

Cahier des charges pour le séquençage du SARS-CoV-2

CNR Virus des infections respiratoires (dont la grippe), 21/01/2021.

Les éléments ci-après sont les prérequis définis par le CNR pour un séquençage utile et conforme aux normes attendues en la matière.

1/ Qualité du séquençage

- Séquençage Sanger

Le séquençage doit couvrir l'ensemble du gène S afin d'identifier l'ensemble des mutations (substitutions et délétions) qui caractérisent les variants définis à ce jour. Les mutations du gène S listées ci-après doivent être présentes pour la qualification des variants définis à ce jour :

- Variant 20I/501Y.V1 [Royaume-Uni] : substitutions N501Y, A570D, P681H, T716I, S982A, D1118H et délétions del69-70 (HV), del144/145 (Y) ;
- Variant 20H/501Y.V2 [Afrique du Sud] : substitutions D80A, D215G, E484K, N501Y, A701V éventuellement associées aux substitutions L18F, R246I, K417N et délétion dell242-244L ou substitution L242H ;
- Variant 20J/501Y.V3 [Brésil] : L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y, T1027I.

Cette méthode n'est pas adaptée à la détection de l'émergence potentielle de nouveaux variants.

- Séquençage génome entier par NGS

Ce séquençage quelle que soit la technique utilisée (Illumina, IonTorrent, MinION) est la méthode de référence, et doit dans la mesure du possible couvrir la totalité du génome viral. D'après notre évaluation de méthode de séquençage génome entier SARS-CoV-2 (Charre, Virus Evol, 2020), les critères pour valider une séquence sont les suivants :

- Couverture minimale pour appeler une base : >10X pour Illumina ; > 20X pour MinION. En dessous de ce seuil, la base n'est pas déterminée (ie appelée N).
- Séquence validée si >99% de la séquence est déterminée (ie sans base ambiguë N) avec une profondeur de couverture moyenne de 1000x
- Pour les délétions : il est indispensable de différencier délétion et défaut de couverture. Une délétion est visible sur le fichier d'alignement (cf Charre, Virus Evol, 2020)
- Un contrôle négatif qui a suivi l'ensemble du pipeline depuis l'extraction doit être séquençé à chaque run afin de déterminer la présence ou non de contamination.

Les mutations (substitutions et délétions) sont définies par rapport à la séquence de référence hCoV-19/Wuhan/WIV04/2019 (WIV04) utilisée dans la base de données GISAID (EPI_ISL_402124) (Pillay et al. <https://doi.org/10.2807/1560-7917.ES.2020.25.8.2000097>). Un variant est défini par sa position dans un arbre phylogénétique, par exemple à l'aide de l'outil nextclade ou pangolin.

2/ Habilitation des laboratoires effectuant le séquençage

L'habilitation d'un laboratoire à rendre des résultats de séquençage suppose :

- Le séquençage simultané de 10 échantillons comprenant au moins deux variants (parmi les 501Y.V1, 501Y.V2 et 501Y.V3) par le laboratoire demandeur et le CNR. Pour se faire le laboratoire devra faire parvenir au plus vite les échantillons sélectionnés. Les résultats seront ensuite comparés à ceux du CNR.

3/ Enregistrement dans la base de données GISAID

Le dépôt des séquences, associées aux métadonnées, dans la base de données GISAID doit être effectué par un laboratoire habilité dans les plus brefs délais (après validation par le CNR pour les laboratoires non habilités).

4/ Alerte

En cas de détection de nouvelles mutations susceptibles d'avoir une signification fonctionnelle ou de nouvelles mutations récurrentes, il est demandé au laboratoire, même s'il est habilité, d'envoyer les séquences brutes sans délai au CNR.