



HAL
open science

Séquençage du VIH : apport du NGS en routine et comparaison avec la méthode Sanger

Thomas Lhossein

► **To cite this version:**

Thomas Lhossein. Séquençage du VIH : apport du NGS en routine et comparaison avec la méthode Sanger. Sciences pharmaceutiques. 2021. dumas-03605576

HAL Id: dumas-03605576

<https://dumas.ccsd.cnrs.fr/dumas-03605576v1>

Submitted on 22 Feb 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License



UNIVERSITE DE PICARDIE JULES VERNE
UFR DE PHARMACIE

Thèse n°

Année : 2021

THESE D'EXERCICE
POUR LE
DIPLÔME D'ETAT DE DOCTEUR EN PHARMACIE
Par **M. LHOSSEIN Thomas**

**Séquençage du VIH: Apport du NGS en routine et
comparaison avec la méthode Sanger**

Présentée et soutenue publiquement

le 27 Octobre 2021

PRESIDENT DU JURY : Monsieur le Professeur Gilles DUVERLIE

JUGES : Madame le Professeur Sandrine CASTELAIN
Monsieur le Docteur Jean-Philippe LANOIX

DIRECTEUR DE THESE : Monsieur le Docteur Etienne BROCHOT

Remerciements

Aux membres du jury :

A Monsieur le Professeur Gilles DUVERLIE de me faire l'honneur de juger et présider cette thèse. Veuillez accepter mes sincères remerciements et ma reconnaissance

A Madame le Professeur Sandrine CASTELAIN de me faire l'honneur de siéger à ce jury et de juger cette thèse.

A monsieur le Docteur Jean-Philippe LANOIX – Maitre de conférences des Universités – de me faire l'honneur pour l'intérêt que vous portez sur ce travail en ayant accepté de siéger à ce jury. Je vous remercie de juger ce travail en qualité de chef de service en infectiologie.

A Monsieur le Docteur Etienne BROCHOT qui a dirigé cette thèse, je vous remercie chaleureusement pour votre encadrement, votre disponibilité et efficacité tout au long de ce travail sans quoi rien n'aurait été possible.

A mes collègues :

A Véronique DESCAMPS de m'avoir « coaché » et soutenu pour la technique. Je te remercie chaleureusement d'avoir pris part à l'aventure NGS avec moi.

A Karine SYLVAIN pour ta bonne humeur, ta disponibilité sans faille et ton aide sur la technique Sanger.

A Didier HERENT pour ta gentillesse et l'ensemble de tes précieux conseils qui m'auront permis de mieux cerner les rouages de l'Illumina.

A Cédric USUREAU pour m'avoir initié à la bio-informatique, cette science obscure.

A l'ensemble des techniciens du laboratoire de Virologie. Une équipe très solide en ces temps difficiles sur qui l'on peut compter en toute circonstance.

A tous les biologistes du laboratoire du CHU d'Amiens pour tous ce que vous m'avez appris tout au long de mon cursus de biologie médicale.

Une pensée à Karine DURAND, MCU-PH au CHU de Limoges pour m'avoir donné goût à la biologie moléculaire et de m'avoir fait découvrir le NGS lorsque j'étais externe. Une pensée également à Murat SAHIN pour ta formation folklorique aux techniques de biologie moléculaire.

A Adrien CHAN SUI KO, pour nos franches parties de rigolade au laboratoire ainsi que notre dialogue clinico-biologique de qualité.

A Emeline et Yann pour toutes les soirées clandestines saupoudrées de sushis et d'hamburgers.

A ma promo : Brahim, Justine, Alan, Elodie et Floriane, je vous remercie pour ces années d'internat passées dans la bonne humeur.

A Kenza et Julie, je vous remercie de m'avoir supporté pendant le stage de parasitologie-mycologie.

A ma "team bactério" : Aurélien, Éric, Mamoune, Victor et Sophie pour votre efficacité sans faille et votre sympathie.

A Kélian, pour toutes les parties d'échecs endiablées.

A tous les autres co-internes, que je n'oublie pas...

A mes vieux copains de la fac, Guillaume et Nicolas, rendez-vous « rue de la Soif » à mon retour dans le Limousin !

A ma famille :

A Eliane, Bernard, le GPX Vincent, Sylvie, Emilie et Mila pour leur aide et pour leur présence auprès de moi tout au long de mes études. Petite pensée à Manguine pour son affection.

En mémoire de mes grands-parents Pierre et Yolande qui auraient tellement été heureux d'être présents à cet événement.

Table des Matières

Table des Illustrations	1
Liste des Tableaux	2
Table des Annexes	3
I. Introduction générale	5
A. La diversité génétique du HIV	5
B. Epidémiologie du HIV en France : Etats actuels des chiffres	6
C. Structure et organisation génomique du HIV	9
1. Les enzymes virales :	10
2. Les Gènes viraux régulateurs et accessoires :	12
D. Le diagnostic de la résistance virale :.....	13
1. Les tests phénotypiques :	13
2. Les différentes techniques de génotypage :.....	13
a) La Méthode Sanger :	14
b) La Méthode Illumina®:(32).....	15
c) La technologie 454 de Roche® :	16
d) La technologie Ion torrent® : Life technologies	17
E. Les différentes mutations recherchées	18
F. Impact du sous type du VIH-1 sur la résistance des antirétroviraux :.....	20
G. Intérêt du dosage de l'ADN proviral ? :.....	20
II. Matériels et Méthodes	22
A. Table des échantillons analysés	22
B. Extraction de l'ARN	23
C. La « Workflow » du NGS : Illumina®	24
D. L'Amplification des régions cibles	25
1. La RT-PCR :.....	25
2. Evaluation de la qualité et quantité d'ADN amplifié :	26
3. La Nested-PCR :.....	27
E. Purification des Amplicons	29
1. Principe de la purification :.....	29
2. Protocole de purification :.....	30
F. Quantification, Dilution et pooling	31
G. Préparation de la librairie	32
1. La Fragmentation :.....	32
2. La Réparation et le « A-tailing » :.....	33
3. La ligation des adaptateurs :.....	33
4. Amplification de la librairie :.....	35
H. Contrôle qualité des échantillons et préparation au séquençage	36

I. Le Protocole du séquençage Illumina®	37
J. La technique Sanger ABL.....	39
1. Dilution :	39
2. Purification :	39
3. Marquage et Séquençage :	39
4. Sanger avec les amorces de l'ANRS	40
III. L'analyse bio-informatique.....	41
A. Mise en place de l'environnement de travail :	41
B. L'appel des bases du séquençage :	41
C. Alignement sur le génome de référence :	44
1. La Transformée de Burrows-Wheeler (BWT) :	45
2. Les différents formats de fichiers :	45
3. Critères de qualités d'un alignement :	46
4. Visualisation des lectures :	47
D. L'Appel des Variants:.....	48
E. Analyse des séquences Sanger :	48
IV. Résultats :.....	49
A. Contrôle qualité de l'amplification RT-PCR:	49
B. Détection des mutations d'intérêts	50
C. Détection des différents sous-types	53
1. Les sous-types sur la protéase pour chaque technique	54
2. Les sous-types sur la reverse transcriptase pour chaque technique	54
3. Les sous-types sur l'intégrase pour chaque technique.....	55
4. Comparaison des sous-types détectés dans la technique ANRS	56
D. Comparaison du nombre de mutations d'intérêts entre NGS et Sanger	56
E. Comparaison du nombre de mutations dans la technique ANRS	57
F. Concordance entre les différentes techniques	58
G. L'apport quantitatif du NGS dans la détection des mutations	59
H. Etude du prix de l'analyse pour chaque technique.....	60
V. Discussion	62
BIBLIOGRAPHIE.....	66
ANNEXE.....	71

Table des Illustrations

Figure 1 : Arbre phylogénétique représentant la diversité génétique du HIV1.....	6
Figure 2 : Nombre de sérologies VIH remboursées, France, janvier 2015-juillet 2020 – SNDS.....	7
Figure 3 : Prévalence mondiale des différents sous-types du VIH.....	8
Figure 4 : Prévalence du sous-type de VIH en France.....	8
Figure 5 : Organisation génomique du HIV1.....	9
Figure 6 : Structure moléculaire de la Reverse transcriptase.....	11
Figure 7 : Principe de séquençage Sanger.....	14
Figure 8 : Principe d'après Illumina®.....	15
Figure 9 : Principe schématique de la technologie 454.....	17
Figure 10 : Principe de la technologie Ion Torrent®.....	18
Figure 11 : Principe de l'extraction automatisée sur eMag®.....	23
Figure 12 : les différentes étapes du NGS avec la technologie Illumina®.....	24
Figure 13 : Principe de la purification des Amplicons à l'aide de billes magnétiques.....	29
Figure 14 : Protocole du dosage Qubit®.....	31
Figure 15 : Les différentes étapes de la préparation de la librairie de nos échantillons.....	32
Figure 16 : Principe du Bioanalyser® adapté d'Agilent.....	36
Figure 17 : Préparation de la cartouche avant analyse dans le séquenceur Illumina®.....	38
Figure 18 : L'intérêt du PhiX dans la diversité nucléotidique lors du séquençage illumina®.....	38
Figure 19 : description du format Fastq issue des données de séquençage.....	42
Figure 20 : Distribution des qualités à chaque position de l'ensemble des reads.....	44
Figure 21 : Le principe d'alignement. Mapping des lectures.....	44
Figure 22 : Schéma explicatif simplifié de la transformée de Burrows-Wheeler.....	45
Figure 23 : description du format SAM issue des données de séquençage.....	46
Figure 24 : Visualisation de l'alignement à l'aide du logiciel IGV.....	47
Figure 25 : Description du fichier VCF.....	48
Figure 26 : Résultats des Gels de la première série de NGS.....	49
Figure 27 : Résultats des Gels de la seconde série de NGS.....	49
Figure 28 : Diagrammes représentant les sous-types retrouvés avec la Protéase.....	54
Figure 29 : Diagrammes représentant les sous-types retrouvés avec la Reverse transcriptase.....	55
Figure 30 : Diagrammes représentant les sous-types retrouvés avec l'Intégrase.....	55
Figure 31 : Diagrammes représentant le profil des 10 échantillons testés sur les 4 techniques.....	56
Figure 32 : Comparaison des mutations d'intérêt entre Viroseq et NGS.....	56
Figure 33 : Comparaison des mutations d'intérêt entre Viroseq, ABL et NGS.....	57
Figure 34 : Comparaison des mutations d'intérêt entre les 4 techniques.....	57

Liste des Tableaux

Tableau 1 : Gènes, protéines et leurs fonctions chez HIV-1	10
Tableau 2 : Liste non exhaustive de la résistance du VIH-1 du groupe M aux antirétroviraux	19
Tableau 3 : Liste des échantillons analysés.....	22
Tableau 4 : Mix RT-PCR de la région Reverse transcriptase.....	25
Tableau 5 : Mix RT-PCR de la région Protéase	25
Tableau 6 : Mix RT-PCR de la région Intégrase	26
Tableau 7 : pour la réalisation de la RT PCR / Table pour la réalisation de la RT PCR Intégrase	26
Tableau 8 : PCR-Nested pour la région de la Reverse transcriptase.....	27
Tableau 9 : PCR-Nested pour la région de la Protéase	28
Tableau 10 : PCR-Nested pour la région de l'intégrase	28
Tableau 11 : Protocole Nested-PCR thermocycleur : Integrase.....	28
Tableau 12 : Protocole Nested-PCR thermocycleur : Reverse Transcriptase et Protéase.....	29
Tableau 13 : Composants de la réaction de fragmentation	32
Tableau 14 : Programme de la fragmentation.....	33
Tableau 15 : Composants de la réaction pour la réparation et le « A-tailing »	33
Tableau 16 : Programme du thermocycleur de la réparation et du « A-tailing ».....	33
Tableau 17 : Les composants du pre-mix de ligation.....	34
Tableau 18 : Les réactifs de l'étape de la ligation ainsi que le programme du thermocycleur.	34
Tableau 19 : Liste des combinaisons d'adaptateurs ainsi que leurs séquences.....	34
Tableau 20 : Les différents volumes du protocole pour la 2ème purification.....	35
Tableau 21 : Le mix pour l'amplification de la librairie.....	35
Tableau 22 : Programme de l'amplification de la librairie.....	35
Tableau 23 : Les différents volumes du protocole pour la 3ème purification.....	36
Tableau 24 : Dilution à réaliser en fonction de la taille des bandes sur le gel d'agarose	39
Tableau 25 : Quantité des réactifs à ajouter dans chaque puit et le programme de la réaction	40
Tableau 26 : Listing des Amorces de l'ANRS	40
Tableau 27 : Score Phred et score de confiance	42
Tableau 28 : ASCII. Exemple : Un score Phred de 30	43
Tableau 29 : Liste des mutations d'intérêts retrouvées pour la Protéase.....	51
Tableau 30 : Liste des mutations d'intérêts retrouvées pour la Reverse transcriptase et l'Intégrase.	52
Tableau 31 : Liste des sous-types détectés pour chaque technique et chaque cible.....	53
Tableau 32 : Table de contingence et coefficient de concordance Kappa	58
Tableau 33 : Liste non exhaustive des mutations discordantes pour Viroseq et ABL associée à la quantification NGS.	59
Tableau 34 : Impact de la sensibilité de certaines molécules.....	60
Tableau 35 : Prix brut à l'échantillon et estimation en fonction d'éventuelles remises	61

Table des Annexes

Annexe I : Lignes de code sur commande linux.....	71
Annexe II : Table brute des résultats de séquençage au seuil de 20%.....	72
Annexe III : table des différents sous types avec le pourcentage de similitude des séquences.	81
Annexe IV : Table brute de l'ensemble des mutations trouvées en NGS <20%.	84
Annexe V : Table brute des résultats de séquençage avec la technique ANRS	85
Annexe VI : Position de chaque échantillon sur le gel d'agarose..	87
Annexe VII : Contrôle qualité des librairies : Profil Bioanalyser®	87
Annexe VIII : Contrôle technique par l'instrument Miseq Illumina®	89
Annexe IX : Visualisation du chromatogramme pour l'échantillon 20315149.	90
Annexe X : Exemple d'un chromatogramme de bonne qualité.....	90
Annexe XI : Exemple de compte rendu pour le NGS et le Sanger ABL	91

Présentation de la thèse :

C'est dans le contexte des avancées technologiques dans le domaine de la biologie moléculaire que ce travail se penche sur une technique 'nouvelle génération' de séquençage haut débit (SHD) en virologie, plus particulièrement sur le VIH.

On peut également y voir un intérêt dans un avenir proche, compte tenu de la probabilité d'une forte demande de génotypage des virus émergents.

Après une introduction sur les généralités concernant le VIH et les différentes techniques de SHD disponibles sur le marché, plusieurs notions seront abordées, notamment les aspects techniques, analytiques et bio-informatiques.

Cliniquement, pour la recherche de mutation chez le VIH, c'est le seuil de 20 % qui est utilisé. Avec le SHD, nous essaierons de voir l'intérêt de descendre ce seuil à 3 % (bonne fiabilité de la technique).

L'aspect économique sera également évoqué, en recherchant dans ce but les pistes possibles pour la mise en place d'une technique 'maison'.

Enfin, une problématique se présente au sein du laboratoire de virologie du CHU d'Amiens, le kit ViroSeq ne sera plus commercialisé et deux techniques de séquençage Sanger (ABL et ANRS) seront mises en œuvre afin de les comparer en terme de mutation.

I. Introduction générale

A. La diversité génétique du HIV

Deux rétrovirus ont été identifiés comme agents responsables de l'immunodéficience acquise chez l'être humain. Le VIH-1 (Human immunodeficiency virus) et le VIH-2 de la famille des *Retroviridae* et du genre *lentivirus* sont la résultante de plusieurs transmissions inter-espèces de virus simiens à l'homme.

Chaque transmission inter-espèces a permis une évolution génétique du virus. En effet, on distingue chez le HIV-1, le groupe M (majeur), N, O et P. Le groupe M est divisé en 9 sous-groupes nommés A-D, F-H et J-K. Certains d'entre eux ont également des sous-types : A1-A3 ; F1-F2. Cette nomenclature est le résultat d'une évolution dite « discontinue » par la diffusion de divers variants à des zones géographiques différentes, ce qui laisse une grande diversité génétique du rétrovirus.(1)

Il existe actuellement 95 formes recombinantes circulantes (CRF, Circulating recombinant form)(2) pouvant être responsables d'épidémies comme on peut le voir en Asie du Sud-Est avec le CRF01_AE(3) et CRF02_AG en Afrique de l'Ouest.(4) Ainsi que de formes recombinantes uniques (URF, Unique recombinant form) limitées à un nombre restreint de patients.

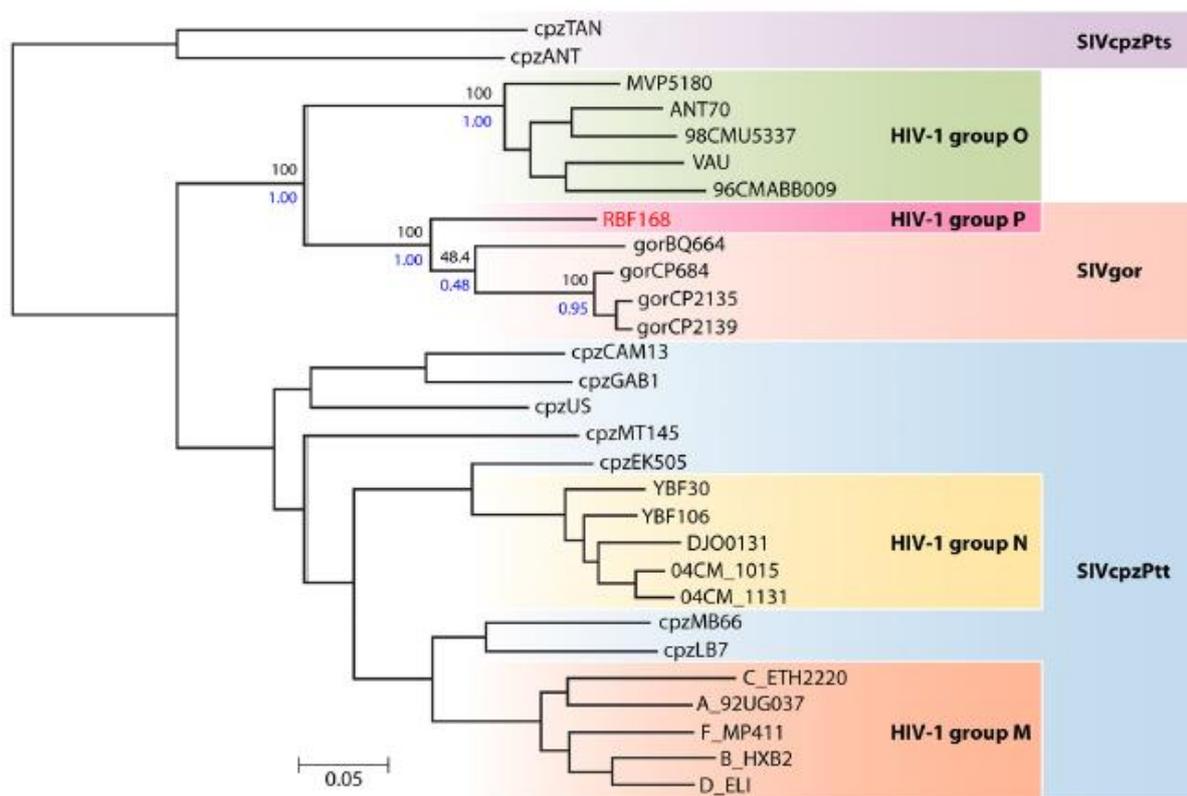


Figure 1 : Arbre phylogénétique représentant la diversité génétique du HIV1 ainsi que leur origine simienne. SIVcpzPts (*Pan troglodytes schweinfurthii*), SIVcpzPtt (*Pan troglodytes troglodytes*), SIVgor (gorilles) (5)

B. Epidémiologie du HIV en France : Etats actuels des chiffres

Selon les dernières données de la santé publique en France (6), l'incidence a globalement diminué entre 2003 et 2014. Une méthode développée par l'Inserm UMRS 1136 indique environ 6000 nouvelles contaminations en 2018 avec un délai médian de 3,6 ans entre la contamination et le diagnostic.

En 2019, on dénombre une activité de dépistage en hausse, de l'ordre de 6,2 millions de sérologies VIH qui ont été réalisées par les laboratoires médicaux. Le nombre de sérologies positives et confirmées est estimé à 11 488 [10 997-11 979](7). On dénote que l'évolution annuelle ne suit aucune tendance particulière et est même plutôt en diminution avec un taux de positivité de 1,9 pour 1000 sérologies réalisées. (8)

Mais malgré une forte activité de dépistage par les laboratoires, le nombre de découvertes de séropositivité n'a pu réellement être calculé pour 2019 par manque de données issues de la déclaration obligatoire de l'infection à VIH. On dénote globalement une sous déclaration plus importantes par rapport aux années précédentes qui peut s'expliquer par la crise sanitaire à SARS-CoV-2 début 2020.

A l'aide d'analyses des chiffres du Système national des données de santé (SNDS), (Figure2), on observe une chute beaucoup plus importante entre mars et avril 2020. Le nombre de sérologies réalisées a diminué de 56%, entre février et avril et ceux dans toutes les tranches d'âges et de sexe. Cette chute n'a malheureusement pas été compensée en juin-Juillet ce qui peut laisser craindre un impact défavorable en terme de prise en charge de patients naïfs ainsi que de la prévention de la transmission par traitement (9) . Par exemple, en 2019 il y a eu 26% de personnes diagnostiquées à un stade avancé de la maladie, donc sans bénéfice d'un traitement précoce.(9)

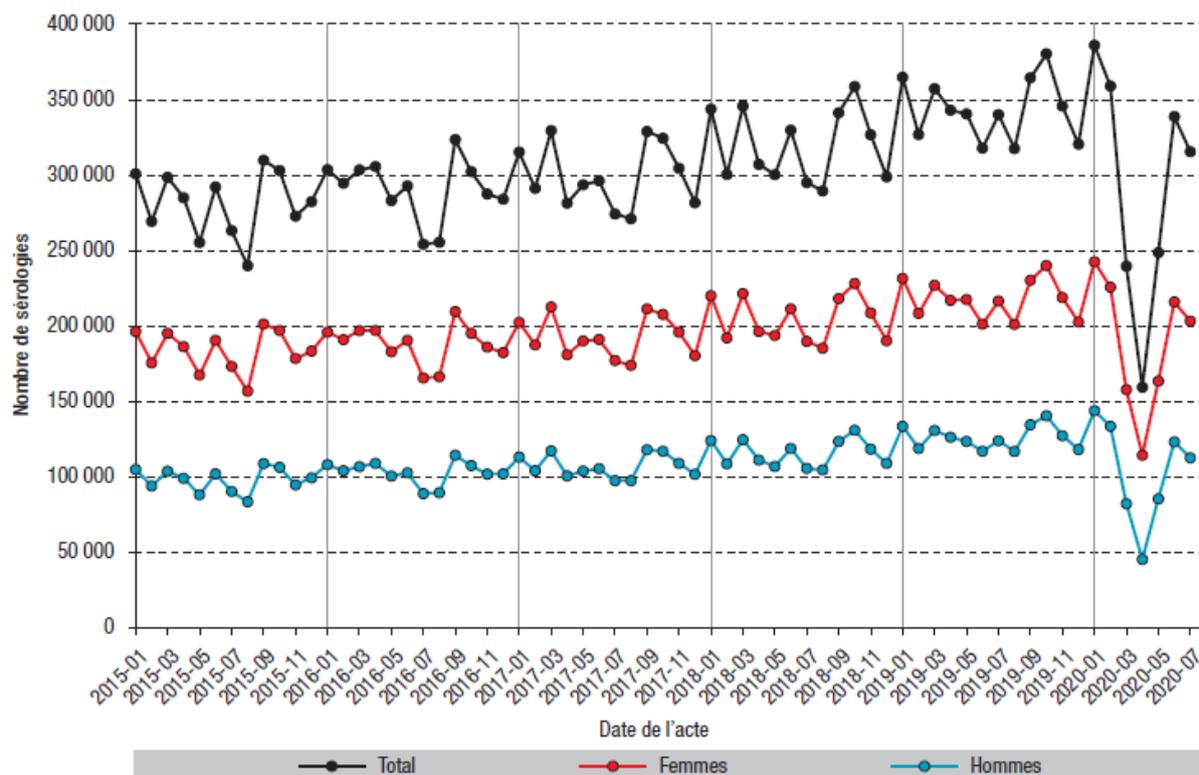


Figure 2 : Nombre de sérologies VIH remboursées, secteur privé, par sexe. France, janvier 2015-juillet 2020 – SNDS

Les diagrammes circulaires suivants indiquent la prévalence de l'infection par les différents sous-types. Les différents chiffres sont issus du bulletin épidémiologique hebdomadaire du suivi du VIH entre 2003 et 2013 par les différents réseaux de surveillance. La moitié des virus non-B isolés en France est la forme recombinante CRF02-AG.(10)

Pour ce qui est des groupes non-M, la prévalence du groupe O est de l'ordre de 0,1% et on note que les groupes N et P sont extrêmement rares, importés du Cameroun et très peu de cas diagnostiqués en France.(11)

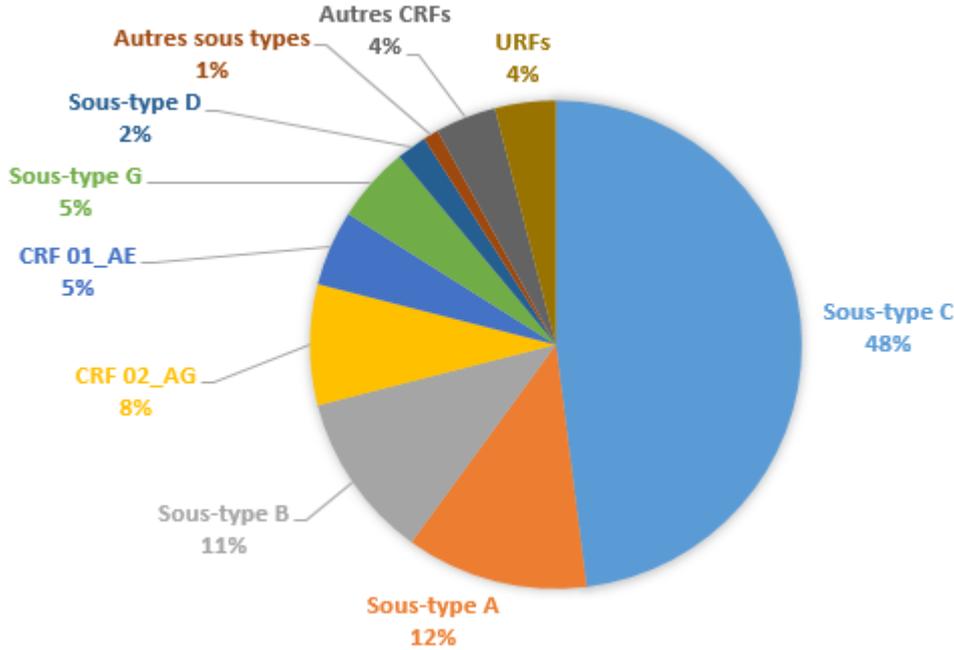


Figure 3 : Prévalence mondiale des différents sous-types du VIH

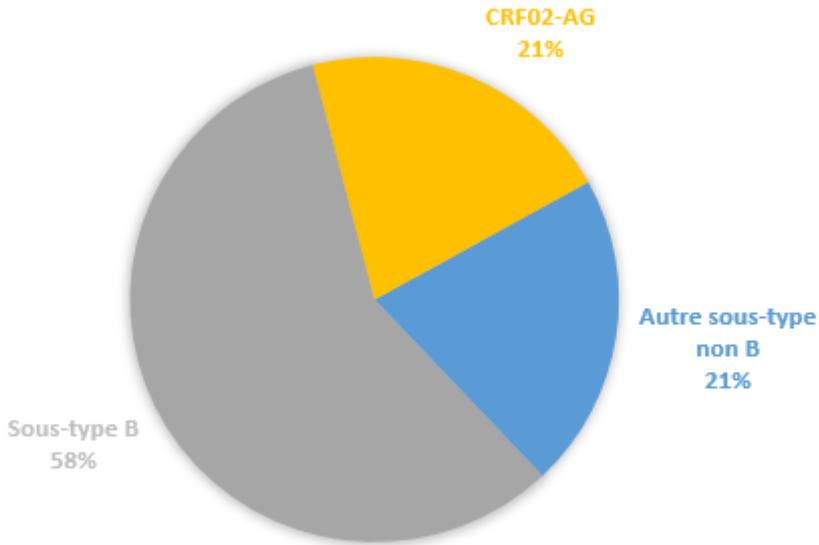


Figure 4 : Prévalence du sous-type de VIH en France

C. Structure et organisation génomique du HIV

L'ADN proviral du HIV-1 intégré dans la cellule hôte est long de 9,7kb. Son génome se compose des gènes gag-pol-env flanqués de deux LTR viraux complets au niveau de chaque extrémité. Un LTR est une séquence terminale longue répétée caractéristique des rétrovirus.

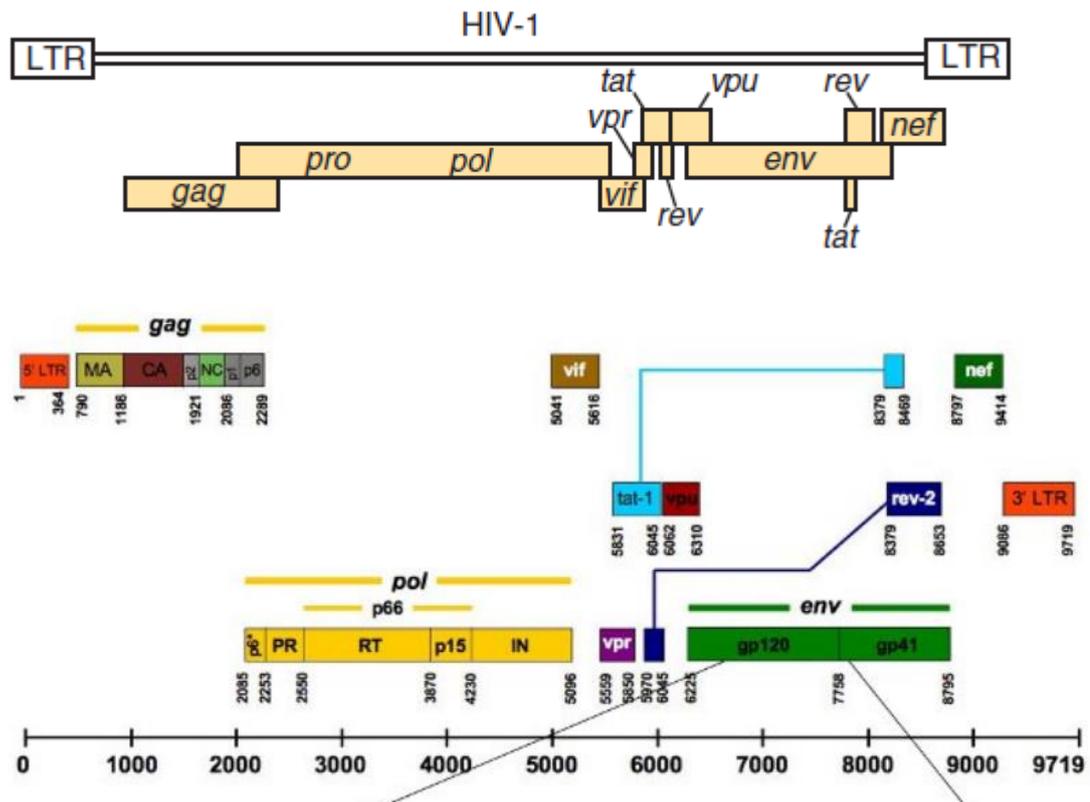


Figure 5 : Organisation génomique du HIV1 ; env, Enveloppe ; gag, antigène spécifique du groupe ; LTR, longue répétition terminale ; nef, facteur de régulation négatif ; ORF, cadres de lecture ouverts ; pol, polymérase ; pro, protéase ; rev, régulateur de l'expression virale ; rex, régulateur de l'expression virale ; tat, trans-activateur ; tax, trans-activateur de la transcription ; vif, virion facteur d'infectivité ; vpr, protéine virale R ; vpu, protéine virale U.

Ces LTR contiennent des séquences de régulation transcriptionnelles ainsi que tout le matériel nécessaire au traitement de l'ARN que ce soit gestion des signaux, sites d'intégration et de conditionnement. La fin de 5' commence avec le gène gag, codant des protéines matricielles ainsi que d'autres protéines principales. Le gène pol quant à lui, code pour des éléments essentiels utiles dans le séquençage et la recherche de mutation et sont : la protéase virale, la RT et l'intégrase. Puis le gène env, responsable du codage de l'enveloppe externe et transmembranaire du virus.

A ces différents gènes de structures, s'ensuit des gènes accessoires codant des protéines régulatrices de la multiplication virale : tat, rev, nef, vpr, vif et vpu. (pour le HIV-1)

Tableau 1 : Gènes, protéines et leurs fonctions chez HIV-1 (Arthur LO, Bess JW Jr, Sowder RC 2nd, et al (12)

Gene	Protéines	Taille (kDa)	Fonctions
gag	P17 P24 P6 P7 P1 P2		Protéine matricielle ; interagit avec la gp41; Protéine essentielle ; se lie à la vpr Nucléocapside ; se lie à l'ARN
pol	Protéase Reverse Transcriptase Intégrase	10 66, 51 32	Clivage protéolytique de Gag et Pol Activité de la polymérase et de la RNase H (p66 seulement) Intégration dans le chromosome
env	gp120 gp41		Enveloppe ; entrée du virus dans la cellule Protéine transmembranaire ; fusion cellulaire
vif	Virion infectivity protein	23	Transmission cellulaire
vpr	Viral protein R	18	Améliore la réplication virale dans les cellules primaires, protéine associée au virion ; arrêt de phase G ₂ /M ; localisation nucléaire
tat	Trans-activator of transcription	14	Principal trans-activateur viral, Suppression immunitaire
rev	Regulator of expression of virion protein	19	Renforce l'expression des ARN non épissés et épissés.
vpu	Viral protein U	15-16	Augmente la libération de virions des cellules, Régulation négative des CD4 et CMH de type I
nef	Negative regulator factor	27	Inhibe ou améliore la réplication virale en fonction de la souche et du type de cellule Régulation négative des CD4 ; CMH de type I Antiapoptose

1. Les enzymes virales :

La Protéase est une protéine de 99 AA de 10 kDa. Elle est entièrement active sous forme de dimère. Elle se clive par le biais d'une réaction autocatalytique à partir de protéines précurseurs au cours du processus d'assemblage du virus. La mutagénèse spécifique au site a démontré que des particules non infectieuses contenant des protéines Gag et Gag-Pol non clivées sont produites si cette enzyme est inactivée. C'est grâce à la similarité de cette protéase virale à d'autres protéases de type aspartique, comme par exemple l'enzyme de conversion de l'angiotensine, qui a grandement facilité la conception de puissants antiviraux, notamment des inhibiteurs de la dimérisation et des molécules se liant au site catalytique actif.(13) -(14)

La Reverse transcriptase (RT) est une ADN polymérase ARN dépendante et ADN dépendante. C'est une enzyme très polyvalente qui synthétise des copies d'ADN à partir d'ARN. Son activité RNase H est indispensable à la réplication virale, elle permet d'hydrolyser le brin d'ARN dans un double brin hybride ADN:ARN. La RT est d'abord clivée par une polyprotéine pour former un homodimère p66, puis un hétérodimère p66/p51 dans un second clivage contenant des extrémités amino-terminales identiques. Cette structure de l'hétérodimère RT a donc révélé son mécanisme enzymatique de la transcription inverse et constitue une base moléculaire de la résistance aux médicaments antiviraux.(15)

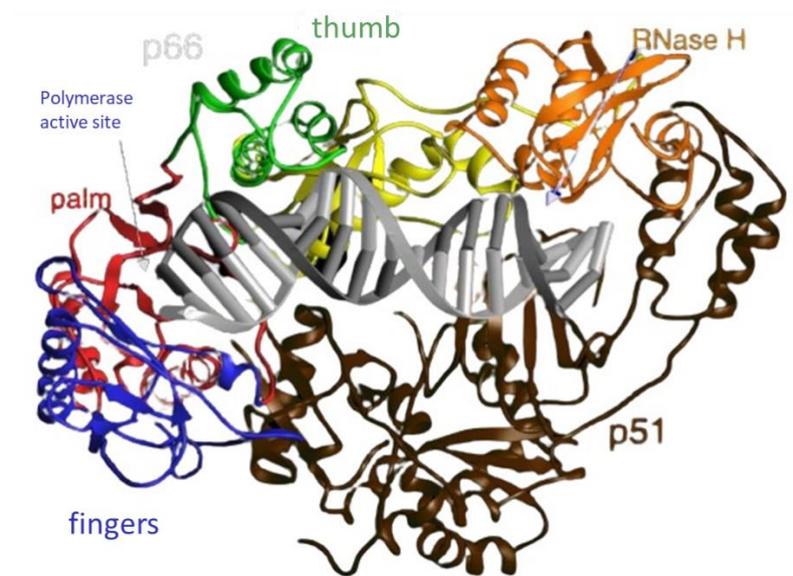


Figure 6 : Structure moléculaire de la Reverse transcriptase (16), 4 domaines moléculaire (doigts, paume, pouce et RNase H) constituant la p66. La fente présente entre p66 et p51 constitue le site catalytique

La RT joue également un rôle majeur en ce qu'il concerne la diversité génétique des rétrovirus. L'enzyme n'est pas fidèle à la perfection, ainsi dans une étude basée sur un modèle mathématique, pour le HIV-1, on retrouve un taux de mauvaise incorporation de l'ordre de 1 : 1700 à 1 : 40000 par nucléotides et par réplication. Pour l'ensemble du génome viral, le taux d'erreur in vivo est estimé à une erreur par cycle de réplication(17). Ces différents variants produits par la RT génèrent donc une diversité des séquences et peut induire une résistance aux antirétroviraux, cependant elle peut également incorporer des analogues de nucléosides dans l'ADN viral plutôt que dans l'ADN cellulaire, ce qui constitue l'efficacité des inhibiteurs de la RT.

L'intégrase d'une longueur de 32kDa et composée de 288 acides aminés, assure la liaison de l'ADN viral double brin afin de l'intégrer dans le génome de la cellule hôte. L'intégrase fait partie d'un complexe catalyseur responsable du clivage de l'ADN viral ainsi que de la ligature à l'ADN de la cellule

hôte. Une fois l'ADN viral bien intégré, le provirus est considéré comme génétiquement stable pendant toute la vie de la cellule(18). Cette enzyme nécessaire à la réplication virale constitue une cible pour les antirétroviraux, les inhibiteurs de l'intégrase, tel le raltégravir.

2. Les Gènes viraux régulateurs et accessoires :

On compte en plus des gènes de structures, 6gènes accessoires : rev (régulation de l'expression virale), tat (transactivateur de la transcription), vpr (protéine virale R), vif (facteur d'infection du virion), vpu (protéine virale U) et nef (impact sur la réplication virale et la diminution des cellules hôtes).

Le gène **tat** augmente considérablement la transcription virale. La petite protéine tat qui en découle, traduite à partir d'un transcrite de 3 exons, dont le premier étant non codant. C'est une protéine nucléaire, d'une taille variable, de 14 à 16 kDa (86 à 102 acides aminés). (19) La protéine tat se lie ensuite à une région d'activation TAR que l'on retrouve dans les ARNm du VIH-1. Il a pour rôle de stabiliser l'ARNm et de participer grandement à son allongement. Il possède 3 domaines fonctionnels riche en cystéine et hautement conservés qui pourrait constituer une bonne cible thérapeutique pour inhiber la réplication virale mais il été démontré que d'autres facteurs cellulaires pouvaient permettre l'expression du virus sans cette protéine tat.(20)

Le gène **rev**, d'un poids moléculaire de 19kDa et d'une taille variable entre 106 et 123 AA, code pour une protéine sérine phosphorylée. Son expression entraîne l'accumulation de protéines structurales codées par les ARNm Gag, Pol, Env, mais aussi Vpy, Vpr et Vif. (21) En revanche, les protéines Tat et Nef sont exprimés de manière efficace sans l'aide de la protéine Rev. Il se compose de 2 domaines fonctionnels, une région conservée riche en arginine, utile dans la liaison spécifique de l'ERR, et un domaine de multimérisation.(22)

Le produit du gène **nef** est une protéine myristoylée de 206 acides aminés qui s'insère dans la face interne de la membrane plasmique. Sa protéine a de nombreuses propriétés qui contribuent à l'immunosuppression. Son activité est cruciale pour la virulence et l'évasion immunitaire. Il régule à la baisse, l'expression de surface des CD4 dans les cellules infectées et semblerai jouer un rôle dans l'infectivité, mais son mécanisme précis n'est pas encore élucidé et nécessite l'étude du rôle de divers co-facteurs. (23)

Le gène **Vif** est une protéine virale de 193 acides aminés de 23 à 27kDa. Il est nécessaire pour une transmission efficace du virus et agit comme antagoniste de APOBEC3G (ARNm de l'apolipoprotéine B, enzyme-catalytique, polypeptide-like 3G)(24) ,une désoxycytidine désaminase qui

agit sur l'ADN viral nouvellement synthétisé pour induire sa dégradation et son hypermutation et qui peut représenter un mécanisme général de défense antirétrovirale.(25) APOBEC3G peut également inhiber la transcription inverse.

Le gène **Vpr** est une protéine de 96 acides aminés traduite à partir d'un ARNm simple épissé. Il est présent en abondance dans le virion mature et est associé à la protéine de capsid. Son action permet l'incorporation et l'exportation de protéines virales.(26) Une étude suggère également qu'il pourrait jouer un rôle dans la régulation de l'expression des gènes (rôle nucléaire) ainsi qu'un arrêt de la progression du cycle cellulaire à l'interface G2/M. (27)

La **Vpu** est une protéine de membrane intégrale amphipathique. Les 27 premiers acides aminés sont hydrophobes, tandis que le reste de cette petite protéine de 16 kDa de 81 acides aminés est hydrophile (28). Il est associé comme Nef à la dégradation rapide de CD4 qui peut en partie éliminer les complexes intracellulaires CD4-gp160 qui interfèrent avec la production du virus. (29)

D. Le diagnostic de la résistance virale :

La résistance du VIH aux différentes molécules est évaluée par deux approches différentes : Les tests phénotypiques et génotypiques.

1. Les tests phénotypiques :

Pour la première approche de la résistance, ils se basent sur la production de virus recombinant à partir de l'ARN plasmatique du patient et permettent d'obtenir un antivirogramme.(30) Ils permettent d'obtenir la mesure de la concentration de l'antiviral inhibant 50% et 90% de la multiplication virale (CI50 et CI90).

Ces techniques restent complexes, fastidieuses et coûteuses pour de la routine et nécessitent une infrastructure adéquate. Elles sont de ce fait utilisées dans le cadre de la recherche afin de tester de nouvelles molécules et constituent également une base dans l'identification en première intention, de mutations non connues.

2. Les différentes techniques de génotypage :

Ce sont des techniques permettant d'identifier diverses mutations présentes dans les gènes d'intérêt par séquençage. Après amplification du matériel génétique cible par PCR, le séquençage permet de cartographier la région qui nous intéresse afin d'y découvrir les mutations connues.

La méthode de référence pour le séquençage du VIH, est la méthode Sanger. Cette technique permet de détecter la population virale majoritaire, ce qui veut dire qu'une mutation doit être présente au minimum à 20% (rapport Morlat) et représente donc l'une des limites de la technique.

En comparaison, les nouvelles techniques de séquençages permettraient d'atteindre une profondeur de l'ordre de 0,1-1% (ultradeep sequencing). On peut donc arriver à détecter les populations virales minoritaires. Cette approche semble prometteuse mais doit être étudié en recherche. En pratique clinique, c'est le taux à 20% qui est le plus couramment utilisé.

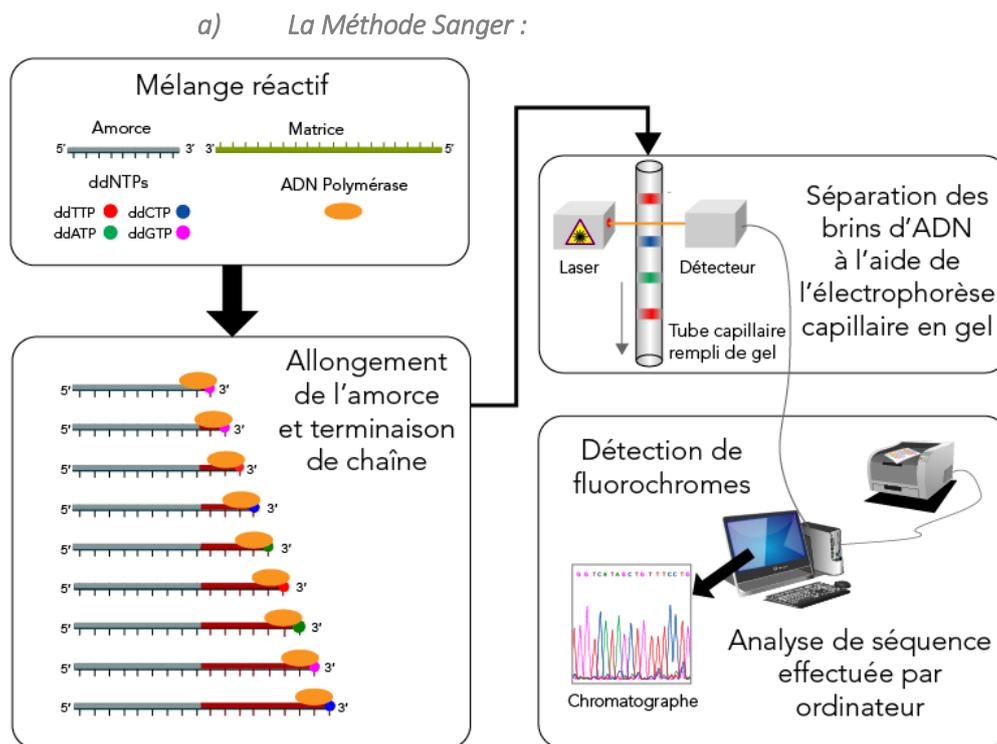


Figure 7 : Principe de séquençage Sanger (31)

La technique utilise l'ADN polymérase afin de synthétiser le brin complémentaire de notre fragment d'ADN. On utilise des didésoxyribonucléotides (ddNTP) marqués par un fluorochrome qui diffèrent des désoxyribonucléotides (dNTP) par l'absence d'un groupement OH en position 3'. Lorsque l'ADN polymérase utilise un ddNTP, l'enzyme n'est plus capable de rajouter un nucléotide à la suite et la synthèse du brin d'ADN s'arrête. La proportion de ddNTP par rapport au dNTP est de l'ordre de 1/10. On réalise par la suite une électrophorèse capillaire. Il s'agit d'une migration des fragments de taille différente à l'aide d'un champ électrique. Chaque fragment étant de tailles différentes et possédant un ddNTP lié à un fluorochrome, il est possible de recueillir le signal et de l'analyser grâce à des outils de bio-informatique. Le résultat final est sous forme d'un électrophorégramme.

b) La Méthode Illumina®:(32)

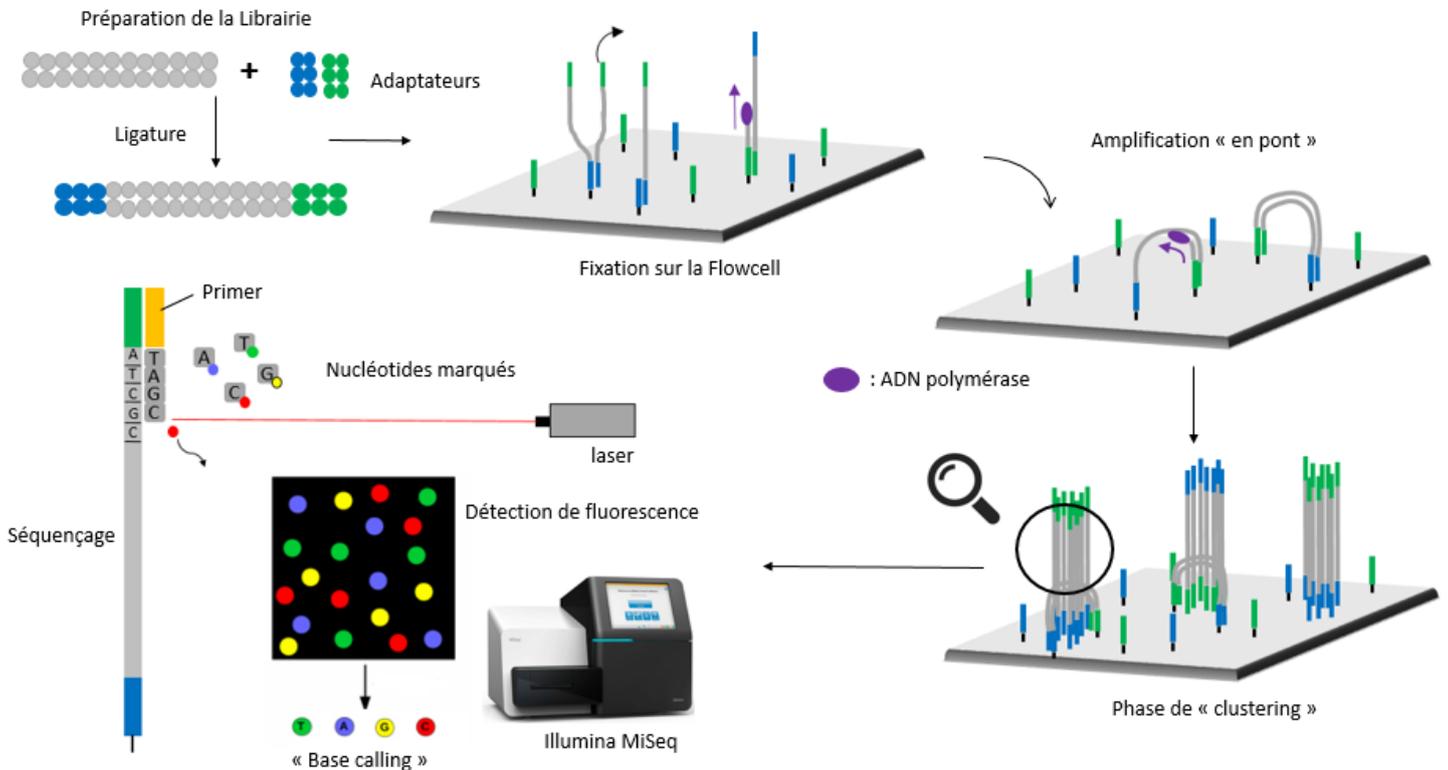


Figure 8 : Principe d'après Illumina®

Lors de la phase de préparation, il est nécessaire d'obtenir de petits fragments d'ADN qui seront ligaturés à certains endroits par des petits oligonucléotides (6 à 8bases). On les appelle des adaptateurs qui font office de double emploi :

- La construction de la librairie avec pour chaque échantillon, sa paire d'adaptateur.
- Hybridation sur la flowcell Illumina® via une petite séquence complémentaire.

L'ADN est ensuite dénaturé et permet en théorie d'avoir le génome fragmenté et composé de ces adaptateurs.

Sur la flow cell, après fixation, la première étape consiste en l'amplification de l'ADN par une polymérisation en chaîne qui va permettre la synthèse du brin complémentaire. Le « clone » synthétisé restera fixé sur la flow cell alors que la seconde partie « originale » non fixée sera lavée hors de la plaque.

La seconde étape consiste en l'amplification « en pont » grâce à l'oligonucléotide du 2^{ème} adaptateur. L'ADN est de nouveau amplifié par une polymérase puis après une dénaturation, nous

allons obtenir les 2 brins sens et anti-sens fixés sur la flow cell. Toute cette procédure va se répéter de nombreuses fois, c'est l'étape de génération de clusters ou « clustering »

Les fragments anti-sens sont ensuite clivés et lavés, ne laissant que les brins sens marquant le début du séquençage. Une amorce va venir se lier aux oligonucléotides et le séquençage va pouvoir s'effectuer à l'aide d'une polymérase et de nucléotides spécifiques marqués par un fluorochrome et possédant un effet terminateur de chaîne. C'est le déroulement d'un cycle qui va se répéter pour chaque base. Ces nucléotides sont ensuite excités par un laser qui émettra un signal fluorescent et sera capté puis intégré informatiquement.

Pour un cluster, tous les brins identiques vont se lire en même temps.

Après la fin de cette première boucle de lecture, le produit est lavé et l'amorce de lecture de l'index 1 est introduite puis hybridée au brin. C'est à ce moment-là que nous avons l'association de l'index à son échantillon.

Pour le MiSeq Illumina®, il va s'opérer à cet instant, le retournement de lecture, c'est-à-dire que toute la procédure va se répéter pour le brin anti-sens. Il y aura la formation d'un nouveau pont avec polymérisation et dénaturation, puis ce sera au tour du brin sens d'être lavé.

Pour chaque échantillon, nous disposeront de milliers de lectures (ou « reads ») Ce nombre de lecture signe la profondeur du séquençage. Plus on a de lectures et plus le séquençage sera sensible. A la fin de l'analyse, c'est la génération de fichiers FastQ qui sera analysé en bio-informatique.

c) La technologie 454 de Roche® :

Après la préparation de la librairie, comprenant 2 adaptateurs de part et d'autre de l'ADN simple brin, l'échantillon est mis en contact de billes sur lesquelles sont fixées des amorces complémentaires des adaptateurs. Il s'ensuit une PCR en émulsion, Chaque goutte contient l'unique brin d'ADN fixé sur la bille (microréacteur) et va pouvoir être amplifié. A la fin de cette PCR, la bille sera recouverte de plusieurs millions de molécules d'ADN simple brin identiques.

Il s'ensuit une réaction de pyroséquençage. Les billes sont déposées dans une plaque PTP (Pico Titer Plate) et s'incorporent dans des micro-puits individuels. Chaque nucléotide est ajouté de manière séquentielle et lors d'une incorporation, on obtient une émission lumineuse produite par la luciférase via la libération et transformation d'un pyrophosphate inorganique en ATP.

Les résultats sont présentés sous forme d'un pyrogramme.

Cette technique à l'avantage d'être très assez rapide avec une longueur des reads satisfaisante (de l'ordre de 700pb). Elle est donc intéressante pour les petits génomes, néanmoins le nombre de reads est assez faible par rapport aux autres techniques et présente des difficultés dans le séquençage : présence d'insertions, de délétions et difficultés à séquencer les homopolymères.

Au final, elle présente une plus faible précision que le Sanger et cette technique n'est plus utilisée aujourd'hui. (33)

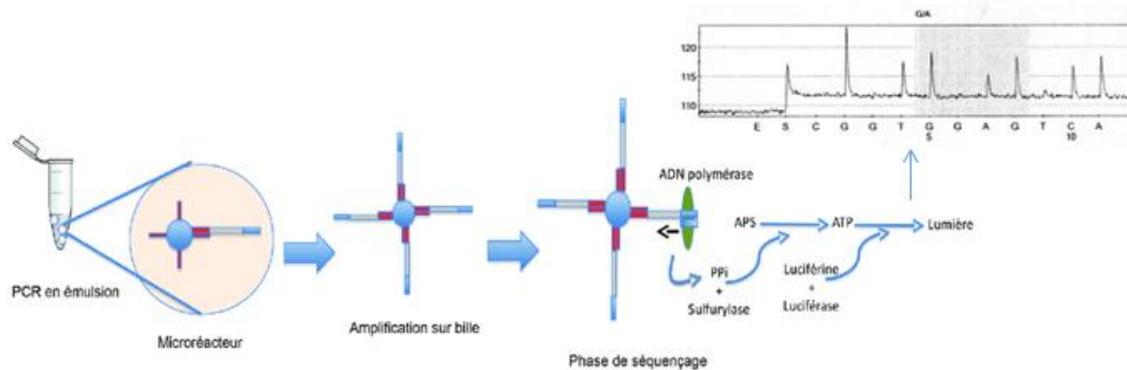


Figure 9 : Principe schématique de la technologie 454 (34)

d) *La technologie Ion torrent® : Life technologies*

L'amplification clonale est également réalisée au cours d'une PCR en émulsion à l'aide de particules sphériques ioniques. Cette technique se différencie des autres par sa méthode de détection. En effet, la matrice de séquençage est chargée sur une puce Ion Torrent composée de circuits semi-conducteurs au fond de chaque puits. Lors de chaque polymérisation de nucléotides, il se produira une libération d'ions H⁺ entraînant une variation de pH détecté par les semi-conducteurs. La transcription est sous forme d'un ionogramme et l'ensemble des données se regroupent dans un fichier .DAT (35)

Cette technique est rapide et présente un bon rapport qualité/prix. Mais elle présente des difficultés dans le séquençage des homopolymères et il existe peu d'outils bio-informatique.

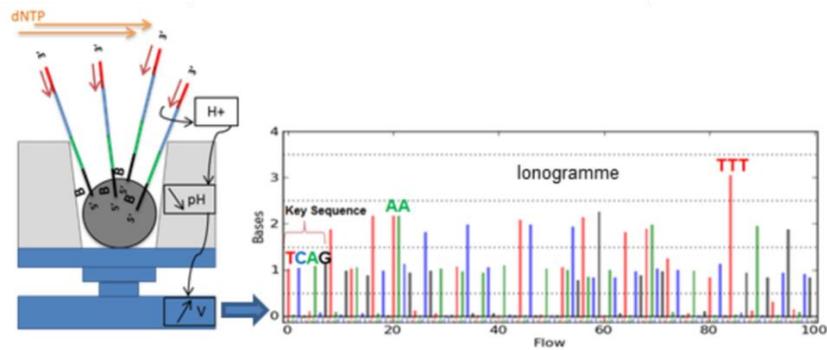


Figure 10 : Principe de la technologie Ion Torrent®

E. Les différentes mutations recherchées

L'étude des résistances (ou test génotypique) par séquençage est une étape clé dans la stratégie thérapeutique du VIH. Elle est indiquée dans plusieurs situations cliniques nécessitant la recherche de mutations sur les gènes de la transcriptase inverse, de la protéase et de l'intégrase, voire la glycoprotéine 120. Les recommandations sont identiques entre un adulte et un enfant :

- Lors de la primo-infection ou avant l'initiation au traitement. (AII)
- Lors d'un échec virologique ($2CV > 50\text{copies/mL}$) (AII)
- Prophylaxie post-exposition (BIII)

Le groupe d'expert Français dirigé par le Pr Morlat et sous l'égide du CNS (Conseil national du sida et des hépatites virales) et de l'ANRS (Agence Nationale de Recherche sur le Sida et les hépatites virales) ont rédigé un rapport expliquant les différentes mutations retrouvées pouvant interférer dans les traitements. Les principales mutations sont résumées dans le tableau ci-dessous. Il existe différents algorithmes d'interprétation de la résistance du VIH-1 comme Stanford, utilisé par les anglophones et l'ANRS v31 en France. La liste exhaustive des mutations d'intérêts est présente sur le site de l'ANRS.(36)

Tableau 2 : Liste non exhaustive de la résistance du VIH-1 du groupe M aux antirétroviraux TAMs = Thymidine Analog Mutations ; AA = Acides aminés

Classe des antirétroviraux	Mécanisme de résistance	Mutations	Molécules touchées
Inhibiteurs nucléosidiques et nucléotidiques de la transcriptase inverse (INTI)	Excision de l'analogue nucléosidique par les mutations appelées TAMs : Favorisent l'accès de l'ATP au site de polymérisation qui réagit avec l'analogue nucléosidique en le détachant de la chaîne d'ADN viral en formation	M41L, D67N, K70R, L210W, T215Y/F et K219Q/E	Résistance progressive à l'ensemble des INTI à des niveaux divers
	La diminution d'incorporation des nucléosides ou nucléotides artificiels au profit de nucléotides naturels	M184V	Lamivudine (3TC), Emtricitabine (FTC)
		K65R/N/E	Ténofovir disopropyl fumarate (TDF), l'abacavir, didanosine
		L74V	Abacavir, Didanosine
Les inhibiteurs non nucléosidiques de la transcriptase inverse (INNTI)	Mutation empêchant le blocage de la transcriptase inverse par fixation au niveau d'une poche hydrophobe proche du site actif de l'enzyme. Une seule mutation entraîne une résistance de haut niveau à l'INNTI et des résistances croisées	K103N	INNTI + Efavirenz, Névirapine
		Y181C	INNTI + Efavirenz, Névirapine, Rilpivirine
		E138K	
		M184I	Association Rilpivirine, Lamivudine, Emtricitabine : Renforcement de la résistance (37)
Inhibiteurs de Protéase	Mutations situées au niveau du site actif de l'enzyme ou à distance de celui-ci. Phénomène graduel avec accumulation progressive. Mutation au niveau des sites de clivage du gène gag : Résistance in vitro (38)	I50L	Atazanavir, pas de résistance croisée
		A431V, V362I (gag)	Impact sur de nouvelles molécules en cours de développement : BMS-955176 (39)
		K20I, K70R et L89M	Impact de la réponse virologique sur les VIH-1 non B
Les inhibiteurs de fusion	Mutations du domaine HR1 de la gp41	AA 36 à 45	enfuvirtide
Antagoniste CCR5	Mécanisme d'échappement viral avec émergence de sous-population X4 minoritaire à l'instauration du traitement ou émergence de virus R5 (40)	Analyse génotypique en cours d'études	Maraviroc
Inhibiteurs d'intégrase	Sélection par les 3 profils de mutations majeures	N155H, Y143C/H/R, Q148K/R/H	Raltégravir, Elvitégravir : présentent une résistance croisée quasi absolue (41)

F. Impact du sous type du VIH-1 sur la résistance des antirétroviraux :

Du fait du grand polymorphisme du VIH-1 en termes de sous type. Il est également important de souligner l'impact des sous types minoritaires en termes de résistances. En France, on sait que c'est le sous-type B qui est le plus représenté. Néanmoins, il existe quelques études s'intéressant à d'autres sous-types.

De manière générale pour les sous-types non B, c'est le gène de la protéase qui présente une variabilité et peut donc présenter des substitutions sur certaines cibles clés considérées comme mutations de résistance chez le sous-type B. Concernant l'étravirine et la rilpivirine, c'est environ 10% des sous-types non B qui présente au moins une mutation.(42)

De même pour le sous-type C, des profils de mutations ont été identifiés induisant une résistance de haut niveau aux INNTI. C'est le cas de la mutation V106M. De plus, on constate une émergence de phénotypes résistant au tenofovir chez ce sous-type. [17]

Tous ces phénomènes de mutations liées aux sous-types doivent encore être élucidés en recherche.

G. Intérêt du dosage de l'ADN proviral ? :

La quantification de l'ADN proviral est un marqueur incontournable pour développer des stratégies d'allègement. Cela permet de mesurer la virémie résiduelle dans les réservoirs de l'hôte qui sont nombreux : toutes cellules ou tissus hébergeant les formes intégrées du VIH dont les principaux sont les lymphocytes T CD4+, les monocytes/macrophages puis le tissu lymphoïde secondaire. (43)

Un taux d'ADN-VIH faible est associé au succès virologique lors de stratégie de bithérapie, notamment lors de l'utilisation de 2 INNTI (44). La monothérapie n'étant pas recommandée

Un des problèmes qui peut se poser dans l'allègement thérapeutique, c'est que même avec un ADN proviral au plus bas, il n'est pas possible de prédire une efficacité dans la stratégie.(45)

Il n'en reste pas moins un marqueur quantitatif et informatif sur l'état infectieux du patient et peut être pertinent chez des patients en échec virologique. Néanmoins le génotypage systématique n'est pas recommandé à cause du risque de détecter des mutations non pertinentes ainsi que de ne pas retrouver des mutations de résistance lors de génotypage antérieur.(46) Il manque en effet de standardisation dans les différentes méthodes de génotypage.

Objectifs de la Thèse :

Suite à un arrêt programmé du kit Viroseq®, le laboratoire du CHU d'Amiens est en quête d'une nouvelle technique pour le génotypage des résistances du VIH-1. Le thème principal de cette étude est l'approche NGS via la technologie Illumina®. Ainsi, le but de cette étude est dans un premier temps de tester et d'évaluer dans notre laboratoire la technique NGS. Dans un second temps, nous souhaitons comparer les résultats du NGS en terme de mutations d'intérêts pour les 3 principales cibles mais aussi en terme de sous-types viraux avec les 2 autres techniques Sanger évaluées (ABL® et ANRS) afin de permettre de faire un choix pour le futur du laboratoire de virologie du CHU d'Amiens.

II. Matériels et Méthodes

A. Table des échantillons analysés

Nous avons fait le choix dans le cadre de ce travail, de sélectionner des échantillons avec une charge virale assez importante permettant de nous affranchir de problèmes techniques en terme de quantité, surtout pour les débuts en NGS. Nous avons également sélectionné des échantillons ayant des profils de mutations et des sous-types viraux variés. Les extraits plasmatiques sélectionnés devaient être également assez récents (<6mois) pour éviter toute dégradation d'ARN et fausser les analyses.

Tableau 3 : Liste des échantillons analysés

Echantillon	CV (log copie)	NGS ABL	Sanger ABL	Sanger ANRS	Profil mutations Sanger Viroseq
202761161	5,61	X	X		IP
202801387	4,64	X			IP
202821685	4.61	X	X	X	IP
202942236	3,97	X	X		INTI + IP
203071366	5,4	X	X		INNTI + IP
203151494	6,05	X	X		IP
203370413	4,68	X	X	X	IP
203461390	4,57	X	X		INTI + IP
203511396	5,17	X	X		IP + INI
203530875	Non renseigné	X	X	X	INTI + INNTI + IP
210110842	4,95	X	X		INNTI
210140539	5,22	X	X		IP
210150931	4,15	X	X	X	IP
210261876	2,47	X			INTI + IP
210321716	5,1	X	X		IP
210481981	6,24	X	X	X	Absence
210490293	5,84	X	X		Absence
210490891	5,52	X	X	X	Absence
210551227	4,7	X	X	X	Absence
210671034	5,8	X	X	X	IP
210720484	5,11	X	X	X	IP
Contrôle négatif	Analyse d'un contrôle négatif et positif seulement sur la technique NGS				
Contrôle positif					

B. Extraction de l'ARN

L'extraction de l'ARN se réalise sur le système eMAG®(47), un automate permettant une extraction automatisée des acides nucléiques à partir du plasma pour le VIH.

La technique se base sur la technologie BOOM(48) qui est une méthode d'extraction en phase solide utilisant des billes de silices magnétiques. Après lyse de l'échantillon par un agent chaotrope (GuSCN), les acides nucléiques viennent se fixer sur les billes magnétiques et s'ensuit deux lavages. A la fin du processus, un tampon d'éluion permet le décrochage de ces acides nucléiques et les billes de silices sont séparées. C'est une méthode simple, rapide et fiable. Après extraction, tous les échantillons sont conservés à -80°C dans l'attente de procéder à la suite des manipulations.

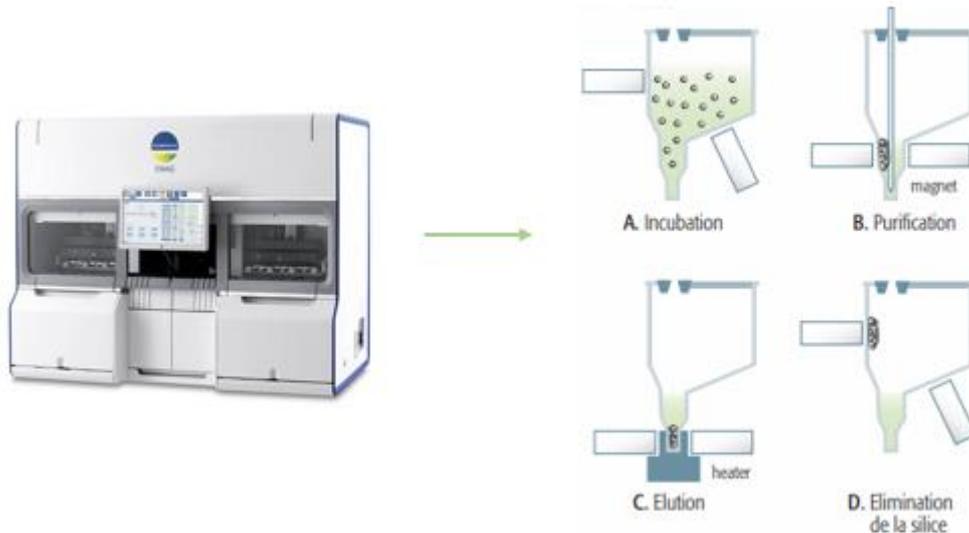


Figure 11 : Principe de l'extraction automatisée sur eMag® A. Après lyse des échantillons, les acides nucléiques sont captés par les particules de silice magnétique. B. Le système d'aimant permet de purifier les acides nucléiques par lavages successifs. C. Le chauffage de la cupule ainsi qu'un tampon d'éluion permet de libérer les acides nucléiques des particules de silice. D. Séparation des particules de silice et récupération de l'éluât

Nous avons besoin de 400µL de plasma au minimum pour obtenir 100µL d'éluât. Lorsque la charge virale de nos échantillons est inférieure à 3 log copies/mL. On utilise 1000µL de plasma avec une éluion à 100µL.

C. La « Workflow » du NGS : Illumina®

Le schéma suivant montre les différentes étapes allant de la préparation de l'échantillon jusqu'au séquençage.

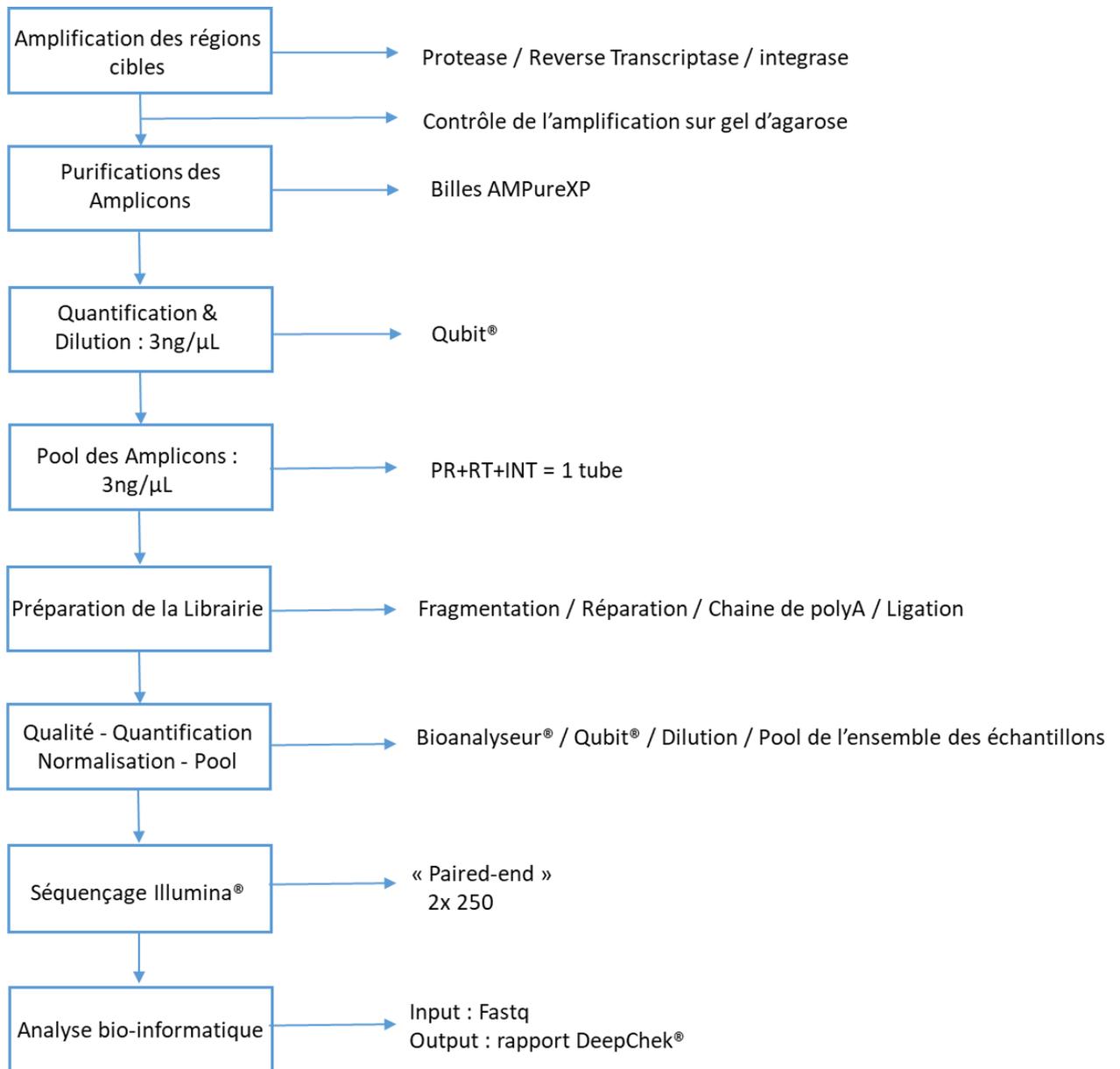


Figure 12: les différentes étapes du NGS avec la technologie Illumina®

D. L'Amplification des régions cibles

1. La RT-PCR :

Après la réalisation de l'extraction de l'ARN, il est nécessaire d'amplifier les régions d'intérêts par RT-PCR. Pour cela, nous disposons avec le kit ABL® de trois « master mix » pour chaque région : Protéase, Reverse transcriptase et Intégrase.

L'ensemble des composants doivent être décongelés 10 minutes avant la préparation du mix et les réactifs doivent être placés directement dans un bac de glace et support de glace pour les tubes tout au long de cette préparation. Il est nécessaire également pour le bon déroulement de la procédure de vortexer (sauf les enzymes) et de centrifuger tous les réactifs avant manipulation et après chaque étape de pipetage. Nous aurons donc pour cette étape, 3 tubes ou 3 puits par échantillon.

Les tables suivantes indiquent les différentes quantités des réactifs à prélever afin de réaliser cette étape :

Tableau 4 : Mix RT-PCR de la région Reverse transcriptase

Réactifs de la RT-PCR	Volume (1N)
RT-PCR Buffer 5X	5 µL
dNTPs 10mM	1 µL
RT-PCR Enzyme Mix	1 µL
RNAsine	0.5 µL
H2O	0.5 µL
RT FOR RT-PCR Primers 10 µM	1.5 µL
RT REV (1) RT-PCR Primers 10 µM	1.5 µL
RT REV (2) RT-PCR Primers 10 µM	1.5 µL
Volume Total de la Réaction	12.5 µL

Quantité x N échantillons

Tableau 5 : Mix RT-PCR de la région Protéase

Réactifs de la RT-PCR	Volume (1N)
RT-PCR Buffer 5X	5 µL
dNTPs 10mM	1 µL
RT-PCR Enzyme Mix	1 µL
RNAsine	0.5 µL
H2O	2 µL
PROT FOR RT-PCR Primers 10 µM	1.5 µL
PROT REV RT-PCR Primers 10 µM	1.5 µL
Volume Total de la Réaction	12.5 µL

Quantité x N échantillons

Tableau 6 : Mix RT-PCR de la région Intégrase

Réactifs de la RT-PCR	Volume (1N)
RT-PCR Buffer 5X	5 µL
dNTPs 10mM	1 µL
RT-PCR Enzyme Mix	1 µL
RNAsine	0.25 µL
H2O	3.25 µL
INT FOR RT-PCR Primers 10 µM	1 µL
INT REV RT-PCR Primers 10 µM	1 µL
Volume Total de la Réaction	12.5 µL

Quantité x N échantillons

Il est à noter que la charge virale du VIH-1 de chaque échantillon doit être détectable et supérieure à 1000 copies/mL pour le bon déroulé de la manipulation. Pour un prélèvement inférieur à 1000copies/mL, une ultracentrifugation peut être réalisée.

Après la préparation de chaque mix, il faut donc distribuer 12.5µL de mix avec 12.5µL d'ARN pour un mélange réactionnel final de 25µL.

Nos échantillons sont prêts à être chargés dans le thermocycleur selon les programmes suivants :

Tableau 7 : pour la réalisation de la RT PCR / Table pour la réalisation de la RT PCR Intégrase

	Temps	Température		Temps	Température
RT step	30 min	50°C	RT step	30 min	50°C
Enzyme activation	15 min	95°C	Enzyme activation	15 min	95°C
X 10 cycles	40 sec	94°C	X 45 cycles	30 sec	94°C
	30 sec	54°C		30 sec	55°C
	1 min	72°C		1 min	72°C
X 35 cycles	30 sec	94°C	Final extension	10 min	72°C
	30 sec	54°C	∞	10°C	
	1 min	72°C (+ 5 sec/cycle)			
Final extension	10 min	72°C			
	∞	10°C			

2. Evaluation de la qualité et quantité d'ADN amplifié :

Il est judicieux de réaliser une migration de nos produits de RT-PCR sur gel d'agarose afin d'avoir une estimation de la qualité, en observant nos bandes d'intérêts à la bonne longueur ainsi que la quantité disponible.

Cela permettra de savoir si notre RT-PCR seule est suffisante pour pouvoir continuer le NGS. Nous avons donc 2 cas de figures à prendre en compte :

- Si on ne retrouve pas nos bandes spécifiques : Cela peut être dû à une mauvaise qualité de nos extraits avec une mauvaise accroche des amorces de PCR. Dans ce cas, il n'y a pas d'intérêt à continuer car on aura une forte probabilité d'échec et une consommation inutile de réactifs
- Si nous retrouvons nos bandes spécifiques avec une quantité faible : nous sommes dans le cadre de la réalisation d'une Nested-PCR afin de refaire une amplification à partir de nos produits de RT-PCR. Cela peut permettre de continuer l'analyse malgré une faible charge virale par exemple.

3. La Nested-PCR :

Cette étape se déroule de la même manière que pour la RT-PCR à la différence que l'on dispose d'un mix et d'un programme de thermocycleur différent pour les différentes cibles. Il utilise donc un 2^{ème} couple d'amorces qui permettra au final la diminution de produits non spécifiques tout en amplifiant le matériel génétique présent. En effet, il permet d'obtenir des quantités détectables visibles lors du contrôle qualité par électrophorèse sur gel.

Pour les charges virales élevées, cette étape n'est pas obligatoire, tant que nos régions ont bien été amplifiées et vérifiées sur un gel.

Tableau 8 : PCR-Nested pour la région de la Reverse transcriptase

Réactifs de Nested-PCR	Volume (1n)
Nested Buffer 10X	2.5 µL
dNTP 10mM	0.5 µL
Nested PCR Enzyme mix	0.13 µL
H2O	17.37 µL
RT FOR Nested PCR primers 10µM	0.5 µL
RT REV (1) Nested primers 10µM	0.5 µL
RT REV (2) Nested primers 10µM	0.5µL
Volume Total de la Réaction	22 µL

Quantité x N échantillons

Tableau 9 : PCR-Nested pour la région de la Protéase

Réactifs de Nested-PCR	Volume (1n)
Nested Buffer 10X	2.5 µL
dNTP 10mM	0.5 µL
Nested PCR Enzyme mix	0.13 µL
H2O	17.87 µL
PROT FOR Nested PCR primers 10µM	0.5 µL
PROT REV Nested primers 10µM	0.5 µL
Volume Total de la Réaction	22 µL

Quantité x N échantillons

Tableau 10 : PCR-Nested pour la région de l'intégrase

Réactifs de Nested-PCR	Volume (1n)
Nested Buffer 10X	5 µL
dNTP 10mM	1 µL
Nested PCR Enzyme	1 µL
H2O	13 µL
INT FOR Nested PCR primers 10µM	1 µL
INT REV Nested primers 10µM	1 µL
Volume Total de la Réaction	22 µL

Quantité x N échantillons

Ensuite, il faut distribuer 3µL de produit de PCR pour un total de 25µL par échantillon avant de lancer la réaction dans les thermocycleurs selon le bon programme :

Tableau 11 : Protocole Nested-PCR thermocycleur : Intégrase

	Temps	Température
Enzyme activation	15 min	95°C
X 35 cycles	30 sec	94°C
	30 sec	55°C
	30 sec	72°C
Final extension	10 min	72°C
	∞	10°C

un tampon d'éluion (rôle de l'H₂O) permet de décrocher le matériel génétique des billes magnétiques et nous pouvons transférer l'éluat purifié dans une nouvelle plaque.

2. Protocole de purification :

- Centrifuger la plaque contenant les produits de PCR pendant **30sec** à 900 x g.
- Dans une nouvelle plaque, ajouter dans chaque puits **30 µL** d'H₂O « Nuclease free » et transférer soigneusement **20µL** de produits PCR (ou de Nested-PCR)
- Vortexer le flacon de billes AMPure XP pendant **20 sec** afin de bien les remettre en suspension, puis ajouter 30µL de billes dans chaque puits.
- Mélanger en pipetant de haut en bas au moins 12 fois afin de rendre homogène les puits.
- Incuber la plaque **10min** à température ambiante.
- Placer la plaque sur le support magnétique et laisser incuber **3 à 5min** à température ambiante jusqu'à que le surnageant soit clair.
- Toujours sur le support magnétique, retirer et jeter soigneusement le surnageant sans pipeter les billes.
- Retirer la plaque du support magnétique et ajouter 150µL d'éthanol à 80% **fraichement préparé**. Bien homogénéiser
- Placer la plaque sur le support magnétique et laissez incuber pendant 30 sec.
- Toujours sur le support magnétique, retirer et jeter le surnageant sans toucher aux billes.
- Répétez l'étape de lavage à l'éthanol une **2nd fois**.
- Sécher les billes à température ambiante pendant **3-5min**. Ne pas laisser sur-sécher les billes.
- Retirer la plaque du support magnétique et remettre en suspension dans **40µL** d'H₂O « Nucléase free ». Incuber à température ambiante pendant **2min** pour l'éluion.
- Replacer la plaque sur le support magnétique, laissez incuber pendant **2 min** et récupérez 35µL de surnageant de chaque puits dans une nouvelle plaque PCR 96 puits.
- Scellez la plaque et conservez à -20°C en attendant de procéder à la quantification.

F. Quantification, Dilution et pooling

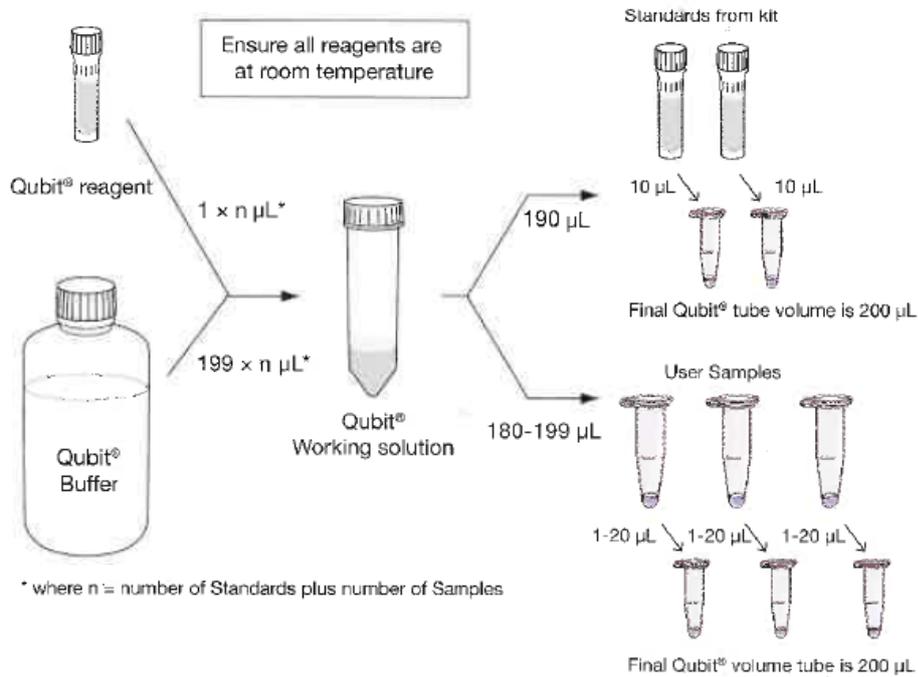


Figure 14 : Protocole du dosage Qubit[®]

Le Qubit[®](49) est un fluorimètre permettant la quantification de l'ADN (ARN et protéines également selon le kit). Il utilise des colorants fluorescents (Dyes) qui viennent se fixer spécifiquement sur l'ADN. L'intensité du signal de fluorescence mesuré par l'appareil est directement proportionnelle à la concentration d'ADN dans la solution. Pour cela, il se réfère à une gamme étalonnée en concentration d'ADN connu (Standards 1 et 2). Pour chaque échantillon, nous avons utilisé 2µL de nos produits de PCR purifiés additionnés de 198µL de solution de travail.

Nous avons donc une concentration précise en ADN pour nos échantillons, ce qui va permettre de « normaliser » par le biais de dilutions ($C1V1 = C2V2$) tous nos échantillons à une concentration de 3ng/µL.

Nous pouvons ensuite « pooler » pour chaque patient, 5µL des différents échantillons de RT/PR/INT dans un seul puit pour une quantité finale de 15µL.

G. Préparation de la librairie

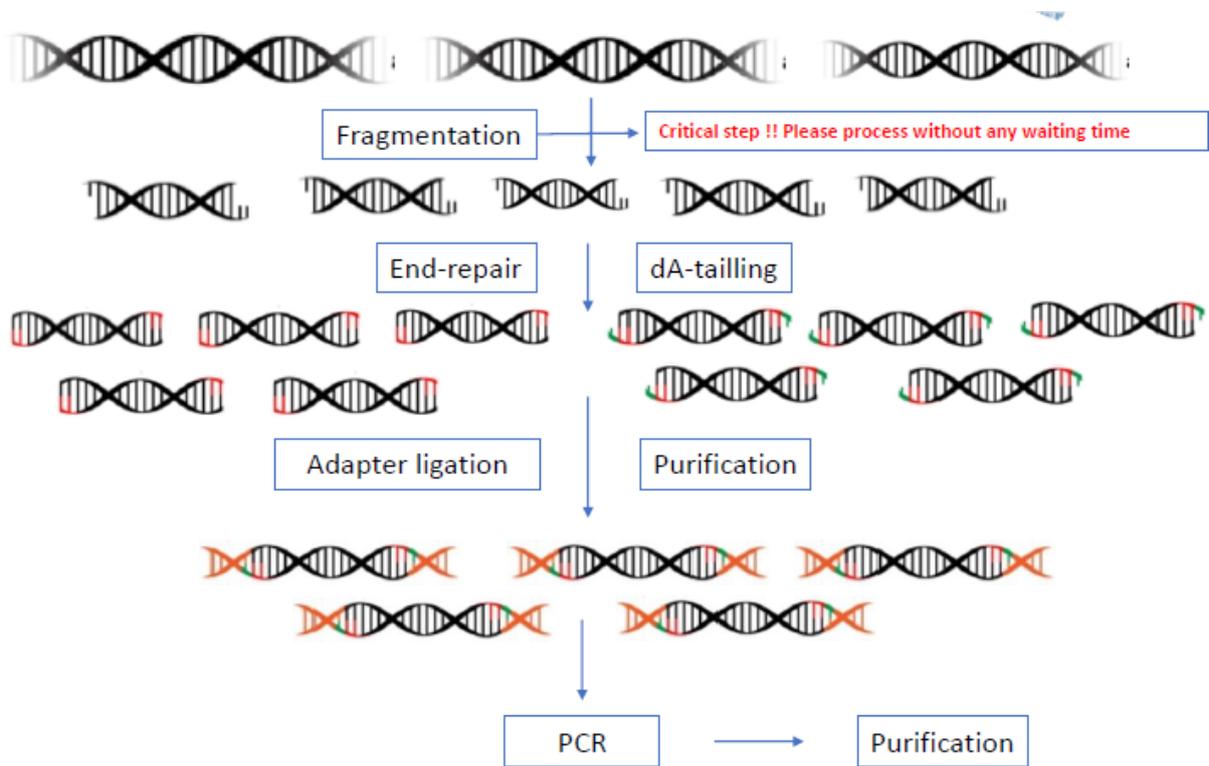


Figure 15 : Les différentes étapes de la préparation de la librairie de nos échantillons.

1. La Fragmentation :

Cette étape permet de réduire la taille de notre région d'ADN en fragments uniformes afin de ne pas avoir des brins d'ADN de trop grandes tailles qui pourraient être difficile à séquencer. Elle permet également de diversifier grandement notre panel de séquences, le « clustering » en sera facilité sur l'Illumina® du fait de cette diversité. Le MiSeq d'Illumina® effectuera un retournement après la lecture de 251pb. Cette procédure a donc été construite pour faire du 251pb x 2 avec le retournement.

Pour cela, il faut préparer une solution de réaction contenant notre ADN et les enzymes de fragmentation puis réaliser le protocole du thermocycleur (pré-refroidi à 4°C et ne pas refermer le couvercle chauffant). Ne pas oublier de vortexer avant passage dans le thermocycleur.

Tableau 13: Composants de la réaction de fragmentation

Réactifs	Volume
ADN	8.75µL
Frag Buffer	1.25µL
Frag Enzyme	2.5µL
Volume final	12.5µL

Pour 1 échantillon

Tableau 14: Programme de la fragmentation

Temps	Température
5 min	4°C
20 min	37°C
∞	4°C

Ce moment est crucial et il est nécessaire de transférer directement nos produits sur de la glace et de passer immédiatement à l'étape suivante.

2. La Réparation et le « A-tailing » :

Suite à la fragmentation, la stratégie de « l'A-tailing » permet d'ajouter une base adénine en surplomb via cette réaction et permettra aux adaptateurs contenant une seule base de thymine de s'apparier avec les fragments d'ADN.

Dans le même tube que celui dans lequel la fragmentation enzymatique a été effectuée, préparer le mélange réactionnel suivant :

Tableau 15 : Composants de la réaction pour la réparation et le « A-tailing »

Réactifs	Volume
ADN Fragmenté	12.5µL
ER and AT buffer	1.75µL
ER and AT Enzyme mix	0.75µL
Volume final	15µL

Pour 1 échantillon

Utiliser le couvercle chauffant cette fois-ci. Et passer immédiatement à l'étape de la ligation sans temps d'attente.

Tableau 16 : Programme du thermocycleur de la réparation et du « A-tailing »

Temps	Température
30 min	65°C
∞	4°C

3. La ligation des adaptateurs :

Une enzyme ligase va venir relier de manière covalente les adaptateurs aux fragments d'ADN pour venir former notre bibliothèque. Ces adaptateurs permettent de fixer les séquences sur la « flowcell » et contiennent également des codes-barres (Index) pour identifier les échantillons et permettre le multiplexage. Nous sommes dans une configuration de « Dual-index ». Chaque échantillon aura un index aux 2 extrémités.

Pour cette étape, il est nécessaire de réaliser un Pre-mix de ligation :

Tableau 17 : Les composants du pre-mix de ligation

Réactifs	Volume
Eau de qualité PCR	1.25µL
Ligation Buffer	7.5µL
DNA ligase	2.5µL
Volume final	11.25µL

Pour 1 échantillon

Tableau 18: Les réactifs de l'étape de la ligation ainsi que le programme du thermocycleur.

Réactifs	Volume
ER and AT reaction product	15µL
Ligation premix	11.25µL
Adaptateurs (1-96)	1.25µL
Volume final	27.5µL

Temps	Température
30 min	20°C

Tableau 19: Liste des combinaisons d'adaptateurs ainsi que leurs séquences. (Kit Trugene DNA CD Indexes®)

	D701	D702	D703	D704	D705	D706	D707	D708	D709	D710	D711	D712
D501	1	2	3	4	5	6	7	8	9	10	11	12
D502	13	14	15	16	17	18	19	20	21	22	23	24

Index	Séquence
D701	ATTACTCG
D702	TCCGGAGA
D703	CGCTCATT
D704	GAGATTCC
D705	ATTCAGAA
D706	GAATTCGT
D707	CTGAAGCT
D708	TAATGCGC
D709	CGGCTATG
D710	TCCGCGAA
D711	TCTCGCGC
D712	AGCGATAG
D501	TATAGCCT
D502	ATAGAGGC

Adaptateurs	Séquence
Sens	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
Anti-sens	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

Immédiatement après l'étape de ligation, il est obligatoire de réaliser la 2^{ème} purification de la manipulation. C'est le même protocole décrit précédemment à la différence d'une variation des différents volumes.

Tableau 20: Les différents volumes du protocole pour la 2^{ème} purification

Réactifs pour purification	Volume
Produit de réaction de la ligation	27.5µL
Billes magnétiques	27.5µL
Ethanol 80%	200µL
H2O pour l'éluion	12µL
Surnageant à récupérer	10µL

4. Amplification de la librairie :

Afin d'avoir une quantité suffisante de nos produits contenant des adaptateurs pour chaque échantillons.

Tableau 21: Le mix pour l'amplification de la librairie

Réactifs	Volume
2X PCR Mix	12.5µL
10X PCR Primer Mix	2.5µL
Librairie	10µL
Volume final	25µL

Tableau 22: Programme de l'amplification de la librairie

Temps	Température	Cycles
45 sec	98°C	1
15 sec	98°C	8
30 sec	60°C	
30 sec	72°C	
1 min	72°C	1
∞	4°C	1

Après l'amplification, en fonction de l'optimisation de la manipulation, 2 choix possibles : soit congeler les prélèvements à 4°C ou bien -20°C si le temps d'attente avant la suite est de plus de 72h.

Il faudra réaliser la 3^{ème} purification de la manipulation.

Tableau 23: Les différents volumes du protocole pour la 3ème purification

Réactifs pour purification	Volume
Produit de réaction de la ligation	25µL
Billes magnétiques	25µL
Ethanol 80%	200µL
H2O pour l'élution	32µL
Surnageant à récupérer	30µL

H. Contrôle qualité des échantillons et préparation au séquençage.

Le « Agilent High Sensitivity DNA kit »(50) constitue une technique standardisée pour le contrôle qualité des échantillons permettant de mesurer et de quantifier de manière fiable la taille et la quantité d'ADN fragmenté, d'ARN, de protéines ou de bibliothèques de séquençage d'ADN.

Les échantillons sont déposés dans des micropuits d'une « Chip » et se déplacent dans les micro-canaux par différence de potentiel. Chaque échantillon va aboutir dans un canal de séparation, ce qui va permettre une séparation haute résolution par électrophorèse capillaire. L'ADN est alors séparé par taille et détecté avec précision grâce à la fluorescence d'un agent intercalant.

Cette étape nous permet donc de juger de la qualité de nos manipulations précédentes dont la fragmentation et les purifications. Nous attendons en moyenne, une taille optimale de l'ordre de 300pb qui signifiera une bonne préparation des bibliothèques.

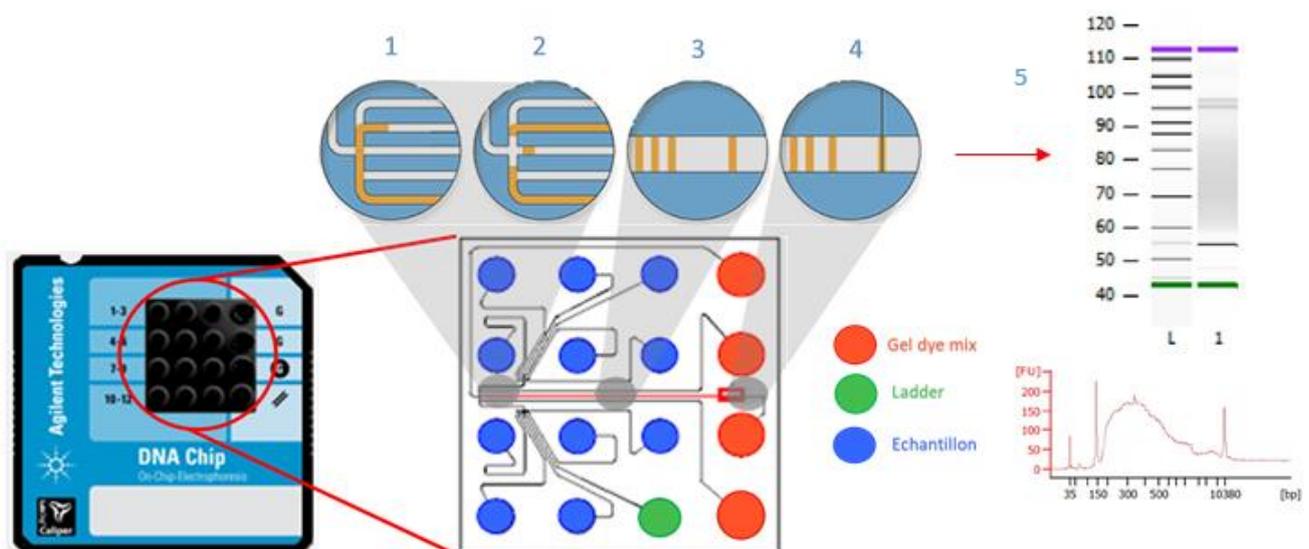


Figure 16: Principe du Bioanalyser® adapté d'Agilent (49) : 1. Les échantillons déposés dans chaque puits vont se déplacer par le biais d'une différence de potentiel dans les micro-canaux. 2. Les échantillons vont être injectés dans le canal de séparation. 3. Séparation en fonction de leurs tailles par électrophorèse capillaire. 4. Détection par la présence de fluorescence d'un agent intercalant. 5. Analyse traduite sous forme d'un gel et d'un électrophorégramme.

Après cette étape du bioanalyser®, il reste une étape de quantification au Qubit® « High sensitivity ». Les librairies amplifiées et purifiées doivent être quantifiées pour une normalisation à **2ng/μL**.

Au final, un pool de **10μL** de chaque librairie normalisée sera réalisé avant de passer à l'étape de séquençage.

I. Le Protocole du séquençage Illumina®

Afin de réaliser le séquençage, nous avons notre pool à la concentration de 2ng/μL et il est nécessaire de diluer encore ce prélèvement pour ne pas surcharger la cartouche de séquençage et la flow cell. Les 2ng/μL correspondent à 10nM. La première étape consiste à diluer l'échantillon à 4nM puis à dénaturer l'ADN à l'aide d'un tampon NaOH (0,2M) fraîchement préparé. Le tampon HT1 permet de stopper cette dénaturation et permet en même temps d'avoir notre librairie à la concentration de 20pM qui sera encore diluée au demi.

Le PhiX à 10nM fourni dans le kit Illumina® est une librairie d'une taille moyenne de 500pb avec une composition de base équilibrée (45%GC et 55%AT). Il permet d'équilibrer la fluorescence émise lors du séquençage en augmentant la diversité des séquences à analyser. En effet dans une librairie de faible diversité, un canal mesurant la fluorescence d'une base peut être plus intense que pour les autres canaux. Le séquenceur peut donc avoir des difficultés dans la détection de l'emplacement des clusters et l'appel de bases. Au final, il permet d'augmenter la qualité du séquençage et fait office de contrôle interne également.(51)

$$2ng = \frac{2 \frac{ng}{\mu L} (\text{cubit concentration})}{660 (\text{molecule}) \times 300 (\text{bp average size})} \times 10^6 = 10.10 \text{ nM} \approx 10 \text{ nM}$$

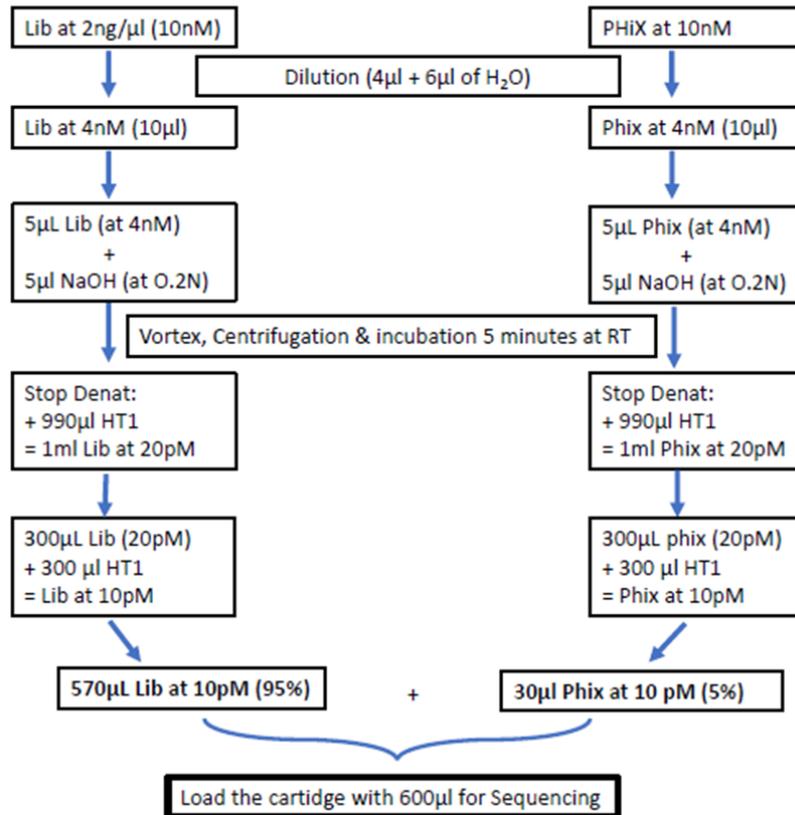


Figure 17 : Préparation de la cartouche avant analyse dans le séquenceur Illumina®

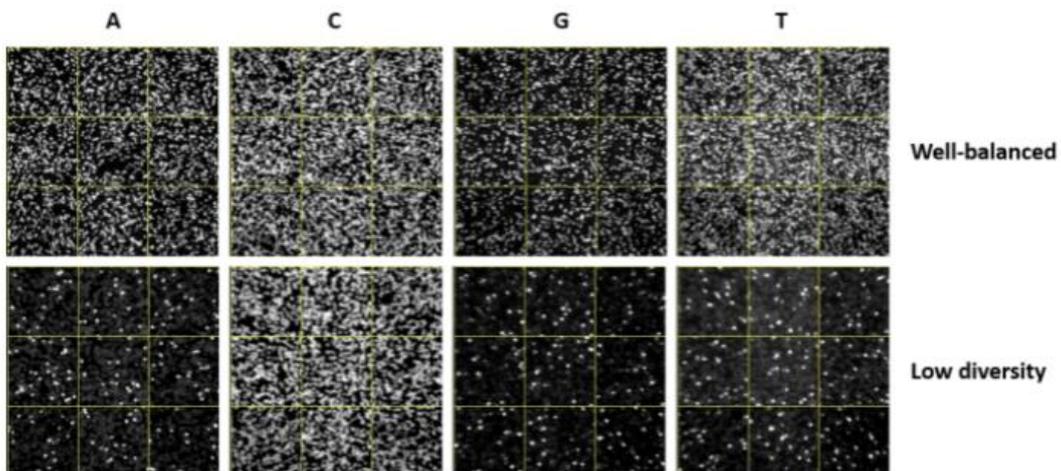


Figure 18: L'intérêt du PhiX dans la diversité nucléotidique lors du séquençage illumina®

J. La technique Sanger ABL

Quelques étapes sont communes avec le NGS comme l'extraction et la RT-PCR (mêmes amorces). Nous avons donc 3 puits pour chaque échantillon (1 pour chaque cible). Pour le séquençage, ce sont les mêmes amorces diluées qui sont utilisées. La suite des étapes seront décrites dans cette partie.

1. Dilution :

Afin d'optimiser le séquençage Sanger, une étape de dilution est nécessaire suite à la réaction de RT-PCR. Pour cela, nous procédons à une estimation semi quantitative sur gel d'agarose de l'intensité des différentes bandes d'ADN. Les tableaux suivant indiquent la dilution à faire pour chaque échantillon. Dans le cas où nous ne détectons pas de bandes, il faudra utiliser les amorces Nested-PCR afin d'amplifier les cibles.

Tableau 24 : Dilution à réaliser en fonction de la taille des bandes sur le gel d'agarose

Tailles des bandes (pb)	Quantité d'ADN (ng)	Quantité d'ADN (ng)	Dilution
100-200	1-3	> 100	1/20
200-500	3-10	Entre 60 et 100	1/10
500-1000	5-20	Entre 30 et 60	1/5
1000-2000	10-40	Entre 10 et 30	1/2
>200	20-50	< 10	Pas de dilution

2. Purification :

Afin d'éliminer les primers, les dNTPs et impuretés de la RT-PCR, nous réalisons une purification de type enzymatique. On mélange 5µL de produits de PCR associé à 2µL du kit de purification ABL puis dans un thermocycleur, nous laissons incuber 4min à 37°C pour activer les enzymes puis 10min à 90°C pour les inhiber.

3. Marquage et Séquençage :

Le Prémix DeepDye v3.1 rr (BigDye Terminator) permet d'effectuer la réaction de séquençage. Dans ce prémix, on retrouve une ADN polymérase, des dNTPs et des ddNTPs marqués avec effet terminateur de chaîne. Cette réaction va générer des tailles d'ADN différentes.

Les produits de PCR sens et anti-sens, doivent être analysés dans un puit de manière indépendante. Pour 1 échantillon, on a besoin de 6 puits.

Tableau 25 : Quantité des réactifs à ajouter dans chaque puit et le programme de la réaction

Réactifs	Quantité (µL)
DeepDye v3.1 rr Premix	1
5x Sequencing Buffer	1.5
Produits de PCR	1
Primer (3.2 - 5pMol)	1
H2O	5.5
Volume final	10

Etapes	Température (°C)	Temps (min:sec)
Dénaturation	96	0 :45
25 cycles	96	0 :10
	50	0 :05
	60	4 :00
Fin du programme	4	∞

Suite à cette réaction, une purification sur des billes de résine est indispensable pour enlever les excédents de dNTPs, ddNTPS et primers. Un premix doit être réalisé en mélangeant après avoir vortexé, 49.5µL de ixpure Activator et 11µL de ixPure Resin (quantité pour 1 puit). Il est ensuite nécessaire de rajouter seulement 55µL dans le puit puis de vortexer 30min à 2600 rpm puis centrifuger 1min à 1000g. 20µL du surnageant sont ensuite transférés dans une nouvelle plaque.

La plaque enfin prête peut être lancée dans l'électrophorèse capillaire. On obtiendra pour chaque cible sens et anti-sens, un fichier AB1.

4. Sanger avec les amorces de l'ANRS

La technique est identique au Sanger ABL® avec pour seule différence, l'utilisation des amorces de l'ANRS :

Tableau 26 : Listing des Amorces de l'ANRS

Protéase	
PCR	5'PROT1: 5'-TAA TTT TTT AGG GAA GAT CTG GCC TTC C-3' 3'PROT1: 5'-GCA AAT ACT GGA GTA TTG TAT GGA TTT TCA GG-3'
Nested-PCR et Séquençage	5'PROT2: 5'- TCA GAG CAG ACC AGA GCC AAC AGC CCC A-3' 3'PROT2: 5'- AAT GCT TTT ATT TTT TCT TCT GTC AAT GGC-3'
Reverse Transcriptase	
PCR	MJ3 : 5'-AGT AGG ACC TAC ACC TGT CA-3' MJ4: 5'-CTG TTA GTG CTT TGG TTC CTC T-3'
Nested-PCR et Séquençage	A35: 5'-TTG GTT GCA CTT TAA ATT TTC CCA TTA GTC CTA TT-3' NE135: 5'-CCT ACT AAC TTC TGT ATG TCA TTG ACA GTC CAG CT-3'
Intégrase	
PCR	INPS1 : 5'- TAG TAG CCA GCT GTG ATA AAT GTC- 3' INPR8: 5'-TTC CAT GTT CTA ATC CTC ATC CTG-3'
Nested-PCR et Séquençage	INPS3 : 5'- GAA GCC ATG CAT GGA CAA G- 3' INPR9: 5'-ATC CTC ATC CTG TCT ACT TGC C -3'

III. L'analyse bio-informatique

A. Mise en place de l'environnement de travail :

La bio-informatique est la discipline critique dans le domaine du séquençage et particulièrement le séquençage haut débit. Il est nécessaire d'avoir quelques connaissances pour pouvoir comprendre et réussir à visualiser les résultats de ses propres analyses. Cette complexité dans l'analyse informatique a permis l'émergence de différentes Pipelines mises à jour régulièrement en fonction des avancées de la recherche scientifique (ici l'ANRS) et possédant une interface graphique intuitive permettant un visuel et une simplicité d'utilisation. C'est le cas de la plateforme DeepCheck mise en place par la société ABL.

Cette partie a pour objectif de comprendre les différentes étapes de bases afin d'arriver aux résultats de séquençage qui nous intéresse.

Afin de réaliser les analyses de manière autonome il est plus simple de travailler sous le système d'exploitation Linux avec l'utilisation de différents outils informatiques et la ligne de commande Bash.

Pour les possesseurs de Windows, il existe la plateforme web gratuite de gestion de données Galaxy(52) possédant un grand nombre d'outils et qui a pour objectif de rendre accessible la bio-informatique sans connaissances de programmation. Néanmoins, les tests réalisés sur cette plateforme ne sont pas forcément concluants et il est plus judicieux d'installer Linux sur machine virtuelle (logiciel Virtual Box(53)) ou en Dual-boot.

B. L'appel des bases du séquençage :

L'identification des bases (ou Base-calling) avec la technologie illumina® est l'obtention des séquences nucléotidiques à partir de signaux de fluorescence. A l'issue de ce séquençage, les index des adaptateurs permettent un démultiplexage ce qui permet d'attribuer à chaque patient ses séquences. L'ensemble des reads se regroupent dans un fichier Fastq .Comme nous sommes dans la situation d'un séquençage « paired-end » avec un couple d'index de chaque côté du brin d'ADN, nous disposons de 2 fichiers Fastq sens et anti-sens (R1 et R2) pour chacun de nos échantillons.

Tableau 28: ASCII. Exemple : Un score Phred de 30 correspond au symbole 30+33 =63. Dans la table, c'est le symbole « ? »

Decimal	Hexadecimal	Binary	Octal	Char	Decimal	Hexadecimal	Binary	Octal	Char	Decimal	Hexadecimal	Binary	Octal	Char
0	0	0	0	[NULL]	48	30	110000	60	0	96	60	1100000	140	.
1	1	1	1	[START OF HEADING]	49	31	110001	61	1	97	61	1100001	141	a
2	2	10	2	[START OF TEXT]	50	32	110010	62	2	98	62	1100010	142	b
3	3	11	3	[END OF TEXT]	51	33	110011	63	3	99	63	1100011	143	c
4	4	100	4	[END OF TRANSMISSION]	52	34	110100	64	4	100	64	1100100	144	d
5	5	101	5	[ENQUIRY]	53	35	110101	65	5	101	65	1100101	145	e
6	6	110	6	[ACKNOWLEDGE]	54	36	110110	66	6	102	66	1100110	146	f
7	7	111	7	[BELL]	55	37	110111	67	7	103	67	1100111	147	g
8	8	1000	10	[BACKSPACE]	56	38	111000	70	8	104	68	1101000	150	h
9	9	1001	11	[HORIZONTAL TAB]	57	39	111001	71	9	105	69	1101001	151	i
10	A	1010	12	[LINE FEED]	58	3A	111010	72	:	106	6A	1101010	152	j
11	B	1011	13	[VERTICAL TAB]	59	3B	111011	73	;	107	6B	1101011	153	k
12	C	1100	14	[FORM FEED]	60	3C	111100	74	<	108	6C	1101100	154	l
13	D	1101	15	[CARRIAGE RETURN]	61	3D	111101	75	=	109	6D	1101101	155	m
14	E	1110	16	[SHIFT OUT]	62	3E	111110	76	>	110	6E	1101110	156	n
15	F	1111	17	[SHIFT IN]	63	3F	111111	77	?	111	6F	1101111	157	o
16	10	10000	20	[DATA LINK ESCAPE]	64	40	1000000	100	@	112	70	1110000	160	p
17	11	10001	21	[DEVICE CONTROL 1]	65	41	1000001	101	A	113	71	1110001	161	q
18	12	10010	22	[DEVICE CONTROL 2]	66	42	1000010	102	B	114	72	1110010	162	r
19	13	10011	23	[DEVICE CONTROL 3]	67	43	1000011	103	C	115	73	1110011	163	s
20	14	10100	24	[DEVICE CONTROL 4]	68	44	1000100	104	D	116	74	1110100	164	t
21	15	10101	25	[NEGATIVE ACKNOWLEDGE]	69	45	1000101	105	E	117	75	1110101	165	u
22	16	10110	26	[SYNCHRONOUS IDLE]	70	46	1000110	106	F	118	76	1110110	166	v
23	17	10111	27	[ENG OF TRANS. BLOCK]	71	47	1000111	107	G	119	77	1110111	167	w
24	18	11000	30	[CANCEL]	72	48	1001000	110	H	120	78	1111000	170	x
25	19	11001	31	[END OF MEDIUM]	73	49	1001001	111	I	121	79	1111001	171	y
26	1A	11010	32	[SUBSTITUTE]	74	4A	1001010	112	J	122	7A	1111010	172	z
27	1B	11011	33	[ESCAPE]	75	4B	1001011	113	K	123	7B	1111011	173	{
28	1C	11100	34	[FILE SEPARATOR]	76	4C	1001100	114	L	124	7C	1111100	174	
29	1D	11101	35	[GROUP SEPARATOR]	77	4D	1001101	115	M	125	7D	1111101	175	}
30	1E	11110	36	[RECORD SEPARATOR]	78	4E	1001110	116	N	126	7E	1111110	176	~
31	1F	11111	37	[UNIT SEPARATOR]	79	4F	1001111	117	O	127	7F	1111111	177	[DEL]
32	20	100000	40	[SPACE]	80	50	1010000	120	P					
33	21	100001	41	!	81	51	1010001	121	Q					
34	22	100010	42	"	82	52	1010010	122	R					
35	23	100011	43	#	83	53	1010011	123	S					
36	24	100100	44	\$	84	54	1010100	124	T					
37	25	100101	45	%	85	55	1010101	125	U					
38	26	100110	46	&	86	56	1010110	126	V					
39	27	100111	47	'	87	57	1010111	127	W					
40	28	101000	50	{	88	58	1011000	130	X					
41	29	101001	51	}	89	59	1011001	131	Y					
42	2A	101010	52	*	90	5A	1011010	132	Z					
43	2B	101011	53	+	91	5B	1011011	133	[
44	2C	101100	54	,	92	5C	1011100	134	\					
45	2D	101101	55	-	93	5D	1011101	135]					
46	2E	101110	56	.	94	5E	1011110	136	^					
47	2F	101111	57	/	95	5F	1011111	137	_					

Enfin, il existe l'outil indispensable FastQC (57) permettant de visualiser différents graphiques analysant la qualité des données du Fastq. Il est à noter une baisse de la qualité en fin de la séquence montre qu'il peut y avoir des erreurs sur un long fragment. Une bonne qualité des reads permet une bonne analyse bio-informatique.

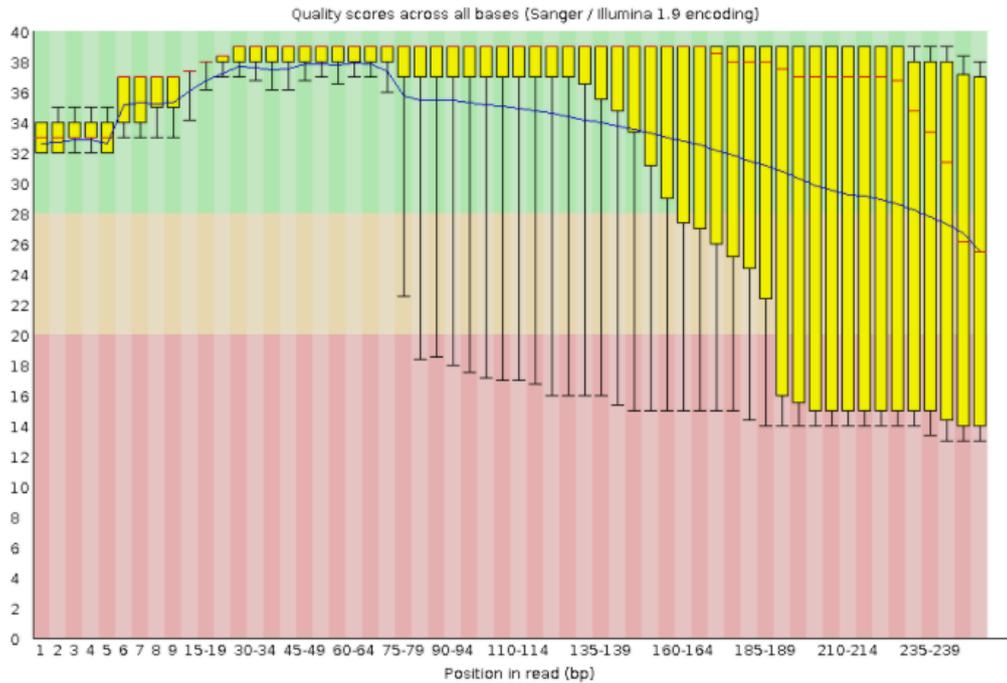


Figure 20 : Distribution des qualités à chaque position de l'ensemble des reads issue des données de séquençage. Le trait rouge central est la valeur médiane. Une barre jaune représente la fourchette interquartile (25-75%), les moustaches représentent 10% et 90% des points. Le trait bleu représente la qualité moyenne.

C. Alignement sur le génome de référence :

L'alignement ou « mapping » est le fait de correspondre nos séquences obtenues à notre génome de référence. Pour cela, il est nécessaire de passer par des algorithmes d'alignement. L'outil le plus utilisé en diagnostic est BWA(58), suivi de Bowtie2. Ils intègrent une transformée de Burrows-Wheeler dans leurs programmes. Les résultats obtenus après alignement se retrouvent sous un format SAM qui sera ensuite converti en fichier BAM.

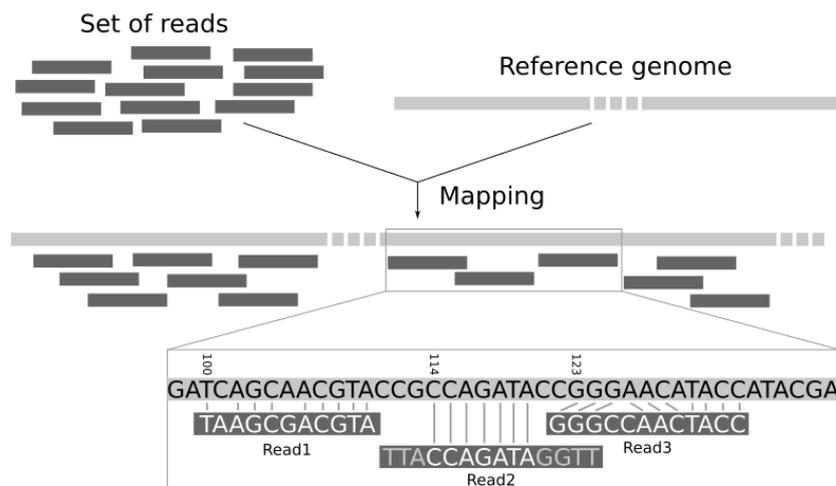


Figure 21 : Le principe d'alignement. Mapping des lectures (format Fastq) au génome de référence (format Fasta) grâce à un outil d'alignement (BWA)

1. La Transformée de Burrows-Wheeler (BWT) :

Inventé par M.Burrows en 1994(59), cet algorithme se présente sous forme de commande informatique, utilisé en génomique permettant une facilité de calcul par l'ordinateur lors du mapping des « reads ». C'est une méthode de transformation réversible d'une séquence qui va permettre d'optimiser l'espace de stockage et la compression de données via d'autres logiciels.

Prenons l'exemple d'une séquence simple. La transformée fonctionne comme suit :

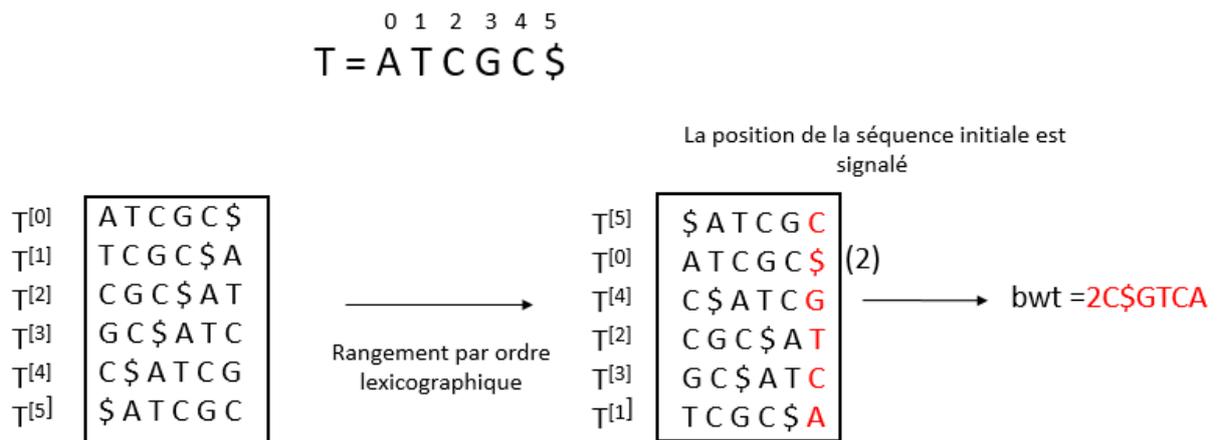


Figure 22: Schéma explicatif simplifié de la transformée de Burrows-Wheeler d'après M.Burrows(58).

La séquence est réorganisée sous forme d'une matrice avec une rotation puis rangée par ordre lexicographique. Chaque position étant renseignée, il est possible de décoder le bwt pour retrouver la séquence initiale.

L'intérêt de la transformée de Burrows-Wheeler est d'augmenter la probabilité d'une succession de plusieurs caractères identiques et d'avoir une réorganisation des données sous forme d'une bibliothèque (chaque séquence étant indexé).

En comparaison d'un algorithme dit « naïf » qui consiste à aligner chaque read l'un après l'autre sur le génome de référence. Le temps de calcul devient plus important pour des milliers de reads et considérablement allongé (de l'ordre de plusieurs mois) pour le séquençage d'un génome humain.

A cet algorithme, est associé en génétique, d'autres logiciels beaucoup plus complexes, permettant d'intégrer une « tolérance » afin de pouvoir mapper et d'accepter des substitutions. Ces logiciels prennent en compte également la qualité des reads.

2. Les différents formats de fichiers :

Les logiciels BWA et Bowtie2 produisent en sortie un fichier BAM/SAM contenant les séquences alignées à partir d'une entrée Fastq et d'un génome de référence en Fasta.

Dans un fichier SAM (Sequence Alignment Map), ouvrable dans un bloc note, il figure les résultats de l'alignement, la séquence originale, des scores de qualité ainsi que d'autres informations.

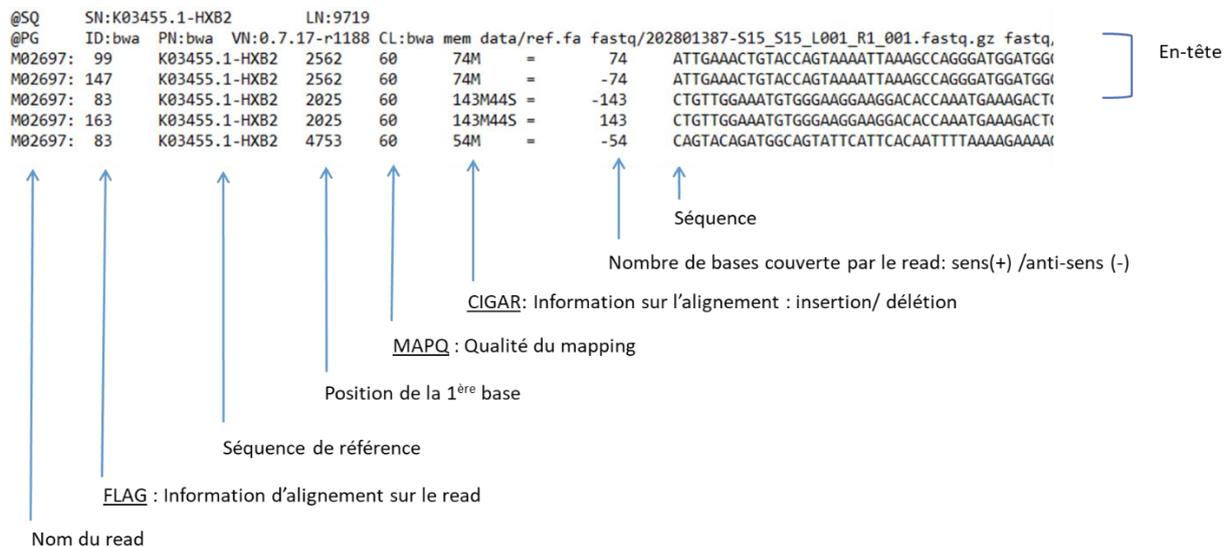


Figure 23: description du format SAM issue des données de séquençage.

Le fichier BAM (Binary Alignment Map) est le format de fichier le plus répandu pour le séquençage. C'est un fichier binaire compressé beaucoup plus petit en taille disposant des mêmes informations que le fichier SAM. Il a besoin d'un fichier index « .bai » afin d'être lu plus rapidement par les systèmes informatiques.

Il est à noter que le séquenceur Illumina® est en mesure de fournir directement des fichiers BAM si on lui renseigne la séquence de référence. C'est une procédure qui peut être mise en place avec l'aide d'un ingénieur d'application pour charger la séquence.

Pour convertir un fichier SAM en BAM, c'est le logiciel SAMtools (60) qui est utilisé.

3. Critères de qualités d'un alignement :

Plusieurs critères de qualité sont générés à l'issu d'un alignement et peuvent être vérifiés. Ils permettent d'assurer que les données produites sont analysables. Les résultats peuvent être intéressants dans une démarche de qualité du séquenceur et sont également nécessaires lors de l'étape de l'appel des variants.

L'indice FLAG dans un fichier SAM: codé sous un format binaire indique si le read est mappé ou non, en sens ou anti-sens, et s'il constitue une paire avec un second read dans un séquençage paired-end.

Le score MAPQ (Mapping Quality) dépend directement de l'aligneur utilisé et est égal à $-10\log_{10}(p)$ avec p , probabilité estimée par l'outil d'alignement que le read soit aligné au mauvais endroit. Par exemple, pour une valeur de 30, l'aligneur estime qu'il y a une chance sur mille que le read ne soit pas correctement aligné.

Le code CIGAR (Compact Idiosyncratic Gapped Alignment) donne des informations sur l'alignement des différentes bases à l'intérieur du read. S'il correspond à la séquence de référence, s'il y a une insertion, délétion ou substitution.

On retrouve également dans le fichier SAM, le score Phred identique au fichier Fastq.

On peut citer par exemple, le logiciel Picard tools(61) qui permet d'interpréter ces différentes données de qualités.

4. Visualisation des lectures :

Afin de visualiser l'alignement de nos reads, il existe le logiciel indispensable et gratuit : IGV (Integrative Genomics Viewer) (62). Il prend en charge de nombreux formats, mais classiquement, on chargera le génome de référence sous format Fasta, puis un fichier BAM et son index « .bai ».

Le logiciel intègre différents outils comme par exemple IGVtools permettant de calculer la couverture d'un génome de référence à partir d'un fichier BAM.

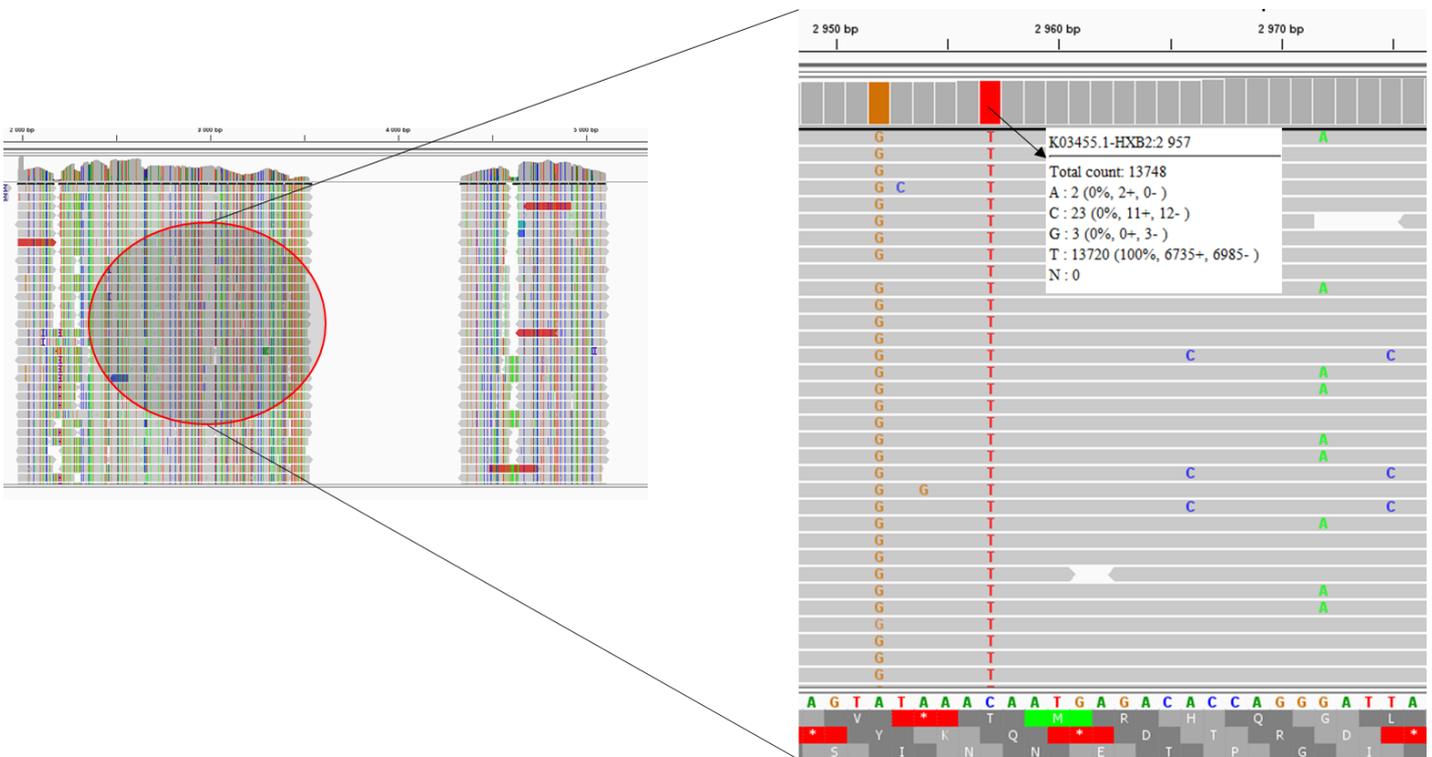


Figure 24: Visualisation de l'alignement à l'aide du logiciel IGV. Les données sont issues du séquençage.

D. L'Appel des Variants:

Cette étape consiste à énumérer tous les variants détectés. Les outils SAMtools (bcftools) ou Varscan permettent d'obtenir un fichier texte VCF (Variant Call Format)

```

##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##bcftoolsVersion=1.10.2+htslib-1.10.2-3
##bcftoolsCommand=mpileup -O b -o result/variant_raw.bcf -f data/ref.fa result/sorted_output.bam
##reference=file://data/ref.fa
##contig=<ID=K03455.1-HXB2,length=9719>
##ALT=<ID=*,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of raw reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of raw reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",Version="3">
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SGB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=ICB,Number=1,Type=Float,Description="Inbreeding Coefficient Binomial test (bigger is better)">
##INFO=<ID=HOB,Number=1,Type=Float,Description="Bias in the number of HOMs number (smaller is better)">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes for each ALT allele, in the same order as listed">
##INFO=<ID=AN,Number=1,Type=Integer,Description="Total number of alleles in called genotypes">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward , ref-reverse, alt-forward and alt-reverse bases">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Average mapping quality">
##bcftools_callVersion=1.10.2+htslib-1.10.2-3
##bcftools_callCommand=call --ploidy 1 -m -v -o result/variant.vcf result/variant_raw.bcf; Date=Wed Jun 23 10:56:32 2021
#CHROM      POS      ID      REF      ALT      QUAL      FILTER      INFO      FORMAT      result/sorted_output.bam
K03455.1-HXB2  1572    .      T      C      30.4183    .      DP=2;SGB=-0.379885;MQ0F=0;AC=1;AN=1;DP4=0,0,1,0;MQ=60    GT:PL    1:60,0
K03455.1-HXB2  1588    .      A      G      30.4183    .      DP=2;SGB=-0.379885;MQ0F=0;AC=1;AN=1;DP4=0,0,1,0;MQ=60    GT:PL    1:60,0
K03455.1-HXB2  1590    .      C      T      30.4183    .      DP=2;SGB=-0.379885;MQ0F=0;AC=1;AN=1;DP4=0,0,1,0;MQ=60    GT:PL    1:60,0
K03455.1-HXB2  1596    .      A      G      30.4183    .      DP=2;SGB=-0.379885;MQ0F=0;AC=1;AN=1;DP4=0,0,1,0;MQ=60    GT:PL    1:60,0

```

Figure 25 : Description du fichier VCF. Les données sont issues du séquençage. Chaque ligne correspond à une variation. Les informations d'intérêts mentionnées dans ce fichier sont la position, la base de référence, la base alternative et un score qualité.

La dernière étape de l'analyse bio-informatique consiste à annoter les variations pour l'interprétation biologique. En effet, ce sont les mutations d'intérêts décrites par l'ANRS qui nous intéresse. Il est intéressant de signaler que l'on peut de manière plus poussé, appliquer différents filtres de qualité ou limite de détection (3% pour la technique ABL).

E. Analyse des séquences Sanger :

Pour une analyse de manière autonome, le logiciel Seqscape®(63) permet de faire un alignement de séquence à partir d'un fichier Fasta de référence ainsi que des séquences AB1. Ce logiciel utilise l'algorithme dynamique Smith-Waterman(64) qui calcule un score d'alignement à partir d'une matrice de similarité.

Dans le cadre de ce travail, nous avons à disposition 2 pipelines permettant l'analyse des séquences. DeepCheck ABL ainsi que Geno2pheno pour les amorces ANRS.

En annexe les différentes lignes de codes sur console de commande linux pour arriver aux résultats.

IV. Résultats :

A. Contrôle qualité de l'amplification RT-PCR:

A partir des figures 26 et 27, nous pouvons visualiser des bandes d'intérêts pour chaque cible après la RT-PCR. L'échantillon n°18 et 21 ont bénéficiés d'une Nested-PCR à cause d'une absence de bandes avec la première paire d'amorce.

En annexe se trouve la table de concordance pour chaque numéro.

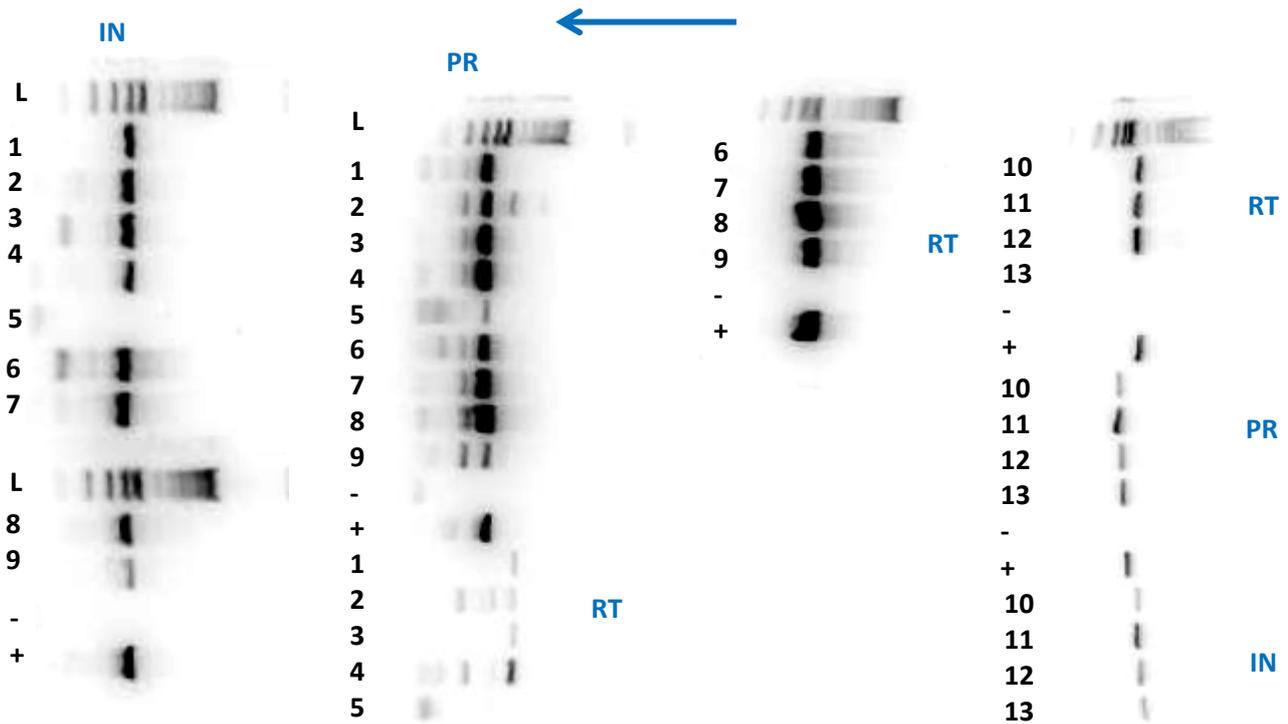


Figure 26: Résultats des Gels de la première série de NGS

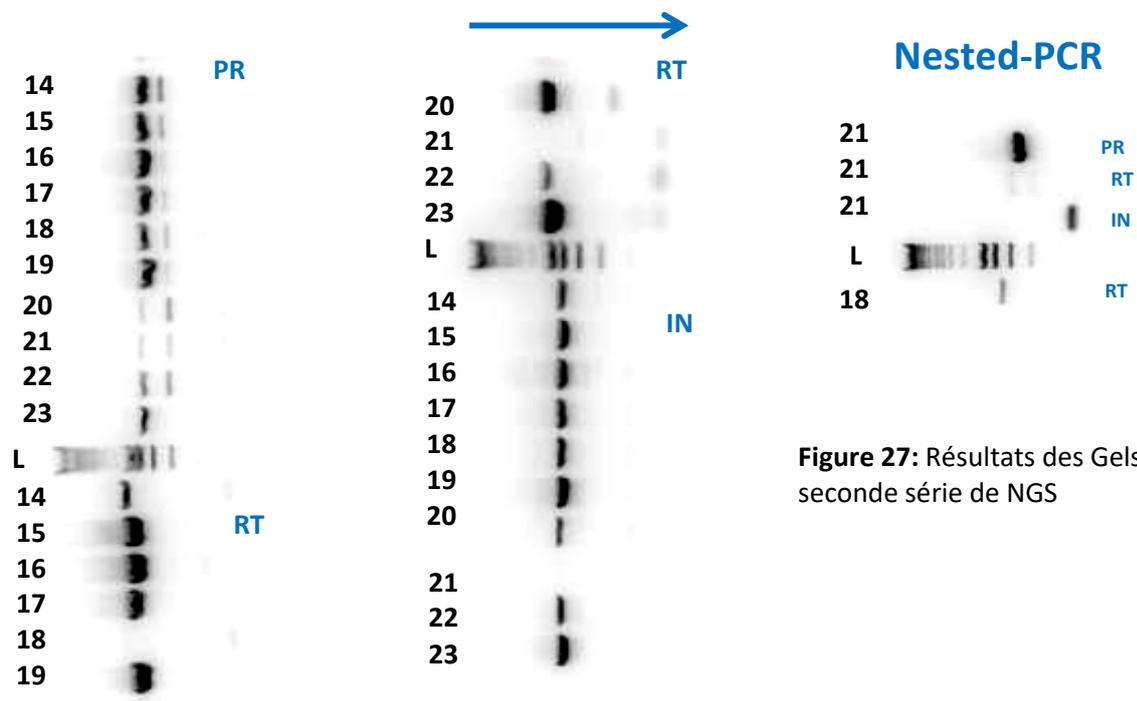


Figure 27: Résultats des Gels de la seconde série de NGS

B. Détection des mutations d'intérêts

Les deux tableaux suivants ont pour but d'exposer les mutations d'intérêts retrouvées pour chaque technique. Cela va permettre visuellement de faire un comparatif rapide pour identifier les mutations discordantes pour chaque technique. Les données NGS <20% sont également représentées, ce qui expliquera la détection de certaines mutations uniquement en NGS. Dans notre série d'échantillons, nous avons compté 13 mutations d'intérêt différentes pour la Protéase, 13 pour la reverse transcriptase et 4 pour l'intégrase.

Par soucis d'optimisation de la table, certaines mutations sont regroupées dans des cases. Pour la protéase, la mutation 10V a été retrouvée pour l'échantillon 210150931 et 210671034. Concernant la Reverse transcriptase, la mutation 69S pour 210551227 et 69N pour 210671034. La 98G pour 210110842 et la 215S pour 202942236. Et pour finir avec l'intégrase, l'échantillon 203151494 présente la double mutation 74I/M et seulement 203511396 présente la mutation 97A.

En annexe, nous avons représenté les tables brutes comprenant l'ensemble des résultats au seuil de 20% mais aussi <20%.

C. Détection des différents sous-types

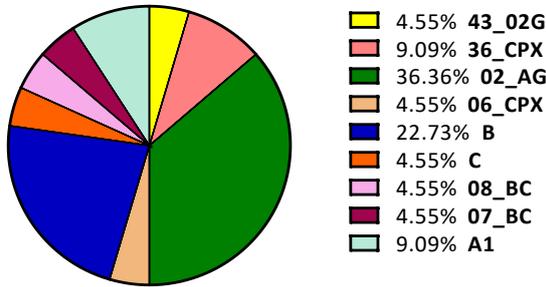
La table suivante regroupe l'ensemble de sous-types détectés en fonction de la cible et de la technique pour chaque échantillon.

Tableau 31: Liste des sous-types détectés pour chaque technique et chaque cible. **V** : Viroseq ; **S** : Sanger ABL ; **N** : NGS ; **A** : Sanger ANRS

	PROTEASE				REVERSE TRANSCRIPTASE				INTEGRASE			
	V	S	N	A	V	S	N	A	V	S	N	A
202761161	G	G	43_02G		G	G	43_02G		43_02G	G	09_CPX	
202801387	36_CPX		36_CPX		02_AG		02_AG		02_AG		02_AG	
202821685	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	B	02_AG	02_AG	02_AG	30_0206	02_AG
202942236	02_AG	G	06_CPX		02_AG	B	Echec		02_AG	G	02_AG	
203071366	B	B	B		B	B	B		B	B	B	
203151494	C	C	C		57_BC	C	C		BC	C	31_BC	
203370413	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG
203461390	02_AG	A1	02_AG		02_AG	G	02_AG		02_AG	G	02_AG	
203511396	02_AG	B	02_AG		94_CPX	D	28_BF		94_CPX	G	02_AG	
203530875	02_AG	02_AG	02_AG	02_AG	02_AG	C	31_BC	02_AG	02_AG	02_AG	02_AG	02_AG
210110842	B	B	B		B	B	B		B	B	39_BF	
210140539	A	D	A1		A1	Echec	22_01A		A1	Echec	A1	
210150931	C	C	08_BC	C	C	C	C	C	C	C	08_BC	C
210261876	F1		02_AG		F1		K		F1		F1	
210321716	02_AG	02_AG	02_AG		02_AG	02_AG	02_AG		02_AG	02_AG	02_AG	
210481981	B	B	07_BC	B	B	B	B	B	B	B	B	B
210490293	B	B	B		B	B	B		B	B	B	
210490891	B	B	B	B	B	B	B	B	B	B	28_BF	B
210551227	B	B	B	B	B	B	B	B	B	B	B	B
210671034	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	30_0206	02_AG	02_AG	02_AG	02_AG	02_AG
210720484	36_CPX	02_AG	36_CPX	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG	02_AG
C +	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1

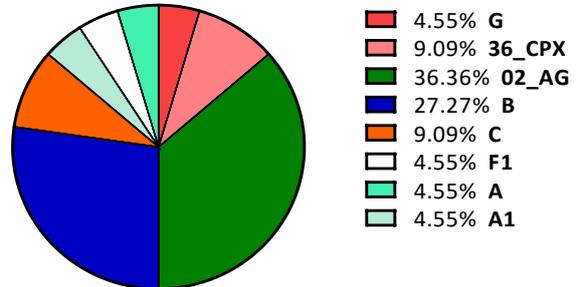
1. Les sous-types sur la protéase pour chaque technique

Sous-type NGS (Protéase)



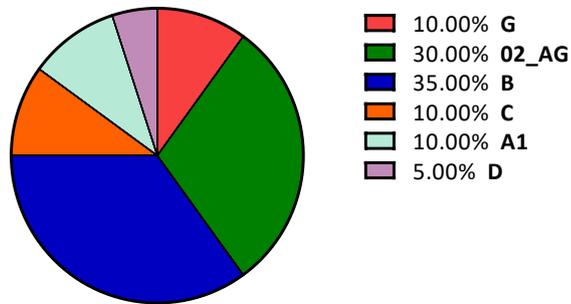
Total=22

Sous-type Viroseq (Protease)



Total=22

Sous-type Sanger ABL (Protéase)

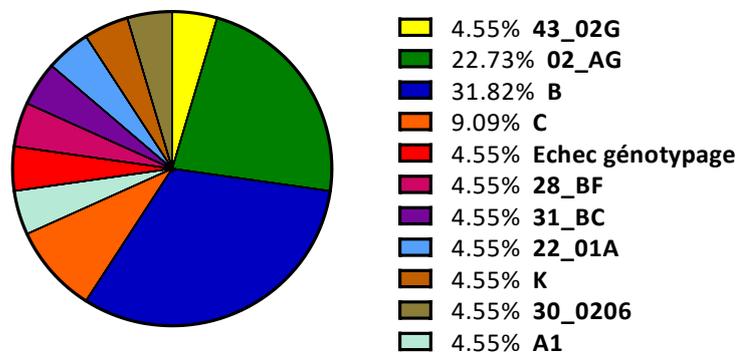


Total=20

Figure 28 : Diagrammes représentant les sous-types retrouvés avec la Protéase

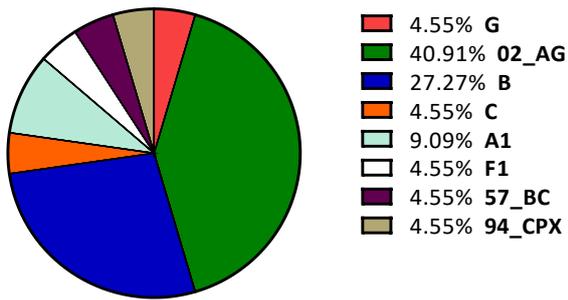
2. Les sous-types sur la reverse transcriptase pour chaque technique

Sous-type NGS (Reverse Transcriptase)



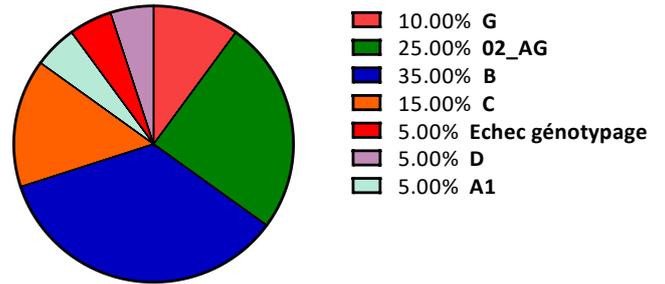
Total=22

Sous-type Viroseq (Reverse Transcriptase)



Total=22

Sous-type Sanger ABL (Reverse Transcriptase)

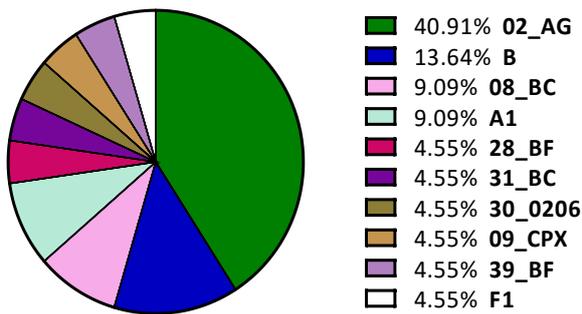


Total=20

Figure 29: Diagrammes représentant les sous-types retrouvés avec la Reverse transcriptase

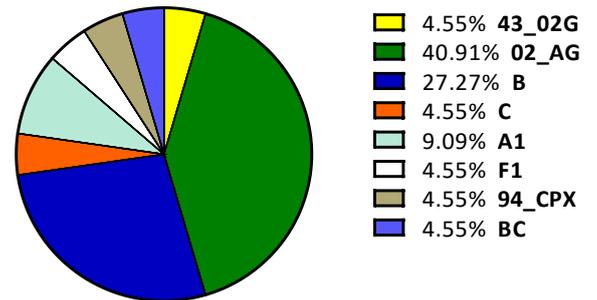
3. Les sous-types sur l'intégrase pour chaque technique

Sous-type NGS (Intégrase)



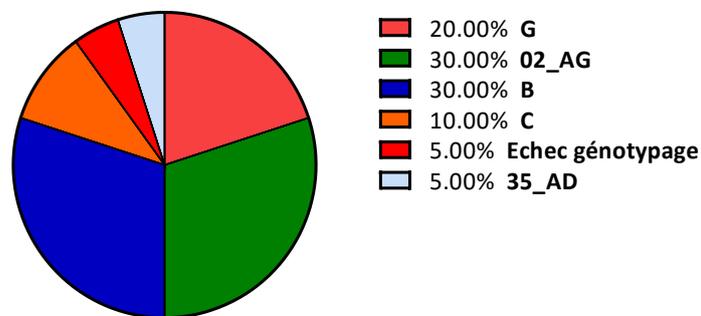
Total=22

Sous-type Viroseq (Intégrase)



Total=22

Sous-type Sanger ABL (Intégrase)



Total=20

Figure 30 : Diagrammes représentant les sous-types retrouvés avec l'Intégrase

4. Comparaison des sous-types détectés dans la technique ANRS

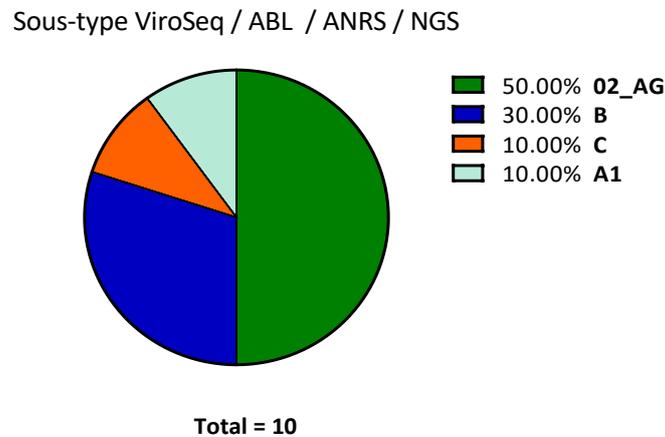


Figure 31 : Diagrammes représentant le profil des 10 échantillons testés sur les 4 techniques. Le profil est identique pour chaque cible.

D. Comparaison du nombre de mutations d'intérêts entre NGS et Sanger

Les graphiques suivants, vont permettre de comparer chaque technique pour le nombre de mutations détectées. Ils permettent également de mettre en évidence le surplus de mutations détectées lorsque le seuil du NGS est abaissé.

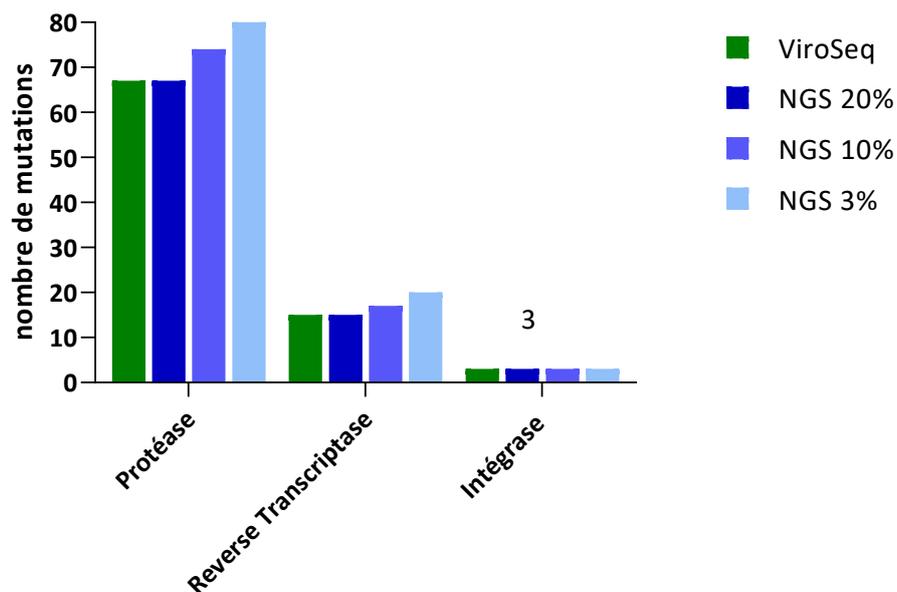


Figure 32 : Comparaison des mutations d'intérêt entre Viroseq et NGS sur nos 22 échantillons

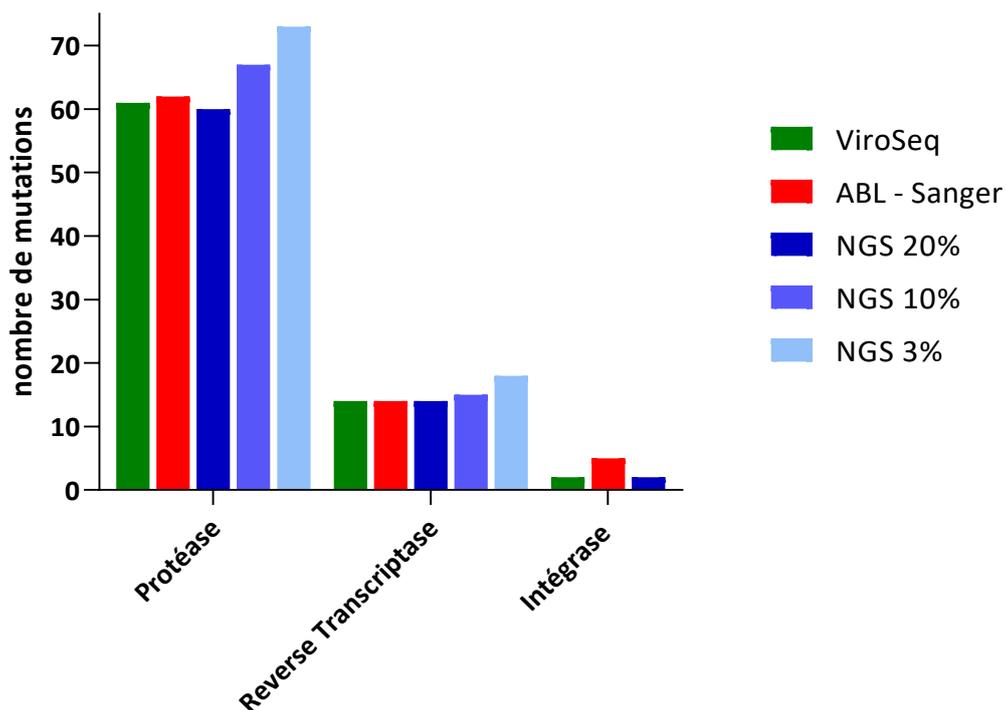


Figure 33 : Comparaison des mutations d'intérêt entre Viroseq, ABL et NGS sur les 20 échantillons

E. Comparaison du nombre de mutations dans la technique ANRS

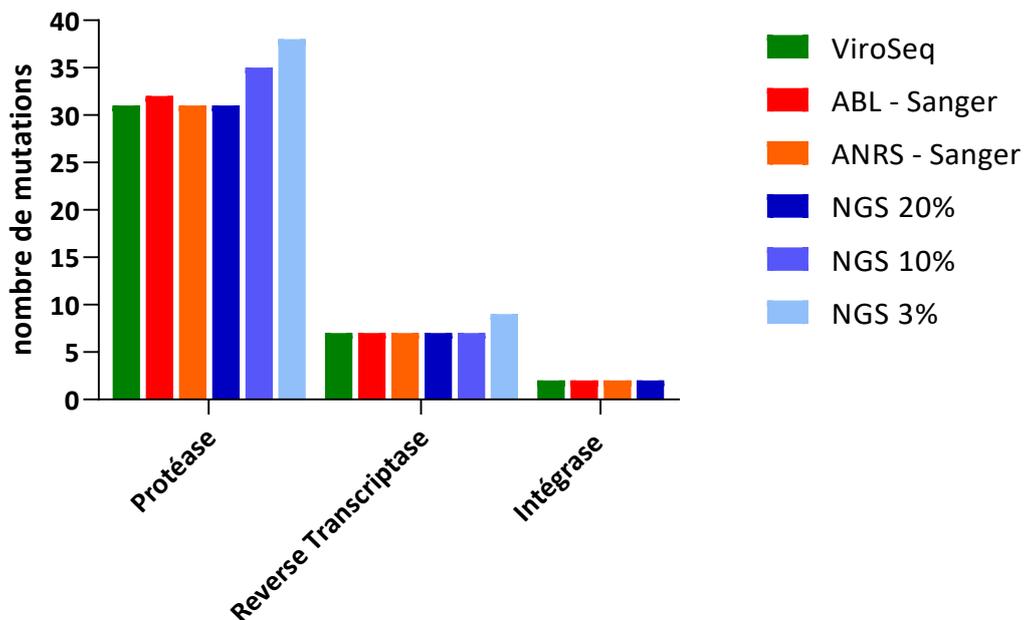


Figure 34: Comparaison des mutations d'intérêt entre les 4 techniques pour les 10 échantillons disponibles.

F. Concordance entre les différentes techniques

En ce qu'il concerne la détection des mutations d'intérêt entre les différentes techniques, la concordance a été évaluée à l'aide du coefficient Kappa. Les différents tableaux de contingence ont été créés en fonction du nombre d'échantillon qui diffère selon la technique et du nombre de mutations décrites par l'ANRS v31. J'ai également choisi le seuil de 20% de la technique NGS afin de pouvoir la comparer avec les méthodes Sanger. Pour chaque technique, le total correspond à 27 mutations d'intérêt différentes. Chaque tableau est adapté en fonction du nombre d'échantillons disponibles pour chaque technique.

Tableau 32: Table de contingence et coefficient de concordance Kappa

	ViroSeq		Coefficient κ	
	Présence (+)	Absence (-)		
NGS 20%	Présence (+)	83	2	85
	Absence (-)	1	508	510
		84	510	594
				0,979181
NGS 20%	Sanger ABL			
	Présence (+)	75	2	77
	Absence (-)	6	457	463
		81	459	540
Sanger ABL	ViroSeq			
	Présence (+)	77	4	81
	Absence (-)	1	458	459
		78	462	540
NGS 20% /Viroseq	Sanger ANRS			
	Présence (+)	40	0	40
	Absence (-)	0	230	230
		40	230	270
ABL	Sanger ANRS			
	Présence (+)	40	1	41
	Absence (-)	0	229	229
		40	230	270

G. L'apport quantitatif du NGS dans la détection des mutations

L'avantage du NGS est qu'il nous apporte une information quantitative sur l'ensemble des mutations qu'elles soient d'intérêt ou non. Nous pouvons ainsi avoir un regard sur certaines mutations discordantes entre les différentes techniques. La table 33 montre des mutations présentes dans les résultats Viroseq ou ABL qui sont peu ou pas détectées sur la technique NGS. Concernant la table 34, l'apport de mutations <20% nous informe qu'il peut y avoir des différences d'interprétations dans la résistance des antirétroviraux.

Tableau 33 : Liste non exhaustive des mutations discordantes pour Viroseq et ABL associée à la quantification NGS.

Echantillons	Mutations discordantes entre Viroseq et NGS (quantification de la mutation par NGS (%))
202761161	RT : 137K (0.04)
202801387	RT : 39P (0.17)
202821685	RT : 174H (2.86) ; 207G (3.57)
202942236	IN : 91S (0.03)
210110842	PR : 77A (0.28)
210150931	IN : 189R (0.12)
210261876	PR : 10I (0.06) ; 13V (0.22) ; 65D (0.01) IN : 157K (0.15) ; 247E (0.23) ; 259I (0.14)
210671034	PR : 12P (0.04) ; 12Q (non détecté) ; 12K (0.16) RT : 82N (0.21)
210720484	PR : 8P (0.01) ; 21Q (non détecté) ; 92E (0.01)
	Mutations d'intérêts discordantes avec le NGS (quantification NGS en %)
202942236	PR (ViroSeq et ABL): 16E (13.04) ; 69R (10.53) RT : 215S (ABL) : échec NGS
203151494	IN (ABL) : 74I/M (non détecté) 230R (0.02)

Tableau 34 : Impact de la sensibilité de certaines molécules par la détection de mutations d'intérêts
<20% ; **S** : Sensible ; **I** : Intermédiaire ; **R** : Résistant

Echantillons et molécules impactées		Interprétation de la résistance en fonction du seuil			Mutations en cause (%)
		20%	10%	3%	
210110842	Doravirine	S	S	I	RT : 98G (3.89)
	Etravirine	I	I	R	
210150931	Atazanavir/r	S	I	I	PR : 10I (10.28)
	Lopinavir/r	S	I	I	PR : 10I, 20M (10.38), 63P (10.3)
	Tipranavir/r	I	R	R	PR : 69R (7.42), 89M (11.42)
210490293	Atazanavir/r	S	S	I	PR : 60E (5.89), 71T (5.37)
210490891	Tipranavir/r	S	S	I	PR : 36I (3.14), 89M (3.14)

H. Etude du prix de l'analyse pour chaque technique

Dans l'estimation du prix, il a été pris en compte seulement les différents kits ou matériels permettant la technique ainsi que les logiciels d'interprétations. Tout le matériel « de base » d'un laboratoire n'a pas été pris en compte comme par exemple les pipettes multicanaux, cônes de pipettes, gels, plaque magnétique 96 puits et la main d'œuvre. Le prix des différents automates ne sera également pas évoqué. Enfin, j'ai calculé le prix pour une série de 5 patients et à l'année afin d'avoir une approche réaliste vis-à-vis du CHU d'Amiens.

Tableau 35: Prix brut à l'échantillon et estimation en fonction d'éventuelles remises pour les principaux réactifs et matériels pour chaque technique.

Technique	Viroseq	Sanger ABL	NGS	Sanger ANRS
Matériels et prix	Kit RT PROT - 225,15 € Kit INT - 256,98 € Exosap inclus ➔ Remise de 60% Logiciel Smartgene – Licence à 9500 €/an	Kit RT PROT - 69,04 € Kit INT - 50,21 € Sequencing V2 Purif 25,66 € ➔ Remise de 40% Prix logiciels Viroscore + DeepCheck : Licence à 6990 €/an Remisé à 1048 €/an	Kit RT PROT - 69,04 € Kit INT - 50,21€ Nano Kit V2 - 13,92 € Billes - 5,32 € Qubit® - 2,06 € DeepChek Assay NGS Library preparation - 45,83 € ➔ Remise de 30% Bioanalyzer® (BioA) – 57€ Prix logiciels : Viroscore + DeepCheck : Licence à 6990 €/an Remisé 1048 €/an	Amorces 3,6 € ➔ Remise de 10% RT PCR Qiagen One step - 13,8 € Nested PCR - 3,81 € Purif exosap - 3,09 € ➔ Remise de 5% Big Dye Terminator - 43,5 € ➔ Remise de 18% X Terminator - 36,27 € Logiciel Smartgene – Licence à 9500 €/an
Somme brute	577,12 €	214,82 €	256,28 € (sans BioA) - 313,28 € (avec BioA)	199,07 €
Somme remisé	287,85 €	145,10 €	183,11€ (sans BioA) - 240,11 € (avec BioA)	190,73 €
Série de 5 patients (remisé)	1439,25 €	725,5 €	1179,95 € (sans BioA) - 1464,95 € (avec BioA)	953,63 €
Prix à l'année pour 100 échantillons. (remisé)	28785 €	14510 €	23599 € (sans BioA) - 29299€ (avec BioA)	19073 €

V. Discussion

Dans la perspective de choisir une nouvelle technique pour le laboratoire de virologie du CHU d'Amiens, nous avons étudié une technique NGS (ABL) ainsi que 3 techniques Sanger : le kit Viroseq actuellement en fonctionnement qui va subir un arrêt de commercialisation, le kit Sanger ABL et la technique maison comprenant les amorces de l'ANRS.

Plusieurs éléments vont être exposés en prenant compte des forces et faiblesses de chacune des techniques afin de permettre de faire le meilleur choix pour l'avenir du laboratoire.

Le premier critère le plus important est la détection fiable des mutations d'intérêt. En effet, les résultats concernant le coefficient Kappa retrouvent globalement une excellente concordance ($K > 94\%$) entre chaque technique au seuil de 20%. (Table32) Les différences en termes de mutations d'intérêt sont minimales et il est intéressant de constater qu'elles n'influencent pas le rendu global des résistances des antirétroviraux pour chacune des techniques étudiées. Il est néanmoins important de souligner le faible nombre d'échantillons pour la technique ANRS qui peut donc constituer un biais. ($K = 1$).

Quelques exceptions dans notre étude, il concerne l'échantillon 203151494 qui possède 2 mutations sur l'intégrase (74I/M et 230R) avec le sanger ABL, non retrouvées sur les autres techniques. (Table30) Il s'agit manifestement d'une erreur d'interprétation de la part du logiciel qui vient fausser les résultats. (Annexe IX) Il est important pour le biologiste dans sa prise de décision de pouvoir observer les chromatogrammes et de constater que l'échantillon dispose d'une mauvaise couverture entre les brins sens et anti-sens. Pour cet échantillon, le logiciel rend une résistance sur l'ensemble des inhibiteurs de l'intégrase. Le logiciel DeepCheck®/ViroScore par son ergonomie permet néanmoins de vérifier et constater les mutations d'intérêts. Un autre biais visible concerne l'échantillon 210150931. Classiquement associé au seuil de 20%, nous observons la détection de la mutation 89M sur la protéase avec le kit Sanger ABL®. Cette mutation est quantifiée par le NGS à 11.42%. Nous l'avons observé seulement sur cet échantillon, ce qui explique cet écart sur les histogrammes. (Figure33) Ce résultat est isolé sur un seul échantillon.

L'intérêt du NGS réside dans sa capacité à détecter des mutations à des seuils $< 20\%$. Dans notre étude, nous avons détecté 13 mutations de plus sur l'ensemble des échantillons pour la Protéase et 5 de plus pour la reverse transcriptase. (figure32-34) Il est à noter que le seuil significatif où l'analyse est capable de garantir un niveau de confiance à 99% est de 3%. En dessous, de ce seuil, les milliers de mutations détectées représentent un bruit de fond. Une autre particularité du NGS, est qu'il est tout à fait possible d'analyser des échantillons avec une charge virale $< 3 \log$ copies/mL. Nous observons

également que la détection de mutations d'intérêts inférieure au seuil de 20% peut changer l'interprétation de la résistance des antirétroviraux (Table34). Cependant Il existe encore très peu de données dans la littérature concernant un éventuel impact clinique, et il serait peut être intéressant d'étudier le génome chez des patients naïfs en échec thérapeutique afin d'identifier un éventuel impact clinique qui justifierai l'abaissement de ce seuil.

A propos des sous-types viraux, sur l'ensemble des échantillons, ce sont les sous-types O2_AG et B qui prédominent. Il est difficile de comparer les résultats entre les techniques du fait de multiples paramètres. L'algorithme, l'alignement, la qualité des reads et le logiciel interpréteur peuvent jouer dans la différence de rendu des sous-types viraux. Par exemple, le NGS ne détecte aucun sous type G mais nous constatons qu'il rend beaucoup plus de formes recombinantes que les autres techniques. Un des avantages serait que cette sensibilité et spécificité accrues du NGS pourraient être intéressantes dans un contexte épidémiologique pour suivre l'apparition de variant minoritaire par exemple. En contrepartie, certaines études mentionnent le fait que le séquençage haut débit dans la détection du tropisme viral pourrait avoir un impact clinique lors de la mise en place d'un traitement par Maraviroc®(65). Pour finir, nous ne détectons aucune différence sur les 9 échantillons communs à toutes les techniques en termes de sous-types viraux.

Un des défauts des techniques ABL et ANRS, est que pour chaque cible, nous ne disposons que d'une paire d'amorce. En cas d'échecs de l'analyse, nous ne pourrions pas rendre de résultats. Le kit Viroseq® avait l'avantage de disposer de plusieurs paires d'amorces chevauchantes, ce qui constitue une excellente approche dans le rendu des mutations. Outre l'aspect technique, c'est ce qui peut expliquer le fait que l'on retrouve bien plus de mutations supplémentaires (Annexe2), notamment en 5' et 3' sur les techniques ABL et ANRS. Cet aspect du séquençage n'est pas du tout un problème pour le NGS du fait de la technologie employée avec l'étape du « Clustering ». Mieux encore, elle permet d'obtenir des informations quantitatives sur toutes les mutations détectées. Néanmoins, ce n'est pas une technique exempte de défauts. Le NGS ne couvre pas les positions 319 à 348 AA pour la reverse transcriptase ainsi que les AA <45 et >284 pour l'intégrase. Nous pouvons noter qu'aucunes mutations d'intérêt ne sont présentes à ces positions.

Au regard du prix entre les techniques (Table35), il convient de choisir une solution économique ou du moins, une technique ayant un bon rapport performance/économie. Il faut également souligner l'importance de choisir une technique facilement accréditable (marquage CE) et de disposer de logiciels simples et intuitifs. La technique la moins chère est sans aucun doute le séquençage « maison » avec les amorces de l'ANRS. Néanmoins, l'absence de couverture optimale due

au non chevauchement des séquences fait que les résultats peuvent être de moins bonne qualité et moins précis.

En prenant le cas NGS, nous remarquons que le prix n'est pas excessif par rapport à la technique Viroseq® et de ce fait, représente une solution très intéressante sur le plan performance, avec un possible regard sur la quantification des mutations, une détection plus sensible et l'application de différents critères et filtres de qualité garantissant un niveau de confiance à 99%. Dans le cadre de ce travail, l'analyse peut quand même revenir plus chère que les autres techniques. Le Bioanalyzer® n'étant pas obligatoire, mais conseillé au moins en début de la mise en place de méthode pour vérifier la qualité notre librairie, gonfle le prix. Il faut savoir que j'ai également calculé le prix dans le cadre d'une optimisation totale, c'est-à-dire lorsque nous réalisons l'analyse de 24 échantillons en même temps. De manière réaliste, sur l'année 2019-2020, nous avons réalisé 20 séries de 5 patients, ce qui en fait la technique la plus chère. Il existe tout de même différentes stratégies pour palier à ce faible nombre d'échantillons comme par exemple l'analyse de divers matériels génétiques sur une même cartouche : Sars-Cov2, VHB, VHC, CMV, HSV, HPV, ARN16S, résistance dans la Tuberculose. Les séquences des 3 cibles du VIH que nous devons analyser représentent une part tout à fait négligeable dans la capacité totale (0.5GB) de la cartouche Nano V2. A titre d'exemple, nous pourrions analyser 96 patients sur une même cartouche. A cela s'ajoute tous les problèmes logistiques. La technique est bien plus longue que Sanger (2.5 jours de technique avant de pouvoir lancer un Séquençage) et nécessite une formation pour le technicien. Afin de rendre la technique plus attrayante, il existe tout de même de l'équipement permettant d'automatiser la préparation de librairie et de minimiser le geste humain. Cela réduirait les coûts techniciens et permettrait une bonne reproductibilité et une productivité optimisée. Nous avons également essayé d'instaurer une méthode NGS maison avec un design d'amorces sans succès causé par des erreurs techniques (Absence de fragmentation dans notre protocole). La technique maison aurait pu permettre d'économiser sur les kits de PCR mais il est toujours nécessaire d'utiliser un kit compatible illumina® pour le séquençage sans compter les problèmes d'accréditation de la technique.

Concernant le Sanger ABL, il constitue une très bonne alternative. C'est la technique la moins chère et le logiciel DeepCheck® dispose d'une bonne ergonomie et permet un rendu clair et précis tout en ayant la possibilité de regarder les chromatogrammes de séquençage. Le kit dispose également d'un marquage CE facilitant l'accréditation.

Aujourd'hui, Il n'est plus à démontrer l'apport incontestable du NGS dans les disciplines oncologiques et maladies génétiques avec sa perpétuelle évolution technologique. En revanche, nous pouvons nous poser la question de l'intérêt du NGS en microbiologie. Une nouvelle approche dans le

séquençage haut débit et qui représente une révolution dans le domaine, est l'approche « Shotgun sequencing » ou métagénomique globale qui a pour but de séquencer l'ensemble des acides nucléiques d'un prélèvement. D'un point de vue technologique, elle permet de s'affranchir des approches PCR nécessitant des amorces spécifiques pour une cible. Le process utilise des enzymes de fragmentation, la réalisation d'une librairie et le séquençage massif. En pratique courante, cette approche peut être intéressante en épidémiologie, dans la détection de virus émergent grâce au séquençage « de novo » (sans référence) ou le suivi épidémiologique de zones endémiques. Certaines études suggèrent même qu'il est possible de gagner du temps dans la prise en charge du patient avec une simplification de la prise en charge au laboratoire.(66) Elle permettrait également en bactériologie, de s'affranchir des contraintes de culture et de données phénotypiques, de faire un diagnostic sur la base génomique et de cibler le traitement de manière efficace. Enfin, des travaux en cours vont également essayer de démontrer une efficacité supérieure en comparaison à l'ARN16S dans le cadre d'impasse diagnostique.

Au final, l'apparition de ces nouvelles technologies qui s'apparentent à de la médecine personnalisée dont il reste à définir dans les recommandations, leur place dans la prise en charge du patient. Lorsque nous raisonnons en termes d'urgence, il faut garder à l'esprit qu'un délai minimum de 48 à 72h est nécessaire pour obtenir des reads. Néanmoins, elles représentent sans aucun doute une révolution dans l'expertise globale en microbiologie et infectiologie.

En conclusion, le NGS au sens large est pour moi une technique d'avenir qu'il est intéressant d'aborder en routine au laboratoire. Mais compte tenu d'un faible échantillonnage pour le génotypage du VIH au CHU d'Amiens et de l'ensemble des problèmes logistiques et bio-informatiques, la technique Sanger ABL constitue une bonne alternative au kit Viroseq®, cette technique n'est pas difficile, le logiciel d'interprétation est complet et ergonomique. Plus tard, le Sanger ABL peut être aussi considéré comme un tremplin vers le NGS grâce aux kits d'amorces communes et l'accès aux logiciels d'analyses.

Bibliographie

1. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. *Cold Spring Harb Perspect Med.* 1 sept 2011;1(1):a006841-a006841.
2. Foley BT, Leitner T, Paraskevis D, Peeters M. Primate immunodeficiency virus classification and nomenclature: Review. *Infect Genet Evol.* 1 déc 2016;46:150-8.
3. Feng Y, He X, Hsi JH, Li F, Li X, Wang Q, et al. The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s. *AIDS.* juill 2013;27(11):1793-802.
4. Mir D, Jung M, Delatorre E, Vidal N, Peeters M, Bello G. Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination. *Infect Genet Evol.* 1 déc 2016;46:190-9.
5. Mourez T, Simon F, Plantier J-C. Non-M variants of human immunodeficiency virus type 1. *Clin Microbiol Rev.* juill 2013;26(3):448-61.
6. VIH/sida. Disponible sur: /maladies-et-traumatismes/infections-sexuellement-transmissibles/vih-sida
7. Masson E. Value of a national administrative database to guide public decisions: From the *système national d'information interrégimes de l'Assurance Maladie (SNIIRAM)* to the *système national des données de santé (SNDS)* in France. EM-Consulte. Disponible sur: <https://www.em-consulte.com/article/1140905/value-of-a-national-administrative-database-to-gui>
8. Cazein F, Sommen C, Moutengou E, Laporal S, Ramus C, Bruyand M, et al. Focus. Dépistage du VIH : état des lieux en 2019 et sur les sept premiers mois de l'année 2020. *Bull Epidémiol Hebd.* 2020;(33-34):685-7. http://beh.santepubliquefrance.fr/beh/2020/33-34/2020_33-34_5.html. :3.
9. Cazein F, Sommen C, Pillonel J, Bruyan M, Ramus C, Pichon P, et al. Activité de dépistage du VIH et circonstances de découverte de l'infection à VIH, France 2018. *Bull Epidémiol Hebd.* 2019;(31-32):615-24. http://beh.santepubliquefrance.fr/beh/2019/31-32/2019_31-32_1.html. Disponible sur: http://beh.santepubliquefrance.fr/beh/2019/31-32/2019_31-32_1.html
10. Article - Bulletin épidémiologique hebdomadaire. Disponible sur: http://beh.santepubliquefrance.fr/beh/2015/9-10/2015_9-10_1.html
11. Blanc A, Bonnet F, Brun-Vezinet F, Costagliola D, Dabis F, Delobel P, et al. Groupe d'experts pour la prise en charge du VIH. 2016;18.
12. Arthur LO, Bess JW, Sowder RC, Benveniste RE, Mann DL, Chermann JC, et al. Cellular proteins bound to immunodeficiency viruses: implications for pathogenesis and vaccines. *Science.* 18 déc 1992;258(5090):1935-8.
13. Flexner C. HIV-protease inhibitors. *N Engl J Med.* 30 avr 1998;338(18):1281-92.
14. Roberts NA, Martin JA, Kinchington D, Broadhurst AV, Craig JC, Duncan IB, et al. Rational design of peptide-based HIV proteinase inhibitors. *Science.* 20 avr 1990;248(4953):358-61.

15. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science*. 27 nov 1998;282(5394):1669-75.
16. Sarafianos SG, Marchand B, Das K, Himmel D, Parniak MA, Hughes SH, et al. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol*. 23 janv 2009;385(3):693-713.
17. Lukashov VV, Goudsmit J. HIV heterogeneity and disease progression in AIDS: a model of continuous virus adaptation. *AIDS Lond Engl*. 1998;12 Suppl A:S43-52.
18. Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc Natl Acad Sci*. 1 juill 1993;90(13):6125-9.
19. Fisher AG, Feinberg MB, Josephs SF, Harper ME, Marselle LM, Reyes G, et al. The trans-activator gene of HTLV-III is essential for virus replication. *Nature*. mars 1986;320(6060):367-71.
20. Luznik L, Kraus G, Guatelli J, Richman D, Wong-Staal F. Tat-independent replication of human immunodeficiency viruses. *J Clin Invest*. janv 1995;95(1):328-32.
21. Sodroski J, Goh WC, Rosen C, Dayton A, Terwilliger E, Haseltine W. A second post-transcriptional trans-activator gene required for HTLV-III replication. *Nature*. mai 1986;321(6068):412-7.
22. Malim MH, Böhnlein S, Hauber J, Cullen BR. Functional dissection of the HIV-1 Rev trans-activator—Derivation of a trans-dominant repressor of Rev function. *Cell*. 14 juill 1989;58(1):205-14.
23. Collins KL, Chen BK, Kalams SA, Walker BD, Baltimore D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature*. 22 janv 1998;391(6665):397-401.
24. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*. 8 août 2002;418(6898):646-50.
25. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell*. 13 juin 2003;113(6):803-9.
26. Heinzinger NK, Bukrinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, et al. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc Natl Acad Sci U S A*. 19 juill 1994;91(15):7311-5.
27. Oberste MS, Gonda MA. Conservation of amino-acid sequence motifs in lentivirus Vif proteins. *Virus Genes*. janv 1992;6(1):95-102.
28. Maldarelli F, Chen MY, Willey RL, Strebel K. Human immunodeficiency virus type 1 Vpu protein is an oligomeric type I integral membrane protein. *J Virol*. août 1993;67(8):5056-61.
29. Neil SJD, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*. 24 janv 2008;451(7177):425-30.
30. Wang K, Samudrala R, Mittler J. Antivirogram or PhenoSense: A comparison of their reproducibility and an analysis of their correlation. *Antivir Ther*. 1 nov 2004;9:703-12.

31. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. déc 1977;74(12):5463-7.
32. MiSeq System | Focused power for targeted gene and small genome sequencing. Disponible sur: <https://emea.illumina.com/systems/sequencing-platforms/miseq.html>
33. Balzer S, Malde K, Lanzén A, Sharma A, Jonassen I. Characteristics of 454 pyrosequencing data—enabling realistic simulation with flowsim. *Bioinformatics*. 15 sept 2010;26(18):i420-5.
34. Dortet L, Bonnin R, Naas T. Impact du séquençage d'ADN à haut débit sur la surveillance des épidémies de bactéries multi-résistantes aux antibiotiques. *Feuill Biol*. 1 janv 2017;354.
35. Semiconductor Sequencing Technology - FR [Internet]. Disponible sur: [//www.thermofisher.com/fr/fr/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html](http://www.thermofisher.com/fr/fr/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html)
36. HIV French Resistance - HIV-1 genotypic drug resistance interpretation's algorithms [Internet]. [cité 11 oct 2021]. Disponible sur: <http://www.hivfrenchresistance.org/>
37. Zaharatos GJ, Wainberg MA. Update on rilpivirine: a new potent non-nucleoside reverse transcriptase inhibitor (NNRTI) of HIV replication. *Ann Med*. mai 2013;45(3):236-41.
38. Larrouy L, Chazallon C, Landman R, Capitant C, Peytavin G, Collin G, et al. Gag Mutations Can Impact Virological Response to Dual-Boosted Protease Inhibitor Combinations in Antiretroviral-Naïve HIV-Infected Patients. *Antimicrob Agents Chemother*. juill 2010;54(7):2910-9.
39. Nowicka-Sans B, Protack T, Lin Z, Li Z, Zhang S, Sun Y, et al. Identification and Characterization of BMS-955176, a Second-Generation HIV-1 Maturation Inhibitor with Improved Potency, Antiviral Spectrum, and Gag Polymorphic Coverage. *Antimicrob Agents Chemother*. juill 2016;60(7):3956-69.
40. Recordon-Pinson P, Raymond S, Bellecave P, Marcelin AG, Soulie C, Descamps D, et al. HIV-1 Dynamics and Coreceptor Usage in Maraviroc-Treated Patients with Ongoing Replication. *Antimicrob Agents Chemother*. févr 2013;57(2):930-5.
41. Geretti AM, Armenia D, Ceccherini-Silberstein F. Emerging patterns and implications of HIV-1 integrase inhibitor resistance. *Curr Opin Infect Dis*. déc 2012;25(6):677-86.
42. Lambert-Niclot S, Charpentier C, Storto A, Fofana DB, Soulié C, Fourati S, et al. Prevalence of pre-existing resistance-associated mutations to rilpivirine, emtricitabine and tenofovir in antiretroviral-naïve patients infected with B and non-B subtype HIV-1 viruses. *J Antimicrob Chemother*. juin 2013;68(6):1237-42.
43. Avettand-Fènoël V, Hocqueloux L, Ghosn J, Cheret A, Frange P, Melard A, et al. Total HIV-1 DNA, a Marker of Viral Reservoir Dynamics with Clinical Implications. *Clin Microbiol Rev*. oct 2016;29(4):859-80.
44. Prazuck T, Zucman D, Avettand-Fènoël V, Ducasse E, Bornarel D, Mille C, et al. Long-term HIV-1 virologic control in patients on a dual NRTI regimen. *HIV Clin Trials*. juin 2013;14(3):120-6.
45. Chun T-W, Justement JS, Murray D, Hallahan CW, Maenza J, Collier AC, et al. Rebound of plasma viremia following cessation of antiretroviral therapy despite profoundly low levels of HIV reservoir: implications for eradication. *AIDS Lond Engl*. 27 nov 2010;24(18):2803-8.

46. EACS Guidelines EACSociety. Disponible sur: <https://www.eacsociety.org/guidelines/eacs-guidelines/>
47. EMAG® bioMérieux France. Disponible sur: <https://www.biomerieux.fr/diagnostic-clinique/emagr>
48. Boom R, Sol CJ, Salimans MM, Jansen CL, Wertheim-van Dillen PM, van der Noordaa J. Rapid and simple method for purification of nucleic acids. *J Clin Microbiol.* mars 1990;28(3):495-503.
49. Qubit Fluorometric Quantitation - FR Disponible sur: [//www.thermofisher.com/fr/fr/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html](http://www.thermofisher.com/fr/fr/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit.html)
50. Bioanalyzer Automated Electrophoresis for Sample Quality Control | Agilent [Internet]. Disponible sur: https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems?gclid=CjwKCAjw2ZaGBhBoEiwA8pFP_oVHWxhw2ziyz-2HoOy2_BIMHaVBXpAXn4oylnORV_Xgc85_8e6I4BoCWYgQAvD_BwE&gclidsrc=aw.ds
51. What is the PhiX Control v3 Library and what is its function in Illumina Next Generation Sequencing? Disponible sur: <https://emea.support.illumina.com/bulletins/2017/02/what-is-the-phix-control-v3-library-and-what-is-its-function-in-.html>
52. Galaxy [Internet]. Disponible sur: <https://usegalaxy.org/>
53. Oracle VM VirtualBox [Internet]. Disponible sur: <https://www.virtualbox.org/>
54. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* avr 2010;38(6):1767-71.
55. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* mars 1998;8(3):175-85.
56. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* mars 1998;8(3):186-94.
57. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data Disponible sur: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
58. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl.* 15 juill 2009;25(14):1754-60.
59. Burrows M, Wheeler DJ. A block-sorting lossless data compression algorithm. :24.
60. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl.* 15 août 2009;25(16):2078-9.
61. Picard Tools - By Broad Institute [Internet]. Disponible sur: <https://broadinstitute.github.io/picard/>
62. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* janv 2011;29(1):24-6.

63. SeqScape™ Software v3.0, Initial License [Internet]. Disponible sur: <https://www.thermofisher.com/order/catalog/product/4474978>
64. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol.* 25 mars 1981;147(1):195-7.
65. Swenson LC, Chui CKS, Brumme CJ, Chan D, Woods CK, Mo T, et al. Genotypic analysis of the V3 region of HIV from virologic nonresponders to maraviroc-containing regimens reveals distinct patterns of failure. *Antimicrob Agents Chemother.* déc 2013;57(12):6122-30.
66. Audebert C, Hot D, Lemoine Y, Caboche S. Le séquençage haut-débit - Vers un diagnostic basé sur la séquence complète du génome de l'agent infectieux. *médecine/sciences.* 1 déc 2014;30(12):1144-51.

Annexes

Annexe I : Lignes de code sur commande linux de l'alignement des séquences à l'appel des variants.

```
1 Analyse bioinformatique sur Linux :
2
3 1- Installer les outils :
4
5 $ sudo apt update
6 $ sudo apt install bwa
7 $ sudo apt install samtools
8 $ sudo apt install bcftools
9
10 2- Pour simplification :
11 dossier de travail : Bioinformatique
12 référence : HXB2.txt
13 Fastq : S1R1.fastq, S1R2.fastq
14
15 3- ligne de commande :
16
17 $ cd ~/Bioinformatique
18 $ bwa index HXB2.txt # index de la référence
19 $ bwa mem HXB2.txt S1R1.fastq S1R2.fastq > alignement.sam # alignement
20 $ samtools view alignement.sam -o alignement.bam # sam en bam
21 $ samtools sort alignement.bam -o alignement.sorted.bam # tri dans le fichier bam
22 $ samtools index alignement.sorted.bam # index du bam (.bai)
23 $ bcftools mpileup -o raw.bcf -f HXB2.txt alignement.sorted.bam # génère un fichier avec la couverture pour chaque base
24 $ bcftools call --ploidy 1 -m -v -o variants.vcf raw.bcf # détection des SNPs
25 $ vcfutils.pl varFilter variants.vcf > final_variants.vcf # filtre et report des SNPs dans le fichier VCF
```

Annexe II : Table brute des résultats de séquençage au seuil de 20% (légende en dessous)

Echantillons	Sanger ViroSeq			NGS ABL			Sanger ABL		
	PR	RT	IN	PR	RT	IN	PR	RT	IN
202761161(+)	3I, 10I , 11A, 13V, 14R, 20I, 35Q, 36I , 37D, 41K, 57K, 63T, 64M, 67E, 69R , 70R, 72T, 82I, 89M	35T, 60I, 122P, 123S, 135K , 137K 173T, 174K, 177E, 178M, 200A, 207E, 211K, 214F, 245Q, 248N, 250E, 272S, 277K, 291D, 292I, 293V, 294Q	50I, 72I, 112V, 123S, 124S, 125A, 127K, 134N, 136T, 167D/V , 193E, 201I, 206S, 232D, 234I, 255N, 267T/I/S/L/F 283G	3I, 10I , 11A, 13V, 14R, 20I, 35Q, 36I , 37D, 41K, 57K, 63T, 64M, 67E, 69R , 70R, 72T, 82I, 89M	35T, 60I, 122P, 123S, 173T, 174K, 177E, 178M, 200A, 207E, 211K, 214F, 245Q, 248N, 250E, 272S, 277K, 291D, 292I, 293V, 294Q	50I, 72I, 112V, 123S, 124S, 125A, 127K, 134N, 136T, 193E, 201I, 206S, 232D, 234I , 254S, 255N, 283G	3I, 10I , 11A, 13V, 14R, 20I, 35Q, 36I , 37D, 41K, 57K, 63T, 64M, 67E, 69R , 70R, 72T, 82I, 89M	35T, 60I, 122P, 123S, 173T, 174K, 177E, 178M, 200A, 207E, 211K, 214F, 245Q, 248N, 250E, 272S, 277K, 291D, 292I, 293V, 294Q, 315D , 321S	50I, 72I, 95R , 96G , 97I, 100V , 112V, 123S, 124S, 125A, 127K, 134N, 136T, 193E, 201I, 206S, 229Y , 230P , 231S , 232D, 233S , 234I , 237R , 329G , 241Q , 254S, 255N, 261T/R, 283G
202801387(+)	3I, 13V, 14R, 17I, 17E , 20I, 20V, 35D, 36I , 37N, 41K, 64M, 69K , 89M	35T, 36D, 39P , 39I , 39L, 60I, 64R, 90I , 122K, 135V, 162A, 173T, 174K, 177E, 200A, 207E, 211K, 245Q, 272A, 277K, 286A, 287R, 291D, 292I, 293V, 294T	72I, 101I, 112V, 119G, 122I, 123S, 125A, 127K, 134N, 136T, 195T, 201I, 203M, 206S, 232D, 234I	3I, 13V, 14R, 17I, 20I, 20V, 35D, 36I , 37N, 41K, 64M, 69K , 89M	35T, 36D, 39L, 60I, 64R, 90I , 122K, 135V, 162A, 173T, 174K, 177E, 200A, 207E, 211K, 245Q, 272A, 277K, 286A, 287R, 291D, 292I, 293V, 294T	72I, 101I, 112V, 119G, 122I, 123S, 125A, 127K, 134N, 136T, 195T, 201I, 203M, 206S, 232D, 234I, 268L , 270H , 283G			
202821685(-)	3I, 13V, 14R, 17E, 20I, 35D, 36I , 37N, 41K,	5V, 30N , 35T, 36A , 60I, 122K, 123N , 174H , 177E, 196E, 200A, 207G , 211K,	57G, 101I, 112V, 123S, 124G, 125A, 127K, 134N,	3I, 13V, 14R, 17E, 20I, 35D, 36I , 37N, 41K, 64M, 69K , 89M	5V, 35T, 49R , 60I, 122K, 123E , 135T , 162A , 162C , 173T , 174K , 177E, 196E,	57G, 101I, 112V, 123S, 124G, 125A, 127K, 134N,	3I, 13V, 14R, 17E, 20I, 35D, 36I , 37N,	2insY, 3Y, 4L, 5V (Echec analyse)	57G, 101I, 112V, 123S, 124G, 125A, 127K, 134N,

	64M, 69K , 89M	214F, 245K , 276I , 277K, 283I , 293V, 297V	135V, 136T, 201I, 206S, 211Q, 218I, 232D, 234I, 256E		200A, 207E , 211K, 214F, 245Q , 248D , 272A , 275Q , 277K, 286A , 293V	135V, 136T, 201I, 206S, 211Q, 218I, 232D, 234I, 256E, 278A , 283G	41K, 64M, 69K , 89M		135V, 136T, 201I, 206S, 211Q, 218I, 232D, 234I, 256E, 278A , 283G
202942236 (-) (echec NGS)	3I, 12I, 13A 16E , 19P, 20P, 22V, 36I , 37N, 41K, 63Q, 68E, 69R , 89M	35T, 36D, 39N, 60I, 103N , 162A, 173T, 174K, 177E, 178M, 195L, 200A, 207E, 214F, 228L, 228F, 238T, 245Q, 248D, 250E, 251D, 272A, 277K, 291D, 292I, 293V, 294T	91S , 112V, 119P, 123S, 125A, 127K, 134N, 135V, 136T, 195C, 201I, 206S, 232D, 234I,	3I, 13V , 15V , 17E , 20I , 36I , 37D , 37N, 41K, 62V , 64M , 69K , 89M	Echec analyse – 49R	112V, 119P, 123S, 125A, 127K, 134N, 135V, 136T, 195C, 201I, 206S, 232D, 234I, 275V , 283G	3I, 12I, 13A 16E , 19P, 20P, 22V, 36I , 37N, 41K, 63Q, 68E, 69R , 89M	35T, 36D, 39N, 60I, 89G , 103N , 162A, 173T, 174K, 177E, 178M, 195L, 200A, 207E, 214F, 215S 238T, 245Q, 248D, 250E, 251D, 272A, 277K, 281N , 292I/D, 293V, 294T, 303F , 310Q , 317L , 324N , 327E	112V, 119P, 123S, 125A, 127K, 134X , 135L , 136H/Q , 195C, 201I, 206S, 232D, 234I, 275V , 283G
203071366 (+)	3I, 10V , 12S, 14R, 16E , 35D, 37N, 63H, 72V, 77I, 89M	6D, 35I, 48E, 106I , 118I, 122K, 135T, 138A , 172K, 174K, 179I , 200I, 211K, 277K, 283I, 311R	72I, 101I, 112A, 123S, 124N, 125A, 127K, 200L, 201I, 230N, 232D, 240R	3I, 10V , 12P , 12S, 14R, 16E , 35D, 37N, 63H, 72V, 77I, 89M	6D, 35I, 48E, 106I , 118I, 122K, 135T, 138A , 172K, 174K, 179I , 200I, 211K, 277K, 283I, 311R	72I, 101I, 112A, 123S, 124N, 125A, 127K, 200L, 201I, 230N, 232D, 240R	3I, 10V , 12P , 12S, 14R, 16E , 35D, 37N, 63H, 72V, 77I, 89M	6D, 35I, 48E, 96Q , 106I , 118I, 122K, 135T, 138A , 172K, 174K, 179I , 200I, 211K, 277K, 283I, 321S	72I, 101I, 112A, 123S, 124N, 125A, 127K, 200L, 201I, 230N, 232D, 240R
203151494 (+)	3I, 12S, 15V, 16E , 19T, 20R , 36I , 37D, 41K,	35T, 36A, 39E, 48T, 121H, 173T, 174K, 177E, 178L, 196M, 196V, 200A, 202V,	72I, 101I, 112I, 112V, 113V, 123S, 127K, 136Q,	3I, 12S, 15V, 16E , 19T, 20R , 36I , 37D, 41K,	35T, 36A, 39E, 48T, 121H, 173T, 174K, 177E, 178L, 196M, 196V, 200A, 202V,	72I, 101I, 112I, 112V, 113V, 123S, 127K, 136Q,	3I, 12S, 15V, 16E , 19T, 20R , 36I , 37D,	35T, 36A, 39E, 40K , 41R , 48T, 49N , 50W , 77L , 78W/K/X , 121H,	72I, 74I/M , 95R , 96G , 97I , 101I, 112I, 123S,

	60E, 62V, 63C, 69K, 93L	207E, 211K, 214F, 243T, 245Q, 250E, 275R, 286A, 291D, 293V	201I, 218I, 232D, 234I	60E, 62V, 63C, 69K, 93L	207E, 211K, 214F, 243T, 245Q, 250E, 275R, 286A, 291D, 293V	201I, 218I, 232D, 234I, 265V, 269K, 278A	41K, 60E, 62V, 63C, 69K, 93L	168V/I, 173T, 174K, 177E, 178L, 196M, 196V, 196L, 200A, 202V, 207E, 211K, 214F, 243T, 245Q, 250E, 275R, 286A, 291D, 293V	127K, 136Q, 167E, 173L/X/M, 201I, 218I, 228K, 229Y, 230R, 232D, 233S, 234I, 242F, 250L, 251V, 255R, 257V, 258T, 261S, 264R, 265V, 269K, 278A
203370413 (+)	3I, 13V, 14R, 16E, 20I, 35D, 36I, 37N, 41K, 64L, 69K, 89M	35T, 122K, 135V, 142V, 162A, 169Q, 173T, 174N, 177E, 200A, 207E, 211K, 245Q, 272A, 281R, 293V	59E, 106A, 112V, 119P, 123S, 125A, 127K, 134N, 135V, 136T, 201I, 206S, 232D, 234I, 283G	3I, 13V, 14R, 16E, 20I, 35D, 36I, 37N, 41K, 64L, 69K, 89M	35T, 122K, 135V, 142V, 162A, 169Q, 173T, 174N, 177E, 200A, 207E, 211K, 245Q, 272A, 281R, 293V	59E, 106A, 112V, 119P, 123S, 125A, 127K, 134N, 135V, 136T, 201I, 206S, 232D, 234I, 283G	3I, 13V, 14R, 16E, 20I, 35D, 36I, 37N, 41K, 64L, 69K, 89M	35T, 122K, 135V, 142V, 162A, 169Q, 173T, 174N, 177E, 200A, 207E, 211K, 245Q, 272A, 281R, 293V, 306K, 307T, 308V, 312G, 315D	59E, 106A, 112V, 119P, 123S, 125A, 127K, 134N, 135V, 136T, 201I, 206S, 232D, 234I, 283G
203461390 (+)	3I, 13V, 17E, 20I, 36I, 37N, 41K, 64M, 69K, 70R, 89M	20R, 35T, 39L, 43S, 49R, 68G, 122P, 135V, 138A, 162A, 169A, 173T, 174K, 175H, 177E, 196E, 200A, 207E, 211K, 214F, 245K, 245Q, 245R, 272A, 277K, 281R, 286A, 291D,	50T, 72I, 101I, 112I, 123S, 125A, 126A, 127K, 134N, 136T, 163A, 201I, 206S, 229E, 232D, 234I,	3I, 13V, 17E, 20I, 36I, 37N, 41K, 64M, 69K, 70R, 89M	20R, 35T, 39L, 43S, 49R, 68G, 122P, 135V, 138A, 162A, 169A, 173T, 174K, 175H, 177E, 196E, 200A, 207E, 211K, 214F, 245K, 272A, 277K, 281R, 286A,	50T, 72I, 101I, 112I, 123S, 125A, 126A, 127K, 134N, 136T, 163A, 201I, 206S, 229E, 232D, 234I,	3I, 13V, 17E, 20I, 36I, 37N, 41K, 64M, 69K, 70R, 89M	Echec analyse	50T, 53E, 72I, 101I, 112I, 123S, 125A, 126A, 127K, 134N, 136T, 163A, 201I, 206S, 229E, 232D, 234I, 256E,

		292I, 293V, 294T, 314E	256E, 269K, 281M, 283G		291D, 292I, 293V, 294T, 314E	256E, 269K, 281M, 283G			269K, 281M, 283G
203511396 (+)	3I, 10I , 13V, 15V, 19P, 20I, 36I , 37N, 41K, 62V, 63P , 89M	49R, 121H, 162N, 177G, 211K, 214F, 272S, 283I, 292I, 293V, 311R	72I, 97A , 101I, 112V, 123S, 124T, 125A, 127K, 134N, 135V, 136T, 206S, 211Q, 232D, 234I	3I, 10I , 13V, 15V, 19P, 20I, 36I , 37N, 41K, 62V, 63P , 89M	49R, 121H, 162N, 177G, 211K, 214F, 272S, 283I, 292I, 293V, 311R	72I, 97A , 101I, 112V, 123S, 124T, 125A, 127K, 134N, 135V, 136T, 206S, 211Q, 232D, 234I	3I, 10I , 13V, 15V, 19P, 20I, 36I , 37N, 41K, 62V, 63P , 89M	49R, 121H, 162N, 177G, 211K, 214F, 244insG , 245M , 254A , 272S, 283I, 292I, 293V, 296K , 299G/D/V , 300R , 311R, 317L/F	72I, 97A , 101I, 112V, 123S, 124T, 125A, 127K, 134K/X/E/Q , 136T, 206S, 211Q, 232D, 234I
203530875 (++)	3I, 13V, 14R, 17E, 20I, 36I , 37N, 43R, 63S, 69K , 89M	35T, 39D, 39N, 40D, 49R, 60I, 123S, 135R, 162A, 169D, 172K, 174K, 177E, 178M, 181C , 184V , 196E, 200A, 203G, 207D, 214F, 221Y , 245Q, 248D, 272A, 277K, 286A, 291D, 292I, 293V	63I, 74I , 84M, 91E, 95K, 101I, 112I, 119P, 122I, 123S, 125A, 127K, 134N, 136T, 201I, 206S, 215N, 218I, 232D, 234I, 265V, 278A, 280S	3I, 13V, 14R, 17E, 20I, 36I , 37N, 43R, 63S, 69K , 89M	35T, 39D, 39N, 40D, 49R, 60I, 123S, 135R, 162A, 169D, 172K, 174K, 177E, 178M, 181C , 184V , 196E, 200A, 203G, 207D, 214F, 221Y , 245Q, 248D, 272A, 277K, 286A, 291D, 292I, 293V	63I, 74I , 84M, 91E, 95K, 101I, 112I, 119P, 122I, 123S, 125A, 127K, 134N, 136T, 201I, 206S, 215N, 218I, 232D, 234I, 265V , 269K, 278A, 280S	3I, 13V, 14R, 17E, 20I, 36I , 37N, 43R, 63S, 69K , 89M	35T, 39D, 39N, 40D, 49R, 60I, 123S, 135R, 162A, 169D, 172K, 174K, 177E, 178M, 181C , 184V , 196E, 200A, 203G, 207D, 214F, 221Y , 245Q, 248D, 272A, 277K, 286A, 291D, 292I, 293V, 313A , 315D	63I, 74I , 84M, 91E, 95K, 101I, 112I, 119P, 122I, 123S, 125A, 127K, 134N, 136T, 201I, 206S, 215N, 218I, 232D, 234I, 265V , 269K, 278A, 280S
210110842 (++) (- ABL)	3I, 15V, 77A	101E , 108I, 122K, 123E, 135T, 142V, 162A, 177E, 178L,	101I, 119P, 123S, 124T, 127K, 173R,	3I, 15V	101E , 108I, 122K, 123E, 135T, 142V, 162A, 177E, 178L,	101I, 119P, 123S, 124T, 127K, 173R,	3I, 15V	101E , 106L , 108I, 118L , 119S , 121Y , 122K, 123E,	101I, 119P, 123S, 124T, 127K, 173R,

		179I , 196E, 214F, 277K, 286P, 297K	216H, 222T, 232D		179I , 196E, 214F, 277K, 286P, 297K	216H, 222T, 232D		135T, 138D , 139I , 142V, 162A, 163R , 170S , 172K , 179I , 183N , 190X , 193X , 194K , 196E/ D/H/Q , 197H , 200R/K , 201Q , 203K , 206S , 208L/P , 214F, 216P , 217L , 218N , 224X , 240K , 242P , 243S , 277K, 286P, 297K	216H, 222T, 232D
210140539 (+)	3I, 13V, 14R, 35D, 36I , 37N, 41K, 57K, 60E , 61E, 63S, 69K , 89M	11T, 35T, 39N, 40D, 121Y, 123G, 173S, 177E, 207A, 211S, 214F, 245Q, 251N, 272A, 277K, 286A, 291D, 293V, 312N	50I, 50T , 101I, 112I, 123S, 125A, 127K, 134N, 136Q, 167E, 201I, 222T, 232D, 234I, 254K, 256E, 283G	3I, 13V, 14R, 35D, 36I , 37N, 41K, 57K, 60E , 61E, 63S, 69K , 89M	11T, 35T, 39N, 40D, 121Y, 123G, 173S, 177E, 207A, 211S, 214F, 245Q, 251N, 272A, 277K, 286A, 291D, 293V, 312N	50I, 101I, 112I, 112V , 123S, 125A, 127K, 134N, 134S , 136Q, 160T , 167E, 201I, 222T, 232D, 234I, 234V , 254K, 256E, 283G	3I, 13V, 14R, 35D, 36I , 37N, 41K, 57K, 60E , 61E, 63S, 69K , 89M	Echec	Echec
210150931 (+)	3I, 12S, 15V, 16E , 19T, 20R , 35K, 36I , 37D, 37H , 41K,	35T, 36A, 39E, 48T, 90I , 121H, 173T, 174K, 177E, 196V, 202V, 207E, 211K, 214F, 243T, 245Q,	72I, 101I, 112V, 113V, 119R, 123S, 127K, 136Q, 167E, 189R , 201I, 218I,	3I, 12S, 15V, 16E , 19T, 20R , 35K, 36I , 37D, 37N, 41K, 63C, 69K , 93L	35T, 36A, 39E, 48T, 90I , 121H, 173T, 174K, 177E, 196V, 202V, 207E, 211K, 214F, 226L , 243T,	72I, 101I, 112V, 113V, 119R, 123S, 127K, 136Q, 167E, 201I, 218I, 232D,	3I, 12S, 15V, 16E , 19T, 20R , 35K, 36I , 37D, 37N, 41K, 63C,	35T, 36A, 39E, 48T, 90I , 121H, 173T, 174K, 177E, 196V, 202V, 207E, 211K, 214F,	72I, 101I, 112V, 113V, 123S, 127K, 136Q, 167E, 201I, 218I, 232D, 234I,

	63C, 69K , 72V , 93L	250E, 275R, 286A, 286V , 291D, 293V	232D, 234I, 265V, 269K, 278A, 281M		245Q, 250E, 275R, 286A, 291D, 293V	234I, 265V, 269K, 278A, 281M	69K, 89M , 93L	243T, 245Q, 250E, 275R, 286A, 291D, 293V, 307S , 309N, 310S , 315D	265V, 269K, 278A, 281M
210261876 (-)	3I, 10I, 13V , 16E , 20V, 22V, 36I , 37N, 41K, 63T, 65D , 89M	20R, 35T, 39M, 43N , 122K, 123E, 135V , 162C, 173I , 174K 178L, 184I , 200A, 207E, 214F, 245Q , 248D, 278H, 286A , 291N, 291D, 292I , 293V, 297A, 310I	72I, 101I, 112I, 123S, 125A, 127K, 136Q, 157K , 181L, 201I, 218I, 232D, 234V, 247E , 256E, 259I	3I, 12I, 13A , 16E, 19P, 20V , 22V, 36I , 37N, 41K, 63Q, 68E , 69R, 89M	162C, 173T, 177E , 196E , 200A, 207E, 211K , 214F, 272A	72I, 101I, 112I, 123S, 125A, 127K, 136Q, 181L, 201I, 218I, 232D, 234V, 256E			
210321716 (+)	3I, 12K, 13V, 14R, 17E, 20I, 36I , 37D, 64M, 69K , 70R, 89M , 92H	35T, 39A, 39S, 135V, 162A, 173T, 174K, 177E, 200A, 207E, 214F, 243S, 245E, 272A, 277K, 281R, 286A, 293V, 294T	72I, 101I, 123S, 125A, 127K, 134N, 135V, 136T, 163E, 201I, 206S, 232D, 234I, 254K, 269K, 283G	3I, 12K, 13V, 14R, 17E, 20I, 36I , 37D, 64M, 69K, 70R, 89M	35T, 39A, 39S, 90I , 121H , 135V, 162A, 173T, 174K, 177E, 200A, 207E, 214F, 243S, 245E, 272A, 277K, 281R, 286A, 293V, 294T	72I, 101I, 123S, 125A, 127K, 134N, 135V, 136T, 163E, 201I, 206S, 232D, 234I, 254K, 269K, 283G	3I, 12K, 13V, 14R, 17E, 20I, 36I , 37D, 64M, 69K , 70R, 87S , 89M	35T, 39A, 39S, 121H , 135V, 162A, 173T, 174K, 177E, 200A, 207E, 214F, 243S, 245E, 272A, 277K, 281R, 286A, 293V, 294T, 312G	72I, 101I, 123S, 125A, 127K, 134X/E/Q/K , 136T, 163E, 201I, 206S, 232D, 235C, 254K, 269K, 283G
210481981-1 (+)	3I, 14R, 35D, 37N, 41K, 63P, 71V , 77I, 77M, 82I, 93L	60I, 122K, 123E, 135T, 177G , 200A, 211N, 214F, 245E, 272A, 277K, 297K, 311R, 312D, 317A	45V, 50I, 112A, 123S, 124T, 125A, 127K, 172I, 218S, 232D	3I, 14R, 35D, 37N, 41K, 63P , 71V , 77I, 77M, 82I, 93L	60I, 122K, 123E, 135T, 200A, 211N, 214F, 245E, 272A, 277K, 297K, 311R, 312D, 313Q , 317A	45V, 50I, 112A, 123S, 124T, 125A, 127K, 172I, 218S, 232D	3I, 14R, 35D, 37N, 41K, 63P , 71V , 77I, 77M, 82I, 93L	4S , 60I, 122K, 123E, 135T, 200A, 211N, 214F, 245E, 272A, 277K, 297K, 312D, 317A, 321S	50I, 53L, 58S , 61R, 64E , 67L, 72G , 112A, 123S, 124T, 125A, 127K, 172I, 218S, 232D

210481981-2 (2ème passage NGS)				3I, 14R, 35D, 37N, 41K, 63P , 71V , 77I, 77M, 82I, 93L	60I, 122K, 123E, 135T, 200A, 211N, 214F, 245E, 272A, 277K, 297K, 311R, 312D, 313Q , 317A	45V, 50I, 112A, 123S, 124T, 125A, 127K, 172I, 218S, 232D			
210490293 (+)	3I, 12R, 19Q, 35D, 37E, 37N, 37D , 37G , 37S , 37R , 37K , 41K, 63P , 77I, 93L	48T, 121H, 135T, 194G, 196E, 200E, 200V, 202V, 204D, 211K, 214F, 272A, 277K, 286A, 288S, 293V, 294T	50I, 72I, 101I, 123S, 125A, 127K, 135V, 156N, 167E, 181L, 201I, 232D, 254S, 255G	3I, 12R, 19Q, 35D, 37E, 37N, 41K, 63P , 77I, 93L	48T, 121H, 135T, 194G, 196E, 200E, 200V, 202V, 204D, 211K, 214F, 272A, 277K, 286A, 288S, 293V, 294T	50I, 72I, 101I, 123S, 125A, 127K, 135V, 156N, 167E, 181L, 201I, 232D, 254S, 255G	3I, 12R, 19Q, 35D, 37E/N/D/R/K, 41K, 63P , 77I, 93L	48T, 121H, 135T, 194G, 196E, 200E, 200V, 202V, 204D, 211K, 214F, 272A, 277K, 286A, 288S, 293V, 294T, 312G , 315D	50I, 72I, 101I, 123S, 125A, 127K, 135V, 156N, 167E, 181L, 201I, 232D, 254S, 255G
210490891 (+)	3I, 13V, 35D, 37H, 41K, 63P	122K, 162C, 177E, 211K, 214F, 245I, 286A, 288S, 293V, 297T, 304E	50I, 71Q, 72I, 119R, 123S, 124T, 125A, 127K, 193E, 201I, 206S, 232D, 234V, 255D	3I, 13V, 35D, 37H, 41K, 63P	122K, 162C, 177E, 211K, 214F, 245I, 286A, 288S, 293V, 297T, 304E	50I, 71Q, 72I, 119R, 123S, 124T, 125A, 127K, 193E, 201I, 206S, 232D, 234V, 255D, 279G	3I, 13V, 35D, 37H, 41K, 63P	122K, 162C, 177E, 211K, 214F, 245I, 286A, 288S, 293V, 297T, 305R , 315D	50I, 71Q, 72I, 119R, 123S, 124T, 125A, 127K, 193E, 201I, 206S, 232D, 234V, 255G
210551227 (+)	3I, 12P, 19I, 45R, 60E , 62V, 63S, 64V, 77I, 93L	69S , 83K, 135T, 162C, 178M , 196E, 207N, 211K, 214F, 224D, 245M, 277K , 293V	57N, 72I, 101I, 123S, 124T, 127K, 201I, 205S, 206S, 232D, 256E, 266E , 284I	3I, 12P, 19I, 45R, 60E , 62V, 63S, 64V, 77I, 93L	69S , 83K, 135T, 162C, 196E, 207N, 211K, 214F, 224D, 245M, 293V, 317A	57N, 72I, 101I, 123S, 124T, 127K, 201I, 205S, 206S, 232D, 256E	3I, 12P, 19I, 25Y , 34X , 45R, 60E , 62V, 63S, 64V, 77I, 87I , 93L	1S , 4S , 69S , 83K, 135T, 162C, 196E, 207N, 211K, 214F, 224D, 245M, 276S/I/G/R , 293V, 313A , 315D , 319N	57N, 72I, 101I, 123S, 124T, 127K, 201I, 205S, 206S, 232D, 256E

210671034-1 (+)	3I, 10V , 12A, 12E, 12P , 12Q , 12K , 13V, 15V, 16E , 20I, 35D, 36I , 37N, 41K, 63P , 69K , 72T, 89M	35T, 36D, 39K, 43Q, 49R, 60I, 82N , 106I , 123E, 135T, 142V, 162A, 169L, 173T, 174E, 177E, 200A, 207E, 211K, 245Q, 272A, 286A, 291D, 292I, 293V, 294T	72I, 101I, 103R, 112V, 123S, 125A, 127K, 134N, 135V, 136T, 163A, 201I, 208L, 218I, 232D, 234I, 279G	3I, 10V , 12A, 12E, 13V, 15V, 16E , 20I, 35D, 36I , 37N, 41K, 63P , 69K , 72T, 89M	35T, 36D, 39K, 43Q, 49R, 60I, 106I , 123E, 135T, 142V, 162A, 169L, 173T, 174E, 177E, 200A, 207E, 211K, 245Q, 272A, 286A, 291D, 292I, 293V, 294T	72I, 101I, 103R, 112V, 123S, 125A, 127K, 134N, 135V, 136T, 163A, 201I, 208L, 218I, 232D, 234I, 279G	3I, 10V , 12A/E/P/K /A/ S, 13V, 15V, 16E , 20I, 35D, 36I , 37N, 41K , 63P , 69K , 72T, 89M	35T, 36D, 39K, 43Q, 49R, 60I, 86E , 106I , 123E, 135T, 142V, 162A, 169L, 173T, 174E, 177E, 200A, 207E, 211K, 245Q, 272A, 286A, 291D, 292I, 293V, 294T, 313L , 315D , 319D	63Q , 64V , 72I, 95R , 101I, 103R, 112V, 123S, 125A, 127K, 134N, 135V, 136T, 163A, 201I, 208L, 218I, 232D, 234I, 279G
210671034-2 (2ème passage NGS)				3I, 10V , 12E, 13V, 15V, 16E , 20I, 35D, 36I , 37N, 41K, 63P , 69K , 72T, 89M	35T, 36D, 39K, 43Q, 49R, 60I, 106I , 122K , 123E, 135T, 142V, 162A, 169L, 173T, 174E, 177E, 200A, 207E, 211K, 245Q, 272A, 286A, 291D, 292I, 293V, 294T	72I, 101I, 103R, 112V, 123S, 125A, 127K, 134N, 135V, 136T, 163A, 201I, 208L, 218I, 232D, 234I, 279G			
210720484 (+)	3I, 8P , 13V, 16E , 20I, 21Q , 36I , 37N, 67G, 69K , 89M , 92E	34F , 35T, 36D, 123E, 135T, 162A, 165I, 173T, 177E, 196E, 200A, 207E, 211K, 214F, 245Q, 248D, 272A, 275Q, 286A, 291D, 292I, 293V, 294T, 297R, 312D	72I, 74I , 96D, 101I, 106A, 112V, 123S, 125A, 127K, 134N, 136T, 163Q, 201I, 206S, 211R, 232D, 234I, 269K	3I, 13V, 16E , 20I, 36I , 37N, 69K , 89M	35T, 36D, 123E, 135T, 162A, 165I, 173T, 177E, 196E, 200A, 207E, 211K, 214F, 245Q, 248D, 272A, 275Q, 286A, 291D, 292I, 293V, 294T, 297R, 312D	72I, 74I , 96D, 101I, 106A, 112V, 123S, 125A, 127K, 134N, 136T, 163Q, 201I, 206S, 211R, 232D, 234I, 269K	3I, 13V, 16E , 20I, 36I , 37N, 69K , 89M	35T, 36D, 123E, 135T, 162A, 165I, 173T, 177E, 196E, 200A, 207E, 211K, 214F, 245Q, 248D, 272A, 275Q, 280X , 286A, 291D, 292I,	55A , 56G , 72I, 74I , 96D, 101I, 106A, 112V, 123S, 125A, 127K, 134N, 136T, 163Q, 201I, 206S, 211R, 232D, 234I, 269K

								293V, 294T, 297R, 306S , 307G, 312D/H	
Contrôle positif (+)	3I, 10M, 13V, 15V, 35D, 36I , 37N, 41K, 45R, 57K, 63I, 69K , 89M	20R, 35T, 39M, 60I, 122K, 123S, 135R, 162Y, 173S, 174K, 177E, 179I , 207A, 211S, 245Q, 250E, 251N, 272A, 277K, 286A, 293V, 294T, 317A	60V, 84M, 90S, 112I, 113V, 123S, 124S, 125A, 127K, 136Q, 167E, 201I, 211R, 232D, 234I	3I, 10M, 13V, 15V, 35D, 36I , 37N, 41K, 45R, 57K, 63I, 69K , 89M	20R, 35T, 39M, 60I, 122K, 123S, 135R, 162Y, 173S, 174K, 177E, 179I , 207A, 211S, 245Q, 250E, 251N, 272A, 277K, 286A, 293V, 294T, 317A	60V, 84M, 90S, 112I, 113V, 123S, 124S, 125A, 127K, 136Q, 167E, 201I, 211R, 232D, 234I, 275V , 283G, 284G	3I, 10M, 13V, 15V, 35D, 36I , 37N, 41K, 45R, 57K, 63I, 69K , 89M	20R, 35P , 39M, 60I, 122K, 123S, 135R, 162Y, 173S, 174K, 177E, 179I 207A, 211S, 228F , 245Q, 250E, 251N, 272A, 277K, 286A, 293V, 294T, 302insP , 303W/G, 306K , 307Q, 308G , 309R	60V, 84M, 90S, 112I, 113V, 123S, 124S, 125A, 127K, 136Q, 167E, 201I, 211R, 232D, 234I, 275V, 283G, 284G

Rouge: mutation d'intérêt décrite par l'ANRS 31v ; **Bleu** : Mutations retrouvées seulement en Sanger Viroseq ; **Violet** : Mutations retrouvées en NGS et +/- en Sanger ABL par rapport au Sanger Viroseq ; **Vert** : mutations retrouvées en Sanger et NGS à <20% ; **Orange** : mutations retrouvées seulement en Sanger ABL Les Symboles « + » et « - » indiquent une estimation personnelle à l'aide de graphique concernant la qualité de la couverture des différentes régions d'ADN.

Echantillons	Sanger ViroSeq			NGS ABL (% de similitude)			Sanger ABL (% de similitude)		
	PR	RT	INT	PR	RT	IN	PR	RT	IN
202761161	G (94)	43_02G (93) et G (93)	G(95)	43_02G (91.92)	43_02G (92.28)	09_CPX (94.88)	G (96)	G (53)	G (81)
202801387	CRF36_cpx (96)	02_AG (95)	CRF02_AG (96)	36_CPX (97.31)	02_AG (94.91)	02_AG (95.98)			
202821685	CRF02_AG (97)	02_AG (92) et CRF95_02 B (92)	CRF02_AG (98)	02_AG (95.96)	B (87.48)	30_0206 (96.54)	02_AG (99)	02_AG (89)	02_AG (100)
202942236	CRF02_AG (93)	CRF02_AG (94)	CRF02_AG (97)	06_CPX (85.53)	Echec	02_AG (96.51)	G (100)	B (94)	G (100)
203071366	B (95)	B (94)	B (89)	B (93.94)	B (93.69)	B (94.74)	B (100)	B (72)	B (100)
203151494	C (94)	CRF57_BC et C (93)	A2C, C ET BC (94)	C (91.58)	C (91.04)	31_BC (94.6)	C (100)	C (98)	C (100)
203370413	02_AG (99)	02_AG (94)	02_AG (97)	02_AG (96.97)	02_AG (93.42)	02_AG (95.84)	02_AG (100)	02_AG (100)	02_AG (100)
203461390	02_AG (98)	02_AG (95)	02_AG (97)	02_AG (97.64)	02_AG (94.03)	02_AG (96.39)	A1 (72)	G (98)	G (100)
203511396	02_AG (95)	CRF94_cpx ET B (97)	CRF94_cpx ET 02_AG (97)	02_AG (94.95)	28_BF (95.18)	02_AG (95.84)	B (100)	D (98)	G(99)
203530875	02_AG (96)	02_AG ET C (92)	02_AG (95)	02_AG (95.62)	31_BC (91.99)	02_AG (93.91)	02_AG (100)	C (90)	02_AG (98)
210110842	B (98)	B ET CRF28_BF (96)	B (96)	B (97.98)	B (95.03)	39_BF (96.26)	B (76)	B (100)	B (100)
210140539	A (96)	A1 ET 22_01A1 (94)	A1 ET 22_01A1 (96)	A1 (94.59)	22_01A1 (94.19)	A1 (94.3)	D (99)	Echec	Echec

210150931	C (93)	C ET CRF57_BC (93)	A1CD ET C (91)	08_BC (90.2)	C (92.29)	08_BC (95.15)	C (100)	C (100)	C (100)
210261876	F1 (98)	F1 (95)	F1 (96)	02_AG (92.93)	K (98.84)	F1 (95.71)			
210321716	02_AG (96)	02_AG (94)	02_AG (97)	02_AG (95.29)	02_AG (93.53)	02_AG (95.43)	02_AG (99)	02_AG (100)	02_AG (100)
210481981-1	B (96)	B (96)	B (96)	07_BC (94.61)	B (95.57)	B (96.26)	B (66)	B (100)	B (100)
210481981-2				07_BC (94.61) (=)	B (95.57) (=)	B (96.54)			
210490293	B (95)	B (94)	B (95)	B (93.94)	B (93.03)	B (95.01)	B (73)	B (100)	Echec
210490891	B (97)	B (95)	B (95)	B (94.61)	B (94.29)	28_BF (94.18)	B (100)	B (100)	B (100)
210551227	B (95)	B (95)	B (94)	B (93.6)	B (95.28)	B (96.12)	B (100)	B (98)	B (100)
210671034-1	CRF02_AG (94)	CRF02_A G (93)	CRF02_ AG (96)	02_AG (92.93)	0206 (92.98)	02_AG (96.12)	02_AG (77)	02_AG (85)	02_AG (92)
210671034-2				02_AG (92.93)	30_0206 (92.98)	02_AG (96.26)			
210720484	36_CPX ET CRF02_AG (95)	CRF02_A G (94)	CRF02_ AG (96)	36_CPX (96.97)	02_AG (93.54)	02_AG (95.71)	02_AG (100)	02_AG (100)	02_AG (100)
Contrôle (+)	A1	A1	A1	A1 (95.96)	A1 (93.03)	A1 (95.97)	A1 (95.96)	A1 (93.14)	35_AD (95.67)

Annexe III : table des différents sous types avec le pourcentage de similitude des séquences.

Echantillons		NGS ABL : Mutations <20% (Prévalence en %)	Nombre de Mutations <3%
202761161	PR	Pas de mutations	584
	RT	135V (8.86) ; 178V (8.78) ; 200E (15.43) ; 250K (3.05)	1184
	IN	124N (7.83) ; 208M (10.42) ; 254S (9.98)	1277
202801387	PR	17E (15.57)	418
	RT	275R (7.23)	1258
	IN	191L (3.91) ; 278N (4.16)	1110
202821685	PR	16E (3.1)	562
	RT	36D (18.52) ; 52T (14.71) ; 83K (15.63) ; 135R (16.67) ; 135V (13.89) ; 142V (13.16) ; 169D (14.63) ; 169Q (12.2)	312
	IN	68V (12.24)	636
202942236	PR	60E (19.23) ; 63P (19.24)	Pas de détection
	RT	Echec analyse	
	IN	51Y (4.4)	1688
203071366	PR	12A (5.21) ; 12N (5.45) ; 19V (15.53)	450
	RT	5V (3.05) ; 22R (6.74) ; 39A (11.07) ; 48K (5.19) ; 200A (14.42) ; 276I (16.8) ; 292I (11.16)	1187
	IN	Pas de mutations	1023
203151494	PR	63A (5.79) ; 72T (3.2)	565
	RT	13R (11.8) ; 121Y (5.61) ; 142T (9.57) ; 196E (9.08) ; 292I (7.88)	1291
	IN	84M (4.18) ; 283G (5.48)	1087
203370413	PR	17E (12.02)	331
	RT	24* (3.77) ; 27A (4.74) ; 122R (8.43) ; 223R (7.57) ; 262E (3.1)	1133
	IN	81T (8.13) ; 82S (8.5) ; 219R (3.74) ; 221H (6.83)	1033
203461390	PR	19I (10.02)	526
	RT	39S (5.11) ; 175Y (12.96) ; 245Q (19.26) ; 245R (12.41) ; 250E (5.65)	795
	IN	Pas de mutations	1085
203511396	PR	Pas de mutations	538
	RT	244V (9.27) ;	1295
	IN	127T (3.13) ; 255R (7.94) ; 259I (19.88)	1107
203530875	PR	Pas de mutations	479
	RT	20R (8.13) ; 162T (15.01) ; 193F (3.2) ;	1326
	IN	79I (6.09) ; 81T (6.41)	586
210110842	PR	36I (12.38) ; 37N (18.48) ; 63S (6.24) ; 63T (11.41)	593
	RT	20R (16.84) ; 82R (4.31) ; 98G (3.89) ; 101Q (5.73) ; 189I (11.05) ; 195L (6.29) ; 211A (5.7) ; 211K (6.98) ; 245E (4.56)	831
	IN	72I (3.23) ; 181L (6.42) ; 204V (3.29) ; 222H (13.5) ; 253E (3.6)	1400
210140539	PR	Pas de mutations	482
	RT	36D (3.09) ; 60I (3.65)	429
	IN	Pas de mutations	372
210150931	PR	10I (10.28) ; 12I (9.72) ; 13A (9.38) ; 13V (11.53) ; 19P (14.88) ; 20I (10.93) ; 20M (10.38) ; 20V (6.83) ; 22V (6.53) ; 34G (7.01) ; 37A (8.45) ; 37H (5.31) ; 39S (4.74) ; 55E (3.45) ; 62V (9.14) ;	194

		63P (10.3) ; 63Q (8.13) ; 68E (6.88) ; 69R (7.42) ; 70R (6.53) ; 72V (15.06) ; 89M (11.42)	
	RT	226L (3.17) ; 257V (9.87) ; 281R (3.92) ; 286V (15.97)	547
	IN	81T (18.91) ; 84M (6.71)	793
210261876	PR	8* (3.65)	563
	RT	17N (13.39)	61
	IN	60V (4.22) ; 275V (9.35)	570
210321716	PR	37E (13.84)	338
	RT	36D (3.76) ; 60I (10.36) ; 123E (8.97) ; 123N (17.94) ; 178M (6.76) ; 202V (5.5) ; 211K (10.72) ; 291D (16.82) ; 292I (15.28)	801
	IN	165I (4.16) ; 181L (13.29) ; 182V (3.6) ; 256E (3.93)	872
210481981-1	PR	Pas de mutations	453
	RT	177G (14.12) ; 207R (12.21) ; 211H (7.51) ; 211K (13.93) ; 211S (3.29) ; 211T (11.11) ; 313Q (11.69)	775
	IN	201I (7.04)	321
210481981-2	PR	Pas de mutations	608
	RT	177G (12.28) ; 207R (10.31) ; 211H (6.72) ; 211K (11.47) ; 211T (11.17) ; 313Q (8.49)	1113
	IN	201I (3.2)	903
210490293	PR	37D (11.15) ; 60E (5.89) ; 62V (12.7) ; 63H (3.78) ; 71T (5.37)	464
	RT	39A (3.38) ; 90I (10.57) ; 200A (4.94)	590
	IN	Pas de mutations	1093
210490891	PR	36I (3.14) ; 37N (5.76) ; 62V (4.35) ; 64V (10.22) ; 89M (3.14)	402
	RT	35I (3.13) ; 121Y (3.38) ; 204K (8.55) ;	947
	IN	101I (3.79) ; 113V (3.02) ; 218I (4.14) ; 234I (4.66) ; 254K (4.15)	446
210551227	PR	Pas de mutations	356
	RT	35M (5.58) ; 162Y (3.24) ; 178M (16.82) ; 207T (10.34) ; 277K (15.09)	1028
	IN	62L (4.38) ; 112A (9.25)	548
210671034-1	PR	14R (5.07) ; 57K (4.92) ; 70R (4.4)	407
	RT	69N (6.15) ; 102R (9.28) ; 207K (15.12)	900
	IN	173R (4.59) ; 203M (10.63)	512
210671034-2	PR	12A (19.8) ; 14R (4.76) ; 57K (5.42) ; 70R (3.59)	405
	RT	69N (6.6) ; 102R (10.47) ; 135V (5.75) ; 122K (8.64) ; 207K (19.23)	215
	IN	203M (10.89) ; 259I (3.75)	1017
210720484	PR	70R (13.89)	595
	RT	202V (6.72)	1401
	IN	135V (3.08)	643

Annexe IV: Table brute de l'ensemble des mutations trouvées en NGS <20%.

Annexe V : Table brute des résultats de séquençage avec la technique ANRS

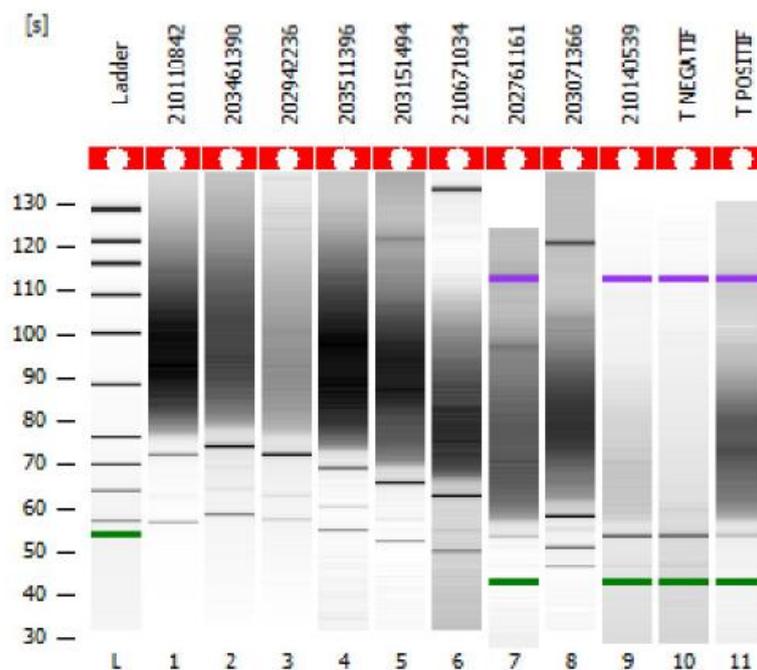
Echantillons	Sanger ANRS			Sous – type
	PR	RT	IN	
202821685(-)	3I, 13V, 14R, 17E, 20I, 35D, 36I, 37N, 41K, 64M, 69K, 89M	Del79_99, 5V, 35T, 36A, 60I, 76E, 77G, 104Q, 122K, 174H, 196E, 200A, 207G, 211K, 214F, 239R, 245K, 260F/L	Del 13_48 ,57G, 101I, 112V, 123S, 124G, 125A, 127K, 134N, 135V, 136T, 201I, 206S, 211Q, 218I, 232D, 234I, 256E, 278A	02_AG
203370413 (+)	3I, 13V, 14R, 16E , 20I, 35D, 36I , 37N, 41K, 64L, 69K , 89M	20R, 35T, 122K, 135V, 142V, 162A, 169Q, 173T, 174N, 177E, 200A, 207E, 211K, 245Q, 258R, 259V, 261G/V, 262R, 264F	Del 8_42, 44Q/R, 52A, 53Q/R, 59E, 96E/G, 106A, 112V, 119P, 123S, 125A, 127K, 134N, 135V, 136T, 187K/R, 201I, 206S, 232D, 234I, 287E/G, 288D/E	02_AG
203530875 (++)	3I, 13V, 14R, 17E, 20I, 36I , 37N, 43R, 63S, 69K , 89M	31P, 35T, 39D, 39N, 40D, 49R, 60I, 123S, 135R, 162A, 169D, 172K, 174K, 177E, 178M, 181C , 184V , 196E, 200A, 203G, 207D, 214F, 221Y , 245Q, 248D	37E, 38P, 43G, 63I, 74I , 84M, 91E, 95K, 101I, 112I, 119P, 122I, 123S, 125A, 127K, 134N, 136T, 201I, 206S, 215N, 218I, 232D, 234I, 265V , 269K, 272E, 273N, 274R, 275W, 276Q, 277R, 278*, 279*, 280S, 281G, 282K, 283*, 284T, 285D/E, 286G/I/M/R/V, 287E/K	02_AG
210150931 (+)	3I, 12S, 15V, 16E , 19T, 20R , 35K, 36I , 37D, 37N, 41K, 63C, 69K , 72V, 93L	35T, 36A, 39E, 48T, 90I , 120F, 121G, 122C, 123T, 124L, 125N, 126F, 127H/R/S/Y, 128I, 129S, 131I, 132E, 133T, 134G, 135P/T, 136V, 137K, 138E/L/V/*, 139K, 173T, 174K, 177E, 196V, 202V, 207E, 211K, 214F, 243T, 245Q	Sub 1_8, Del 11_44, 45I/V, 72I, 101I, 104I, 112V, 113V, 123S, 127K, 136Q, 164K, 167E, 169D, 180G, 187K, 201I, 218I, 232D, 234I, 265V, 269K, 278A, 279S, 280R, 281Q, 282D, 283E, 284D, 285N, 286K/N/T, 287L/M/R/W	C
210481981	3I, 14R, 35D, 37N, 41K, 63P , 71V , 77I, 77M, 82I, 93L	60I, 122K, 123E, 135T, 200A, 211K/N, 214F, 245E, 250E, 257K, 259S, 260*, 261L, 263N	38P, 41E, 43G, 44R, 45V, 50H/K/N/Q, 112A, 123S, 124T, 125A, 127K, 172I, 218S, 232D, 271*, 277V, 278M, 279I, 280V, 281W, 282Q, 283V, 284E, 285G, 286L/M/R/W, 287E/G/K/R	B

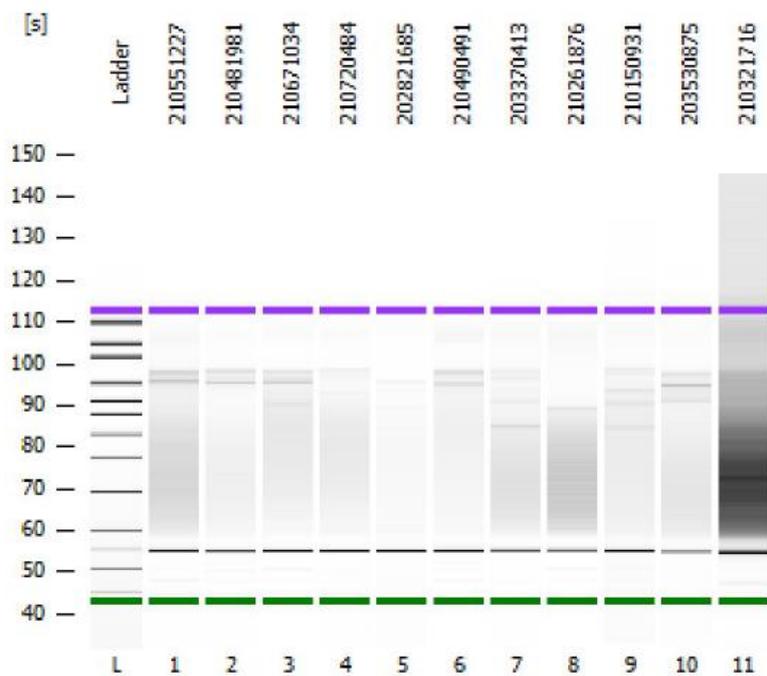
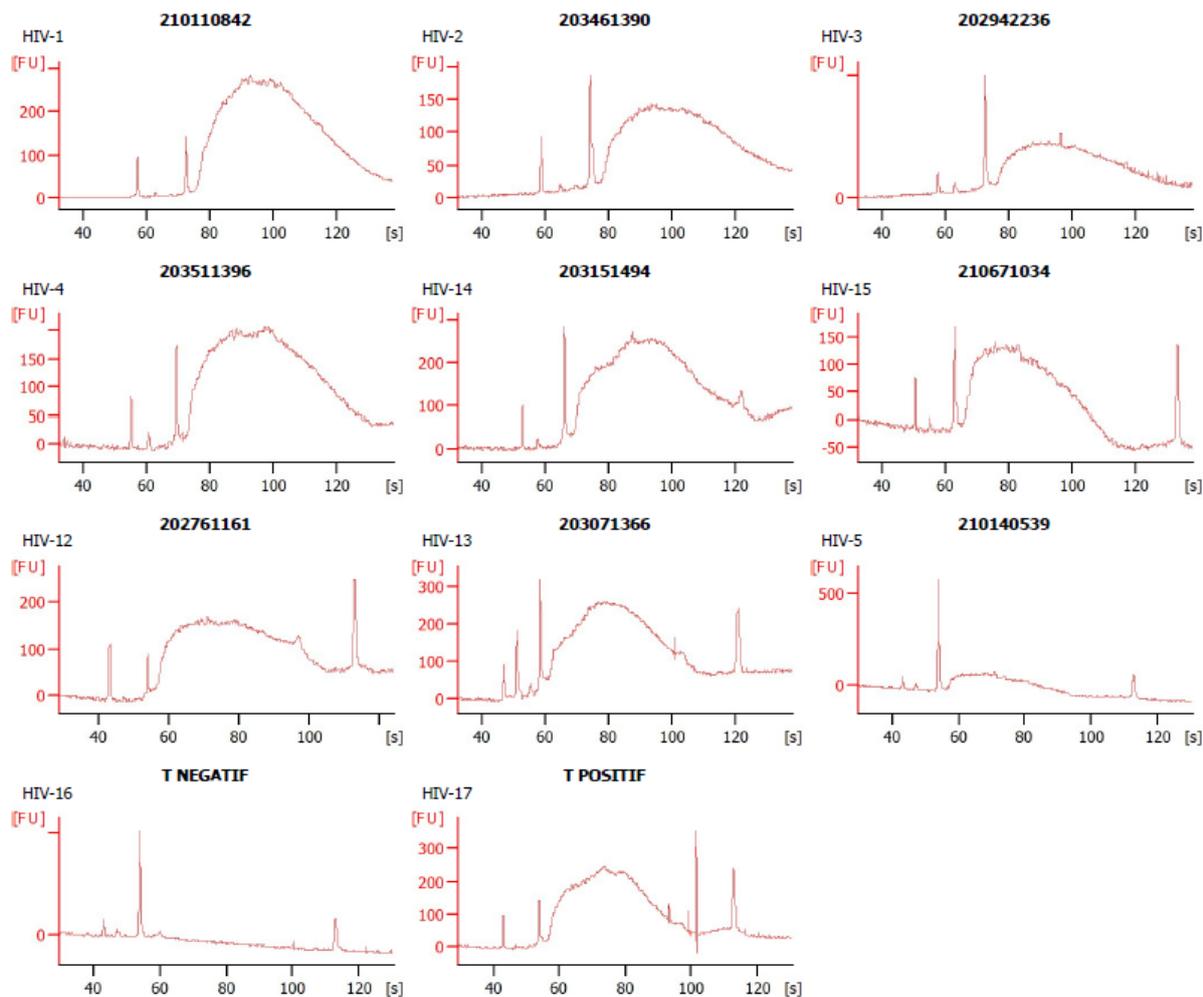
210490891	3I, 13V, 35D, 37H, 41K, 63P	60I, 122K, 123E, 135T, 177E, 200A, 211K/N, 214F, 244M, 245M, 250E, 257K, 259S, 260*, 261L, 263N	37L/E/*, 38P, 40F/L/R/W, 44K/R, 46N, 50I/K, 59A, 71Q, 72I, 101F, 119R, 123S, 124T, 125A, 127K, 193E, 201I, 206S, 232D, 234V, 255G, 278Q, 279V, 280D, 281R, 282M, 283R, 284I, 285K, 286H, 287W, 288R/S	B
210551227 (+)	3I, 12P, 19I, 45R, 60E , 62V, 63S, 64V, 77I, 93L	69S , 83K, 135T, 162C, 196E, 207N, 211K, 214F, 224D, 245M, 250E, 269R, 260F/S, 261C/K/N/R/S/W/Y/*, 262*, 264N, 265*	Sub 1_9, Del10_45, 46E/Q/*, 57N, 72I, 101I, 123S, 124T, 127K, 201I, 205S, 206S, 232D, 256E	B
210671034	3I, 10V , 12A/E/K, 13V, 15V, 16E , 20I, 35D, 36I , 37N, 41K, 63P , 69K , 72T, 89M	35T, 36D, 39K, 43Q, 49R, 60I, 106I , 123E, 135T, 142V, 162A, 169L, 173T, 174E, 177E, 192E, 194*, 200A, 202V, 203A, 207E, 211K, 245Q,	Sub1_7, Del 8_44, 45F/I/V, 48*, 72I, 101I, 103R, 112V, 123S, 125A, 127K, 134N, 135V, 136T, 163A, 201I, 208L, 218I, 232D, 234I, 272E, 273N, 274I/K/M/N/R/S/T, 276D, 277A, 278S, 280D, 281C, 282V, 283A, 284S, 285R, 286Q, 287D, 288G/W	02_AG
210720484	3I, 13V, 16E , 20I, 36I , 37N, 69K , 89M	35T, 36D, 72K, 123E, 135T, 162A, 165I/K, 173T, 177E, 196E, 200A, 207E, 211K, 214F, 245Q, 248D, 250E, 255H	40F/L/R, 41N, 43G, 44R, 72I, 74I , 96D, 101I, 106A, 112V, 123S, 125A, 127K, 134N, 136T, 163Q, 201I, 206S, 211R, 232D, 234I, 269K, 271*, 272E, 274*, 275L, 276C, 278K, 279*, 280T, 281G, 282W, 284I, 285E, 286H, 287G, 288R/S/T	02_AG
Contrôle Positif	3I, 10M, 13V, 15V, 35D, 36I , 37N, 41K, 45R, 57K, 63I, 69K , 89M	20R, 35T, 36E/K, 39M, 68A/G/S/T, 72K, 75L/V, 76N, 78K, 83K/R, 91H/Q, 100*, 110E, 117*, 118F, 121N, 122K, 123S/T, 125K/R, 127F/Y, 131P/T, 134I 135R/K, 145H/Q, 146S/Y, 158A/P, 159L, 161H/Q, 162Y, 169K, 172K, 173S, 174K, 177E, 179I , 182H, 207A, 211S, 249E, 250A, 251N, 252C, 253H, 254D, 255Y	60V, 84M, 90S, 112I, 113V, 123S, 124S, 125A, 127K, 136Q, 167E, 201I, 211R, 232D, 234I	A1

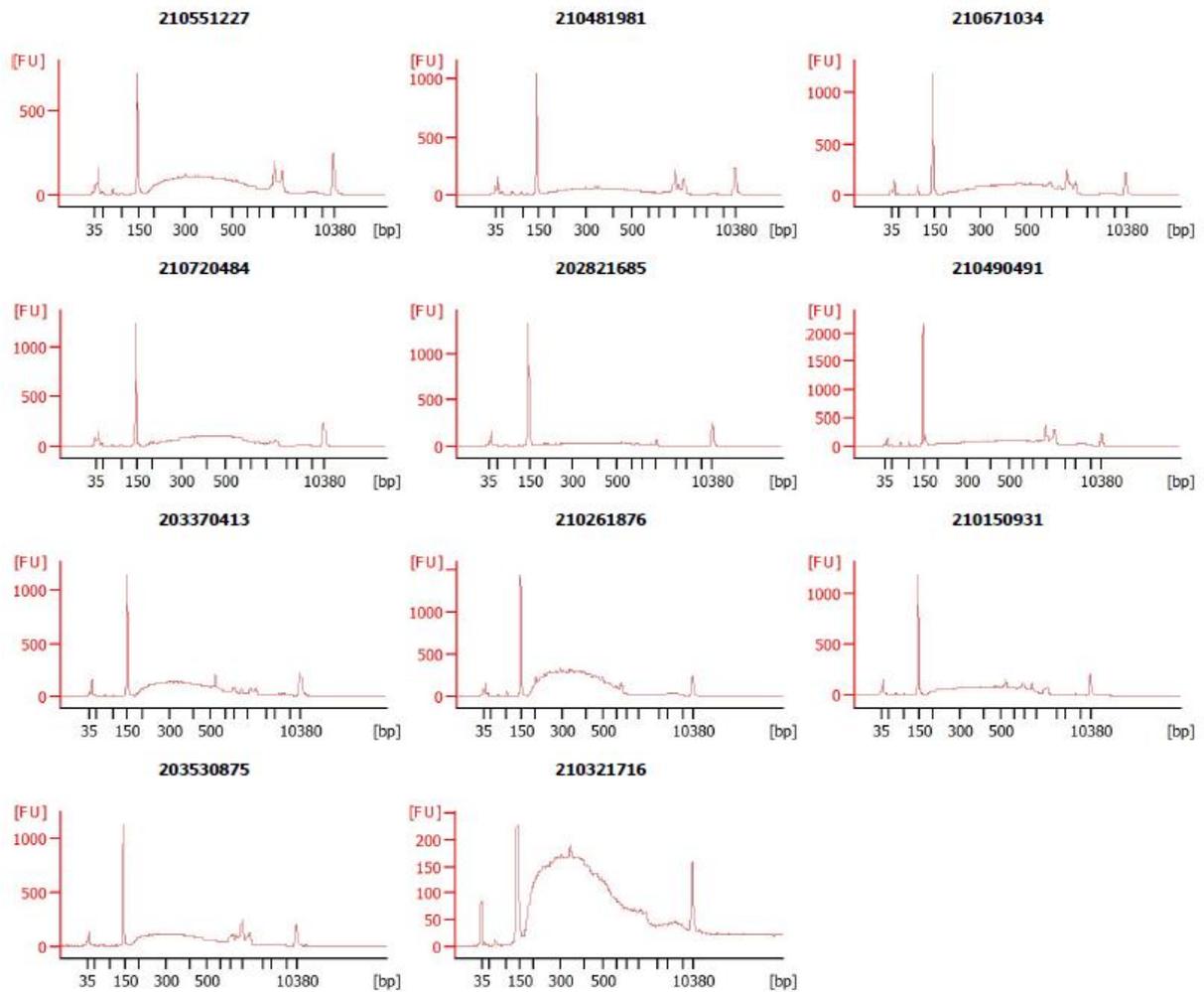
Annexe VI: Position de chaque échantillon sur le gel d'agarose. L'échantillon 210481981 et 210671034 ont été passé en doublon dans 2 séries différentes de NGS.

Position	Echantillons	
1	210110842	14 210551227
2	203461390	15 210481981
3	202942236	16 210671034
4	203511396	17 210720484
5	210140539	18 202821685
6	210321716	19 210490891
7	210481981	20 203370413
8	210490293	21 210261876
9	202801387	22 210150931
10	202761161	23 203530875
11	203071366	
12	203151494	
13	210671034	
-	Contrôle négatif	
+	Contrôle positif	

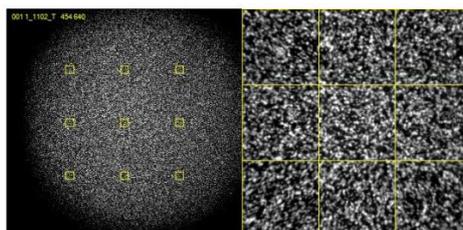
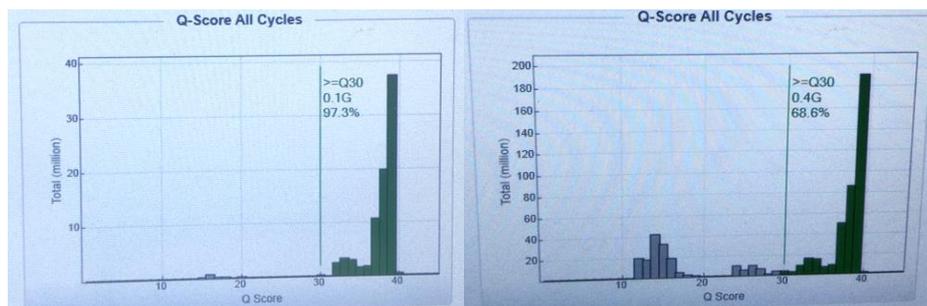
Annexe VII: Contrôle qualité des librairies : Profil Bioanalyser®





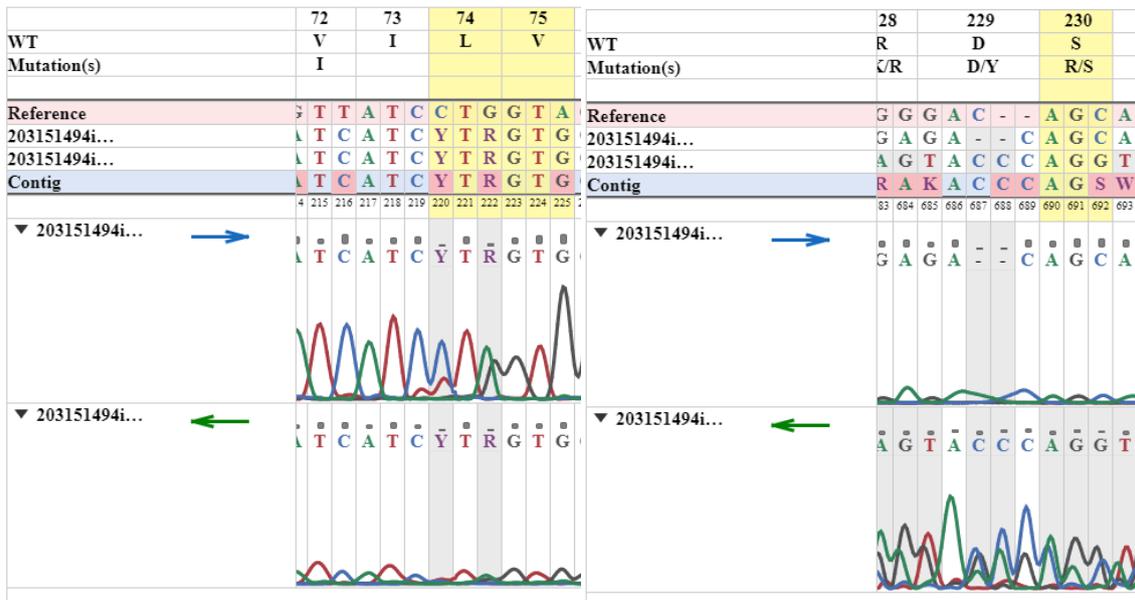


Annexe VIII : Contrôle technique par l'instrument Miseq Illumina® : Pourcentage des « clusters passing filter » pour les deux cartouches NGS et visualisation des capteurs pour estimer la densité des clusters



*

Annexe IX : Visualisation du chromatogramme pour l'échantillon 20315149 discordant dans la technique ABL par rapport aux autres techniques.



Annexe X: Exemple d'un chromatogramme de bonne qualité



Annexe XI: Exemple de compte rendu pour le NGS et le Sanger ABL



Patient / Sample Information		Sample		Physician / Project Information	
Patient ID		Alternative ID		Name	
Name		Date Collected	10/12/2020	Institution	
DOB		Date Reported	07/06/2021	Address	
Gender		Sequencing platform	Illumina - MiSeq		
Viral Load		Software version	2.0		
Sample ID	203530875	Expert system	2.3		

SUBTYPING				COVERED POSITIONS			
PROT	02_AG (95.62 %)	RT	31_BC (91.99 %)	PROT	1→99	RT	1→318
INT	02_AG (93.91 %)			INT	45→284		

[RT, PROT, INT] : Subtype B [K294S](#) was used as the reference sequence for the alignment (using DWA v0.7.15 alignment tool). Subtyping determination performed through homology testing of a 20% consensus sequence generated from all the reads mapped to the particular region and compared to an updated set of reference sequences.

ANTIRETROVIRAL DRUG	INTERPRETATION	DEEPCHEK® HIV (ANRS 31)	
Generic name	Assessment	Resistance mutations >20.00%	Resistance mutations between >3% and <20.00%
N	Atazanavir/r	S	
	Darunavir/r 600/100 mg BID	S	
	Darunavir/r 800/100 mg QD	S	
	Lopinavir/r	S	
	Tipranavir/r	R	
NRTI	Abacavir	I	
	Emtricitabine	R	
	Lamivudine	R	
	Tenofovir	S	
	Zidovudine	S	
NNRTI	Doravirine	I	
	Efavirenz	R	
	Etravirine	R	
	Islatavir	R	
	Nevirapine	R	
Rilpivirine	R		
INI	Bictegravir	S	
	Cabotegravir	S	
	Dolutegravir 50mg BID	S	
	Dolutegravir 50mg QD	S	
	Elvitegravir	S	
Raltegravir	S		

S	Susceptible	I	Possible resistance	R	Resistance
---	-------------	---	---------------------	---	------------

HIV Reverse Transcriptase mutations

Position	Mutation	20.00%	10.00%	5.00%	3.00%	Prevalence %	Q-Score	Mutational load (cp/mL)
20	K→R			✓	✓	8.13	38	0
35	V→T	✓	✓	✓	✓	99.09	39	0
39	T→D	✓	✓	✓	✓	36.53	39	0
	T→N	✓	✓	✓	✓	62.24	39	0
40	E→D	✓	✓	✓	✓	61.28	39	0
49	K→R	✓	✓	✓	✓	98.5	39	0
60	V→I	✓	✓	✓	✓	99.19	39	0
123	D→S	✓	✓	✓	✓	98.25	39	0
135	I→R	✓	✓	✓	✓	99.36	39	0
162	S→A	✓	✓	✓	✓	84.54	39	0
	S→T		✓	✓	✓	15.01	39	0
169	E→D	✓	✓	✓	✓	98.89	39	0
172	R→K	✓	✓	✓	✓	98.72	39	0
174	Q→K	✓	✓	✓	✓	99.04	39	0
177	D→E	✓	✓	✓	✓	99.51	39	0
178	I→M	✓	✓	✓	✓	97.87	39	0
181	Y→C	✓	✓	✓	✓	97.75	39	0
184	M→V	✓	✓	✓	✓	64	39	0
193	L→F				✓	3.2	38	0
196	G→E	✓	✓	✓	✓	98.67	39	0
200	T→A	✓	✓	✓	✓	99.39	38	0
203	E→G	✓	✓	✓	✓	23.9	38	0
207	Q→D	✓	✓	✓	✓	98.36	39	0
214	L→F	✓	✓	✓	✓	96.63	38	0
221	H→Y	✓	✓	✓	✓	74.08	38	0
245	V→Q	✓	✓	✓	✓	99.27	39	0
248	E→D	✓	✓	✓	✓	98.88	39	0
272	P→A	✓	✓	✓	✓	99.55	39	0
277	R→K	✓	✓	✓	✓	99.04	39	0
286	T→A	✓	✓	✓	✓	99.26	38	0
291	E→D	✓	✓	✓	✓	98.8	39	0
292	V→I	✓	✓	✓	✓	94.5	39	0
293	I→V	✓	✓	✓	✓	99.7	38	0

Fragment between amino acid 319 and 348 could not be amplified

Subtype B [K03455](#) was used as the reference sequence for the alignment (using BWA v0.7.15 alignment tool).

HIV Protease mutations

Position	Mutation	20.00%	10.00%	5.00%	3.00%	Prevalence %	Q-Score	Mutational load (cp/mL)
3	V→I	✓	✓	✓	✓	99.26	38	0
13	I→V	✓	✓	✓	✓	98.84	39	0
14	K→R	✓	✓	✓	✓	83.66	39	0
17	G→E	✓	✓	✓	✓	98.02	39	0
20	K→I	✓	✓	✓	✓	98.09	39	0
36	M→I	✓	✓	✓	✓	98.82	39	0
37	S→N	✓	✓	✓	✓	99.18	39	0
43	K→R	✓	✓	✓	✓	97.69	39	0
63	L→S	✓	✓	✓	✓	97.78	39	0
69	H→K	✓	✓	✓	✓	98.23	39	0
89	L→M	✓	✓	✓	✓	98.94	39	0

Subtype B [K03455](#) was used as the reference sequence for the alignment (using BWA v0.7.15 alignment tool).

Mutations of interest based on ANRS 31 (Bold red text)



Insufficient number of sequences to guarantee, at the 99% confidence level, that all mutations with the given threshold frequency have been found at that position.

HIV Integrase mutations

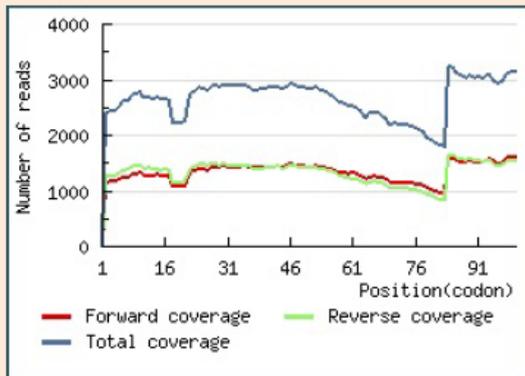
Position	Mutation	20.00%	10.00%	5.00%	3.00%	Prevalence %	Q-Score	Mutational load (cp/mL)
63	L→I	✓	✓	✓	✓	97.89	39	0
74	L→I	✓	✓	✓	✓	98.6	39	0
79	V→I			✓	✓	6.09	37	0
81	S→T			✓	✓	6.41	38	0
84	I→M	✓	✓	✓	✓	84.89	39	0
91	A→E	✓	✓	✓	✓	97.03	39	0
95	Q→K	✓	✓	✓	✓	97.72	39	0
101	L→I	✓	✓	✓	✓	99.24	39	0
112	T→I	✓	✓	✓	✓	97.53	39	0
119	S→P	✓	✓	✓	✓	97.48	38	0
122	T→I	✓	✓	✓	✓	97.92	39	0
123	G→S	✓	✓	✓	✓	99.45	38	0
125	T→A	✓	✓	✓	✓	99.03	38	0
127	R→K	✓	✓	✓	✓	99.59	38	0
134	G→N	✓	✓	✓	✓	98.36	39	0
136	K→T	✓	✓	✓	✓	98.12	39	0
201	V→I	✓	✓	✓	✓	99.59	38	0
206	T→S	✓	✓	✓	✓	98.42	38	0
215	K→N	✓	✓	✓	✓	98.07	39	0
218	T→I	✓	✓	✓	✓	99.15	39	0
232	N→D	✓	✓	✓	✓	99.06	38	0
234	L→I	✓	✓	✓	✓	99.43	39	0
265	A→V	✓	✓	✓	✓	97.64	39	0
269	R→K	✓	✓	✓	✓	23.33	39	0
278	D→A	✓	✓	✓	✓	99.33	38	0
280	C→S	✓	✓	✓	✓	99.31	38	0

Subtype B [K03455](#) was used as the reference sequence for the alignment (using BWA v0.7.15 alignment tool).

Mutations of interest based on ANRS 31 (Bold red text)

Coverage

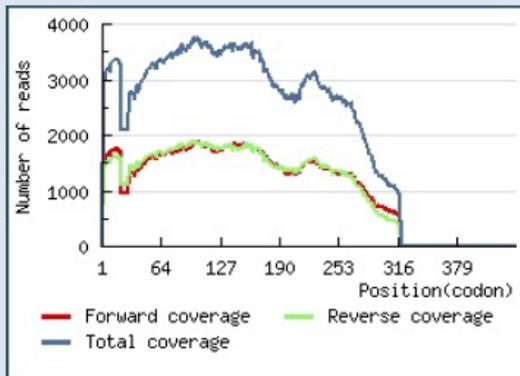
Protease (PROT)



Subtype B [K03455](#) was used as the reference sequence for the alignment (using BWA v0.7.15 alignment tool).

Threshold	Minimum sequences for 99% confidence	Low covered position
20.00%	23	
10.00%	46	
5.00%	92	
3.00%	153	

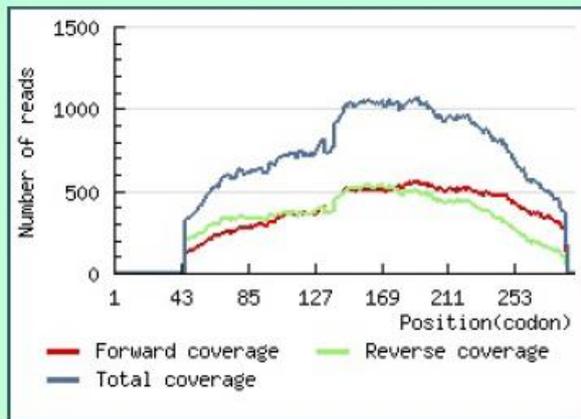
Reverse Transcriptase (RT)



Subtype B [K03455](#) was used as the reference sequence for the alignment (using BWA v0.7.15 alignment tool).

Threshold	Minimum sequences for 99% confidence	Low covered position
20.00%	23	319→440
10.00%	46	319→440
5.00%	92	319→440
3.00%	153	319→440

Integrase (INT)



Subtype B [K03455](#) was used as the reference sequence for the alignment (using BWA v0.7.15 alignment tool).

Threshold	Minimum sequences for 99% confidence	Low covered position
20.00%	23	1→44,285→289
10.00%	46	1→44,285→289
5.00%	92	1→44,285→289
3.00%	153	1→44,285→289

Patient/Sample information

ABL/TE ID	Analysis RT-PROT-INT	Type of report Corrected
Patient name	Clinician name	Data report edition 14/06/2021,14:36:38
Patient birth date	Genotyping method DeepChekA® SingleRound	Data report version 12.4
Patient viral load Missing_data	Type of sample Plasma	
Your sample ID 203530875	Date of sample	
Your patient ID	Date of sequencing	

Subtype:

Reverse transcriptase C (90 % similarity) Protease 02_AG (+count_legend_subtyping+) (100 % similarity) Integrase G (check for 02_AG) (98 % similarity)

All Mutations Detected (HXB2 reference Sequence) Resistance mutations in bold based on ANRS 31

Reverse transcriptase V35T, T39N/D, E40D, K49R, V60I, D123S, I135R, S162A, E169D, R172K, Q174K, D177E, I178M, **Y181C**, M184V, G196E, T200A, E203wt/G, Q207D, L214F, **H221D/Y**wt/N, V245Q, E248D, P272A, R277K, T286A, E291D, V292I, 293insV, P313A/wt, H315D/wt

Protease V3I, I13V, K14R, G17E, K20I, **M36I**, S37N, K43R, L63S, **H69K**, L89M

Integrase del33, A33wt/G, del34, del35, del36, del37, S39H/R/N/wt, L63I, **L74I**, I84M, A91E, Q95K, L101I, T112I, S119P, T122I, G123S, T125A, R127K, G134N, K136T, V201I, T206S, K215N, T218I, N232D, L234I, A265V, D278A, C280S

Genotypic interpretation (ANRS-31)

Class	Drug		ANRS III 31 12-2020	
NRTI	Zidovudine	S		
	Lamivudine	R	M184V	
	Emtricitabine	R	M184V	
	Abacavir	I	M184V	
	Tenofovir	S		
	Islatavir	R	M184V	
NNRTI	Nevirapine	R	Y181C	
	Efavirenz	R	Y181C	
	Etravirine	R	Y181C, H221Y	
	Rilpivirine	R	Y181C, H221Y	
PI	Doravirine	I	Y181C, H221Y	
	Lopinavir/r	S		
	Atazanavir/r	S		
	Tipranavir/r *	R	M36I, H69K, L89M	
	Darunavir/r 600/100 mg BID	S		
	Darunavir/r 800/100 mg QD	S		
	Raltegravir	S		
	Elvitegravir	S		
	II	Dolutegravir 50mg BID	S	
		Dolutegravir 50mg QD	S	
Cabotegravir		S		
Bictegravir		S		

S

ANRS III
Sensible

LHOSSEIN Thomas

**Thèse d'exercice pour le diplôme d'Etat de Docteur en
Pharmacie**

-

Université de Picardie Jules Verne

**Séquençage du VIH: Apport du NGS en routine et
comparaison avec la méthode Sanger**

Résumé :

Le thème principal de ce travail est l'approche NGS dans un contexte où les avancées technologiques dans le domaine de la biologie moléculaire sont en perpétuelle évolution. Le kit Viroseq®, actuellement utilisé au CHU d'Amiens va subir un arrêt programmé. Il est alors nécessaire de trouver une solution pour pouvoir continuer le séquençage du VIH. Dans un premier temps, le but sera d'évaluer la technique NGS via la technologie Illumina® que ce soit sur le plan technique ou analytique, puis dans un second temps, de comparer les résultats en termes de mutations d'intérêt pour les 3 principales cibles du VIH ainsi que la détection de sous-types avec les 2 autres techniques Sanger évaluées (ABL® et ANRS). L'ensemble des données analysées seront discutées pour permettre la prise de décision au sein du laboratoire de virologie du CHU d'Amiens.

Mots clés : VIH, NGS, Illumina, Sanger, bio-informatique, Mutations d'intérêts.