing gained by direct analysis of the naturally processed class I MHC associated peptide repertoire', *Molecular Immunology*, 39(3–4), pp. 127–137. Available at: [https://doi.org/10.1016/s0161-5890\(02\)00096-2](https://doi.org/10.1016/s0161-5890(02)00096-2).

Engelman, A., Mizuuchi, K. and Craigie, R. (1991) 'HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer', *Cell*, 67(6), pp. 1211–1221. Available at: [https://doi.org/10.1016/0092-8674\(91\)90297-c](https://doi.org/10.1016/0092-8674(91)90297-c).

Engelman, A.N. and Singh, P.K. (2018) 'Cellular and molecular mechanisms of HIV-1 integration targeting', *Cellular and molecular life sciences : CMLS*, 75(14), pp. 2491–2507. Available at: <https://doi.org/10.1007/s00018-018-2772-5>.

Erhard, F. *et al.* (2018) 'Improved Ribo-seq enables identification of cryptic translation events', *Nature Methods*, 15(5), pp. 363–366. Available at: <https://doi.org/10.1038/nmeth.4631>.

Erhard, F. *et al.* (2020) 'Identification of the Cryptic HLA-I Immunopeptidome', *Cancer Immunology Research*, 8(8), pp. 1018–1026. Available at: <https://doi.org/10.1158/2326-6066.CIR-19-0886>.

- Erickson, F.L. and Hannig, E.M. (1996) 'Ligand interactions with eukaryotic translation initiation factor 2: role of the gamma-subunit', *The EMBO journal*, 15(22), pp. 6311–6320.
- Espert, L. *et al.* (2006) 'Autophagy is involved in T cell death after binding of HIV-1 envelope proteins to CXCR4', *The Journal of Clinical Investigation*, 116(8), pp. 2161–2172. Available at: <https://doi.org/10.1172/JCI26185>.
- Espert, L. and Biard-Piechaczyk, M. (2009) 'Autophagy in HIV-induced T cell death', *Current Topics in Microbiology and Immunology*, 335, pp. 307–321. Available at: https://doi.org/10.1007/978-3-642-00302-8_15.
- Esquivel, F., Yewdell, J. and Bennink, J. (1992) 'RMA/S cells present endogenously synthesized cytosolic proteins to class I-restricted cytotoxic T lymphocytes.', *Journal of Experimental Medicine*, 175(1), pp. 163–168. Available at: <https://doi.org/10.1084/jem.175.1.163>.
- Eyerich, S. *et al.* (2009) 'Th22 cells represent a distinct human T cell subset involved in epidermal immunity and remodeling', *The Journal of Clinical Investigation*, 119(12), pp. 3573–3585. Available at: <https://doi.org/10.1172/JCI40202>.
- Falk, K. *et al.* (1991) 'Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules', *Nature*, 351(6324), pp. 290–296. Available at: <https://doi.org/10.1038/351290a0>.
- Fang, P. *et al.* (2004) 'A nascent polypeptide domain that can regulate translation elongation', *Proceedings of the National Academy of Sciences*, 101(12), pp. 4059–4064. Available at: <https://doi.org/10.1073/pnas.0400554101>.
- Faridi, P. *et al.* (2018) 'A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands', *Science Immunology*, 3(28), p. eaar3947. Available at: <https://doi.org/10.1126/sciimmunol.aar3947>.
- Fernandes, J.D., Booth, D.S. and Frankel, A.D. (2016) 'A structurally plastic ribonucleoprotein complex mediates post-transcriptional gene regulation in HIV-1', *Wiley interdisciplinary reviews. RNA*, 7(4), pp. 470–486. Available at: <https://doi.org/10.1002/wrna.1342>.
- Ferrari, G. *et al.* (2017) 'Humoral and Innate Antiviral Immunity as Tools to Clear Persistent HIV Infection', *The Journal of Infectious Diseases*, 215(Suppl 3), pp. S152–S159. Available at: <https://doi.org/10.1093/infdis/jiw555>.
- Ferre, A.L. *et al.* (2009) 'Mucosal immune responses to HIV-1 in elite controllers: a potential correlate of immune control', *Blood*, 113(17), pp. 3978–3989. Available at: <https://doi.org/10.1182/blood-2008-10-182709>.
- Ferre, A.L. *et al.* (2010) 'HIV controllers with HLA-DRB1*13 and HLA-DQB1*06 alleles have strong, polyfunctional mucosal CD4+ T-cell responses', *Journal of Virology*, 84(21), pp. 11020–11029. Available at: <https://doi.org/10.1128/JVI.00980-10>.
- Ferrington, D.A. and Gregerson, D.S. (2012) 'Immunoproteasomes: Structure, Function, and Antigen Presentation', *Progress in molecular biology and translational science*, 109, pp. 75–112. Available at: <https://doi.org/10.1016/B978-0-12-397863-9.00003-1>.
- Fetten, J.V., Roy, N. and Gilboa, E. (1991) 'A frameshift mutation at the NH2 terminus of the nucleoprotein gene does not affect generation of cytotoxic T lymphocyte epitopes', *Journal of Immunology (Baltimore, Md.: 1950)*, 147(8), pp. 2697–2705.
- Fijałkowska, D. *et al.* (2017) 'eIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs', *Nucleic Acids Research*, 45(13), pp. 7997–8013. Available at: <https://doi.org/10.1093/nar/gkx469>.
- Finkel, T.H. *et al.* (1995) 'Apoptosis occurs predominantly in bystander cells and not in productively infected cells of HIV- and SIV-infected lymph nodes', *Nature Medicine*, 1(2), pp. 129–134. Available at: <https://doi.org/10.1038/nm0295-129>.
- Finkel, Y. *et al.* (2021) 'The coding capacity of SARS-CoV-2', *Nature*, 589(7840), pp. 125–130. Available at: <https://doi.org/10.1038/s41586-020-2739-1>.

- Finkel, Y., Stern-Ginossar, N. and Schwartz, M. (2018) 'Viral Short ORFs and Their Possible Functions', *Proteomics*, 18(10), p. e1700255. Available at: <https://doi.org/10.1002/pmic.201700255>.
- Firth, A.E. and Brierley, I. (2012) 'Non-canonical translation in RNA viruses', *Journal of General Virology*, 93(7), pp. 1385–1409. Available at: <https://doi.org/10.1099/vir.0.042499-0>.
- Fischer, W. *et al.* (2010) 'Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing', *PLoS One*, 5(8), p. e12303. Available at: <https://doi.org/10.1371/journal.pone.0012303>.
- Fletcher, K. *et al.* (2018) 'The WD40 domain of ATG16L1 is required for its non-canonical role in lipidation of LC3 at single membranes', *The EMBO journal*, 37(4), p. e97840. Available at: <https://doi.org/10.15252/embj.201797840>.
- Fouillot, N. *et al.* (1993) 'Translation of the hepatitis B virus P gene by ribosomal scanning as an alternative to internal initiation', *Journal of Virology*, 67(8), pp. 4886–4895. Available at: <https://doi.org/10.1128/JVI.67.8.4886-4895.1993>.
- Frank, J. (2003) 'Toward an understanding of the structural basis of translation', *Genome biology*, 4(12). Available at: <https://doi.org/10.1186/gb-2003-4-12-237>.
- Frank, J. and Agrawal, R.K. (2000) 'A ratchet-like inter-subunit reorganization of the ribosome during translocation', *Nature*, 406(6793), pp. 318–322. Available at: <https://doi.org/10.1038/35018597>.
- Frankel, A.D. and Young, J.A. (1998) 'HIV-1: fifteen proteins and an RNA', *Annual Review of Biochemistry*, 67, pp. 1–25. Available at: <https://doi.org/10.1146/annurev.biochem.67.1.1>.
- Fraser, C.S. (2015) 'Quantitative studies of mRNA recruitment to the eukaryotic ribosome', *Biochimie*, 114, pp. 58–71. Available at: <https://doi.org/10.1016/j.biochi.2015.02.017>.
- Freed, E.O. (2015) 'HIV-1 assembly, release and maturation', *Nature reviews. Microbiology*, 13(8), pp. 484–496. Available at: <https://doi.org/10.1038/nrmicro3490>.
- Fresno, M., Jiménez, A. and Vázquez, D. (1977) 'Inhibition of translation in eukaryotic systems by harringtonine', *European Journal of Biochemistry*, 72(2), pp. 323–330. Available at: <https://doi.org/10.1111/j.1432-1033.1977.tb11256.x>.
- Fukazawa, Y. *et al.* (2015) 'B cell follicle sanctuary permits persistent productive simian immunodeficiency virus infection in elite controllers', *Nature Medicine*, 21(2), pp. 132–139. Available at: <https://doi.org/10.1038/nm.3781>.
- Gagnon, M. *et al.* (2021) 'Potentiation of B2 receptor signaling by AltB2R, a newly identified alternative protein encoded in the human bradykinin B2 receptor gene', *Journal of Biological Chemistry*, 296. Available at: <https://doi.org/10.1016/j.jbc.2021.100329>.
- Galperin, M. *et al.* (2018) 'CD4+ T cell-mediated HLA class II cross-restriction in HIV controllers', *Science Immunology*, 3(24), p. eaat0687. Available at: <https://doi.org/10.1126/sciimmunol.aat0687>.
- García-Lerma, J.G. *et al.* (2008) 'Prevention of rectal SHIV transmission in macaques by daily or intermittent prophylaxis with emtricitabine and tenofovir', *PLoS medicine*, 5(2), p. e28. Available at: <https://doi.org/10.1371/journal.pmed.0050028>.
- Garrido, C. *et al.* (2018) 'Interleukin-15-Stimulated Natural Killer Cells Clear HIV-1-Infected Cells following Latency Reversal Ex Vivo', *Journal of Virology*, 92(12), pp. e00235-18. Available at: <https://doi.org/10.1128/JVI.00235-18>.
- Garrison, K.E. *et al.* (2009) 'Transcriptional errors in human immunodeficiency virus type 1 generate targets for T-cell responses', *Clinical and vaccine immunology: CVI*, 16(9), pp. 1369–1371. Available at: <https://doi.org/10.1128/CI.00410-08>.

- Gaudin, C. *et al.* (2005) 'Structure of the RNA signal essential for translational frameshifting in HIV-1', *Journal of Molecular Biology*, 349(5), pp. 1024–1035. Available at: <https://doi.org/10.1016/j.jmb.2005.04.045>.
- Geissler, R., Golbik, R.P. and Behrens, S.-E. (2012) 'The DEAD-box helicase DDX3 supports the assembly of functional 80S ribosomes', *Nucleic Acids Research*, 40(11), pp. 4998–5011. Available at: <https://doi.org/10.1093/nar/gks070>.
- Gendron, K. *et al.* (2011) 'The activity of the HIV-1 IRES is stimulated by oxidative stress and controlled by a negative regulatory element', *Nucleic Acids Research*, 39(3), pp. 902–912. Available at: <https://doi.org/10.1093/nar/gkq885>.
- Giodini, A. and Cresswell, P. (2008) 'Hsp90-mediated cytosolic refolding of exogenous proteins internalized by dendritic cells', *The EMBO journal*, 27(1), pp. 201–211. Available at: <https://doi.org/10.1038/sj.emboj.7601941>.
- Glickman, M.H. and Ciechanover, A. (2002) 'The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction', *Physiological Reviews*, 82(2), pp. 373–428. Available at: <https://doi.org/10.1152/physrev.00027.2001>.
- Godot, V. *et al.* (2020) 'TLR-9 agonist and CD40-targeting vaccination induces HIV-1 envelope-specific B cells with a diversified immunoglobulin repertoire in humanized mice', *PLoS pathogens*, 16(11), p. e1009025. Available at: <https://doi.org/10.1371/journal.ppat.1009025>.
- Gonzalez, C. *et al.* (2014) 'Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors', *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(33), pp. 10924–10936. Available at: <https://doi.org/10.1523/JNEUROSCI.0084-14.2014>.
- Gonzalez-Galarza, F.F. *et al.* (2020) 'Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools', *Nucleic Acids Research*, 48(D1), pp. D783–D788. Available at: <https://doi.org/10.1093/nar/gkz1029>.
- Goonetilleke, N. *et al.* (2009) 'The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection', *The Journal of Experimental Medicine*, 206(6), pp. 1253–1272. Available at: <https://doi.org/10.1084/jem.20090365>.
- Goulder, P. *et al.* (1997) 'Co-evolution of human immunodeficiency virus and cytotoxic T-lymphocyte responses', *Immunological Reviews*, 159(1), pp. 17–29. Available at: <https://doi.org/10.1111/j.1600-065X.1997.tb01004.x>.
- Grabar, S. *et al.* (2009) 'Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French Hospital Database on HIV', *AIDS (London, England)*, 23(9), pp. 1163–1169. Available at: <https://doi.org/10.1097/QAD.0b013e32832b44c8>.
- Grabmeier-Pfistershammer, K. *et al.* (2011) 'Identification of PD-1 as a unique marker for failing immune reconstitution in HIV-1-infected patients on treatment', *Journal of Acquired Immune Deficiency Syndromes (1999)*, 56(2), pp. 118–124. Available at: <https://doi.org/10.1097/QAI.0b013e3181fbab9f>.
- Gray, C.M. *et al.* (1999) 'Frequency of class I HLA-restricted anti-HIV CD8+ T cells in individuals receiving highly active antiretroviral therapy (HAART)', *Journal of Immunology (Baltimore, Md.: 1950)*, 162(3), pp. 1780–1788.
- Gray, E.S. *et al.* (2011) 'The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection', *Journal of Virology*, 85(10), pp. 4828–4840. Available at: <https://doi.org/10.1128/JVI.00198-11>.
- Griffin, J.P., Chu, R. and Harding, C.V. (1997) 'Early endosomes and a late endocytic compartment generate different peptide-class II MHC complexes via distinct processing mechanisms', *Journal of Immunology (Baltimore, Md.: 1950)*, 158(4), pp. 1523–1532.
- Grivel, J.C., Malkevitch, N. and Margolis, L. (2000) 'Human immunodeficiency virus type 1 induces apoptosis in CD4(+) but not in CD8(+) T cells in ex vivo-infected human lymphoid tissue', *Journal of Virology*, 74(17), pp. 8077–8084. Available at: <https://doi.org/10.1128/jvi.74.17.8077-8084.2000>.

- Gros, M. and Amigorena, S. (2019) 'Regulation of Antigen Export to the Cytosol During Cross-Presentation', *Frontiers in Immunology*, 10. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00041> (Accessed: 6 March 2023).
- Guce, A.I. *et al.* (2013) 'HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism', *Nature Structural & Molecular Biology*, 20(1), pp. 90–98. Available at: <https://doi.org/10.1038/nsmb.2460>.
- Guenther, U.-P. *et al.* (2018) 'The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs', *Nature*, 559(7712), pp. 130–134. Available at: <https://doi.org/10.1038/s41586-018-0258-0>.
- Guerrero, S. *et al.* (2015) 'HIV-1 Replication and the Cellular Eukaryotic Translation Apparatus', *Viruses*, 7(1), pp. 199–218. Available at: <https://doi.org/10.3390/v7010199>.
- Guydosh, N.R. and Green, R. (2014) 'Dom34 rescues ribosomes in 3' untranslated regions', *Cell*, 156(5), pp. 950–962. Available at: <https://doi.org/10.1016/j.cell.2014.02.006>.
- Hahn, B.H. *et al.* (1984) 'Molecular cloning and characterization of the HTLV-III virus associated with AIDS', *Nature*, 312(5990), pp. 166–169. Available at: <https://doi.org/10.1038/312166a0>.
- Hallenberger, S. *et al.* (1992) 'Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160', *Nature*, 360(6402), pp. 358–361. Available at: <https://doi.org/10.1038/360358a0>.
- Hammer, G.E. *et al.* (2007) 'In the absence of aminopeptidase ERAAP, MHC class I molecules present many unstable and highly immunogenic peptides', *Nature Immunology*, 8(1), pp. 101–108. Available at: <https://doi.org/10.1038/ni1409>.
- Hamoudi, M. *et al.* (2013) 'Genetic diversity of the highly variable V1 region interferes with Human Immunodeficiency Virus type 1 envelope functionality', *Retrovirology*, 10(1), p. 114. Available at: <https://doi.org/10.1186/1742-4690-10-114>.
- Hanada, K.-I., Yewdell, J.W. and Yang, J.C. (2004) 'Immune recognition of a human renal cancer antigen through post-translational protein splicing', *Nature*, 427(6971), pp. 252–256. Available at: <https://doi.org/10.1038/nature02240>.
- Hann, S.R. *et al.* (1988) 'A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas', *Cell*, 52(2), pp. 185–195. Available at: [https://doi.org/10.1016/0092-8674\(88\)90507-7](https://doi.org/10.1016/0092-8674(88)90507-7).
- Hanson, G. and Collier, J. (2018) 'Codon optimality, bias and usage in translation and mRNA decay', *Nature reviews. Molecular cell biology*, 19(1), pp. 20–30. Available at: <https://doi.org/10.1038/nrm.2017.91>.
- Harari, A. *et al.* (2004) 'Skewed representation of functionally distinct populations of virus-specific CD4 T cells in HIV-1-infected subjects with progressive disease: changes after antiretroviral therapy', *Blood*, 103(3), pp. 966–972. Available at: <https://doi.org/10.1182/blood-2003-04-1203>.
- Hartman, I.Z. *et al.* (2010) 'A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes', *Nature Medicine*, 16(11), pp. 1333–1340. Available at: <https://doi.org/10.1038/nm.2248>.
- Hartman, T.R. *et al.* (2006) 'RNA helicase A is necessary for translation of selected messenger RNAs', *Nature Structural & Molecular Biology*, 13(6), pp. 509–516. Available at: <https://doi.org/10.1038/nsmb1092>.
- Hassan, C. *et al.* (2014) 'Accurate quantitation of MHC-bound peptides by application of isotopically labeled peptide MHC complexes', *Journal of Proteomics*, 109, pp. 240–244. Available at: <https://doi.org/10.1016/j.jprot.2014.07.009>.
- Havenar-Daughton, C. *et al.* (2016) 'Direct Probing of Germinal Center Responses Reveals Immunological Features and Bottlenecks for Neutralizing Antibody Responses to HIV Env Trimer', *Cell Reports*, 17(9), pp. 2195–2209. Available at: <https://doi.org/10.1016/j.celrep.2016.10.085>.

- Haynes, B.F. *et al.* (2012) 'Immune-correlates analysis of an HIV-1 vaccine efficacy trial', *The New England Journal of Medicine*, 366(14), pp. 1275–1286. Available at: <https://doi.org/10.1056/NEJMoa1113425>.
- Haynes, B.F. *et al.* (2023) 'Strategies for HIV-1 vaccines that induce broadly neutralizing antibodies', *Nature Reviews. Immunology*, 23(3), pp. 142–158. Available at: <https://doi.org/10.1038/s41577-022-00753-w>.
- Hecht, A. *et al.* (2017) 'Measurements of translation initiation from all 64 codons in *E. coli*', *Nucleic Acids Research*, 45(7), pp. 3615–3626. Available at: <https://doi.org/10.1093/nar/gkx070>.
- Heinkelein, M. *et al.* (1996) 'Lysis of human immunodeficiency virus type 1 antigen-expressing cells by CD4 and CD8 T cells *ex vivo*', *The Journal of Infectious Diseases*, 174(1), pp. 209–213. Available at: <https://doi.org/10.1093/infdis/174.1.209>.
- Hellen, C.U.T. (2018a) 'Translation Termination and Ribosome Recycling in Eukaryotes', *Cold Spring Harbor Perspectives in Biology*, 10(10), p. a032656. Available at: <https://doi.org/10.1101/cshperspect.a032656>.
- Hellen, C.U.T. (2018b) 'Translation Termination and Ribosome Recycling in Eukaryotes', *Cold Spring Harbor Perspectives in Biology*, 10(10), p. a032656. Available at: <https://doi.org/10.1101/cshperspect.a032656>.
- Herr, W. *et al.* (1999) 'Identification of naturally processed and HLA-presented Epstein–Barr virus peptides recognized by CD4+ or CD8+ T lymphocytes from human blood', *Proceedings of the National Academy of Sciences*, 96(21), pp. 12033–12038. Available at: <https://doi.org/10.1073/pnas.96.21.12033>.
- Hershko, A. *et al.* (1980) 'Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis', *Proceedings of the National Academy of Sciences of the United States of America*, 77(4), pp. 1783–1786. Available at: <https://doi.org/10.1073/pnas.77.4.1783>.
- Hersperger, A.R. *et al.* (2010) 'Perforin expression directly *ex vivo* by HIV-specific CD8 T-cells is a correlate of HIV elite control', *PLoS pathogens*, 6(5), p. e1000917. Available at: <https://doi.org/10.1371/journal.ppat.1000917>.
- van Heuvel, Y. *et al.* (2022) 'Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine', *Toxins*, 14(2), p. 138. Available at: <https://doi.org/10.3390/toxins14020138>.
- Hickman, H.D. *et al.* (2003) 'Cutting Edge: Class I Presentation of Host Peptides Following HIV Infection¹', *The Journal of Immunology*, 171(1), pp. 22–26. Available at: <https://doi.org/10.4049/jimmunol.171.1.22>.
- Hijazi, K. *et al.* (2011) 'DC-SIGN increases the affinity of HIV-1 envelope glycoprotein interaction with CD4', *PLoS One*, 6(12), p. e28307. Available at: <https://doi.org/10.1371/journal.pone.0028307>.
- Hilliker, A. *et al.* (2011) 'The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex', *Molecular Cell*, 43(6), pp. 962–972. Available at: <https://doi.org/10.1016/j.molcel.2011.08.008>.
- Hinnebusch, A.G. (2014) 'The scanning mechanism of eukaryotic translation initiation', *Annual Review of Biochemistry*, 83, pp. 779–812. Available at: <https://doi.org/10.1146/annurev-biochem-060713-035802>.
- Hladik, F. and McElrath, M.J. (2008) 'Setting the stage: host invasion by HIV', *Nature Reviews. Immunology*, 8(6), pp. 447–457. Available at: <https://doi.org/10.1038/nri2302>.
- Hoof, I. *et al.* (2012) 'Proteome sampling by the HLA class I antigen processing pathway', *PLoS computational biology*, 8(5), p. e1002517. Available at: <https://doi.org/10.1371/journal.pcbi.1002517>.
- Hraber, P. *et al.* (2014) 'Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection', *AIDS (London, England)*, 28(2), pp. 163–169. Available at: <https://doi.org/10.1097/QAD.000000000000106>.
- Hu, W.-S. and Hughes, S.H. (2012) 'HIV-1 Reverse Transcription', *Cold Spring Harbor Perspectives in Medicine*, 2(10), p. a006882. Available at: <https://doi.org/10.1101/cshperspect.a006882>.

- Hu, Z. *et al.* (2021) 'Personal Neoantigen Vaccines Induce Persistent Memory T Cell Responses and Epitope Spreading in Patients with Melanoma', *Nature medicine*, 27(3), pp. 515–525. Available at: <https://doi.org/10.1038/s41591-020-01206-4>.
- Hühn, P. *et al.* (1997) 'Characterization of the autoantigen La (SS-B) as a dsRNA unwinding enzyme', *Nucleic Acids Research*, 25(2), pp. 410–416. Available at: <https://doi.org/10.1093/nar/25.2.410>.
- Hunt, D.F. *et al.* (1992) 'Characterization of Peptides Bound to the Class I MHC Molecule HLA-A2.1 by Mass Spectrometry', *Science*, 255(5049), pp. 1261–1263. Available at: <https://doi.org/10.1126/science.1546328>.
- Iborra, F.J., Jackson, D.A. and Cook, P.R. (2004) 'The case for nuclear translation', *Journal of Cell Science*, 117(Pt 24), pp. 5713–5720. Available at: <https://doi.org/10.1242/jcs.01538>.
- Iglesias, M.C. *et al.* (2011) 'Escape from highly effective public CD8+ T-cell clonotypes by HIV', *Blood*, 118(8), pp. 2138–2149. Available at: <https://doi.org/10.1182/blood-2011-01-328781>.
- Imai, T. *et al.* (2011) 'Heat shock protein 90 (HSP90) contributes to cytosolic translocation of extracellular antigen for cross-presentation by dendritic cells', *Proceedings of the National Academy of Sciences of the United States of America*, 108(39), pp. 16363–16368. Available at: <https://doi.org/10.1073/pnas.1108372108>.
- 'IN DANGER: UNAIDS Global AIDS Update 2022' (2022).
- Ingolia, N.T. *et al.* (2009a) 'Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling', *Science (New York, N.Y.)*, 324(5924), pp. 218–223. Available at: <https://doi.org/10.1126/science.1168978>.
- Ingolia, N.T. *et al.* (2009b) 'Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling', *Science (New York, N.Y.)*, 324(5924), pp. 218–223. Available at: <https://doi.org/10.1126/science.1168978>.
- Ingolia, N.T. *et al.* (2012) 'The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments', *Nature Protocols*, 7(8), pp. 1534–1550. Available at: <https://doi.org/10.1038/nprot.2012.086>.
- Ingolia, N.T. *et al.* (2014) 'Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes', *Cell Reports*, 8(5), pp. 1365–1379. Available at: <https://doi.org/10.1016/j.celrep.2014.07.045>.
- Ingolia, N.T., Hussmann, J.A. and Weissman, J.S. (2019) 'Ribosome Profiling: Global Views of Translation', *Cold Spring Harbor Perspectives in Biology*, 11(5), p. a032698. Available at: <https://doi.org/10.1101/cshperspect.a032698>.
- Ingolia, N.T., Lareau, L.F. and Weissman, J.S. (2011) 'Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes', *Cell*, 147(4), pp. 789–802. Available at: <https://doi.org/10.1016/j.cell.2011.10.002>.
- Irigoyen, N. *et al.* (2018) 'Ribosome profiling of the retrovirus murine leukemia virus', *Retrovirology*, 15(1), p. 10. Available at: <https://doi.org/10.1186/s12977-018-0394-5>.
- Ivanov, I.P. *et al.* (2010) 'Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1)', *Proceedings of the National Academy of Sciences of the United States of America*, 107(42), pp. 18056–18060. Available at: <https://doi.org/10.1073/pnas.1009269107>.
- Jacks, T. *et al.* (1988) 'Characterization of ribosomal frameshifting in HIV-1 gag-pol expression', *Nature*, 331(6153), pp. 280–283. Available at: <https://doi.org/10.1038/331280a0>.
- Jacobson, S. *et al.* (1988) 'Recognition of intracellular measles virus antigens by HLA class II restricted measles virus-specific cytotoxic T lymphocytes', *Annals of the New York Academy of Sciences*, 540, pp. 352–353. Available at: <https://doi.org/10.1111/j.1749-6632.1988.tb27096.x>.

- Jäger, S. *et al.* (2011) 'Global landscape of HIV-human protein complexes', *Nature*, 481(7381), pp. 365–370. Available at: <https://doi.org/10.1038/nature10719>.
- Jan, C.H., Williams, C.C. and Weissman, J.S. (2014) 'Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling', *Science*, 346(6210), p. 1257521. Available at: <https://doi.org/10.1126/science.1257521>.
- Jang, S.K. *et al.* (1988) 'A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation', *Journal of Virology*, 62(8), pp. 2636–2643. Available at: <https://doi.org/10.1128/JVI.62.8.2636-2643.1988>.
- Janich, P. *et al.* (2015) 'Ribosome profiling reveals the rhythmic liver transcriptome and circadian clock regulation by upstream open reading frames', *Genome Research*, 25(12), pp. 1848–1859. Available at: <https://doi.org/10.1101/gr.195404.115>.
- Janssen, E.M. *et al.* (2003) 'CD4+ T cells are required for secondary expansion and memory in CD8+ T lymphocytes', *Nature*, 421(6925), pp. 852–856. Available at: <https://doi.org/10.1038/nature01441>.
- Jean-Jean, O. *et al.* (1989) 'Expression mechanism of the hepatitis B virus (HBV) C gene and biosynthesis of HBe antigen', *Virology*, 170(1), pp. 99–106. Available at: [https://doi.org/10.1016/0042-6822\(89\)90356-5](https://doi.org/10.1016/0042-6822(89)90356-5).
- Jean-Jean, O., Cassan, M. and Rousset, J.-P. (1993) 'L'initiation de la traduction chez les eucaryotes, source de diversification et de modulation de l'expression des gènes'.
- Jeger-Madiot, R., Heredia, M. and Graff-Dubois, S. (2019) 'Germinal centers B-cell reaction and T follicular helper cells in response to HIV-1 infection', *Current opinion in HIV and AIDS*, 14(4), pp. 246–252. Available at: <https://doi.org/10.1097/COH.0000000000000557>.
- Jekle, A. *et al.* (2003) 'In vivo evolution of human immunodeficiency virus type 1 toward increased pathogenicity through CXCR4-mediated killing of uninfected CD4 T cells', *Journal of Virology*, 77(10), pp. 5846–5854. Available at: <https://doi.org/10.1128/jvi.77.10.5846-5854.2003>.
- Ji, Z. *et al.* (2015) 'Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins', *eLife*, 4, p. e08890. Available at: <https://doi.org/10.7554/eLife.08890>.
- Jiang, T. *et al.* (2019) 'Tumor neoantigens: from basic research to clinical applications', *Journal of Hematology & Oncology*, 12(1), p. 93. Available at: <https://doi.org/10.1186/s13045-019-0787-5>.
- Joag, V.R. *et al.* (2016) 'Identification of preferential CD4+ T-cell targets for HIV infection in the cervix', *Mucosal Immunology*, 9(1), pp. 1–12. Available at: <https://doi.org/10.1038/mi.2015.28>.
- Johannsen, W. (1909) *Elemente der exakten erblichkeitslehre.: Deutsche wesentlich erweiterte ausgabe in fünfundzwanzig vorlesungen*. Jena: G. Fischer.
- Johnson, S. *et al.* (2015) 'Cooperativity of HIV-Specific Cytolytic CD4 T Cells and CD8 T Cells in Control of HIV Viremia', *Journal of Virology*, 89(15), pp. 7494–7505. Available at: <https://doi.org/10.1128/JVI.00438-15>.
- Jones, R.B. *et al.* (2008) 'Tim-3 expression defines a novel population of dysfunctional T cells with highly elevated frequencies in progressive HIV-1 infection', *The Journal of Experimental Medicine*, 205(12), pp. 2763–2779. Available at: <https://doi.org/10.1084/jem.20081398>.
- Jones, R.B. and Walker, B.D. (2016) 'HIV-specific CD8+ T cells and HIV eradication', *The Journal of Clinical Investigation*, 126(2), pp. 455–463. Available at: <https://doi.org/10.1172/JCI80566>.
- Julg, B. and Barouch, D. (2021) 'Broadly neutralizing antibodies for HIV-1 prevention and therapy', *Seminars in Immunology*, 51, p. 101475. Available at: <https://doi.org/10.1016/j.smim.2021.101475>.
- Juno, J.A. *et al.* (2017) 'Cytotoxic CD4 T Cells—Friend or Foe during Viral Infection?', *Frontiers in Immunology*, 8. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2017.00019> (Accessed: 10 April 2023).

- Jürgens, L. and Wethmar, K. (2022) 'The Emerging Role of uORF-Encoded uPeptides and HLA uLigands in Cellular and Tumor Biology', *Cancers*, 14(24), p. 6031. Available at: <https://doi.org/10.3390/cancers14246031>.
- Kalams, S.A. et al. (1999) 'Levels of human immunodeficiency virus type 1-specific cytotoxic T-lymphocyte effector and memory responses decline after suppression of viremia with highly active antiretroviral therapy', *Journal of Virology*, 73(8), pp. 6721–6728. Available at: <https://doi.org/10.1128/JVI.73.8.6721-6728.1999>.
- Kapp, L.D. and Lorsch, J.R. (2004) 'GTP-dependent recognition of the methionine moiety on initiator tRNA by translation factor eIF2', *Journal of Molecular Biology*, 335(4), pp. 923–936. Available at: <https://doi.org/10.1016/j.jmb.2003.11.025>.
- Karacostas, V. et al. (1993) 'Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles', *Virology*, 193(2), pp. 661–671. Available at: <https://doi.org/10.1006/viro.1993.1174>.
- Kaufmann, D.E. et al. (2007) 'Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction', *Nature Immunology*, 8(11), pp. 1246–1254. Available at: <https://doi.org/10.1038/ni1515>.
- Kearse, M.G. et al. (2019) 'Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors', *Genes & Development*, 33(13–14), pp. 871–885. Available at: <https://doi.org/10.1101/gad.324715.119>.
- Kearse, M.G. and Wilusz, J.E. (2017) 'Non-AUG translation: a new start for protein synthesis in eukaryotes', *Genes & Development*, 31(17), pp. 1717–1731. Available at: <https://doi.org/10.1101/gad.305250.117>.
- Kervevan, J. and Chakrabarti, L.A. (2021) 'Role of CD4+ T Cells in the Control of Viral Infections: Recent Advances and Open Questions', *International Journal of Molecular Sciences*, 22(2), p. 523. Available at: <https://doi.org/10.3390/ijms22020523>.
- Ketteler, R. (2012) 'On programmed ribosomal frameshifting: the alternative proteomes', *Frontiers in Genetics*, 3, p. 242. Available at: <https://doi.org/10.3389/fgene.2012.00242>.
- Khatter, H. et al. (2015) 'Structure of the human 80S ribosome', *Nature*, 520(7549), pp. 640–645. Available at: <https://doi.org/10.1038/nature14427>.
- Khurana, E. et al. (2016) 'Role of non-coding sequence variants in cancer', *Nature Reviews. Genetics*, 17(2), pp. 93–108. Available at: <https://doi.org/10.1038/nrg.2015.17>.
- Kim, J. et al. (2021) 'Current approaches to HIV vaccine development: a narrative review', *Journal of the International AIDS Society*, 24 Suppl 7(Suppl 7), p. e25793. Available at: <https://doi.org/10.1002/jia2.25793>.
- Kim, S.Y. et al. (1989) 'Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression', *Journal of Virology*, 63(9), pp. 3708–3713. Available at: <https://doi.org/10.1128/JVI.63.9.3708-3713.1989>.
- Klatzmann, D., Barré-Sinoussi, F., et al. (1984) 'Selective tropism of lymphadenopathy associated virus (LAV) for helper-inducer T lymphocytes', *Science (New York, N.Y.)*, 225(4657), pp. 59–63. Available at: <https://doi.org/10.1126/science.6328660>.
- Klatzmann, D., Champagne, E., et al. (1984) 'T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV', *Nature*, 312(5996), pp. 767–768. Available at: <https://doi.org/10.1038/312767a0>.
- Klein, F. et al. (2013) 'Antibodies in HIV-1 vaccine development and therapy', *Science (New York, N.Y.)*, 341(6151), pp. 1199–1204. Available at: <https://doi.org/10.1126/science.1241144>.
- Kniepert, A. and Groettrup, M. (2014) 'The unique functions of tissue-specific proteasomes', *Trends in Biochemical Sciences*, 39(1), pp. 17–24. Available at: <https://doi.org/10.1016/j.tibs.2013.10.004>.

- Komar, A.A. (2016) 'The Yin and Yang of codon usage', *Human Molecular Genetics*, 25(R2), pp. R77–R85. Available at: <https://doi.org/10.1093/hmg/ddw207>.
- Koopmann, J.O. *et al.* (1996) 'Translocation of long peptides by transporters associated with antigen processing (TAP)', *European Journal of Immunology*, 26(8), pp. 1720–1728. Available at: <https://doi.org/10.1002/eji.1830260809>.
- Kotturi, M.F. *et al.* (2008) 'Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance', *Journal of Immunology (Baltimore, Md.: 1950)*, 181(3), pp. 2124–2133. Available at: <https://doi.org/10.4049/jimmunol.181.3.2124>.
- Koup, R.A. *et al.* (1994) 'Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome', *Journal of Virology*, 68(7), pp. 4650–4655. Available at: <https://doi.org/10.1128/JVI.68.7.4650-4655.1994>.
- Kozak, M. (1978) 'How do eucaryotic ribosomes select initiation regions in messenger RNA?', *Cell*, 15(4), pp. 1109–1123. Available at: [https://doi.org/10.1016/0092-8674\(78\)90039-9](https://doi.org/10.1016/0092-8674(78)90039-9).
- Kozak, M. (1984) 'Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin', *Nucleic Acids Research*, 12(9), pp. 3873–3893. Available at: <https://doi.org/10.1093/nar/12.9.3873>.
- Kozak, M. (1987) 'Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes', *Molecular and Cellular Biology*, 7(10), pp. 3438–3445. Available at: <https://doi.org/10.1128/mcb.7.10.3438-3445.1987>.
- Kozak, M. (1989) 'The scanning model for translation: an update', *The Journal of Cell Biology*, 108(2), pp. 229–241. Available at: <https://doi.org/10.1083/jcb.108.2.229>.
- Kreiter, S. *et al.* (2015) 'Mutant MHC class II epitopes drive therapeutic immune responses to cancer', *Nature*, 520(7549), pp. 692–696. Available at: <https://doi.org/10.1038/nature14426>.
- Krowka, J.F. *et al.* (1989) 'Lymphocyte proliferative responses to human immunodeficiency virus antigens in vitro', *The Journal of Clinical Investigation*, 83(4), pp. 1198–1203. Available at: <https://doi.org/10.1172/JCI114001>.
- Krummheuer, J. *et al.* (2007) 'A minimal uORF within the HIV-1 vpu leader allows efficient translation initiation at the downstream env AUG', *Virology*, 363(2), pp. 261–271. Available at: <https://doi.org/10.1016/j.virol.2007.01.022>.
- La Gruta, N.L. *et al.* (2018) 'Understanding the drivers of MHC restriction of T cell receptors', *Nature Reviews Immunology*, 18(7), pp. 467–478. Available at: <https://doi.org/10.1038/s41577-018-0007-5>.
- Labaronne, E. *et al.* (2022) 'Extensive uORF translation from HIV-1 transcripts conditions DDX3 dependency for expression of main ORFs and elicits specific T cell immune responses in infected individuals'. bioRxiv, p. 2022.04.29.489990. Available at: <https://doi.org/10.1101/2022.04.29.489990>.
- Lai, M.-C., Lee, Y.-H.W. and Tarn, W.-Y. (2008) 'The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control', *Molecular Biology of the Cell*, 19(9), pp. 3847–3858. Available at: <https://doi.org/10.1091/mbc.e07-12-1264>.
- Lambotte, O. *et al.* (2005) 'HIV controllers: a homogeneous group of HIV-1-infected patients with spontaneous control of viral replication', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 41(7), pp. 1053–1056. Available at: <https://doi.org/10.1086/433188>.
- Lambotte, O. *et al.* (2009) 'Heterogeneous neutralizing antibody and antibody-dependent cell cytotoxicity responses in HIV-1 elite controllers', *AIDS (London, England)*, 23(8), pp. 897–906. Available at: <https://doi.org/10.1097/QAD.0b013e328329f97d>.

- Lambotte, O. (2012) 'Les patients HIV controllers', *médecine/sciences*, 28(2), pp. 172–178. Available at: <https://doi.org/10.1051/medsci/2012282015>.
- Landais, E. and Moore, P.L. (2018) 'Development of broadly neutralizing antibodies in HIV-1 infected elite neutralizers', *Retrovirology*, 15(1), p. 61. Available at: <https://doi.org/10.1186/s12977-018-0443-0>.
- Laumont, C.M. *et al.* (2016) 'Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames', *Nature Communications*, 7, p. 10238. Available at: <https://doi.org/10.1038/ncomms10238>.
- Laumont, C.M. *et al.* (2018) 'Noncoding regions are the main source of targetable tumor-specific antigens', *Science Translational Medicine*, 10(470), p. eaau5516. Available at: <https://doi.org/10.1126/scitranslmed.aau5516>.
- Laumont, C.M. and Perreault, C. (2018) 'Exploiting non-canonical translation to identify new targets for T cell-based cancer immunotherapy', *Cellular and molecular life sciences: CMLS*, 75(4), pp. 607–621. Available at: <https://doi.org/10.1007/s00018-017-2628-4>.
- Lauria, F. *et al.* (2018) 'riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data', *PLoS Computational Biology*, 14(8), p. e1006169. Available at: <https://doi.org/10.1371/journal.pcbi.1006169>.
- Lautwein, A. *et al.* (2002) 'Inflammatory stimuli recruit cathepsin activity to late endosomal compartments in human dendritic cells', *European Journal of Immunology*, 32(12), pp. 3348–3357. Available at: [https://doi.org/10.1002/1521-4141\(200212\)32:12<3348::AID-IMMU3348>3.0.CO;2-S](https://doi.org/10.1002/1521-4141(200212)32:12<3348::AID-IMMU3348>3.0.CO;2-S).
- Laverdure, S. *et al.* (2012) 'HIV-1 antisense transcription is preferentially activated in primary monocyte-derived cells', *Journal of Virology*, 86(24), pp. 13785–13789. Available at: <https://doi.org/10.1128/JVI.01723-12>.
- Law, G.L. *et al.* (2001) 'Polyamine Regulation of Ribosome Pausing at the Upstream Open Reading Frame of S-Adenosylmethionine Decarboxylase *', *Journal of Biological Chemistry*, 276(41), pp. 38036–38043. Available at: <https://doi.org/10.1074/jbc.M105944200>.
- Lee, Sooncheol *et al.* (2012) 'Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution', *Proceedings of the National Academy of Sciences*, 109(37), pp. E2424–E2432. Available at: <https://doi.org/10.1073/pnas.1207846109>.
- Lelouard, H. *et al.* (2007) 'Regulation of translation is required for dendritic cell function and survival during activation', *The Journal of Cell Biology*, 179(7), pp. 1427–1439. Available at: <https://doi.org/10.1083/jcb.200707166>.
- Lepage, S. and Lapointe, R. (2006) 'Melanosomal targeting sequences from gp100 are essential for MHC class II-restricted endogenous epitope presentation and mobilization to endosomal compartments', *Cancer Research*, 66(4), pp. 2423–2432. Available at: <https://doi.org/10.1158/0008-5472.CAN-05-2516>.
- Leppek, K., Das, R. and Barna, M. (2018) 'Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them', *Nature Reviews. Molecular Cell Biology*, 19(3), pp. 158–174. Available at: <https://doi.org/10.1038/nrm.2017.103>.
- Leslie, A.J. *et al.* (2004) 'HIV evolution: CTL escape mutation and reversion after transmission', *Nature Medicine*, 10(3), pp. 282–289. Available at: <https://doi.org/10.1038/nm992>.
- Leung, C.S. *et al.* (2010) 'Nuclear location of an endogenously expressed antigen, EBNA1, restricts access to macroautophagy and the range of CD4 epitope display', *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), pp. 2165–2170. Available at: <https://doi.org/10.1073/pnas.0909448107>.
- Leung, C.S.K. (2015) 'Endogenous Antigen Presentation of MHC Class II Epitopes through Non-Autophagic Pathways', *Frontiers in Immunology*, 6. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2015.00464> (Accessed: 2 April 2023).

- Li, C. *et al.* (2021) 'HIV-1 cores retain their integrity until minutes before uncoating in the nucleus', *Proceedings of the National Academy of Sciences of the United States of America*, 118(10), p. e2019467118. Available at: <https://doi.org/10.1073/pnas.2019467118>.
- Li, Q. *et al.* (2009) 'Glycerol monolaurate prevents mucosal SIV transmission', *Nature*, 458(7241), pp. 1034–1038. Available at: <https://doi.org/10.1038/nature07831>.
- Libre, C. *et al.* (2021) 'A Conserved uORF Regulates APOBEC3G Translation and Is Targeted by HIV-1 Vif Protein to Repress the Antiviral Factor', *Biomedicines*, 10(1), p. 13. Available at: <https://doi.org/10.3390/biomedicines10010013>.
- Lichti, C.F. *et al.* (2022) 'Navigating Critical Challenges Associated with Immunopeptidomics-Based Detection of Proteasomal Spliced Peptide Candidates', *Cancer Immunology Research*, 10(3), pp. 275–284. Available at: <https://doi.org/10.1158/2326-6066.CIR-21-0727>.
- Liepe, J. *et al.* (2016) 'A large fraction of HLA class I ligands are proteasome-generated spliced peptides', *Science (New York, N.Y.)*, 354(6310), pp. 354–358. Available at: <https://doi.org/10.1126/science.aaf4384>.
- Lin, A. and Loré, K. (2017) 'Granulocytes: New Members of the Antigen-Presenting Cell Family', *Frontiers in Immunology*, 8, p. 1781. Available at: <https://doi.org/10.3389/fimmu.2017.01781>.
- Lin, L. *et al.* (2015) 'COMPASS identifies T-cell subsets correlated with clinical outcomes', *Nature Biotechnology*, 33(6), pp. 610–616. Available at: <https://doi.org/10.1038/nbt.3187>.
- Lin, M. and Da, L.-T. (2020) 'Refolding Dynamics of gp41 from Pre-fusion to Pre-hairpin States during HIV-1 Entry', *Journal of Chemical Information and Modeling*, 60(1), pp. 162–174. Available at: <https://doi.org/10.1021/acs.jcim.9b00746>.
- Lindner, R. (2017) 'Invariant Chain Complexes and Clusters as Platforms for MIF Signaling', *Cells*, 6(1), p. 6. Available at: <https://doi.org/10.3390/cells6010006>.
- Linnemann, C. *et al.* (2015) 'High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma', *Nature Medicine*, 21(1), pp. 81–85. Available at: <https://doi.org/10.1038/nm.3773>.
- Linsalata, A.E. *et al.* (2019) 'DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation', *EMBO reports*, 20(9), p. e47498. Available at: <https://doi.org/10.15252/embr.201847498>.
- Lissina, A. *et al.* (2016) 'TCR clonotypes: molecular determinants of T-cell efficacy against HIV', *Current Opinion in Virology*, 16, pp. 77–85. Available at: <https://doi.org/10.1016/j.coviro.2016.01.017>.
- Liu, J. and Gao, G.F. (2011) 'Major Histocompatibility Complex: Intertaction with Peptides'. Available at: <https://doi.org/10.1002/9780470015902.a0000922.pub2>.
- Liu, M.K.P. *et al.* (2013) 'Vertical T cell immunodominance and epitope entropy determine HIV-1 escape', *The Journal of Clinical Investigation*, 123(1), pp. 380–393. Available at: <https://doi.org/10.1172/JCI65330>.
- Liu, S. *et al.* (2012) 'Diphthamide modification on eukaryotic elongation factor 2 is needed to assure fidelity of mRNA translation and mouse development', *Proceedings of the National Academy of Sciences*, 109(34), pp. 13817–13822. Available at: <https://doi.org/10.1073/pnas.1206933109>.
- Liu, S. *et al.* (2019) 'Efficient identification of neoantigen-specific T-cell responses in advanced human ovarian cancer', *Journal for Immunotherapy of Cancer*, 7(1), p. 156. Available at: <https://doi.org/10.1186/s40425-019-0629-6>.
- Liu, Z. *et al.* (2019) 'HIV-1 Antisense Protein of Different Clades Induces Autophagy and Associates with the Autophagy Factor p62', *Journal of Virology*, 93(2), pp. e01757-18. Available at: <https://doi.org/10.1128/JVI.01757-18>.

- Lizée, G. *et al.* (2003) 'Control of dendritic cell cross-presentation by the major histocompatibility complex class I cytoplasmic domain', *Nature Immunology*, 4(11), pp. 1065–1073. Available at: <https://doi.org/10.1038/ni989>.
- Locci, M. *et al.* (2013) 'Human circulating PD-1+CXCR3-CXCR5+ memory Tfh cells are highly functional and correlate with broadly neutralizing HIV antibody responses', *Immunity*, 39(4), pp. 758–769. Available at: <https://doi.org/10.1016/j.immuni.2013.08.031>.
- Los Alamos HIV Databases* (no date). Available at: <https://www.hiv.lanl.gov/> (Accessed: 22 March 2023).
- Loughran, G. *et al.* (2020) 'Unusually efficient CUG initiation of an overlapping reading frame in POLG mRNA yields novel protein POLGARF', *Proceedings of the National Academy of Sciences*, 117(40), pp. 24936–24946. Available at: <https://doi.org/10.1073/pnas.2001433117>.
- Lu, P.D., Harding, H.P. and Ron, D. (2004) 'Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response', *The Journal of Cell Biology*, 167(1), pp. 27–33. Available at: <https://doi.org/10.1083/jcb.200408003>.
- Luciw, P.A. *et al.* (1984) 'Molecular cloning of AIDS-associated retrovirus', *Nature*, 312(5996), pp. 760–763. Available at: <https://doi.org/10.1038/312760a0>.
- Lundegaard, C. *et al.* (2008) 'NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11', *Nucleic Acids Research*, 36(Web Server issue), pp. W509-512. Available at: <https://doi.org/10.1093/nar/gkn202>.
- Machkovech, H.M., Bloom, J.D. and Subramaniam, A.R. (2019) 'Comprehensive profiling of translation initiation in influenza virus infected cells', *PLoS pathogens*, 15(1), p. e1007518. Available at: <https://doi.org/10.1371/journal.ppat.1007518>.
- MacParland, S.A. *et al.* (2016) 'HCV Specific IL-21 Producing T Cells but Not IL-17A Producing T Cells Are Associated with HCV Viral Control in HIV/HCV Coinfection', *PLoS One*, 11(4), p. e0154433. Available at: <https://doi.org/10.1371/journal.pone.0154433>.
- Malandro, N. *et al.* (2016) 'Clonal Abundance of Tumor-Specific CD4(+) T Cells Potentiates Efficacy and Alters Susceptibility to Exhaustion', *Immunity*, 44(1), pp. 179–193. Available at: <https://doi.org/10.1016/j.immuni.2015.12.018>.
- Malarkannan, S. *et al.* (1999) 'Presentation of Out-of-Frame Peptide/MHC Class I Complexes by a Novel Translation Initiation Mechanism', *Immunity*, 10(6), pp. 681–690. Available at: [https://doi.org/10.1016/S1074-7613\(00\)80067-9](https://doi.org/10.1016/S1074-7613(00)80067-9).
- Malim, M.H. *et al.* (1989) 'The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA', *Nature*, 338(6212), pp. 254–257. Available at: <https://doi.org/10.1038/338254a0>.
- Malnati, M.S. *et al.* (1992) 'Processing pathways for presentation of cytosolic antigen to MHC class II-restricted T cells', *Nature*, 357(6380), pp. 702–704. Available at: <https://doi.org/10.1038/357702a0>.
- Malys, N. and McCarthy, J.E.G. (2011) 'Translation initiation: variations in the mechanism can be anticipated', *Cellular and Molecular Life Sciences*, 68(6), pp. 991–1003. Available at: <https://doi.org/10.1007/s00018-010-0588-z>.
- Maness, N.J. *et al.* (2007) 'AIDS virus specific CD8+ T lymphocytes against an immunodominant cryptic epitope select for viral escape', *The Journal of Experimental Medicine*, 204(11), pp. 2505–2512. Available at: <https://doi.org/10.1084/jem.20071261>.
- Maness, N.J. *et al.* (2010) 'Robust, vaccine-induced CD8(+) T lymphocyte response against an out-of-frame epitope', *Journal of Immunology (Baltimore, Md.: 1950)*, 184(1), pp. 67–72. Available at: <https://doi.org/10.4049/jimmunol.0903118>.

- Manoury, B. *et al.* (2003) 'Asparagine endopeptidase can initiate the removal of the MHC class II invariant chain chaperone', *Immunity*, 18(4), pp. 489–498. Available at: [https://doi.org/10.1016/s1074-7613\(03\)00085-2](https://doi.org/10.1016/s1074-7613(03)00085-2).
- Marchingo, J.M. and Cantrell, D.A. (2022) 'Protein synthesis, degradation, and energy metabolism in T cell immunity', *Cellular & Molecular Immunology*, 19(3), pp. 303–315. Available at: <https://doi.org/10.1038/s41423-021-00792-8>.
- Marcilla, M. *et al.* (2016) 'Comparative Analysis of the Endogenous Peptidomes Displayed by HLA-B*27 and Mamu-B*08: Two MHC Class I Alleles Associated with Elite Control of HIV/SIV Infection', *Journal of Proteome Research*, 15(3), pp. 1059–1069. Available at: <https://doi.org/10.1021/acs.jproteome.5b01146>.
- Margulies, D.H. *et al.* (2022) 'Chaperones and Catalysts: How Antigen Presentation Pathways Cope With Biological Necessity', *Frontiers in Immunology*, 13. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.859782> (Accessed: 30 March 2023).
- Mariani, R. *et al.* (2003) 'Species-Specific Exclusion of APOBEC3G from HIV-1 Virions by Vif', *Cell*, 114(1), pp. 21–31. Available at: [https://doi.org/10.1016/S0092-8674\(03\)00515-4](https://doi.org/10.1016/S0092-8674(03)00515-4).
- Marsden, S. *et al.* (2006) 'Unwinding Single RNA Molecules Using Helicases Involved in Eukaryotic Translation Initiation', *Journal of Molecular Biology*, 361(2), pp. 327–335. Available at: <https://doi.org/10.1016/j.jmb.2006.06.016>.
- Martinelli, E. *et al.* (2013) 'The frequency of $\alpha_4\beta_7$ (high) memory CD4⁺ T cells correlates with susceptibility to rectal simian immunodeficiency virus infection', *Journal of Acquired Immune Deficiency Syndromes (1999)*, 64(4), pp. 325–331. Available at: <https://doi.org/10.1097/QAI.0b013e31829f6e1a>.
- Martinez, T.F. *et al.* (2020) 'Accurate annotation of human protein-coding small open reading frames', *Nature Chemical Biology*, 16(4), pp. 458–468. Available at: <https://doi.org/10.1038/s41589-019-0425-0>.
- Martinez-Picado, J. *et al.* (2006) 'Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1', *Journal of Virology*, 80(7), pp. 3617–3623. Available at: <https://doi.org/10.1128/JVI.80.7.3617-3623.2006>.
- Martin-Gayo, E. *et al.* (2017) 'Circulating CXCR5+CXCR3+PD-1^{lo} Tfh-like cells in HIV-1 controllers with neutralizing antibody breadth', *JCI insight*, 2(2), p. e89574. Available at: <https://doi.org/10.1172/jci.insight.89574>.
- Martín-Moreno, A. and Muñoz-Fernández, M.A. (2019) 'Dendritic Cells, the Double Agent in the War Against HIV-1', *Frontiers in Immunology*, 10, p. 2485. Available at: <https://doi.org/10.3389/fimmu.2019.02485>.
- Matsuzaki, J. *et al.* (2014) 'Nonclassical antigen-processing pathways are required for MHC class II-restricted direct tumor recognition by NY-ESO-1-specific CD4(+) T cells', *Cancer Immunology Research*, 2(4), pp. 341–350. Available at: <https://doi.org/10.1158/2326-6066.CIR-13-0138>.
- Matthews, P.C. *et al.* (2008) 'Central role of reverting mutations in HLA associations with human immunodeficiency virus set point', *Journal of Virology*, 82(17), pp. 8548–8559. Available at: <https://doi.org/10.1128/JVI.00580-08>.
- Matthews, P.C. *et al.* (2009) 'HLA Footprints on Human Immunodeficiency Virus Type 1 Are Associated with Interclade Polymorphisms and Intraclade Phylogenetic Clustering', *Journal of Virology*, 83(9), pp. 4605–4615. Available at: <https://doi.org/10.1128/JVI.02017-08>.
- Mayrand, S.M., Schwarz, D.A. and Green, W.R. (1998) 'An alternative translational reading frame encodes an immunodominant retroviral CTL determinant expressed by an immunodeficiency-causing retrovirus', *Journal of Immunology (Baltimore, Md.: 1950)*, 160(1), pp. 39–50.
- McMichael, A. and Klenerman, P. (2002) 'HIV/AIDS. HLA leaves its footprints on HIV', *Science (New York, N.Y.)*, 296(5572), pp. 1410–1411. Available at: <https://doi.org/10.1126/science.1072492>.
- Mellors, J.W. *et al.* (1996) 'Prognosis in HIV-1 infection predicted by the quantity of virus in plasma', *Science (New York, N.Y.)*, 272(5265), pp. 1167–1170. Available at: <https://doi.org/10.1126/science.272.5265.1167>.

- Mendel, G. (1965) 'Versuche über Pflanzenhybriden', in G. Mendel (ed.) *Versuche über Pflanzenhybriden*. Wiesbaden: Vieweg+Teubner Verlag (Ostwalds Klassiker der Exakten Wissenschaften), pp. 21–64. Available at: https://doi.org/10.1007/978-3-663-19714-0_4.
- Mendoza, P. *et al.* (2018) 'Combination therapy with anti-HIV-1 antibodies maintains viral suppression', *Nature*, 561(7724), pp. 479–484. Available at: <https://doi.org/10.1038/s41586-018-0531-2>.
- Meng, G. *et al.* (2000) 'Lamina propria lymphocytes, not macrophages, express CCR5 and CXCR4 and are the likely target cell for human immunodeficiency virus type 1 in the intestinal mucosa', *The Journal of Infectious Diseases*, 182(3), pp. 785–791. Available at: <https://doi.org/10.1086/315790>.
- Merrick, W.C. and Pavitt, G.D. (2018) 'Protein Synthesis Initiation in Eukaryotic Cells', *Cold Spring Harbor Perspectives in Biology*, 10(12), p. a033092. Available at: <https://doi.org/10.1101/cshperspect.a033092>.
- Michael, N.L. *et al.* (1994) 'Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat.', *Journal of Virology*, 68(2), pp. 979–987.
- Migueles, S.A. *et al.* (2000) 'HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors', *Proceedings of the National Academy of Sciences of the United States of America*, 97(6), pp. 2709–2714.
- Migueles, S.A. *et al.* (2002) 'HIV-specific CD8+ T cell proliferation is coupled to perforin expression and is maintained in nonprogressors', *Nature Immunology*, 3(11), pp. 1061–1068. Available at: <https://doi.org/10.1038/ni845>.
- Migueles, S.A. *et al.* (2008) 'Lytic granule loading of CD8+ T cells is required for HIV-infected cell elimination associated with immune control', *Immunity*, 29(6), pp. 1009–1021. Available at: <https://doi.org/10.1016/j.immuni.2008.10.010>.
- Miller, J.S. *et al.* (2022) 'Safety and virologic impact of the IL-15 superagonist N-803 in people living with HIV: a phase 1 trial', *Nature Medicine*, 28(2), pp. 392–400. Available at: <https://doi.org/10.1038/s41591-021-01651-9>.
- Miller, M.A. *et al.* (2015) 'Endogenous antigen processing drives the primary CD4+ T cell response to influenza', *Nature Medicine*, 21(10), pp. 1216–1222. Available at: <https://doi.org/10.1038/nm.3958>.
- Milner, E. *et al.* (2006) 'The turnover kinetics of major histocompatibility complex peptides of human cancer cells', *Molecular & cellular proteomics: MCP*, 5(2), pp. 357–365. Available at: <https://doi.org/10.1074/mcp.M500241-MCP200>.
- Miura, T. *et al.* (2009) 'HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition', *Journal of Virology*, 83(6), pp. 2743–2755. Available at: <https://doi.org/10.1128/JVI.02265-08>.
- Modrow, S. *et al.* (1987) 'Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions', *Journal of Virology*, 61(2), pp. 570–578. Available at: <https://doi.org/10.1128/JVI.61.2.570-578.1987>.
- Moir, S., Chun, T.-W. and Fauci, A.S. (2011) 'Pathogenic mechanisms of HIV disease', *Annual Review of Pathology*, 6, pp. 223–248. Available at: <https://doi.org/10.1146/annurev-pathol-011110-130254>.
- Moldt, B. *et al.* (2022) 'HIV envelope antibodies and TLR7 agonist partially prevent viral rebound in chronically SHIV-infected monkeys', *PLoS pathogens*, 18(4), p. e1010467. Available at: <https://doi.org/10.1371/journal.ppat.1010467>.
- Momburg, F. and Hämmerling, G.J. (1998) 'Generation and TAP-mediated transport of peptides for major histocompatibility complex class I molecules', *Advances in Immunology*, 68, pp. 191–256. Available at: [https://doi.org/10.1016/s0065-2776\(08\)60560-x](https://doi.org/10.1016/s0065-2776(08)60560-x).

Moore, C.B. *et al.* (2002) 'Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level', *Science (New York, N.Y.)*, 296(5572), pp. 1439–1443. Available at: <https://doi.org/10.1126/science.1069660>.

Moore, P.L., Williamson, C. and Morris, L. (2015) 'Virological features associated with the development of broadly neutralizing antibodies to HIV-1', *Trends in Microbiology*, 23(4), pp. 204–211. Available at: <https://doi.org/10.1016/j.tim.2014.12.007>.

Moris, A. *et al.* (2006a) 'Dendritic cells and HIV-specific CD4+ T cells: HIV antigen presentation, T-cell activation, and viral transfer', *Blood*, 108(5), pp. 1643–1651. Available at: <https://doi.org/10.1182/blood-2006-02-006361>.

Moris, A. *et al.* (2006b) 'Dendritic cells and HIV-specific CD4+ T cells: HIV antigen presentation, T-cell activation, and viral transfer', *Blood*, 108(5), pp. 1643–1651. Available at: <https://doi.org/10.1182/blood-2006-02-006361>.

Moris, A., Pereira, M. and Chakrabarti, L. (2019a) 'A role for antibodies in natural HIV control', *Current Opinion in HIV and AIDS*, 14(4), p. 265. Available at: <https://doi.org/10.1097/COH.0000000000000554>.

Moris, A., Pereira, M. and Chakrabarti, L. (2019b) 'A role for antibodies in natural HIV control', *Current opinion in HIV and AIDS*, 14(4), pp. 265–272. Available at: <https://doi.org/10.1097/COH.0000000000000554>.

Mothe, B. *et al.* (2012) 'CTL responses of high functional avidity and broad variant cross-reactivity are associated with HIV control', *PLoS One*, 7(1), p. e29717. Available at: <https://doi.org/10.1371/journal.pone.0029717>.

Mouquet, H. (2014) 'Antibody B cell responses in HIV-1 infection', *Trends in Immunology*, 35(11), pp. 549–561. Available at: <https://doi.org/10.1016/j.it.2014.08.007>.

Mucida, D. *et al.* (2013) 'Transcriptional reprogramming of mature CD4+ helper T cells generates distinct MHC class II-restricted cytotoxic T lymphocytes', *Nature Immunology*, 14(3), pp. 281–289. Available at: <https://doi.org/10.1038/ni.2523>.

Mueller, S.N. and Mackay, L.K. (2016) 'Tissue-resident memory T cells: local specialists in immune defence', *Nature Reviews Immunology*, 16(2), pp. 79–89. Available at: <https://doi.org/10.1038/nri.2015.3>.

Muller, H.J. (1922) 'Variation Due to Change in the Individual Gene', *The American Naturalist*, 56(642), pp. 32–50. Available at: <https://doi.org/10.1086/279846>.

Muller, H.J. (1927) 'ARTIFICIAL TRANSMUTATION OF THE GENE', *Science (New York, N.Y.)*, 66(1699), pp. 84–87. Available at: <https://doi.org/10.1126/science.66.1699.84>.

Muller, H.J. (1966) 'The Gene Material as the Initiator and the Organizing Basis of Life', *The American Naturalist*, 100(915), pp. 493–517.

Muntasell, A. *et al.* (2002) 'HLA-DR4 molecules in neuroendocrine epithelial cells associate to a heterogeneous repertoire of cytoplasmic and surface self peptides', *Journal of Immunology (Baltimore, Md.: 1950)*, 169(9), pp. 5052–5060. Available at: <https://doi.org/10.4049/jimmunol.169.9.5052>.

Münz, C. *et al.* (2000) 'Human CD4(+) T lymphocytes consistently respond to the latent Epstein-Barr virus nuclear antigen EBNA1', *The Journal of Experimental Medicine*, 191(10), pp. 1649–1660. Available at: <https://doi.org/10.1084/jem.191.10.1649>.

Murakami, T. and Ono, A. (2021) 'HIV-1 entry: Duels between Env and host antiviral transmembrane proteins on the surface of virus particles', *Current Opinion in Virology*, 50, pp. 59–68. Available at: <https://doi.org/10.1016/j.coviro.2021.07.005>.

Nair-Gupta, P. *et al.* (2014) 'TLR signals induce phagosomal MHC-I delivery from the endosomal recycling compartment to allow cross-presentation', *Cell*, 158(3), pp. 506–521. Available at: <https://doi.org/10.1016/j.cell.2014.04.054>.

Namy, O. *et al.* (2004) 'Reprogrammed genetic decoding in cellular gene expression', *Molecular Cell*, 13(2), pp. 157–168. Available at: [https://doi.org/10.1016/s1097-2765\(04\)00031-0](https://doi.org/10.1016/s1097-2765(04)00031-0).

- Namy, O. *et al.* (2006) 'A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting', *Nature*, 441(7090), pp. 244–247. Available at: <https://doi.org/10.1038/nature04735>.
- Ndhlovu, Z.M. *et al.* (2015) 'Magnitude and Kinetics of CD8+ T Cell Activation during Hyperacute HIV Infection Impact Viral Set Point', *Immunity*, 43(3), pp. 591–604. Available at: <https://doi.org/10.1016/j.immuni.2015.08.012>.
- Neefjes, J. *et al.* (2011) 'Towards a systems understanding of MHC class I and MHC class II antigen presentation', *Nature Reviews. Immunology*, 11(12), pp. 823–836. Available at: <https://doi.org/10.1038/nri3084>.
- Newman, Z.R. *et al.* (2016) 'Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9', *Proceedings of the National Academy of Sciences*, 113(10), pp. E1362–E1371. Available at: <https://doi.org/10.1073/pnas.1518976113>.
- Nguyen Quang, N. *et al.* (2020) 'Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection', *Retrovirology*, 17(1), p. 25. Available at: <https://doi.org/10.1186/s12977-020-00533-1>.
- Nicol, M.R., Adams, J.L. and Kashuba, A.D. (2013) 'HIV PrEP Trials: The Road to Success', *Clinical investigation*, 3(3), p. 10.4155/cli.12.155. Available at: <https://doi.org/10.4155/cli.12.155>.
- Niesler, B. *et al.* (2001) 'Association between the 5' UTR variant C178T of the serotonin receptor gene HTR3A and bipolar affective disorder', *Pharmacogenetics*, 11(6), pp. 471–475. Available at: <https://doi.org/10.1097/00008571-200108000-00002>.
- Nurk, S. *et al.* (2022) 'The complete sequence of a human genome', *Science*, 376(6588), pp. 44–53. Available at: <https://doi.org/10.1126/science.abj6987>.
- Ochsenbauer, C. *et al.* (2012) 'Generation of Transmitted/Founder HIV-1 Infectious Molecular Clones and Characterization of Their Replication Capacity in CD4 T Lymphocytes and Monocyte-Derived Macrophages', *Journal of Virology*, 86(5), pp. 2715–2728. Available at: <https://doi.org/10.1128/JVI.06157-11>.
- Ocwieja, K.E. *et al.* (2012) 'Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing', *Nucleic Acids Research*, 40(20), pp. 10345–10355. Available at: <https://doi.org/10.1093/nar/gks753>.
- Ogg, G.S. *et al.* (1999) 'Decay kinetics of human immunodeficiency virus-specific effector cytotoxic T lymphocytes after combination antiretroviral therapy', *Journal of Virology*, 73(1), pp. 797–800. Available at: <https://doi.org/10.1128/JVI.73.1.797-800.1999>.
- Okulicz, J.F. *et al.* (2009) 'Clinical outcomes of elite controllers, viremic controllers, and long-term nonprogressors in the US Department of Defense HIV natural history study', *The Journal of Infectious Diseases*, 200(11), pp. 1714–1723. Available at: <https://doi.org/10.1086/646609>.
- Orr, M.W. *et al.* (2020) 'Alternative ORFs and small ORFs: shedding light on the dark proteome', *Nucleic Acids Research*, 48(3), pp. 1029–1042. Available at: <https://doi.org/10.1093/nar/gkz734>.
- Osuna, C.E. *et al.* (2014) 'TCR affinity associated with functional differences between dominant and subdominant SIV epitope-specific CD8+ T cells in Mamu-A*01+ rhesus monkeys', *PLoS pathogens*, 10(4), p. e1004069. Available at: <https://doi.org/10.1371/journal.ppat.1004069>.
- Ott, P.A. *et al.* (2017) 'An immunogenic personal neoantigen vaccine for patients with melanoma', *Nature*, 547(7662), pp. 217–221. Available at: <https://doi.org/10.1038/nature22991>.
- Ouspenskaia, T. *et al.* (2022) 'Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer', *Nature Biotechnology*, 40(2), pp. 209–217. Available at: <https://doi.org/10.1038/s41587-021-01021-3>.
- Oxenius, A. *et al.* (2000) 'Early highly active antiretroviral therapy for acute HIV-1 infection preserves immune function of CD8+ and CD4+ T lymphocytes', *Proceedings of the National Academy of Sciences of the United States of America*, 97(7), pp. 3382–3387. Available at: <https://doi.org/10.1073/pnas.97.7.3382>.

- Paes, W. *et al.* (2019) 'Contribution of proteasome-catalyzed peptide cis-splicing to viral targeting by CD8+ T cells in HIV-1 infection', *Proceedings of the National Academy of Sciences of the United States of America*, 116(49), pp. 24748–24759. Available at: <https://doi.org/10.1073/pnas.1911622116>.
- Palmer, B.E. *et al.* (2002) 'Discordance between frequency of human immunodeficiency virus type 1 (HIV-1)-specific gamma interferon-producing CD4(+) T cells and HIV-1-specific lymphoproliferation in HIV-1-infected subjects with active viral replication', *Journal of Virology*, 76(12), pp. 5925–5936. Available at: <https://doi.org/10.1128/jvi.76.12.5925-5936.2002>.
- Palmer, C. *et al.* (1996) 'Functional Characterization of the V1V2 Region of Human Immunodeficiency Virus Type 1', *Virology*, 220(2), pp. 436–449. Available at: <https://doi.org/10.1006/viro.1996.0331>.
- Paludan, C. *et al.* (2005) 'Endogenous MHC class II processing of a viral nuclear antigen after autophagy', *Science (New York, N.Y.)*, 307(5709), pp. 593–596. Available at: <https://doi.org/10.1126/science.1104904>.
- Pan, R. *et al.* (2015) 'The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel', *Journal of Virology*, 89(15), pp. 8003–8010. Available at: <https://doi.org/10.1128/JVI.00754-15>.
- Papadopoulos, C. *et al.* (2021) 'Intergenic ORFs as elementary structural modules of de novo gene birth and protein evolution', *Genome Research*, 31(12), pp. 2303–2315. Available at: <https://doi.org/10.1101/gr.275638.121>.
- Papadopoulos, C. (2022) *The noncoding genome, a reservoir of genetic novelty*. Paris-Saclay. Available at: <https://theses.hal.science/tel-03563243>.
- Parikh, U.M. *et al.* (2009) 'Complete protection from repeated vaginal simian-human immunodeficiency virus exposures in macaques by a topical gel containing tenofovir alone or with emtricitabine', *Journal of Virology*, 83(20), pp. 10358–10365. Available at: <https://doi.org/10.1128/JVI.01073-09>.
- Park, J. and Morrow, C.D. (1991) 'Overexpression of the gag-pol precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production', *Journal of Virology*, 65(9), pp. 5111–5117. Available at: <https://doi.org/10.1128/JVI.65.9.5111-5117.1991>.
- Parkin, N.T. *et al.* (1988) 'Mutational analysis of the 5' non-coding region of human immunodeficiency virus type 1: effects of secondary structure on translation', *The EMBO journal*, 7(9), pp. 2831–2837. Available at: <https://doi.org/10.1002/j.1460-2075.1988.tb03139.x>.
- Pastor, Y. *et al.* (2022) 'Refining the DC-targeting vaccination for preventing emerging infectious diseases', *Frontiers in Immunology*, 13, p. 949779. Available at: <https://doi.org/10.3389/fimmu.2022.949779>.
- Paul, S. *et al.* (2013) 'HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity', *Journal of Immunology (Baltimore, Md.: 1950)*, 191(12), pp. 5831–5839. Available at: <https://doi.org/10.4049/jimmunol.1302101>.
- Pechmann, S. and Frydman, J. (2013) 'Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding', *Nature Structural & Molecular Biology*, 20(2), pp. 237–243. Available at: <https://doi.org/10.1038/nsmb.2466>.
- Peeters, M., D'Arc, M. and Delaporte, E. (2014) 'The origin and diversity of human retroviruses', *AIDS reviews*, 16(1), pp. 23–34.
- Pelletier, J. and Sonenberg, N. (1988) 'Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA', *Nature*, 334(6180), pp. 320–325. Available at: <https://doi.org/10.1038/334320a0>.
- Pelletier, J., Thomas, G. and Volarević, S. (2018) 'Ribosome biogenesis in cancer: new players and therapeutic avenues', *Nature Reviews Cancer*, 18(1), pp. 51–63. Available at: <https://doi.org/10.1038/nrc.2017.104>.
- Peng, B.J. *et al.* (2018) 'Antisense-Derived HIV-1 Cryptic Epitopes Are Not Major Drivers of Viral Evolution during the Acute Phase of Infection', *Journal of Virology*, 92(19), pp. e00711-18. Available at: <https://doi.org/10.1128/JVI.00711-18>.

- Perdomo-Celis, F. *et al.* (2022) 'Reprogramming dysfunctional CD8+ T cells to promote properties associated with natural HIV control', *The Journal of Clinical Investigation*, 132(11), p. e157549. Available at: <https://doi.org/10.1172/JCI157549>.
- Pereyra, F. *et al.* (2008) 'Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy', *The Journal of Infectious Diseases*, 197(4), pp. 563–571. Available at: <https://doi.org/10.1086/526786>.
- Pereyra, F. *et al.* (2009) 'Persistent low-level viremia in HIV-1 elite controllers and relationship to immunologic parameters', *The Journal of Infectious Diseases*, 200(6), pp. 984–990. Available at: <https://doi.org/10.1086/605446>.
- Pereyra, F. *et al.* (2010) 'The Major Genetic Determinants of HIV-1 Control Affect HLA Class I Peptide Presentation', *Science (New York, N.Y.)*, 330(6010), pp. 1551–1557. Available at: <https://doi.org/10.1126/science.1195271>.
- Petrovas, C. *et al.* (2006) 'PD-1 is a regulator of virus-specific CD8+ T cell survival in HIV infection', *The Journal of Experimental Medicine*, 203(10), pp. 2281–2292. Available at: <https://doi.org/10.1084/jem.20061496>.
- Pinet, V. *et al.* (1995) 'Antigen presentation mediated by recycling of surface HLA-DR molecules', *Nature*, 375(6532), pp. 603–606. Available at: <https://doi.org/10.1038/375603a0>.
- Pink, R.C. *et al.* (2011) 'Pseudogenes: Pseudo-functional or key regulators in health and disease?', *RNA*, 17(5), p. 792. Available at: <https://doi.org/10.1261/rna.2658311>.
- Pishesha, N., Harmand, T.J. and Ploegh, H.L. (2022) 'A guide to antigen processing and presentation', *Nature Reviews Immunology*, 22(12), pp. 751–764. Available at: <https://doi.org/10.1038/s41577-022-00707-2>.
- Pitcher, C.J. *et al.* (1999) 'HIV-1-specific CD4+ T cells are detectable in most individuals with active HIV-1 infection, but decline with prolonged viral suppression', *Nature Medicine*, 5(5), pp. 518–525. Available at: <https://doi.org/10.1038/8400>.
- Plank, T.-D.M. *et al.* (2014) 'Internal translation initiation from HIV-1 transcripts is conferred by a common RNA structure', *Translation (Austin, Tex.)*, 2(1), p. e27694. Available at: <https://doi.org/10.4161/trla.27694>.
- Plank, T.-D.M., Whitehurst, J.T. and Kieft, J.S. (2013) 'Cell type specificity and structural determinants of IRES activity from the 5' leaders of different HIV-1 transcripts', *Nucleic Acids Research*, 41(13), pp. 6698–6714. Available at: <https://doi.org/10.1093/nar/gkt358>.
- PLANT, E.P. and DINMAN, J.D. (2006) 'Comparative study of the effects of heptameric slippery site composition on -1 frameshifting among different eukaryotic systems', *RNA*, 12(4), pp. 666–673. Available at: <https://doi.org/10.1261/rna.2225206>.
- Planz, O. *et al.* (2001) 'A Naturally Processed Rat Major Histocompatibility Complex Class I-associated Viral Peptide as Target Structure of Borna Disease Virus-specific CD8+ T Cells *', *Journal of Biological Chemistry*, 276(17), pp. 13689–13694. Available at: <https://doi.org/10.1074/jbc.M009889200>.
- Platteel, A.C.M. *et al.* (2016) 'CD8+ T cells of Listeria monocytogenes-infected mice recognize both linear and spliced proteasome products', *European Journal of Immunology*, 46(5), pp. 1109–1118. Available at: <https://doi.org/10.1002/eji.201545989>.
- Pneumocystis Pneumonia* --- Los Angeles (1981). Available at: https://www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm (Accessed: 22 March 2023).
- Pope, M. *et al.* (1995) 'Low levels of HIV-1 infection in cutaneous dendritic cells promote extensive viral replication upon binding to memory CD4+ T cells', *The Journal of Experimental Medicine*, 182(6), pp. 2045–2056. Available at: <https://doi.org/10.1084/jem.182.6.2045>.
- Porichis, F. and Kaufmann, D.E. (2011) 'HIV-specific CD4 T cells and immune control of viral replication', *Current opinion in HIV and AIDS*, 6(3), pp. 174–180. Available at: <https://doi.org/10.1097/COH.0b013e3283454058>.

- Portin, P. and Wilkins, A. (2017) 'The Evolving Definition of the Term "Gene"', *Genetics*, 205(4), pp. 1353–1364. Available at: <https://doi.org/10.1534/genetics.116.196956>.
- Poschke, I.C. *et al.* (2020) 'The Outcome of Ex Vivo TIL Expansion Is Highly Influenced by Spatial Heterogeneity of the Tumor T-Cell Repertoire and Differences in Intrinsic In Vitro Growth Capacity between T-Cell Clones', *Clinical Cancer Research*, 26(16), pp. 4289–4301. Available at: <https://doi.org/10.1158/1078-0432.CCR-19-3845>.
- Potter, S.J. *et al.* (2007) 'Preserved central memory and activated effector memory CD4+ T-cell subsets in human immunodeficiency virus controllers: an ANRS EP36 study', *Journal of Virology*, 81(24), pp. 13904–13915. Available at: <https://doi.org/10.1128/JVI.01401-07>.
- Prasad, S., Starck, S.R. and Shastri, N. (2016) 'Presentation of Cryptic Peptides by MHC Class I Is Enhanced by Inflammatory Stimuli', *Journal of Immunology (Baltimore, Md.: 1950)*, 197(8), pp. 2981–2991. Available at: <https://doi.org/10.4049/jimmunol.1502045>.
- Prensner, J.R. *et al.* (2021) 'Non-canonical open reading frames encode functional proteins essential for cancer cell survival', *Nature biotechnology*, 39(6), pp. 697–704. Available at: <https://doi.org/10.1038/s41587-020-00806-2>.
- Products, I. of M. (US) C. to S.H.T.T.B. and B. *et al.* (1995) 'History of the Controversy', in *HIV And The Blood Supply: An Analysis Of Crisis Decisionmaking*. National Academies Press (US). Available at: <https://www.ncbi.nlm.nih.gov/books/NBK232419/> (Accessed: 3 May 2023).
- Purcell, A.W., Ramarathinam, S.H. and Ternette, N. (2019) 'Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics', *Nature Protocols*, 14(6), pp. 1687–1707. Available at: <https://doi.org/10.1038/s41596-019-0133-y>.
- Purcell, D.F. and Martin, M.A. (1993) 'Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity', *Journal of Virology*, 67(11), pp. 6365–6378. Available at: <https://doi.org/10.1128/JVI.67.11.6365-6378.1993>.
- Purnima, A., Ganesan, V. and Eisenlohr, L.C. (2017) 'The elucidation of non-classical MHC class II antigen processing through the study of viral antigens', *Current opinion in virology*, 22, pp. 71–76. Available at: <https://doi.org/10.1016/j.coviro.2016.11.009>.
- Pymm, P. *et al.* (2017) 'MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape', *Nature Structural & Molecular Biology*, 24(4), pp. 387–394. Available at: <https://doi.org/10.1038/nsmb.3381>.
- Qi, J. *et al.* (2020) 'Advances in Developing CAR T-Cell Therapy for HIV Cure', *Frontiers in Immunology*, 11. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00361> (Accessed: 29 March 2023).
- Rabson, A.B. and Martin, M.A. (1985) 'Molecular organization of the AIDS retrovirus', *Cell*, 40(3), pp. 477–480. Available at: [https://doi.org/10.1016/0092-8674\(85\)90189-8](https://doi.org/10.1016/0092-8674(85)90189-8).
- Raiborg, C. and Stenmark, H. (2009) 'The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins', *Nature*, 458(7237), pp. 445–452. Available at: <https://doi.org/10.1038/nature07961>.
- Raj, A. *et al.* (2016) 'Thousands of novel translated open reading frames in humans inferred by ribosome footprint profiling', *eLife*, 5, p. e13328. Available at: <https://doi.org/10.7554/eLife.13328>.
- Rajagopal, V. *et al.* (2012) 'Specific Domains in Yeast Translation Initiation Factor eIF4G Strongly Bias RNA Unwinding Activity of the eIF4F Complex toward Duplexes with 5'-Overhangs *', *Journal of Biological Chemistry*, 287(24), pp. 20301–20312. Available at: <https://doi.org/10.1074/jbc.M112.347278>.
- Ramarathinam, S.H. *et al.* (2018) 'Identification of Native and Posttranslationally Modified HLA-B*57:01-Restricted HIV Envelope Derived Epitopes Using Immunoproteomics', *Proteomics*, 18(12), p. e1700253. Available at: <https://doi.org/10.1002/pmic.201700253>.
- Rammensee, H. *et al.* (1999) 'SYFPEITHI: database for MHC ligands and peptide motifs', *Immunogenetics*, 50(3–4), pp. 213–219. Available at: <https://doi.org/10.1007/s002510050595>.

- Rammensee, H.-G. *et al.* (1999) 'SYFPEITHI: database for MHC ligands and peptide motifs', *Immunogenetics*, 50(3), pp. 213–219. Available at: <https://doi.org/10.1007/s002510050595>.
- Rammensee, H.G., Friede, T. and Stevanović, S. (1995) 'MHC ligands and peptide motifs: first listing', *Immunogenetics*, 41(4), pp. 178–228. Available at: <https://doi.org/10.1007/BF00172063>.
- Ranasinghe, S. *et al.* (2012) 'HIV-Specific CD4 T Cell Responses to Different Viral Proteins Have Discordant Associations with Viral Load and Clinical Outcome', *Journal of Virology*, 86(1), pp. 277–283. Available at: <https://doi.org/10.1128/JVI.05577-11>.
- Regner, M. *et al.* (2001) 'Immunogenicity of two peptide determinants in the cytolytic T-cell response to flavivirus infection: inverse correlation between peptide affinity for MHC class I and T-cell precursor frequency', *Viral Immunology*, 14(2), pp. 135–149. Available at: <https://doi.org/10.1089/088282401750234510>.
- Rehr, M. *et al.* (2008) 'Emergence of polyfunctional CD8+ T cells after prolonged suppression of human immunodeficiency virus replication by antiretroviral therapy', *Journal of Virology*, 82(7), pp. 3391–3404. Available at: <https://doi.org/10.1128/JVI.02383-07>.
- Reis, B.S. *et al.* (2013) 'Mutual expression of the transcription factors Runx3 and ThPOK regulates intestinal CD4+ T cell immunity', *Nature Immunology*, 14(3), pp. 271–280. Available at: <https://doi.org/10.1038/ni.2518>.
- Reits, E. *et al.* (2004) 'A Major Role for TPP1 in Trimming Proteasomal Degradation Products for MHC Class I Antigen Presentation', *Immunity*, 20(4), pp. 495–506. Available at: [https://doi.org/10.1016/S1074-7613\(04\)00074-3](https://doi.org/10.1016/S1074-7613(04)00074-3).
- Rerks-Ngarm, S. *et al.* (2009) 'Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand', *The New England Journal of Medicine*, 361(23), pp. 2209–2220. Available at: <https://doi.org/10.1056/NEJMoa0908492>.
- Rey, F. *et al.* (1986) 'HIV-I and HIV-II double infection in Central African Republic', *Lancet (London, England)*, 2(8520), pp. 1391–1392. Available at: [https://doi.org/10.1016/s0140-6736\(86\)92027-1](https://doi.org/10.1016/s0140-6736(86)92027-1).
- Reynisson, B. *et al.* (2020) 'NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data', *Nucleic Acids Research*, 48(W1), pp. W449–W454. Available at: <https://doi.org/10.1093/nar/gkaa379>.
- Rhie, A. *et al.* (2022) *The complete sequence of a human Y chromosome*. preprint. Genomics. Available at: <https://doi.org/10.1101/2022.12.01.518724>.
- Ricci, E.P. *et al.* (2014) 'Staufen1 senses overall transcript secondary structure to regulate translation', *Nature Structural & Molecular Biology*, 21(1), pp. 26–35. Available at: <https://doi.org/10.1038/nsmb.2739>.
- Rice, A.P. (2017) 'The HIV-1 Tat Protein: Mechanism of Action and Target for HIV-1 Cure Strategies', *Current Pharmaceutical Design*, 23(28), pp. 4098–4102. Available at: <https://doi.org/10.2174/1381612823666170704130635>.
- Rinaldo, C.R. *et al.* (2000) 'Anti-human immunodeficiency virus type 1 (HIV-1) CD8(+) T-lymphocyte reactivity during combination antiretroviral therapy in HIV-1-infected patients with advanced immunodeficiency', *Journal of Virology*, 74(9), pp. 4127–4138. Available at: <https://doi.org/10.1128/jvi.74.9.4127-4138.2000>.
- Robbins, P.F. *et al.* (1996) 'A mutated beta-catenin gene encodes a melanoma-specific antigen recognized by tumor infiltrating lymphocytes', *The Journal of Experimental Medicine*, 183(3), pp. 1185–1192. Available at: <https://doi.org/10.1084/jem.183.3.1185>.
- Robbins, P.F. *et al.* (2013) 'Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells', *Nature Medicine*, 19(6), pp. 747–752. Available at: <https://doi.org/10.1038/nm.3161>.
- Robinson, J. *et al.* (2020) 'IPD-IMGT/HLA Database', *Nucleic Acids Research*, 48(D1), pp. D948–D955. Available at: <https://doi.org/10.1093/nar/gkz950>.

- Roche, P.A. *et al.* (1993) 'Cell surface HLA-DR-invariant chain complexes are targeted to endosomes by rapid internalization.', *Proceedings of the National Academy of Sciences of the United States of America*, 90(18), pp. 8581–8585.
- Roche, P.A. and Cresswell, P. (1991) 'Proteolysis of the class II-associated invariant chain generates a peptide binding site in intracellular HLA-DR molecules', *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), pp. 3150–3154. Available at: <https://doi.org/10.1073/pnas.88.8.3150>.
- Roche, P.A. and Furuta, K. (2015) 'The ins and outs of MHC class II-mediated antigen processing and presentation', *Nature reviews. Immunology*, 15(4), pp. 203–216. Available at: <https://doi.org/10.1038/nri3818>.
- Rock, K.L. *et al.* (2014) 'Re-examining class-I presentation and the DRiP hypothesis', *Trends in Immunology*, 35(4), pp. 144–152. Available at: <https://doi.org/10.1016/j.it.2014.01.002>.
- Rock, K.L., Benacerraf, B. and Abbas, A.K. (1984) 'Antigen presentation by hapten-specific B lymphocytes. I. Role of surface immunoglobulin receptors', *The Journal of Experimental Medicine*, 160(4), pp. 1102–1113. Available at: <https://doi.org/10.1084/jem.160.4.1102>.
- Rock, K.L. and Goldberg, A.L. (1999) 'Degradation of cell proteins and the generation of MHC class I-presented peptides', *Annual Review of Immunology*, 17, pp. 739–779. Available at: <https://doi.org/10.1146/annurev.immunol.17.1.739>.
- Rodnina, M.V. and Wintermeyer, W. (2010) 'The ribosome goes Nobel', *Trends in Biochemical Sciences*, 35(1), pp. 1–5. Available at: <https://doi.org/10.1016/j.tibs.2009.11.003>.
- Rodriguez, A. *et al.* (1999) 'Selective transport of internalized antigens to the cytosol for MHC class I presentation in dendritic cells', *Nature Cell Biology*, 1(6), pp. 362–368. Available at: <https://doi.org/10.1038/14058>.
- Rodriguez-Plata, M.T. *et al.* (2012) 'HIV-1 capture and antigen presentation by dendritic cells: enhanced viral capture does not correlate with better T cell activation', *Journal of Immunology (Baltimore, Md.: 1950)*, 188(12), pp. 6036–6045. Available at: <https://doi.org/10.4049/jimmunol.1200267>.
- Romagnoli, P. and Germain, R.N. (1994) 'The CLIP region of invariant chain plays a critical role in regulating major histocompatibility complex class II folding, transport, and peptide occupancy', *The Journal of Experimental Medicine*, 180(3), pp. 1107–1113. Available at: <https://doi.org/10.1084/jem.180.3.1107>.
- Ronsin, C. *et al.* (1999) 'A Non-AUG-Defined Alternative Open Reading Frame of the Intestinal Carboxyl Esterase mRNA Generates an Epitope Recognized by Renal Cell Carcinoma-Reactive Tumor-Infiltrating Lymphocytes In Situ', *The Journal of Immunology*, 163(1), pp. 483–490. Available at: <https://doi.org/10.4049/jimmunol.163.1.483>.
- van Rooij, N. *et al.* (2013) 'Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma', *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 31(32), pp. e439–442. Available at: <https://doi.org/10.1200/JCO.2012.47.7521>.
- Rosás-Umbert, M. *et al.* (2019) 'Mechanisms of Abrupt Loss of Virus Control in a Cohort of Previous HIV Controllers', *Journal of Virology*, 93(4), pp. e01436–18. Available at: <https://doi.org/10.1128/JVI.01436-18>.
- Rosenberg, A.H. *et al.* (1993) 'Effects of consecutive AGG codons on translation in Escherichia coli, demonstrated with a versatile codon test system', *Journal of Bacteriology*, 175(3), pp. 716–722. Available at: <https://doi.org/10.1128/jb.175.3.716-722.1993>.
- Rosenberg, E.S. *et al.* (1997) 'Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia', *Science (New York, N.Y.)*, 278(5342), pp. 1447–1450. Available at: <https://doi.org/10.1126/science.278.5342.1447>.
- Rosenberg, E.S. *et al.* (2000) 'Immune control of HIV-1 after early treatment of acute infection', *Nature*, 407(6803), pp. 523–526. Available at: <https://doi.org/10.1038/35035103>.

- Rosenberg, S.A. *et al.* (2002) 'Identification of BING-4 cancer antigen translated from an alternative open reading frame of a gene in the extended MHC class II region using lymphocytes from a patient with a durable complete regression following immunotherapy', *Journal of Immunology (Baltimore, Md.: 1950)*, 168(5), pp. 2402–2407. Available at: <https://doi.org/10.4049/jimmunol.168.5.2402>.
- Rötzschke, O. *et al.* (1990) 'Isolation and analysis of naturally processed viral peptides as recognized by cytotoxic T cells', *Nature*, 348(6298), pp. 252–254. Available at: <https://doi.org/10.1038/348252a0>.
- Rouers, A., Klingler, J., *et al.* (2017) 'HIV-Specific B Cell Frequency Correlates with Neutralization Breadth in Patients Naturally Controlling HIV-Infection', *EBioMedicine*, 21, pp. 158–169. Available at: <https://doi.org/10.1016/j.ebiom.2017.05.029>.
- Rouers, A., Jeger-Madiot, R., *et al.* (2017) 'Lymphocytes T folliculaires helper et VIH - Unis pour le meilleur et pour le pire', *médecine/sciences*, 33(10), pp. 878–886. Available at: <https://doi.org/10.1051/medsci/20173310020>.
- Roy, N. *et al.* (2014) 'Mechanisms underlying HIV-1 Vpu-mediated viral egress', *Frontiers in Microbiology*, 5, p. 177. Available at: <https://doi.org/10.3389/fmicb.2014.00177>.
- Ruiz Cuevas, M.V. *et al.* (2021) 'Most non-canonical proteins uniquely populate the proteome or immunopeptidome', *Cell Reports*, 34(10), p. 108815. Available at: <https://doi.org/10.1016/j.celrep.2021.108815>.
- Ruiz-Orera, J. *et al.* (2014) 'Long non-coding RNAs as a source of new peptides', *eLife*, 3, p. e03523. Available at: <https://doi.org/10.7554/eLife.03523>.
- Ruiz-Orera, J. *et al.* (2018) 'Translation of neutrally evolving peptides provides a basis for de novo gene evolution', *Nature Ecology & Evolution*, 2(5), pp. 890–896. Available at: <https://doi.org/10.1038/s41559-018-0506-6>.
- Ruiz-Orera, J. and Albà, M.M. (2019) 'Translation of Small Open Reading Frames: Roles in Regulation and Evolutionary Innovation', *Trends in Genetics*, 35(3), pp. 186–198. Available at: <https://doi.org/10.1016/j.tig.2018.12.003>.
- Rutishauser, R.L. and Trautmann, L. (2022) 'CD8 + T-cell responses in HIV controllers: potential implications for novel HIV remission strategies', *Current opinion in HIV and AIDS*, 17(5), pp. 315–324. Available at: <https://doi.org/10.1097/COH.0000000000000748>.
- Sadanand, S. *et al.* (2018) 'Temporal variation in HIV-specific IgG subclass antibodies during acute infection differentiates spontaneous controllers from chronic progressors', *AIDS (London, England)*, 32(4), pp. 443–450. Available at: <https://doi.org/10.1097/QAD.0000000000001716>.
- Sadegh-Nasseri, S. *et al.* (1994) 'MHC class II function preserved by low-affinity peptide interactions preceding stable binding', *Nature*, 370(6491), pp. 647–650. Available at: <https://doi.org/10.1038/370647a0>.
- Sáez-Cirión, A. *et al.* (2007) 'HIV controllers exhibit potent CD8 T cell capacity to suppress HIV infection ex vivo and peculiar cytotoxic T lymphocyte activation phenotype', *Proceedings of the National Academy of Sciences of the United States of America*, 104(16), pp. 6776–6781. Available at: <https://doi.org/10.1073/pnas.0611244104>.
- Sáez-Cirión, A. *et al.* (2009) 'Heterogeneity in HIV suppression by CD8 T cells from HIV controllers: association with Gag-specific CD8 T cell responses', *Journal of Immunology (Baltimore, Md.: 1950)*, 182(12), pp. 7828–7837. Available at: <https://doi.org/10.4049/jimmunol.0803928>.
- Sáez-Cirión, A. and Sereti, I. (2021) 'Immunometabolism and HIV-1 pathogenesis: food for thought', *Nature Reviews Immunology*, 21(1), pp. 5–19. Available at: <https://doi.org/10.1038/s41577-020-0381-7>.
- Sahin, U. *et al.* (2017) 'Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer', *Nature*, 547(7662), pp. 222–226. Available at: <https://doi.org/10.1038/nature23003>.
- Salazar-Gonzalez, J.F. *et al.* (2008) 'Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing', *Journal of Virology*, 82(8), pp. 3952–3970. Available at: <https://doi.org/10.1128/JVI.02660-07>.

- Salazar-Gonzalez, J.F. *et al.* (2009) 'Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection', *The Journal of Experimental Medicine*, 206(6), pp. 1273–1289. Available at: <https://doi.org/10.1084/jem.20090378>.
- Salzberg, S.L. (2018) 'Open questions: How many genes do we have?', *BMC Biology*, 16(1), p. 94. Available at: <https://doi.org/10.1186/s12915-018-0564-x>.
- Sanchez-Pescador, R. *et al.* (1985) 'Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2)', *Science (New York, N.Y.)*, 227(4686), pp. 484–492. Available at: <https://doi.org/10.1126/science.2578227>.
- Sanderson, F. *et al.* (1996) 'Association between HLA-DM and HLA-DR in vivo', *Immunity*, 4(1), pp. 87–96. Available at: [https://doi.org/10.1016/s1074-7613\(00\)80301-5](https://doi.org/10.1016/s1074-7613(00)80301-5).
- Santos, D.A. *et al.* (2019) 'Cycloheximide can distort measurements of mRNA levels and translation efficiency', *Nucleic Acids Research*, 47(10), pp. 4974–4985. Available at: <https://doi.org/10.1093/nar/gkz205>.
- Saper, M.A., Bjorkman, P.J. and Wiley, D.C. (1991) 'Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution', *Journal of Molecular Biology*, 219(2), pp. 277–319. Available at: [https://doi.org/10.1016/0022-2836\(91\)90567-P](https://doi.org/10.1016/0022-2836(91)90567-P).
- Sarango, G. *et al.* (2022) 'The Autophagy Receptor TAX1BP1 (T6BP) improves antigen presentation by MHC-II molecules', *EMBO reports*, 23(12), p. e55470. Available at: <https://doi.org/10.15252/embr.202255470>.
- Saric, T. *et al.* (2002) 'An IFN-gamma-induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I-presented peptides', *Nature Immunology*, 3(12), pp. 1169–1176. Available at: <https://doi.org/10.1038/ni859>.
- Sauce, D., Elbim, C. and Appay, V. (2013) 'Monitoring cellular immune markers in HIV infection: from activation to exhaustion', *Current opinion in HIV and AIDS*, 8(2), pp. 125–131. Available at: <https://doi.org/10.1097/COH.0b013e32835d08a9>.
- Saveanu, L. *et al.* (2005) 'Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum', *Nature Immunology*, 6(7), pp. 689–697. Available at: <https://doi.org/10.1038/ni1208>.
- Savina, A. *et al.* (2006) 'NOX2 controls phagosomal pH to regulate antigen processing during crosspresentation by dendritic cells', *Cell*, 126(1), pp. 205–218. Available at: <https://doi.org/10.1016/j.cell.2006.05.035>.
- Savoret, J. *et al.* (2020a) 'A Pilot Study of the Humoral Response Against the AntiSense Protein (ASP) in HIV-1-Infected Patients', *Frontiers in Microbiology*, 11, p. 20. Available at: <https://doi.org/10.3389/fmicb.2020.00020>.
- Savoret, J. *et al.* (2020b) 'A Pilot Study of the Humoral Response Against the AntiSense Protein (ASP) in HIV-1-Infected Patients', *Frontiers in Microbiology*, 11, p. 20. Available at: <https://doi.org/10.3389/fmicb.2020.00020>.
- Savoret, J. *et al.* (2021) 'Antisense Transcripts and Antisense Protein: A New Perspective on Human Immunodeficiency Virus Type 1', *Frontiers in Microbiology*, 11. Available at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.625941> (Accessed: 19 April 2023).
- Sbarrato, T. *et al.* (2016) 'A ribosome-related signature in peripheral blood CLL B cells is linked to reduced survival following treatment', *Cell Death & Disease*, 7(6), p. e2249. Available at: <https://doi.org/10.1038/cddis.2016.148>.
- Scheid, J.F. *et al.* (2009) 'Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals', *Nature*, 458(7238), pp. 636–640. Available at: <https://doi.org/10.1038/nature07930>.
- Scheid, J.F. *et al.* (2011) 'Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding', *Science (New York, N.Y.)*, 333(6049), pp. 1633–1637. Available at: <https://doi.org/10.1126/science.1207227>.
- Schieffer, M. *et al.* (2014) 'Induction of Gag-Specific CD4 T Cell Responses during Acute HIV Infection Is Associated with Improved Viral Control', *Journal of Virology*, 88(13), pp. 7357–7366. Available at: <https://doi.org/10.1128/JVI.00728-14>.

- Schlesinger, D. and Elsässer, S.J. (2022) 'Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins', *The FEBS Journal*, 289(1), pp. 53–74. Available at: <https://doi.org/10.1111/febs.15769>.
- Schlötterer, C. (2015) 'Genes from scratch--the evolutionary fate of de novo genes', *Trends in genetics: TIG*, 31(4), pp. 215–219. Available at: <https://doi.org/10.1016/j.tig.2015.02.007>.
- Schlunzen, F. *et al.* (2000) 'Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution', *Cell*, 102(5), pp. 615–623. Available at: [https://doi.org/10.1016/S0092-8674\(00\)00084-2](https://doi.org/10.1016/S0092-8674(00)00084-2).
- Schmitz, J.E. *et al.* (1999) 'Control of viremia in simian immunodeficiency virus infection by CD8+ lymphocytes', *Science (New York, N.Y.)*, 283(5403), pp. 857–860. Available at: <https://doi.org/10.1126/science.283.5403.857>.
- Schmitz, J.F., Ullrich, K.K. and Bornberg-Bauer, E. (2018) 'Incipient de novo genes can evolve from frozen accidents that escaped rapid transcript turnover', *Nature Ecology & Evolution*, 2(10), pp. 1626–1632. Available at: <https://doi.org/10.1038/s41559-018-0639-7>.
- Schneider-Poetsch, T. *et al.* (2010) 'Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin', *Nature Chemical Biology*, 6(3), pp. 209–217. Available at: <https://doi.org/10.1038/nchembio.304>.
- Schneidewind, A. *et al.* (2007) 'Escape from the dominant HLA-B27-restricted cytotoxic T-lymphocyte response in Gag is associated with a dramatic reduction in human immunodeficiency virus type 1 replication', *Journal of Virology*, 81(22), pp. 12382–12393. Available at: <https://doi.org/10.1128/JVI.01543-07>.
- Schröder, M. (2010) 'Human DEAD-box protein 3 has multiple functions in gene regulation and cell cycle control and is a prime target for viral manipulation', *Biochemical Pharmacology*, 79(3), pp. 297–306. Available at: <https://doi.org/10.1016/j.bcp.2009.08.032>.
- Schubert, U. *et al.* (2000) 'Rapid degradation of a large fraction of newly synthesized proteins by proteasomes', *Nature*, 404(6779), pp. 770–774. Available at: <https://doi.org/10.1038/35008096>.
- Schuhmacher, J. *et al.* (2022) 'Simultaneous Identification of Functional Antigen-Specific CD8+ and CD4+ Cells after In Vitro Expansion Using Elongated Peptides', *Cells*, 11(21), p. 3451. Available at: <https://doi.org/10.3390/cells11213451>.
- Schumacher, T.N. and Schreiber, R.D. (2015) 'Neoantigens in cancer immunotherapy', *Science (New York, N.Y.)*, 348(6230), pp. 69–74. Available at: <https://doi.org/10.1126/science.aaa4971>.
- Schwab, S.R. *et al.* (2004) 'Unanticipated antigens: translation initiation at CUG with leucine', *PLoS biology*, 2(11), p. e366. Available at: <https://doi.org/10.1371/journal.pbio.0020366>.
- Schwartz, S., Felber, B.K., Benko, D.M., *et al.* (1990) 'Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1', *Journal of Virology*, 64(6), pp. 2519–2529. Available at: <https://doi.org/10.1128/JVI.64.6.2519-2529.1990>.
- Schwartz, S., Felber, B.K., Fenyő, E.M., *et al.* (1990) 'Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs', *Journal of Virology*, 64(11), pp. 5448–5456. Available at: <https://doi.org/10.1128/JVI.64.11.5448-5456.1990>.
- Scoca, V. and Di Nunzio, F. (2021) 'The HIV-1 Capsid: From Structural Component to Key Factor for Host Nuclear Invasion', *Viruses*, 13(2), p. 273. Available at: <https://doi.org/10.3390/v13020273>.
- SenGupta, D.N. *et al.* (1990) 'Direct evidence for translational regulation by leader RNA and Tat protein of human immunodeficiency virus type 1', *Proceedings of the National Academy of Sciences of the United States of America*, 87(19), pp. 7492–7496. Available at: <https://doi.org/10.1073/pnas.87.19.7492>.
- Senissar, M. *et al.* (2014) 'The DEAD-box helicase Ded1 from yeast is an mRNP cap-associated protein that shuttles between the cytoplasm and nucleus', *Nucleic Acids Research*, 42(15), pp. 10005–10022. Available at: <https://doi.org/10.1093/nar/gku584>.

- Sertznig, H. *et al.* (2018) 'Behind the scenes of HIV-1 replication: Alternative splicing as the dependency factor on the quiet', *Virology*, 516, pp. 176–188. Available at: <https://doi.org/10.1016/j.virol.2018.01.011>.
- Sette, A. *et al.* (1994) 'The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes', *Journal of Immunology (Baltimore, Md.: 1950)*, 153(12), pp. 5586–5592.
- Shabman, R.S. *et al.* (2013) 'An Upstream Open Reading Frame Modulates Ebola Virus Polymerase Translation and Virus Replication', *PLOS Pathogens*, 9(1), p. e1003147. Available at: <https://doi.org/10.1371/journal.ppat.1003147>.
- Sharp, P.M. and Hahn, B.H. (2011) 'Origins of HIV and the AIDS pandemic', *Cold Spring Harbor Perspectives in Medicine*, 1(1), p. a006841. Available at: <https://doi.org/10.1101/cshperspect.a006841>.
- Sharp, S.J. *et al.* (1985) 'Structure and transcription of eukaryotic tRNA genes', *CRC critical reviews in biochemistry*, 19(2), pp. 107–144. Available at: <https://doi.org/10.3109/10409238509082541>.
- Shastri, N. and Gonzalez, F. (1993) 'Endogenous generation and presentation of the ovalbumin peptide/Kb complex to T cells', *Journal of Immunology (Baltimore, Md.: 1950)*, 150(7), pp. 2724–2736.
- Shatsky, I.N. *et al.* (2018) 'Cap-Independent Translation: What's in a Name?', *Trends in Biochemical Sciences*, 43(11), pp. 882–895. Available at: <https://doi.org/10.1016/j.tibs.2018.04.011>.
- Sheehy, A.M. *et al.* (2002) 'Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein', *Nature*, 418(6898), pp. 646–650. Available at: <https://doi.org/10.1038/nature00939>.
- Shehu-Xhilaga, M., Crowe, S.M. and Mak, J. (2001) 'Maintenance of the Gag/Gag-Pol Ratio Is Important for Human Immunodeficiency Virus Type 1 RNA Dimerization and Viral Infectivity', *Journal of Virology*, 75(4), pp. 1834–1841. Available at: <https://doi.org/10.1128/JVI.75.4.1834-1841.2001>.
- Shen, L. *et al.* (2004) 'Important role of cathepsin S in generating peptides for TAP-independent MHC class I crosspresentation in vivo', *Immunity*, 21(2), pp. 155–165. Available at: <https://doi.org/10.1016/j.immuni.2004.07.004>.
- Shi, Z. *et al.* (2017) 'Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide', *Molecular Cell*, 67(1), pp. 71–83.e7. Available at: <https://doi.org/10.1016/j.molcel.2017.05.021>.
- Shiina, T. *et al.* (2009) 'The HLA genomic loci map: expression, interaction, diversity and disease', *Journal of Human Genetics*, 54(1), pp. 15–39. Available at: <https://doi.org/10.1038/jhg.2008.5>.
- Shingai, M. *et al.* (2013) 'Antibody-mediated immunotherapy of macaques chronically infected with SHIV suppresses viraemia', *Nature*, 503(7475), pp. 277–280. Available at: <https://doi.org/10.1038/nature12746>.
- Sidney, J. *et al.* (2013) 'Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture', *Current Protocols in Immunology*, Chapter 18, p. Unit 18.3. Available at: <https://doi.org/10.1002/0471142735.im1803s100>.
- Simek, M.D. *et al.* (2009) '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', *Journal of Virology*, 83(14), pp. 7337–7348. Available at: <https://doi.org/10.1128/JVI.00110-09>.
- Singh, R. and Cresswell, P. (2010) 'Defective cross-presentation of viral antigens in GILT-free mice', *Science (New York, N.Y.)*, 328(5984), pp. 1394–1398. Available at: <https://doi.org/10.1126/science.1189176>.
- Sinnathamby, G. and Eisenlohr, L.C. (2003) 'Presentation by Recycling MHC Class II Molecules of an Influenza Hemagglutinin-Derived Epitope That Is Revealed in the Early Endosome by Acidification¹', *The Journal of Immunology*, 170(7), pp. 3504–3513. Available at: <https://doi.org/10.4049/jimmunol.170.7.3504>.

- Sivro, A. *et al.* (2018) 'Integrin $\alpha 4\beta 7$ expression on peripheral blood CD4+ T cells predicts HIV acquisition and disease progression outcomes', *Science translational medicine*, 10(425), p. eaam6354. Available at: <https://doi.org/10.1126/scitranslmed.aam6354>.
- Sloan, D.D. *et al.* (2015) 'Targeting HIV Reservoir in Infected CD4 T Cells by Dual-Affinity Re-targeting Molecules (DARTs) that Bind HIV Envelope and Recruit Cytotoxic T Cells', *PLoS pathogens*, 11(11), p. e1005233. Available at: <https://doi.org/10.1371/journal.ppat.1005233>.
- Smed-Sørensen, A. *et al.* (2005) 'Differential susceptibility to human immunodeficiency virus type 1 infection of myeloid and plasmacytoid dendritic cells', *Journal of Virology*, 79(14), pp. 8861–8869. Available at: <https://doi.org/10.1128/JVI.79.14.8861-8869.2005>.
- Smith, C.C. *et al.* (2018) 'Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma', *The Journal of Clinical Investigation*, 128(11), pp. 4804–4820. Available at: <https://doi.org/10.1172/JCI121476>.
- Smith, J.E. *et al.* (2014) 'Translation of small open reading frames within unannotated RNA transcripts in *Saccharomyces cerevisiae*', *Cell Reports*, 7(6), pp. 1858–1866. Available at: <https://doi.org/10.1016/j.celrep.2014.05.023>.
- Somers, J., Pöyry, T. and Willis, A.E. (2013) 'A perspective on mammalian upstream open reading frame function', *The International Journal of Biochemistry & Cell Biology*, 45(8), pp. 1690–1700. Available at: <https://doi.org/10.1016/j.biocel.2013.04.020>.
- Sonntag, K. *et al.* (2018) 'Immune monitoring and TCR sequencing of CD4 T cells in a long term responsive patient with metastasized pancreatic ductal carcinoma treated with individualized, neoepitope-derived multipeptide vaccines: a case report', *Journal of Translational Medicine*, 16, p. 23. Available at: <https://doi.org/10.1186/s12967-018-1382-1>.
- Sorokin, I.I. *et al.* (2021) 'Non-Canonical Translation Initiation Mechanisms Employed by Eukaryotic Viral mRNAs', *Biochemistry. Biokhimiia*, 86(9), pp. 1060–1094. Available at: <https://doi.org/10.1134/S0006297921090042>.
- Soto-Rifo, R. *et al.* (2012) 'DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs', *The EMBO journal*, 31(18), pp. 3745–3756. Available at: <https://doi.org/10.1038/emboj.2012.220>.
- de Souza, M.S. *et al.* (2012) 'The Thai phase III trial (RV144) vaccine regimen induces T cell responses that preferentially target epitopes within the V2 region of HIV-1 envelope', *Journal of Immunology (Baltimore, Md.: 1950)*, 188(10), pp. 5166–5176. Available at: <https://doi.org/10.4049/jimmunol.1102756>.
- Spiegel, H. *et al.* (1992) 'Follicular dendritic cells are a major reservoir for human immunodeficiency virus type 1 in lymphoid tissues facilitating infection of CD4+ T-helper cells.', *The American Journal of Pathology*, 140(1), pp. 15–22.
- Starck, S.R. *et al.* (2008) 'A distinct translation initiation mechanism generates cryptic peptides for immune surveillance', *PLoS One*, 3(10), p. e3460. Available at: <https://doi.org/10.1371/journal.pone.0003460>.
- Starck, S.R. *et al.* (2012) 'Leucine-tRNA Initiates at CUG Start Codons for Protein Synthesis and Presentation by MHC Class I', *Science*, 336(6089), pp. 1719–1723. Available at: <https://doi.org/10.1126/science.1220270>.
- Starck, S.R. *et al.* (2016) 'Translation from the 5' untranslated region shapes the integrated stress response', *Science (New York, N.Y.)*, 351(6272), p. aad3867. Available at: <https://doi.org/10.1126/science.aad3867>.
- Stefic, K. *et al.* (2019) 'Impact of HIV-1 Diversity on Its Sensitivity to Neutralization', *Vaccines*, 7(3), p. 74. Available at: <https://doi.org/10.3390/vaccines7030074>.
- Steimle, V. *et al.* (1994) 'Regulation of MHC Class II Expression by Interferon- γ Mediated by the Transactivator Gene CIITA', *Science*, 265(5168), pp. 106–109. Available at: <https://doi.org/10.1126/science.8016643>.
- Steitz, J.A. (1969) 'Polypeptide Chain Initiation: Nucleotide Sequences of the Three Ribosomal Binding Sites in Bacteriophage R17 RNA', *Nature*, 224(5223), pp. 957–964. Available at: <https://doi.org/10.1038/224957a0>.

- Steitz, T.A. (2008) 'A structural understanding of the dynamic ribosome machine', *Nature Reviews Molecular Cell Biology*, 9(3), pp. 242–253. Available at: <https://doi.org/10.1038/nrm2352>.
- Stern-Ginossar, N. *et al.* (2012) 'Decoding human cytomegalovirus', *Science (New York, N.Y.)*, 338(6110), pp. 1088–1093. Available at: <https://doi.org/10.1126/science.1227919>.
- Stopak, K. *et al.* (2003) 'HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability', *Molecular Cell*, 12(3), pp. 591–601. Available at: [https://doi.org/10.1016/s1097-2765\(03\)00353-8](https://doi.org/10.1016/s1097-2765(03)00353-8).
- Su, L.F. *et al.* (2013) 'Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults', *Immunity*, 38(2), pp. 373–383. Available at: <https://doi.org/10.1016/j.immuni.2012.10.021>.
- Suni, M.A. *et al.* (2001) 'CD4(+)CD8(dim) T lymphocytes exhibit enhanced cytokine expression, proliferation and cytotoxic activity in response to HCMV and HIV-1 antigens', *European Journal of Immunology*, 31(8), pp. 2512–2520. Available at: [https://doi.org/10.1002/1521-4141\(200108\)31:8<2512::aid-immu2512>3.0.co;2-m](https://doi.org/10.1002/1521-4141(200108)31:8<2512::aid-immu2512>3.0.co;2-m).
- Sussman, J.K., Simons, E.L. and Simons, R.W. (1996) 'Escherichia coli translation initiation factor 3 discriminates the initiation codon in vivo', *Molecular Microbiology*, 21(2), pp. 347–360. Available at: <https://doi.org/10.1046/j.1365-2958.1996.6371354.x>.
- Svitkin, Y.V., Pause, A. and Sonenberg, N. (1994) 'La autoantigen alleviates translational repression by the 5' leader sequence of the human immunodeficiency virus type 1 mRNA', *Journal of Virology*, 68(11), pp. 7001–7007. Available at: <https://doi.org/10.1128/JVI.68.11.7001-7007.1994>.
- Sze, C.W. and Tan, Y.-J. (2015) 'Viral Membrane Channels: Role and Function in the Virus Life Cycle', *Viruses*, 7(6), pp. 3261–3284. Available at: <https://doi.org/10.3390/v7062771>.
- Takata, H. *et al.* (2022) 'Long-term antiretroviral therapy initiated in acute HIV infection prevents residual dysfunction of HIV-specific CD8+ T cells', *eBioMedicine*, 84. Available at: <https://doi.org/10.1016/j.ebiom.2022.104253>.
- Tanaka, K. (2009) 'The proteasome: Overview of structure and functions', *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, 85(1), pp. 12–36. Available at: <https://doi.org/10.2183/pjab.85.12>.
- Tangye, S.G. *et al.* (2013) 'The good, the bad and the ugly - TFH cells in human health and disease', *Nature Reviews Immunology*, 13(6), pp. 412–426. Available at: <https://doi.org/10.1038/nri3447>.
- Tate, W.P. *et al.* (1995) 'Translational termination efficiency in both bacteria and mammals is regulated by the base following the stop codon', *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 73(11–12), pp. 1095–1103. Available at: <https://doi.org/10.1139/o95-118>.
- Ternette, N. *et al.* (2015) 'Early Kinetics of the HLA Class I-Associated Peptidome of MVA.HIVconsv-Infected Cells', *Journal of Virology*, 89(11), pp. 5760–5771. Available at: <https://doi.org/10.1128/JVI.03627-14>.
- Ternette, N. *et al.* (2016) 'Defining the HLA class I-associated viral antigen repertoire from HIV-1-infected human cells', *European Journal of Immunology*, 46(1), pp. 60–69. Available at: <https://doi.org/10.1002/eji.201545890>.
- Tewari, M.K. *et al.* (2005) 'A cytosolic pathway for MHC class II-restricted antigen processing that is proteasome and TAP dependent', *Nature Immunology*, 6(3), pp. 287–294. Available at: <https://doi.org/10.1038/ni1171>.
- Texier, C. *et al.* (2000) 'HLA-DR Restricted Peptide Candidates for Bee Venom Immunotherapy', *The Journal of Immunology*, 164(6), pp. 3177–3184. Available at: <https://doi.org/10.4049/jimmunol.164.6.3177>.
- Thakur, A., Marler, L. and Hinnebusch, A.G. (2019) 'A network of eIF2 β interactions with eIF1 and Met-tRNAi promotes accurate start codon selection by the translation preinitiation complex', *Nucleic Acids Research*, 47(5), pp. 2574–2593. Available at: <https://doi.org/10.1093/nar/gky1274>.

- Thanaraj, T.A. and Argos, P. (1996) 'Protein secondary structural types are differentially coded on messenger RNA', *Protein Science: A Publication of the Protein Society*, 5(10), pp. 1973–1983. Available at: <https://doi.org/10.1002/pro.5560051003>.
- Thèze, J. *et al.* (2011) 'HIV controllers: a multifactorial phenotype of spontaneous viral suppression', *Clinical Immunology (Orlando, Fla.)*, 141(1), pp. 15–30. Available at: <https://doi.org/10.1016/j.clim.2011.07.007>.
- Thomas, C. and Tampé, R. (2017) 'Proofreading of Peptide—MHC Complexes through Dynamic Multivalent Interactions', *Frontiers in Immunology*, 8. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2017.00065> (Accessed: 2 April 2023).
- Thomas, C. and Tampé, R. (2021) 'MHC I assembly and peptide editing — chaperones, clients, and molecular plasticity in immunity', *Current Opinion in Immunology*, 70, pp. 48–56. Available at: <https://doi.org/10.1016/j.coi.2021.02.004>.
- Tomaras, G.D. *et al.* (2008) 'Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia', *Journal of Virology*, 82(24), pp. 12449–12463. Available at: <https://doi.org/10.1128/JVI.01708-08>.
- Tomaras, G.D. *et al.* (2011) 'Polyclonal B cell responses to conserved neutralization epitopes in a subset of HIV-1-infected individuals', *Journal of Virology*, 85(21), pp. 11502–11519. Available at: <https://doi.org/10.1128/JVI.05363-11>.
- Tran, E. *et al.* (2014) 'Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer', *Science (New York, N.Y.)*, 344(6184), pp. 641–645. Available at: <https://doi.org/10.1126/science.1251102>.
- Tran, E., Robbins, P.F. and Rosenberg, S.A. (2017) "'Final common pathway" of human cancer immunotherapy: targeting random somatic mutations', *Nature Immunology*, 18(3), pp. 255–262. Available at: <https://doi.org/10.1038/ni.3682>.
- Trautmann, L. *et al.* (2006) 'Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction', *Nature Medicine*, 12(10), pp. 1198–1202. Available at: <https://doi.org/10.1038/nm1482>.
- Tsuji, T. *et al.* (2012) 'Heat Shock Protein 90-Mediated Peptide-Selective Presentation of Cytosolic Tumor Antigen for Direct Recognition of Tumors by CD4+ T Cells', *The Journal of Immunology*, 188(8), pp. 3851–3858. Available at: <https://doi.org/10.4049/jimmunol.1103269>.
- Turnbull, E.L. *et al.* (2006) 'HIV-1 epitope-specific CD8+ T cell responses strongly associated with delayed disease progression cross-recognize epitope variants efficiently', *Journal of Immunology (Baltimore, Md.: 1950)*, 176(10), pp. 6130–6146. Available at: <https://doi.org/10.4049/jimmunol.176.10.6130>.
- Tuyishime, M. *et al.* (2021) 'Elimination of SHIV Infected Cells by Combinations of Bispecific HIVxCD3 DART® Molecules', *Frontiers in Immunology*, 12. Available at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.710273> (Accessed: 29 March 2023).
- Uenaka, A. *et al.* (1994) 'Identification of a unique antigen peptide pRL1 on BALB/c RL male 1 leukemia recognized by cytotoxic T lymphocytes and its relation to the Akt oncogene', *The Journal of Experimental Medicine*, 180(5), pp. 1599–1607. Available at: <https://doi.org/10.1084/jem.180.5.1599>.
- Vagner, S., Galy, B. and Pyronnet, S. (2001) 'Irresistible IRES. Attracting the translation machinery to internal ribosome entry sites', *EMBO reports*, 2(10), pp. 893–898. Available at: <https://doi.org/10.1093/embo-reports/kve208>.
- Valle, M. *et al.* (2003) 'Locking and Unlocking of Ribosomal Motions', *Cell*, 114(1), pp. 123–134. Available at: [https://doi.org/10.1016/S0092-8674\(03\)00476-8](https://doi.org/10.1016/S0092-8674(03)00476-8).

- Van Bleek, G.M. and Nathenson, S.G. (1990) 'Isolation of an endogenously processed immunodominant viral peptide from the class I H-2Kb molecule', *Nature*, 348(6298), pp. 213–216. Available at: <https://doi.org/10.1038/348213a0>.
- Vander Lugt, B. *et al.* (2017) 'Transcriptional determinants of tolerogenic and immunogenic states during dendritic cell maturation', *The Journal of Cell Biology*, 216(3), pp. 779–792. Available at: <https://doi.org/10.1083/jcb.201512012>.
- Vanderperre, B. *et al.* (2011) 'An overlapping reading frame in the PRNP gene encodes a novel polypeptide distinct from the prion protein', *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 25(7), pp. 2373–2386. Available at: <https://doi.org/10.1096/fj.10-173815>.
- Vanhamel, J., Bruggemans, A. and Debysers, Z. (2019) 'Establishment of latent HIV-1 reservoirs: what do we really know?', *Journal of Virus Eradication*, 5(1), pp. 3–9.
- Vanhée-Brossollet, C. *et al.* (1995) 'A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals', *Virology*, 206(1), pp. 196–202. Available at: [https://doi.org/10.1016/s0042-6822\(95\)80034-4](https://doi.org/10.1016/s0042-6822(95)80034-4).
- Vanhée-Brossollet, Christine *et al.* (1995) 'A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals', *Virology*, 206(1), pp. 196–202. Available at: [https://doi.org/10.1016/S0042-6822\(95\)80034-4](https://doi.org/10.1016/S0042-6822(95)80034-4).
- Vasudevan, D. *et al.* (2020) 'Translational induction of ATF4 during integrated stress response requires noncanonical initiation factors eIF2D and DENR', *Nature Communications*, 11(1), p. 4677. Available at: <https://doi.org/10.1038/s41467-020-18453-1>.
- Veazey, R.S. *et al.* (1998) 'Gastrointestinal tract as a major site of CD4+ T cell depletion and viral replication in SIV infection', *Science (New York, N.Y.)*, 280(5362), pp. 427–431. Available at: <https://doi.org/10.1126/science.280.5362.427>.
- van der Velden, G.J. *et al.* (2012) 'Upstream AUG Codons in the Simian Immunodeficiency Virus SIVmac239 Genome Regulate Rev and Env Protein Translation', *Journal of Virology*, 86(22), pp. 12362–12371. Available at: <https://doi.org/10.1128/JVI.01532-12>.
- van der Velden, G.J. *et al.* (2013) 'An AUG codon upstream of rev and env open reading frames ensures optimal translation of the simian immunodeficiency virus Env protein', *Virology*, 436(1), pp. 191–200. Available at: <https://doi.org/10.1016/j.virol.2012.11.009>.
- Ventoso, I. *et al.* (2001) 'HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation', *Proceedings of the National Academy of Sciences of the United States of America*, 98(23), pp. 12966–12971. Available at: <https://doi.org/10.1073/pnas.231343498>.
- Verdegaal, E.M.E. *et al.* (2016) 'Neoantigen landscape dynamics during human melanoma-T cell interactions', *Nature*, 536(7614), pp. 91–95. Available at: <https://doi.org/10.1038/nature18945>.
- Verdegaal, E.M.E. and van der Burg, S.H. (2017) 'The Potential and Challenges of Exploiting the Vast But Dynamic Neoepitope Landscape for Immunotherapy', *Frontiers in Immunology*, 8, p. 1113. Available at: <https://doi.org/10.3389/fimmu.2017.01113>.
- Vermund, S.H. and Leigh-Brown, A.J. (2012) 'The HIV Epidemic: High-Income Countries', *Cold Spring Harbor Perspectives in Medicine*, 2(5), p. a007195. Available at: <https://doi.org/10.1101/cshperspect.a007195>.
- de Verteuil, D. *et al.* (2010) 'Deletion of immunoproteasome subunits imprints on the transcriptome and has a broad impact on peptides presented by major histocompatibility complex I molecules', *Molecular & cellular proteomics: MCP*, 9(9), pp. 2034–2047. Available at: <https://doi.org/10.1074/mcp.M900566-MCP200>.
- Vigano, S. *et al.* (2020) 'Cancer and HIV-1 Infection: Patterns of Chronic Antigen Exposure', *Frontiers in Immunology*, 11, p. 1350. Available at: <https://doi.org/10.3389/fimmu.2020.01350>.

- Vigneron, N. *et al.* (2004) 'An Antigenic Peptide Produced by Peptide Splicing in the Proteasome', *Science*, 304(5670), pp. 587–590. Available at: <https://doi.org/10.1126/science.1095522>.
- Vigneron, N. *et al.* (2019) 'Production of spliced peptides by the proteasome', *Molecular Immunology*, 113, pp. 93–102. Available at: <https://doi.org/10.1016/j.molimm.2018.03.030>.
- Villacé, P., Marión, R.M. and Ortín, J. (2004) 'The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs', *Nucleic Acids Research*, 32(8), pp. 2411–2420. Available at: <https://doi.org/10.1093/nar/gkh552>.
- Villegas, A. and Kropinski, A.M. (2008) 'An analysis of initiation codon utilization in the Domain Bacteria - concerns about the quality of bacterial genome annotation', *Microbiology (Reading, England)*, 154(Pt 9), pp. 2559–2661. Available at: <https://doi.org/10.1099/mic.0.2008/021360-0>.
- Vingert, B. *et al.* (2010) 'HIV controller CD4+ T cells respond to minimal amounts of Gag antigen due to high TCR avidity', *PLoS pathogens*, 6(2), p. e1000780. Available at: <https://doi.org/10.1371/journal.ppat.1000780>.
- Vingert, B. *et al.* (2012) 'HIV controllers maintain a population of highly efficient Th1 effector cells in contrast to patients treated in the long term', *Journal of Virology*, 86(19), pp. 10661–10674. Available at: <https://doi.org/10.1128/JVI.00056-12>.
- Wahren, B. *et al.* (1987) 'Characteristics of the specific cell-mediated immune response in human immunodeficiency virus infection', *Journal of Virology*, 61(6), pp. 2017–2023. Available at: <https://doi.org/10.1128/JVI.61.6.2017-2023.1987>.
- Wain-Hobson, S. *et al.* (1985) 'Nucleotide sequence of the AIDS virus, LAV', *Cell*, 40(1), pp. 9–17. Available at: [https://doi.org/10.1016/0092-8674\(85\)90303-4](https://doi.org/10.1016/0092-8674(85)90303-4).
- Walker-Sperling, V.E.K. *et al.* (2022) 'Therapeutic efficacy of combined active and passive immunization in ART-suppressed, SHIV-infected rhesus macaques', *Nature Communications*, 13(1), p. 3463. Available at: <https://doi.org/10.1038/s41467-022-31196-5>.
- Wang, R.F. *et al.* (1996) 'Utilization of an alternative open reading frame of a normal gene in generating a novel human cancer antigen', *The Journal of Experimental Medicine*, 183(3), pp. 1131–1140. Available at: <https://doi.org/10.1084/jem.183.3.1131>.
- Wearsch, P.A. and Cresswell, P. (2007) 'Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer', *Nature Immunology*, 8(8), pp. 873–881. Available at: <https://doi.org/10.1038/ni1485>.
- Webb, G.M. *et al.* (2018) 'The human IL-15 superagonist ALT-803 directs SIV-specific CD8+ T cells into B-cell follicles', *Blood Advances*, 2(2), pp. 76–84. Available at: <https://doi.org/10.1182/bloodadvances.2017012971>.
- Wei, J. *et al.* (2017) 'Varied Role of Ubiquitylation in Generating MHC Class I Peptide Ligands', *Journal of Immunology (Baltimore, Md.: 1950)*, 198(10), pp. 3835–3845. Available at: <https://doi.org/10.4049/jimmunol.1602122>.
- Wei, J. *et al.* (2019) 'Ribosomal Proteins Regulate MHC Class I Peptide Generation for Immunosurveillance', *Molecular Cell*, 73(6), pp. 1162–1173.e5. Available at: <https://doi.org/10.1016/j.molcel.2018.12.020>.
- Weingarten-Gabbay, S. *et al.* (2016) 'Systematic discovery of cap-independent translation sequences in human and viral genomes', *Science*, 351(6270), p. aad4939. Available at: <https://doi.org/10.1126/science.aad4939>.
- Weingarten-Gabbay, S. *et al.* (2021) 'Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs', *Cell*, 184(15), pp. 3962–3980.e17. Available at: <https://doi.org/10.1016/j.cell.2021.05.046>.
- Weinzierl, A.O. *et al.* (2008) 'A cryptic vascular endothelial growth factor T-cell epitope: identification and characterization by mass spectrometry and T-cell assays', *Cancer Research*, 68(7), pp. 2447–2454. Available at: <https://doi.org/10.1158/0008-5472.CAN-07-2540>.

- Weiss, L. *et al.* (1999) 'Restoration of normal interleukin-2 production by CD4+ T cells of human immunodeficiency virus-infected patients after 9 months of highly active antiretroviral therapy', *The Journal of Infectious Diseases*, 180(4), pp. 1057–1063. Available at: <https://doi.org/10.1086/315025>.
- Wen, Y. *et al.* (2009) 'Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis', *Nature Genetics*, 41(2), pp. 228–233. Available at: <https://doi.org/10.1038/ng.276>.
- Wethmar, K. (2014) 'The regulatory potential of upstream open reading frames in eukaryotic gene expression', *WIREs RNA*, 5(6), pp. 765–768. Available at: <https://doi.org/10.1002/wrna.1245>.
- Whisnant, A.W. *et al.* (2020) 'Integrative functional genomics decodes herpes simplex virus 1', *Nature Communications*, 11(1), p. 2038. Available at: <https://doi.org/10.1038/s41467-020-15992-5>.
- White, E. (2022) *Efficacité et Évolution de la réponse lymphocytaire T CD8 dans le contexte de l'infection à VIH-1*.
- Wilson, B.A. *et al.* (2017) 'Young Genes are Highly Disordered as Predicted by the Preadaptation Hypothesis of De Novo Gene Birth', *Nature Ecology & Evolution*, 1(6), pp. 0146–0146. Available at: <https://doi.org/10.1038/s41559-017-0146>.
- Wolf, T. *et al.* (2020) 'Dynamics in protein translation sustaining T cell preparedness', *Nature Immunology*, 21(8), pp. 927–937. Available at: <https://doi.org/10.1038/s41590-020-0714-5>.
- Wölfel, T. *et al.* (1995) 'A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma', *Science (New York, N.Y.)*, 269(5228), pp. 1281–1284. Available at: <https://doi.org/10.1126/science.7652577>.
- Wolin, S.L. and Walter, P. (1988) 'Ribosome pausing and stacking during translation of a eukaryotic mRNA.', *The EMBO Journal*, 7(11), pp. 3559–3569. Available at: <https://doi.org/10.1002/j.1460-2075.1988.tb03233.x>.
- Wolinsky, S.M. *et al.* (1992) 'Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants', *Science (New York, N.Y.)*, 255(5048), pp. 1134–1137. Available at: <https://doi.org/10.1126/science.1546316>.
- World Health Organization (WHO) (no date). Available at: <https://www.who.int/data/gho/data/themes/hiv-aids> (Accessed: 22 March 2023).
- Wright, B.W. *et al.* (2022) 'The dark proteome: translation from noncanonical open reading frames', *Trends in Cell Biology*, 32(3), pp. 243–258. Available at: <https://doi.org/10.1016/j.tcb.2021.10.010>.
- Wu, R.S. and Warner, J.R. (1971) 'Cytoplasmic synthesis of nuclear proteins. Kinetics of accumulation of radioactive proteins in various cell fractions after brief pulses', *The Journal of Cell Biology*, 51(3), pp. 643–652. Available at: <https://doi.org/10.1083/jcb.51.3.643>.
- Wu, T. *et al.* (2019) 'Quantification of epitope abundance reveals the effect of direct and cross-presentation on influenza CTL responses', *Nature Communications*, 10(1), p. 2846. Available at: <https://doi.org/10.1038/s41467-019-10661-8>.
- Xiao, Z. *et al.* (2018) 'De novo annotation and characterization of the translome with ribosome profiling data', *Nucleic Acids Research*, 46(10), p. e61. Available at: <https://doi.org/10.1093/nar/gky179>.
- Yang, O.O. *et al.* (1997) 'Suppression of human immunodeficiency virus type 1 replication by CD8+ cells: evidence for HLA class I-restricted triggering of cytolytic and noncytolytic mechanisms', *Journal of Virology*, 71(4), pp. 3120–3128. Available at: <https://doi.org/10.1128/JVI.71.4.3120-3128.1997>.
- Yang, Y. and Wang, Z. (2019) 'IRES-mediated cap-independent translation, a path leading to hidden proteome', *Journal of Molecular Cell Biology*, 11(10), pp. 911–919. Available at: <https://doi.org/10.1093/jmcb/mjz091>.
- Yang, Z. *et al.* (2015) 'Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling', *Journal of Virology*, 89(13), pp. 6874–6886. Available at: <https://doi.org/10.1128/JVI.00528-15>.

- Yao, R.-W., Wang, Y. and Chen, L.-L. (2019) 'Cellular functions of long noncoding RNAs', *Nature Cell Biology*, 21(5), pp. 542–551. Available at: <https://doi.org/10.1038/s41556-019-0311-8>.
- Yedavalli, V.S.R.K. *et al.* (2004) 'Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function', *Cell*, 119(3), pp. 381–392. Available at: <https://doi.org/10.1016/j.cell.2004.09.029>.
- Yewdell, J. (2002) 'To DRiP or not to DRiP: generating peptide ligands for MHC class I molecules from biosynthesized proteins', *Molecular Immunology*, 39(3), pp. 139–146. Available at: [https://doi.org/10.1016/S0161-5890\(02\)00097-4](https://doi.org/10.1016/S0161-5890(02)00097-4).
- Yewdell, J.W. (2003) 'Hide and Seek in the Peptidome', *Science*, 301(5638), pp. 1334–1335. Available at: <https://doi.org/10.1126/science.1089553>.
- Yewdell, J.W. (2022) 'MHC Class I Immunopeptidome: Past, Present, and Future', *Molecular & cellular proteomics: MCP*, 21(7), p. 100230. Available at: <https://doi.org/10.1016/j.mcpro.2022.100230>.
- Yewdell, J.W., Antón, L.C. and Bennink, J.R. (1996) 'Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules?', *Journal of Immunology (Baltimore, Md.: 1950)*, 157(5), pp. 1823–1826.
- Yewdell, J.W. and Hollý, J. (2020) 'DRiPs get molecular', *Current Opinion in Immunology*, 64, pp. 130–136. Available at: <https://doi.org/10.1016/j.coi.2020.05.009>.
- Yewdell, J.W. and Nicchitta, C.V. (2006) 'The DRiP hypothesis decennial: support, controversy, refinement and extension', *Trends in Immunology*, 27(8), pp. 368–373. Available at: <https://doi.org/10.1016/j.it.2006.06.008>.
- York, I.A. *et al.* (2002) 'The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8–9 residues', *Nature Immunology*, 3(12), pp. 1177–1184. Available at: <https://doi.org/10.1038/ni860>.
- Yoshinaka, Y. *et al.* (1985) 'Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon.', *Proceedings of the National Academy of Sciences*, 82(6), pp. 1618–1622. Available at: <https://doi.org/10.1073/pnas.82.6.1618>.
- Younes, S.-A. *et al.* (2003) 'HIV-1 viremia prevents the establishment of interleukin 2-producing HIV-specific memory CD4+ T cells endowed with proliferative capacity', *The Journal of Experimental Medicine*, 198(12), pp. 1909–1922. Available at: <https://doi.org/10.1084/jem.20031598>.
- Young, S.K. and Wek, R.C. (2016) 'Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response', *The Journal of Biological Chemistry*, 291(33), pp. 16927–16935. Available at: <https://doi.org/10.1074/jbc.R116.733899>.
- Youngblood, B. *et al.* (2013) 'Cutting edge: Prolonged exposure to HIV reinforces a poised epigenetic program for PD-1 expression in virus-specific CD8 T cells', *Journal of Immunology (Baltimore, Md.: 1950)*, 191(2), pp. 540–544. Available at: <https://doi.org/10.4049/jimmunol.1203161>.
- Zaunders, J.J. *et al.* (2004) 'Identification of circulating antigen-specific CD4+ T lymphocytes with a CCR5+, cytotoxic phenotype in an HIV-1 long-term nonprogressor and in CMV infection', *Blood*, 103(6), pp. 2238–2247. Available at: <https://doi.org/10.1182/blood-2003-08-2765>.
- Zaunders, J.J. *et al.* (2005) 'Early proliferation of CCR5(+) CD38(+++) antigen-specific CD4(+) Th1 effector cells during primary HIV-1 infection', *Blood*, 106(5), pp. 1660–1667. Available at: <https://doi.org/10.1182/blood-2005-01-0206>.
- Zhang, G., Hubalewska, M. and Ignatova, Z. (2009) 'Transient ribosomal attenuation coordinates protein synthesis and co-translational folding', *Nature Structural & Molecular Biology*, 16(3), pp. 274–280. Available at: <https://doi.org/10.1038/nsmb.1554>.
- Zhang, J. and Crumpacker, C. (2022) 'HIV UTR, LTR, and Epigenetic Immunity', *Viruses*, 14(5), p. 1084. Available at: <https://doi.org/10.3390/v14051084>.

- Zhang, W., Dunkle, J. and Cate, J.H.D. (2009) 'Structures of the ribosome in intermediate states of ratcheting', *Science (New York, N.Y.)*, 325(5943), pp. 1014–1017. Available at: <https://doi.org/10.1126/science.1175275>.
- Zhang, X. *et al.* (2019) 'Application of mass spectrometry-based MHC immunopeptidome profiling in neoantigen identification for tumor immunotherapy', *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, 120, p. 109542. Available at: <https://doi.org/10.1016/j.biopha.2019.109542>.
- Zheng, C.F. *et al.* (2007) 'Cytotoxic CD4+ T cells use granulysin to kill *Cryptococcus neoformans*, and activation of this pathway is defective in HIV patients', *Blood*, 109(5), pp. 2049–2057. Available at: <https://doi.org/10.1182/blood-2006-03-009720>.
- Zheng, C.F. *et al.* (2008) 'Late expression of granulysin by microbicidal CD4+ T cells requires PI3K- and STAT5-dependent expression of IL-2Rbeta that is defective in HIV-infected patients', *Journal of Immunology (Baltimore, Md.: 1950)*, 180(11), pp. 7221–7229. Available at: <https://doi.org/10.4049/jimmunol.180.11.7221>.
- Zhou, M. *et al.* (2015) 'Non-optimal codon usage influences protein structure in intrinsically disordered regions', *Molecular microbiology*, 97(5), pp. 974–987. Available at: <https://doi.org/10.1111/mmi.13079>.
- Zhu, J., Yamane, H. and Paul, W.E. (2010) 'Differentiation of effector CD4 T cell populations (*)', *Annual Review of Immunology*, 28, pp. 445–489. Available at: <https://doi.org/10.1146/annurev-immunol-030409-101212>.
- Zhu, P. *et al.* (2006) 'Distribution and three-dimensional structure of AIDS virus envelope spikes', *Nature*, 441(7095), pp. 847–852. Available at: <https://doi.org/10.1038/nature04817>.
- Zhu, T. *et al.* (1993) 'Genotypic and phenotypic characterization of HIV-1 patients with primary infection', *Science (New York, N.Y.)*, 261(5125), pp. 1179–1181. Available at: <https://doi.org/10.1126/science.8356453>.
- Ziegler, M.C. *et al.* (2020) 'HIV-1 induced changes in HLA-C*03:04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors', *AIDS (London, England)*, 34(12), pp. 1713–1723. Available at: <https://doi.org/10.1097/QAD.0000000000002596>.
- Zimmerman, E.F. *et al.* (1969) 'Protein synthesis in isolated nuclei and nucleoli of HeLa cells', *Biochemistry*, 8(6), pp. 2636–2644. Available at: <https://doi.org/10.1021/bi00834a058>.
- Zinkernagel, R.M. and Doherty, P.C. (1974) 'Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system', *Nature*, 248(5450), pp. 701–702. Available at: <https://doi.org/10.1038/248701a0>.
- Zinkernagel, R.M. and Doherty, P.C. (1997) 'The discovery of MHC restriction', *Immunology Today*, 18(1), pp. 14–17. Available at: [https://doi.org/10.1016/S0167-5699\(97\)80008-4](https://doi.org/10.1016/S0167-5699(97)80008-4).
- Zolla-Pazner, S. (2004) 'Identifying epitopes of HIV-1 that induce protective antibodies', *Nature Reviews. Immunology*, 4(3), pp. 199–210. Available at: <https://doi.org/10.1038/nri1307>.
- Zolla-Pazner, S. and Cardozo, T. (2010) 'STRUCTURE–FUNCTION RELATIONSHIPS OF HIV-1 ENVELOPE SEQUENCE-VARIABLE REGIONS PROVIDE A PARADIGM FOR VACCINE DESIGN', *Nature reviews. Immunology*, 10(7), pp. 527–535. Available at: <https://doi.org/10.1038/nri2801>.

Titre : Définition du translome du VIH-1 dans le but d'identifier de nouveaux antigènes reconnus par les lymphocytes T

Mots clés : VIH-1, traduction, ribosome profiling, immunopeptidome, lymphocytes T CD4⁺ et CD8⁺, polyfonctionnalité

Résumé : Les récentes avancées dans les méthodes de détection ont remis en question la définition traditionnelle du génome codant en révélant que des milliers de cadres ouverts de lecture non canoniques, tels que les cadres de lecture alternatifs (ARFs), sont capables de coder des polypeptides. Bien que les ARFs aient été identifiés dans plusieurs génomes viraux grâce à la technique du ribosome profiling, tels que ceux du HCMV, KSV et SARS-CoV-2, aucune démonstration n'avait pas été faite sur le génome du VIH-1. En utilisant le ribosome profiling, nous avons identifié 98 ARFs du VIH-1 répartis sur l'ensemble du génome, avec une forte conservation parmi les isolats des clades B et C du VIH-1. Notre étude a révélé qu'au moins 42 ARFs codent des polypeptides viraux, ce qui est attesté par la détection de réponses T spécifiques de 46 peptides dérivés d'ARFs chez des donneurs sous traitement ou contrôlant naturellement l'infection.

Parallèlement, nous avons identifié dans des cellules primaires infectées un ligand de HLA-A*0201 dérivé d'un ARF identifié. Dans l'ensemble, les réponses T spécifiques des peptides dérivés d'ARFs étaient majoritairement médiées par des cellules T CD4⁺ polyfonctionnelles sécrétant au moins 3 cytokines simultanément. Notre étude met en évidence l'immense capacité de codage du VIH-1 et l'intérêt d'étudier la traduction des régions génomiques non canoniques dans les maladies infectieuses. Notre découverte élargit la liste des polypeptides viraux conservés qui pourraient être des cibles potentielles pour les stratégies de vaccination.

Title : Defining the translome of HIV-1 to identify novel conserved T-cell antigens.

Keywords : HIV-1, translation, ribosome profiling, immunopeptidome, CD4⁺ and CD8⁺ T-cells, polyfunctionality

Abstract : Recent advances in detection methods challenged the coding definition revealing that thousands of non-canonical open reading frames, including alternative reading frames (ARFs), encode polypeptides. ARFs have been detected in several viral genomes using ribosome profiling such as HCMV, KSV, and SARS-CoV-2 but no demonstration has been made on HIV-1 genome.

By utilizing ribosome profiling, we uncovered 98 HIV ARFs that are distributed throughout the genome, and highly conserved among HIV-1 clades B and C isolates. Our analysis revealed that at least 42

ARFs encode viral polypeptides, as demonstrated by T-cell responses targeting 46 ARF-derived peptides in donors under treatment or naturally controlling the infection.

At the same time, we identified on primary infected cells a ligand of HLA-A*0201 derived from an identified ARF. Overall, the ARF-specific T-cell responses were mediated by polyfunctional CD4⁺ T-cells secreting at least 3 cytokines simultaneously. Our study highlights the immense translation landscape of HIV-1 genome and the interest to study hidden alternative translated genomic regions in infectious diseases. Our discovery expands the list of conserved viral polypeptides that might be potential targets for vaccine strategies.