

Séquençage complet du génomme entier avec Illumina Complete Long Read Prep, Human

Une performance
exceptionnelle dans
les régions complexes,
difficiles à cartographier



Introduction

Le séquençage de nouvelle génération (SNG) permet aux scientifiques de déchiffrer le génome afin de mieux comprendre la biologie. La chimie de séquençage par synthèse (SBS, Sequencing by Synthesis) éprouvée d'Illumina, utilisée de concert avec l'analyse secondaire primée DRAGEN^{MC}, permet d'obtenir des données de séquençage du génome entier (WGS, Whole-Genome Sequencing) avec une précision exceptionnelle¹. Une infime fraction des régions géniques peut pourtant bénéficier de longueurs de lecture plus importantes pour obtenir une résolution et une cartographie précise de ces régions complexes. De nombreuses solutions de séquençage à lectures longues ont toutefois des flux de travail complexes, des exigences élevées en matière d'entrées d'ADN et des résultats extrêmement variables²⁻⁵.

Illumina Complete Long Reads simplifie le séquençage à lectures longues et le rend accessible aux chercheurs en génomique. Illumina Complete Long Read Prep, Human est le premier produit qui repose sur cette chimie novatrice de lecture longue. Ce test de WGS haute performance utilise un flux de travail de SNG standard pour produire des séquences contiguës de lecture longue sur le NovaSeq^{MC} 6000 System et la série NovaSeq X (figure 1). Le protocole de préparation de la librairie en une journée, très efficace, s'adapte aisément aux études à haut débit et ne nécessite que 10 ng d'entrée d'ADN sans extraction spécifique, cisaillement, ni sélection de taille.

Cette note technique démontre les performances exceptionnelles d'Illumina Complete Long Read Prep, Human, avec une large gamme d'entrées d'ADN, de qualité d'ADN et de types d'échantillons pour un WGS complet et extrêmement précis pour l'analyse des variants germinaux.

Méthodes

L'échantillon de référence HG002/NA24385 a été obtenu auprès du consortium Genome in a Bottle (GIAB) sous forme d'ADN génomique (ADNg) purifié. Séparément, des échantillons de sang et de salive ont été obtenus et l'ADNg a été extrait à l'aide de trousse disponibles sur le marché avec une extraction standard ou de poids moléculaire élevé, en suivant les instructions du fabricant.

Préparation de librairies

Les librairies marquées ont été préparées à partir de quantités variables d'ADNg d'entrée à l'aide d'Illumina Complete Long Read Prep, Human (Illumina, référence n° 20089108). En résumé, les longs fragments d'ADN à molécule unique subissent un marquage enzymatique avec des formes uniques (ou « points de repère »). Les librairies non marquées, utilisées lors de l'analyse pour produire de longues lectures contiguës représentant le fragment à molécule unique d'origine, ont été préparées à l'aide d'Illumina DNA PCR-Free Prep, Tagmentation (Illumina, référence n° 20041794).

Séquençage

Les librairies marquées et non marquées ont été séquencées lors d'analyses distinctes sur le NovaSeq 6000 System à une longueur de lecture de 2 × 150 pb.

Analyse des données

L'assemblage et la fourniture des lectures marquées ont été effectués à l'aide de l'application Illumina Complete Long Read WGS sur BaseSpace^{MC} Sequence Hub. Le programme Integrative Genomics Viewer (IGV) a été utilisé pour l'analyse et la visualisation des données.



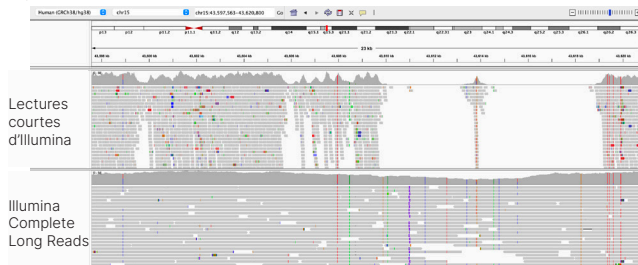
Figure 1 : Flux de travail d'Illumina Complete Long Read Prep, Human – Illumina Complete Long Read Prep, Human offre un flux de travail rationalisé composé de quatre étapes : extraction d'ADN, préparation des librairies, séquençage et analyse des données.

a. Bien qu'il soit possible d'avoir des entrées aussi faibles que 10 ng, Illumina recommande d'utiliser 50 ng d'ADN

Des données de haute qualité dans les régions complexes

ILLUMINA Complete Long Read Prep, Human fournit une couverture dans les régions où les lectures courtes ne peuvent pas être cartographiées sans ambiguïté. Cela permet le séquençage contigu de gènes et de pseudogènes associés à la maladie ayant des problèmes de cartographie connus avec des lectures courtes uniquement, notamment *STRC*, (figure 2A), *SEMG1* (figure 2B) et *SULTIA1* (figure 2C).

A. Gène *STRC*



B. Gène *SEMG1*



C. Gène *SULTIA1*



Figure 2 : La couverture complète permet une détection précise de plusieurs types de variants dans les régions complexes – ILLUMINA Complete Long Read Prep, Human permet un séquençage précis des gènes associés à la maladie dans les régions génomiques complexes, notamment (A) *STRC*, (B) *SEMG1* et (C) *SULTIA1*.

De plus, ILLUMINA Complete Long Read Prep, Human fournit des blocs de phase considérablement plus grands que la taille de gène type. Cela permet une couverture uniforme du locus d'antigène leucocytaire humain (HLA, Human Leukocyte Antigen), une région hautement polymorphe avec des niveaux élevés d'homologie de séquence, de pseudogènes, de variabilité et d'allèles. Bien qu'il soit difficile de séquençer avec précision, ILLUMINA Complete Long Read Prep, Human offre une couverture uniforme de l'ensemble de la région du HLA et permet la mise en phase du génome pour visualiser et distinguer clairement les allèles entre les chromosomes maternels et paternels (figure 3).

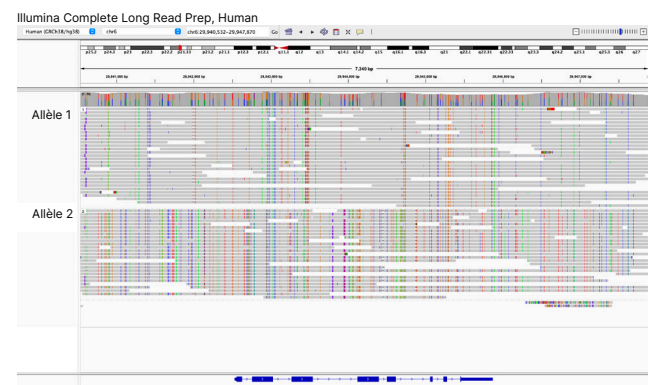


Figure 3 : Une couverture uniforme permet une mise en phase claire des haplotypes – ILLUMINA Complete Long Read Prep, Human fournit une couverture complète de la région du HLA et permet une mise en phase claire et précise des allèles dans la région du HLA.

Pour une évaluation plus complète de la précision de l'appel des variants, ILLUMINA Long Read Prep, Human a été comparé à une plateforme tierce à lectures longues et aux lectures courtes d'ILLUMINA lors du défi PrecisionFDA Truth Challenge V2. Parrainé par PrecisionFDA, le consortium GIAB et le National Institute of Standards and Technology (NIST), ce défi a été lancé pour évaluer la performance du pipeline d'appel de petits variants au sein d'un cadre de référence commun, en mettant l'accent sur les régions « difficiles à cartographier ». À l'aide de cet ensemble de données, ILLUMINA Complete Long Read Prep, Human démontre une précision exceptionnelle et fait mieux que la plateforme tierce à lectures longues en obtenant un score F1 de 99,87 % (figure 4).

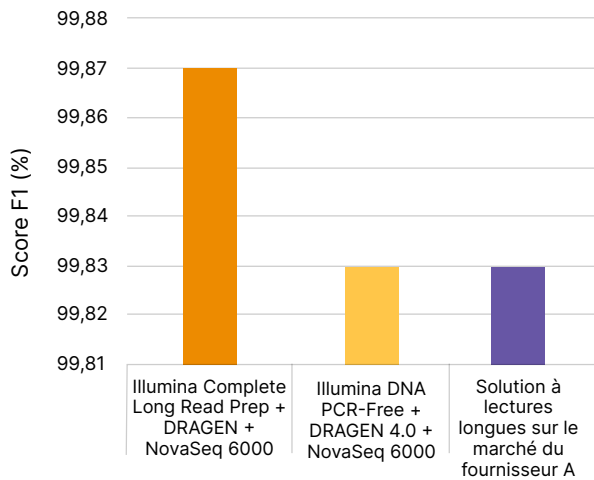


Figure 4 : Un nouvel étalon de précision – Illumina Complete Long Read Prep, Human offre une précision sans précédent pour l'appel des variants, qui est mesurée par le score F1 (%), reflétant la précision et le rappel pour le WGS¹.

Une performance fiable pour un éventail de quantités d'entrées d'ADN

La performance d'Illumina Complete Long Read Prep, Human a été évaluée sur un éventail de quantités d'entrées d'ADN. Les bibliothèques ont été préparées à partir de 10, 25, 100, 600 et 1 200 ng d'ADN en trois exemplaires. Les résultats du séquençage ont démontré que les bibliothèques préparées à partir de seulement 5 ng d'entrée affichaient des indicateurs de haute qualité, notamment la valeur N50 et le bloc de mise en phase N50 (figure 5). Bien qu'Illumina recommande une entrée d'ADN de 50 ng, des entrées inférieures peuvent être utilisées.

Prise en charge de la variation de la qualité de l'ADN

Pour simuler l'ADN endommagé, de l'ADNg de haute qualité a été soumis au cisaillement à l'aide d'un sonicateur à ultrasons Covaris pendant une ou trois secondes. La qualité de l'ADN a été évaluée à l'aide d'un analyseur de fragments (figure 6). Les bibliothèques ont été préparées à l'aide de l'entrée d'ADNg cisailé et d'Illumina Complete Long Read Prep, Human. Les résultats du séquençage ont montré que l'ADN cisailé pendant une seconde fournissait des données de haute qualité et permettait un appel des variants précis (figure 7). En outre, l'ADNg d'entrée qui a été soumis à des cycles croissants de congélation et de décongélation a produit des données de haute qualité (figure 8).

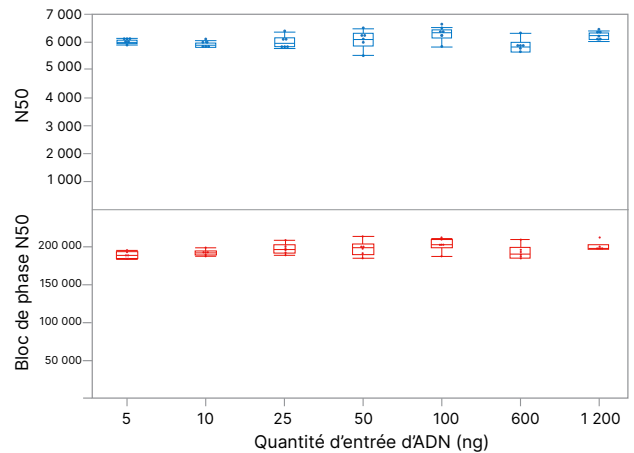


Figure 5 : Performance de haute qualité sur un vaste éventail d'entrées d'ADN – Illumina Complete Long Read Prep, Human produit des bibliothèques de haute qualité préparées à partir d'entrées d'ADN de 5 ng à 1 200 ng (en trois exemplaires) et génère une qualité de données similaire pour la valeur N50 et le bloc de phase N50. N50 est définie en tant que longueur de séquence du contig le plus court à 50 % de la longueur totale de l'ensemble. Elle peut également être utilisée en tant que mesure des blocs de phase.

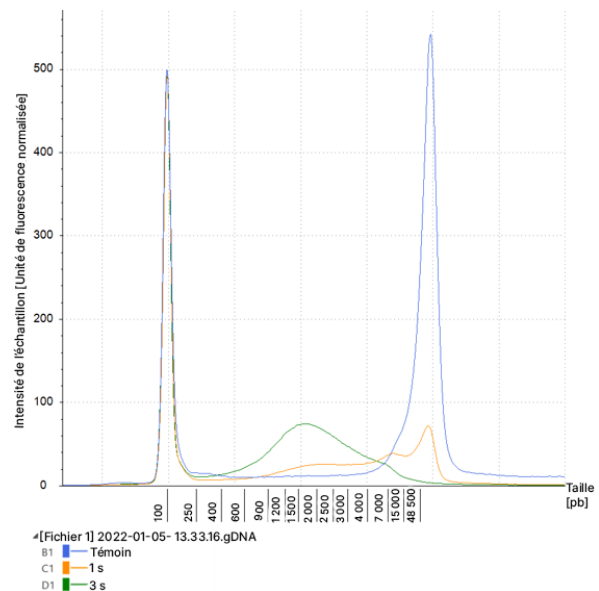


Figure 6 : Qualité de l'ADN cisailé – L'analyse de l'ADN sur un analyseur de fragments montre une réduction de la qualité et une augmentation de la durée de cisaillement. Les échantillons d'ADN ont été soumis au cisaillement de Covaris pendant 1 ou 3 secondes et la qualité a été mesurée à l'aide d'un analyseur de fragments.

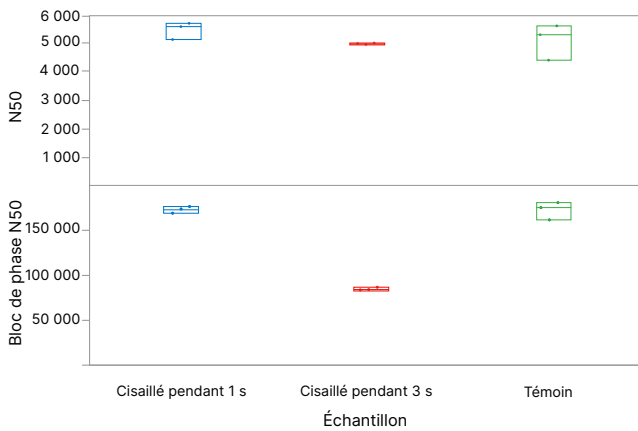


Figure 7 : Indicateurs de séquençage de l'ADN cisailé – L'ADN cisailé pendant 1 et 3 secondes a été utilisé comme entrée pour Illumina Complete Long Read Prep, Human. Les bibliothèques obtenues ont généré des données de séquençage à lectures longues de haute qualité, telles que mesurées par la valeur N50 et le bloc de phase N50.

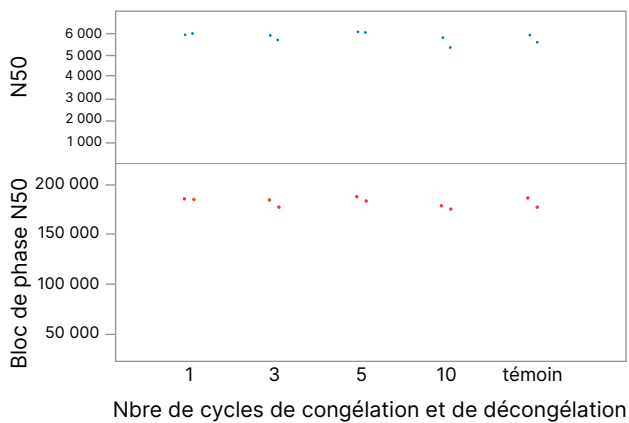


Figure 8 : Indicateurs de séquençage de l'ADN soumis à des cycles de congélation et de décongélation – L'ADN soumis à 1, 3, 5 et 10 cycles de congélation et de décongélation a été utilisé comme entrée pour Illumina Complete Long Read Prep, Human. Les bibliothèques obtenues ont généré des données de séquençage à lectures longues de haute qualité, telles que mesurées par la valeur N50 et le bloc de phase N50.

Prise en charge de différents types d'échantillons

Illumina Complete Long Read Prep, Human prend en charge divers types d'échantillons, notamment le sang et la salive. Une fois l'extraction effectuée à l'aide d'une trousse standard ou de poids moléculaire élevé (HMW, High Molecular Weight), la qualité de l'ADN a été évaluée à l'aide d'un analyseur de fragments (figure 9). Les résultats du séquençage ont montré que l'ADN du sang et de la salive produisait des données de haute qualité (figure 10). Il convient de noter qu'il n'y avait aucune différence de performance entre l'ADN extrait avec une trousse standard et une trousse HMW (figure 10).

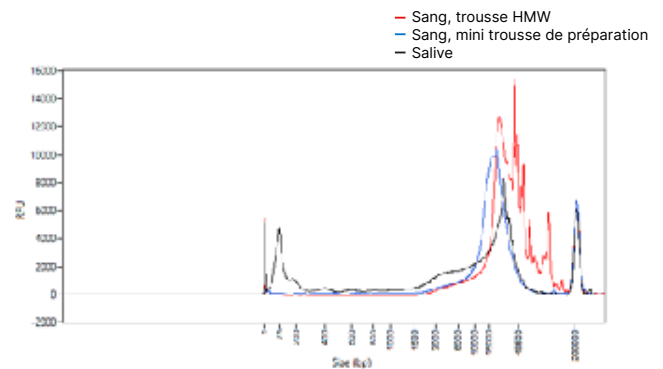


Figure 9 : Qualité de l'ADN du sang et de la salive – L'ADN extrait du sang (HMW, mini) ou de la salive a été analysé à l'aide d'un analyseur de fragments.

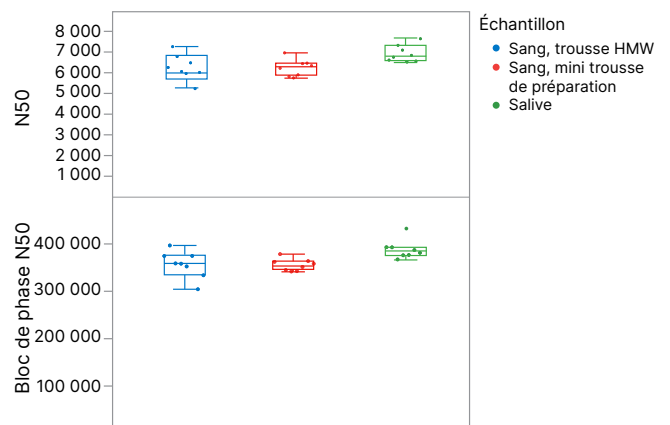


Figure 10 : Performances de haute qualité avec différents types d'échantillons – Les bibliothèques d'Illumina Complete Long Read Prep, Human générées à partir de sang (HMW, mini) et de salive fournissent des données de haute qualité, telles que mesurées par la valeur N50 et le bloc de phase N50.

Résumé

La chimie de séquençage à lectures longues fournit des renseignements supplémentaires sur les difficultés des régions du génome difficiles à surmonter avec le SNG à lectures courtes. Illumina Complete Long Reads simplifie le séquençage à lectures longues et le rend accessible aux chercheurs en génomique. Illumina Complete Long Read Prep, le premier produit basé sur cette nouvelle chimie, offre un flux de travail familier rationalisé et une synergie avec la chimie de SBS d'Illumina et l'analyse DRAGEN. Cette chimie hautement novatrice permet d'effectuer un séquençage à lectures courtes et longues sur un seul instrument, rendant ainsi le SNG à lectures longues accessible aux laboratoires de génomique. Illumina Complete Long Read Prep, Human démontre des performances fiables avec de l'ADN de quantité et de qualité variables provenant de différentes sources d'échantillons. Il en résulte une solution de WGS humain hautement évolutive et précise.

En savoir plus

[Illumina Complete Long Read Prep, Human](#)

Références

1. Mehio R, Ruehle M, Catreux S, et al. DRAGEN Wins at Precision- FDA Truth Challenge V2 Showcase Accuracy Gains from Alt-aware Mapping and Graph Reference Genomes. illumina.com/science/genomics-research/dragen-wins-precisionfda-challenge-showcase-accuracy-gains.html. Consulté le 12 janvier 2023.
2. Pacific Biosciences. Preparing DNA for PacBio HiFi sequencing—Extraction and quality control. pacb.com/wp-content/uploads/Technical-Note-Preparing-DNA-for-PacBio-HiFi-Sequencing-Extraction-and-Quality-Control.pdf. Publié en 2022. Consulté le 12 janvier 2023.
3. Pacific Biosciences. Preparing whole genome and metagenome libraries using SMRTbell prep kit 3.0. pacb.com/wp-content/uploads/Procedure-checklist-Preparing-whole-genome-and-metagenome-libraries-using-SMRTbell-prep-kit-3.0.pdf. Publié en 2022. Consulté le 12 janvier 2023.
4. Oxford Nanopore Technologies. Ligation Sequencing Kit. store.nanoporetech.com/us/ligation-sequencing-kit110.html. Consulté le 12 janvier 2023.
5. Pacific Biosciences. Low Yield Troubleshooting Guide. pacb.com/wp-content/uploads/Guide-Low-Yield-Troubleshooting.pdf. Publié en 2018. Consulté le 12 janvier 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-01421 FRA v1.0