

Flux de travail d'enrichissement pour la détection du coronavirus à l'aide des systèmes de SNG d'Illumina

Détection et caractérisation
hautement sensibles des virus
respiratoires courants, y compris
les souches du coronavirus



La surveillance virale avec le SNG

Les infections virales sont une préoccupation majeure en matière de santé mondiale, de nouveaux variants de maladies infectieuses émergeant constamment. L'épidémie d'un nouveau coronavirus (SRAS-CoV-2) en 2019 qui a commencé à Wuhan, en Chine, et s'est rapidement propagée à plusieurs pays est un bon exemple des risques potentiels associés à une pandémie virale mondiale. Les coronavirus (CoV) sont une grande famille de virus qui peuvent infecter l'homme, provoquant des maladies respiratoires allant du rhume à des maladies plus graves, comme le syndrome respiratoire du Moyen-Orient (MERS-CoV) et le syndrome respiratoire aigu sévère (SRAS-CoV). L'Organisation mondiale de la santé (OMS) a désigné la maladie associée au SRAS-CoV-2 (COVID-19) comme une urgence de santé publique de portée internationale, soulignant la nécessité d'une détection rapide et précise du virus¹.

Le séquençage de nouvelle génération (SNG) offre une nouvelle façon efficace de cribler les échantillons et de détecter les virus sans nécessiter de connaissances préalables sur l'agent infectieux². L'enrichissement de la cible est une méthode de séquençage qui capture les régions génomiques d'intérêt en hybridant les acides nucléiques sur des sondes biotinylées spécifiques à la cible. L'enrichissement de la cible par capture hybride permet une détection hautement sensible, sans nécessiter la profondeur de lecture élevée requise pour le séquençage métagénomique aléatoire. De plus, le flux de travail de SNG d'enrichissement de la cible permet d'obtenir des données de séquençage quasi complètes des cibles et la réalisation, entre autres, de l'analyse des variants pour l'évolution virale ou la surveillance virale³. Par rapport à d'autres méthodes de séquençage ciblé, telles que le séquençage des amplicons, l'enrichissement par capture hybride permet d'obtenir des panels de sondes beaucoup plus larges avec un profilage plus complet des régions cibles. En outre,

les oligo-sondes utilisées pour les protocoles de capture hybride restent efficaces même dans les régions hautement mutagènes, ce qui permet de cibler les virus en évolution rapide comme les virus à ARN.

Flux de travail d'enrichissement viral

Cette note d'application met en évidence un flux de travail rationalisé pour la détection et l'analyse du coronavirus à l'aide d'Illumina DNA Prep with Enrichment Library Preparation Kit, des panels de ciblage viral, du séquençage éprouvé d'Illumina et de l'analyse simplifiée des données (figure 1). Ce flux de travail est destiné à enrichir les cibles d'ADN et d'ARN viraux à partir de l'extraction d'acide nucléique total. Le protocole commence par l'extraction d'échantillons d'acide nucléique et la transcription inverse ultérieure de l'ARN extrait en ADNc double brin. La préparation des bibliothèques est effectuée à l'aide d'Illumina DNA Prep with Enrichment with IDT for Illumina DNA/RNA UD Indexes. Dans ce protocole, l'ADN et l'ADNc subissent une tagmentation, un nettoyage et une amplification de préenrichissement. Après l'amplification, jusqu'à 12 échantillons peuvent être regroupés pour une réaction d'enrichissement à l'aide d'un panel d'oligos qui ciblent les séquences virales. L'hybridation des sondes est suivie de la capture des sondes, de l'amplification d'enrichissement, de la quantification et du séquençage.

Méthodes

Préparation des échantillons

Pour démontrer les performances du panel d'enrichissement de la cible virale, une souche d'échantillon de coronavirus désactivé, la souche CoV OC43 de Microbiologics (QC Sets

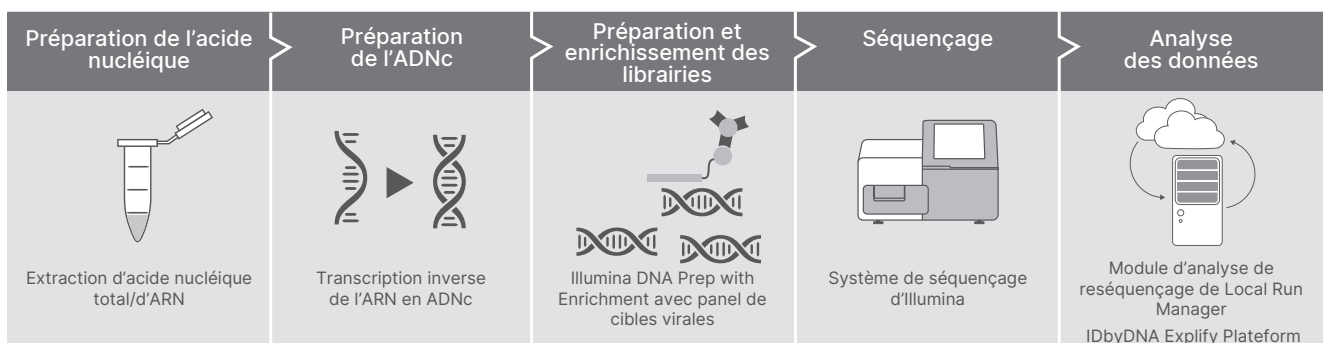


Figure 1 : Flux de travail d'enrichissement pour la détection des coronavirus – Le flux de travail de SNG rationalisé pour la détection des coronavirus intègre la préparation des échantillons, la préparation des bibliothèques, l'enrichissement de la cible, le séquençage et l'analyse des données.

and Panels: Helix Elite, référence n° 8217), a été analysée ([tableau 1](#)). L'échantillon de culture virale du coronavirus a été extrait à l'aide de QIAGEN QIAmp Viral Mini Kit (QIAGEN, référence n° 52904) dans un environnement de laboratoire BSL2. Cent-cinquante (150) ng d'ARN extrait ont été soumis à une transcription inverse en ADNc à l'aide de deux flux de travail différents : l'un dérivé des réactifs TruSeq RNA d'Illumina et l'autre utilisant Thermo Fisher Scientific Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific, référence n° K2561). Pour reproduire les résultats d'un séquençage d'échantillons de patients, l'échantillon viral a également été ajouté à 95 % d'ARN de référence humaine universelle (UHR, Universal Human Reference) (Agilent Technologies, référence n° 74000) et soumis à une transcription inverse en ADNc en utilisant les méthodes décrites ci-dessus ([tableau 1](#)).

De plus, un regroupement de virus respiratoires contenant quatre virus à ARN et un virus à ADN ([tableau 2](#)) a été extrait à l'aide de QIAGEN AllPrep PowerViral DNA/RNA Kit (QIAGEN, référence n° 28000-50) dans un environnement de laboratoire BSL2. Le volume maximal d'acide nucléique total a été soumis à une transcription inverse en ADNc (sans traitement à la DNase) à l'aide d'un flux de travail dérivé des réactifs TruSeq RNA d'Illumina. Pour reproduire les résultats d'un séquençage d'échantillons de patients avec plusieurs agents pathogènes présents, l'échantillon de regroupement de virus a également été ajouté à 95 % d'ARN de référence humaine universelle (UHR) (Agilent Technologies, référence n° 74000) et soumis à une transcription inverse en ADNc ([tableau 1](#)).

Tableau 1 : Composition des échantillons viraux pour analyse

Échantillon	Composition ^a	Génome de référence
CoVOC43	150 ng d'ARN de CoV OC43	AY391777.1 Coronavirus humain OC43
CoVOC43_95UHR	7,5 ng d'ARN de CoV OC43 et 142,5 ng d'ARN UHR	AY391777.1 Coronavirus humain OC43
ViralPool	Volume maximal d'acide nucléique total du regroupement de virus (8,5 µl)	Consulter le tableau 2
ViralPool_95UHR	5 % d'acide nucléique total du regroupement de virus et 95 % d'ARN UHR par volume	Consulter le tableau 2

a. L'entrée minimale recommandée d'ARN/acide nucléique total pour la transcription inverse est de 10 ng. Pour de meilleurs résultats, la transcription inverse doit être effectuée sur des échantillons d'acide nucléique récemment extraits.

Tableau 2 : Composition du regroupement de virus

Virus	Type d'acides nucléiques	Génome de référence
Virus grippal A (H1N1)	ARN	Grippe A/ Michigan/45/2015
Virus grippal B	ARN	Grippe B/ Colorado/06/2017
Virus parainfluenza humain 3	ARN	NC_001796.2
VRS B9320	ARN	AY353550.1
Adénovirus 7	ADN	AY594255.1

Préparation des bibliothèques

Les bibliothèques prêtes pour le séquençage ont été préparées avec Illumina DNA Prep with Enrichment (Illumina, référence n° 20025524) et IDT for Illumina DNA UD Indexes (Illumina, référence n° 20027213)* à l'aide de l'ADNc de l'échantillon de CoV (CoVOC43), de l'échantillon de regroupement de virus (ViralPool) et des deux échantillons ajoutés (CoVOC43_95UHR et ViralPool_95UHR). L'entrée d'ADN totale recommandée pour la tagmentation est de 10 à 1 000 ng.

Après l'amplification, les échantillons ont été divisés en regroupements distincts pour l'enrichissement en fonction du type d'échantillon (CoVOC43, CoVOC43_95UHR, ViralPool et ViralPool_95HR). Les réactions d'enrichissement ont été effectuées avec Respiratory Virus Oligos Panel (Illumina, référence n° 20042472), qui comprend environ 7 800 sondes conçues pour détecter les virus respiratoires, les souches récentes du virus de la grippe et le SRAS-CoV-2, ainsi que des sondes humaines qui servent de contrôles positifs ([tableau 3](#)). Après l'enrichissement ([figure 2](#)), les bibliothèques préparées ont été quantifiées, regroupées et chargées sur le MiSeq^{MC} System pour le séquençage.

* La référence n° 20027213 est désormais obsolète, elle est remplacée par Illumina DNA/RNA UD Indexes Set A, Tagmentation (96 index, 96 échantillons), référence n° 20091654.

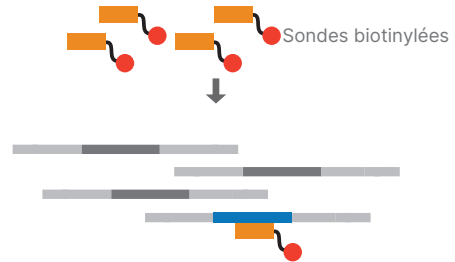
Tableau 3 : Virus ciblés par Respiratory Virus Oligos Panel

Coronavirus humain 229E	Virus parainfluenza humain 1
Coronavirus humain NL63	Virus parainfluenza humain 2
Coronavirus humain OC43	Virus parainfluenza humain 3
Coronavirus humain HKU1	Virus parainfluenza humain 4a
SRAS-CoV-2	Métapneumovirus humain (CAN97-83)
Adénovirus humain B1	Virus respiratoire syncytial (type A)
Adénovirus humain C2	Virus respiratoire syncytial humain 9320 (type B)
Adénovirus humain E4	Virus grippal A (A/Porto Rico/8/1934(H1N1))
Bocavirus humain 1 (primate bocaparvovirus 1, isolat st2)	Virus grippal A (A/Corée/426/1968(H2N2))
Bocavirus humain 2c PK, isolat PK-5510	Virus grippal A (A/New York/392/2004(H3N2))
Bocavirus humain 3	Virus grippal A (A/oie/Guangdong/1/1996(H5N1))
Souche NI du bocavirus humain 4 HBoV4-NI-385	Virus grippal A (A/Zhejiang/DTID-ZJU01/2013(H7N9))
Polyomavirus KI Stockholm 60	Virus grippal A (A/Hong Kong/1073/99(H9N2))
Polyomavirus WU	Virus grippal A (A/Texas/50/2012(H3N2))
Parechovirus humain de type 1	Virus grippal A PicoBank/HPeV1/a (A/Michigan/45/2015(H1N1))
Parechovirus humain 6	Virus grippal B (B/Lee/1940)
Rhinovirus humain A89	Virus grippal B (B/Wisconsin/01/2010)
Rhinovirus humain C (souche 024)	Virus grippal B (B/Brisbane/60/2008)
Rhinovirus humain B14	Virus grippal B (B/Colorado/06/2017)
Souche d'entérovirus humain C104 : AK11	Virus grippal B (B/Washington/02/2019)
Entérovirus humain C109, isolat NICA08-4327	Gènes humains de régulation

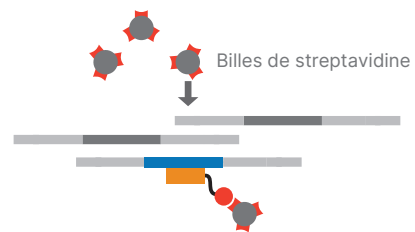
A. Regroupement des librairies Illumina DNA Prep with Enrichment



B. Hybridation des sondes biotinylées aux régions ciblées



C. Capture à l'aide des billes de streptavidine



D. Éluion à partir des billes

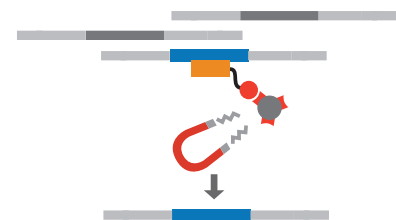


Figure 2 : Chimie d'enrichissement – Les librairies préparées subissent un enrichissement avec un panel de cibles virales, suivi d'une amplification, d'une quantification et d'un séquençage.

Séquençage

Les librairies préparées peuvent être séquencées sur n'importe quel instrument Illumina. Les systèmes de paillasse iSeq^{MC} 100, MiniSeq^{MC} et MiSeq conviennent parfaitement en raison des faibles exigences de lecture pour ces échantillons. Dans cette étude, les librairies préparées à partir des échantillons viraux ont été dénaturées et diluées à une concentration de chargement finale de 10 pM, conformément au Guide de dénaturation et de dilution des librairies pour le système MiSeq (document n° 15039740 v10) et séquencées sur le MiSeq System à une longueur de lecture de 2 × 151 pb à l'aide des réactifs MiSeq v3 (Illumina, référence n° MS-102-3001).

Les aliquotes de l'ensemble des librairies ont également été séquencées avant l'hybridation des sondes pour déterminer la modification de facteur de la réaction d'enrichissement.

Le titre viral, la qualité de l'ARN et le nombre de lectures par échantillon ont un impact sur le nombre de lectures spécifiques au virus obtenues et la couverture du génome viral. En règle générale, la recommandation de lecture pour ce flux de travail est de 500 000 lectures par échantillon, mais ces chiffres peuvent varier et il ne s'agit là que d'un point de départ recommandé.

Analyse des données

Pour l'analyse automatique sur instrument, les analyses de séquençage ont été configurées dans Local Run Manager à l'aide du module d'analyse de reséquençage. Ce module permet la saisie de tous les renseignements sur l'analyse et la sélection du génome de référence pour l'alignement ultérieur de la séquence. Les utilisateurs peuvent téléverser les génomes de référence directement sur l'instrument, facilitant ainsi la personnalisation. L'analyse est lancée automatiquement une fois le séquençage terminé afin que les utilisateurs puissent interpréter les résultats dans les plus brefs délais. Le module de reséquençage fournit des données sur l'alignement, la couverture et les petits variants, ainsi que des fichiers FASTQ, BAM et VCF à utiliser dans d'autres pipelines d'analyse de données, au besoin (figure 3).

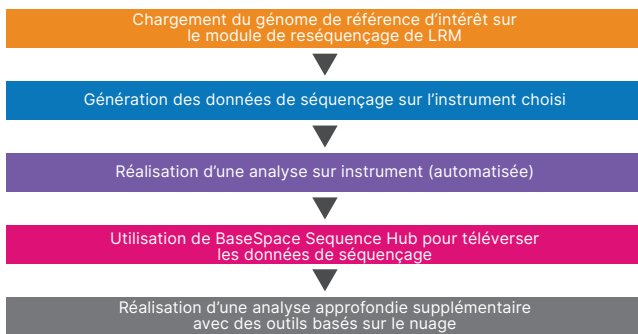


Figure 3 : Flux de travail d'analyse d'enrichissement – Les analyses de séquençage sont configurées dans Local Run Manager et l'analyse des données se poursuit automatiquement une fois l'analyse terminée. Une analyse supplémentaire avec des applications tierces est disponible, le cas échéant.

IDbyDNA Explify Platform est une solution facile à utiliser pour l'analyse approfondie des données qui comprend un contrôle de la qualité (CQ) des données fiable, une interprétation normalisée des résultats, des bases de données soigneusement conservées et la génération de rapports personnalisés. L'analyse des données est basée sur les k-mers et les étapes d'alignement, notamment la détection des virus au niveau des protéines, augmentant ainsi la capacité à identifier de nouveaux virus et des virus

hautement divergents. La plateforme est accessible dans BaseSpace^{MC} Sequence Hub.

Résultats

Après la préparation des librairies et le séquençage, Local Run Manager Resequencing Analysis Module (v2.5.56.11) a été utilisé pour aligner chaque échantillon sur le génome de référence OC43 du coronavirus, comme décrit dans le [Guide du flux de travail du module d'analyse de reséquençage de Local Run Manager](#).

Transcription inverse et détection du CoV avec le module de reséquençage de Local Run Manager

Dans cette étude, deux trousse de transcription inverse différentes ont été utilisées pour générer de l'ADNc pour la préparation des librairies : une trousse de Thermo Fisher Scientific disponible sur le marché et un flux de travail dérivé des réactifs TruSeq RNA d'Illumina (voir la section Méthodes pour plus de détails). Pour les échantillons de CoVOC43, les résultats ont montré que le pourcentage d'alignement par rapport au génome de référence était > 99 % pour la méthode de transcription inverse d'Illumina et > 97 % pour la méthode de Thermo Fisher Scientific (figure 4). Le pourcentage d'alignement des échantillons ajoutés était > 64 % et > 19 % pour Respiratory Virus Oligos Panel, quelle que soit la méthode de transcription inverse utilisée (figure 4).

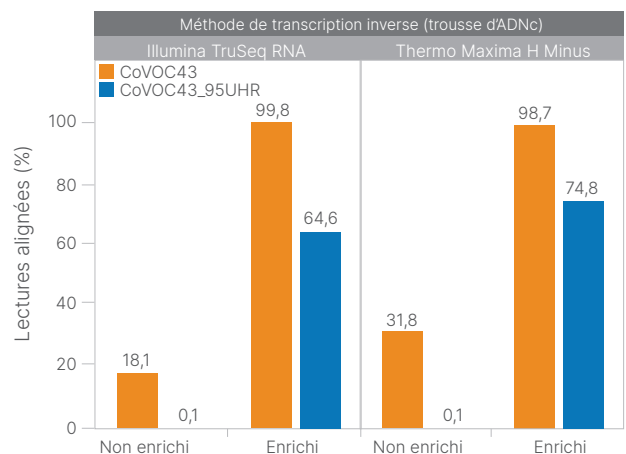


Figure 4 : Performance équivalente avec différentes trousse d'ADNc – Les pourcentages de lectures alignées pour les échantillons viraux soumis à une transcription inverse avec les deux trousse d'ADNc différentes (voir la section Méthodes) sont représentés avant (non enrichies) et après (enrichies) l'enrichissement avec Respiratory Virus Oligos Panel.

Comme mesure de la réussite de la capture de la cible et de la réaction d'hybridation, les pourcentages de lectures alignées des échantillons avant et après enrichissement ont été comparés pour calculer la modification de facteur d'enrichissement. Les résultats ont montré des augmentations significatives du facteur d'enrichissement pour les échantillons de CoV analysés, quelle que soit la méthode de transcription inverse utilisée. Cela était particulièrement vrai pour les échantillons CoVOC43_95UHR, qui ont montré un enrichissement > 640 fois pour Respiratory Virus Oligos Panel (tableau 4).

La couverture moyenne fournit un aperçu de la profondeur de couverture pour chaque base du génome de référence, identifiant ainsi les régions qui n'ont peut-être pas été séquencées. L'enrichissement avec des sondes spécifiques à la cible virale a considérablement augmenté la couverture moyenne de la séquence virale pour tous les échantillons, quelle que soit la méthode de transcription inverse utilisée (tableau 5).

Tableau 4 : Indicateurs d'enrichissement du Respiratory Virus Oligo Panel

Échantillon	Illumina TruSeq RNA		Thermo Maxima H Minus	
	CoVOC43	CoVOC43_95UHR	CoVOC43	CoVOC43_95UHR
Facteur d'enrichissement	5,5×	646×	3,1×	748×

Tableau 5 : Couverture moyenne avec Respiratory Virus Oligos Panel^a

Trousse d'ADNc	CoVOC43		CoVOC43_95UHR	
	Non enrichi	Enrichi	Non enrichi	Enrichi
Illumina TruSeq RNA	427,1	3 283,5	2,6	2 437,3
Thermo Maxima H Minus	987,0	3 861,4	4,2	3 035,0

a. Les indicateurs de couverture moyenne ont été normalisés à 1 million de lectures par échantillon.

Analyse avec IDbyDNA Explify Platform

La plateforme Explify a identifié le coronavirus ajouté (figure 5) et les cinq virus du regroupement de virus (figure 6). La plateforme Explify a fourni des séquences consensus génomiques virales, des courbes de couverture et a démontré sa capacité à détecter les coinfections avec d'autres virus, bactéries, champignons ou parasites (figure 5 et figure 6).

La plateforme Explify a également démontré que les données d'enrichissement du CoVOC43_95UHR du Respiratory Virus Oligos Panel affichaient une couverture équivalente aux données de séquençage métagénomique aléatoire (figure 7),

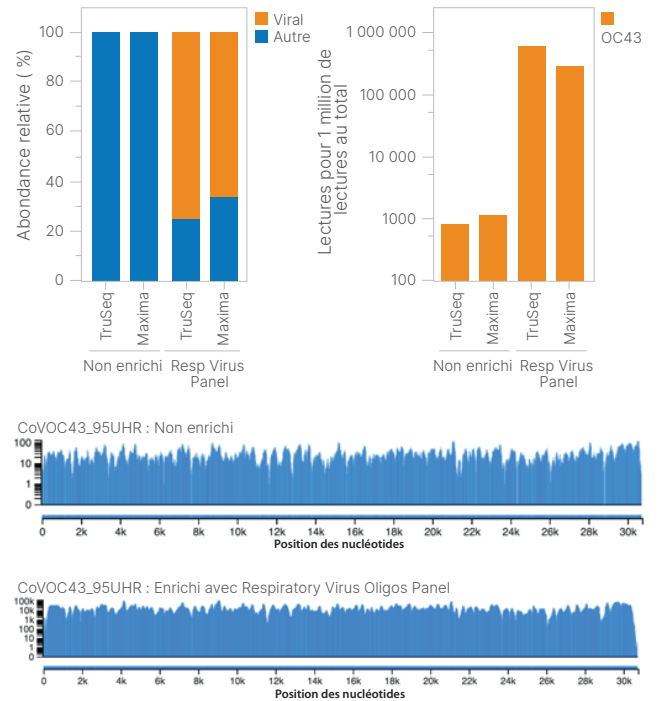
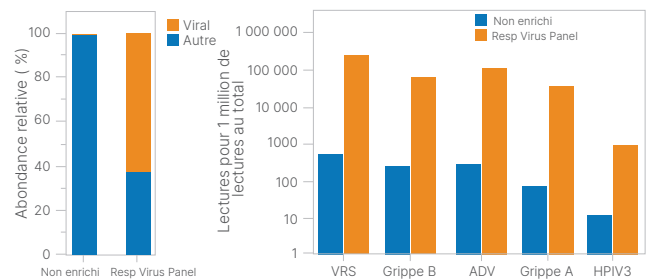


Figure 5 : Identification de OC43 – OC43 a été identifié, quelle que soit la trousse d'ADNc utilisée, avec une couverture génomique élevée. Les données ont été sous-échantillonnées à 1 million de lectures par échantillon.



MD	Alert	Pathogen	Comment	Organism Name	Evidence	Type	% Coverage	ANI	Median Depth
				Human respiratory syncytial virus B [®]	13,658	RNA	100.0%	98.8%	22,444
				Influenza A virus (H1N1) [®]	16,950	RNA	99.3%	97.6%	2,337
				Influenza B virus [®]	3,687	RNA	99.2%	98.1%	8,457
				Human adenovirus 7 [®]	198	RNA	99.1%	99.7%	3,833
				Human parainfluenza virus 3 [®]	3	RNA	85.6%	95.0%	15

Figure 6 : Caractérisation des regroupements de virus – Toutes les souches virales d'ARN et d'ADN ont été identifiées dans le regroupement de virus. Le regroupement de virus n'a pas été vérifié par des techniques orthogonales, il n'y a donc aucune garantie d'une représentation équivalente de chaque virus. Les données ont été sous-échantillonnées à 1 million de lectures par échantillon.

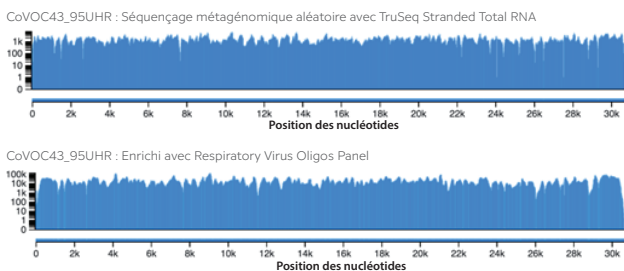
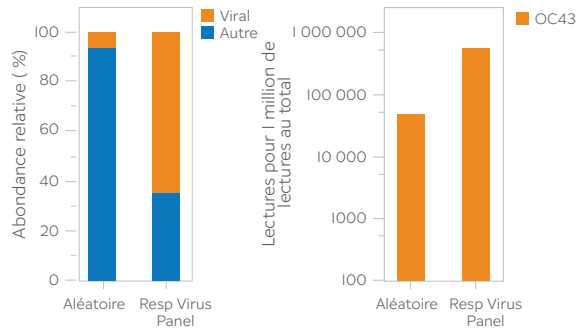


Figure 7 : Performances équivalentes avec la méthode de séquençage aléatoire – L’enrichissement a montré des performances et une couverture équivalentes à celles du séquençage métagénomique aléatoire.

démontrant que l’enrichissement avec les regroupements de sondes spécifiques à la cible virale n’introduisait pas de biais de couverture par rapport à la méthode de séquençage aléatoire.

Résumé

L’identification et la caractérisation des virus émergents sont essentielles à l’amélioration de la santé publique. Dans ces situations, le SNG est une méthode puissante de détection à grande échelle pour identifier les virus connus et émergents. L’utilisation d’Illumina DNA Prep

with Enrichment avec des panels qui ciblent les agents pathogènes viraux permet aux chercheurs d’obtenir des données génomiques qui peuvent confirmer la présence du CoV et des analyses avancées telles que le génotypage et l’analyse des variants. Sa conception « agnostique » permet une identification généralisée des virus pathogènes dans tous les types d’échantillons d’intérêt et l’utilisation d’index doubles uniques réduit le risque d’enjambement des index à partir d’échantillons multiplex. Ce flux de travail facile à suivre permet la détection et la caractérisation des épidémies d’agents pathogènes, comme le nouveau SRAS-CoV-2.

En savoir plus

[Méthodes de séquençage de virus](#)

Références

1. World Health Organization. [COVID 19 Public Health Emergency of International Concern \(PHEIC\) Global research and innovation forum: towards a research roadmap](#). Publié le 12 février 2020. Consulté le 9 septembre 2023.
2. Bulcha B. [Review on viral metagenomics and its future perspective in zoonotic and arboviral disease surveillance](#). *J Biol Agric Healthcare*. 2017;7(21):35–41.
3. Gaudin M, Desnues C. [Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases](#). *Front Microbiol*. 2018;9:2924. Publié le 27 novembre 2018. doi:10.3389/fmicb.2018.02924
4. Illumina. [Comprehensive workflow for detecting coronavirus using Illumina benchtop systems](#). Publié en 2020. Consulté le 29 août 2023.



Numéro sans frais aux États-Unis : + (1) 800 809-4566 | Téléphone : + (1) 858 202-4566
 techsupport@illumina.com | www.illumina.com

© 2023 Illumina, Inc. Tous droits réservés. Toutes les marques de commerce sont la propriété d’Illumina, Inc. ou de leurs détenteurs respectifs. Pour obtenir des renseignements sur les marques de commerce, consultez la page www.illumina.com/company/legal.html.
 M-GL-02146 FRA v1.0