

# Conception de panel personnalisé pour Illumina Complete Long Read Prep with Enrichment, Human

Enrichissement ciblé à lectures  
longues hautement flexible pour  
les génomes humains

## Introduction

Lors de la réalisation du séquençage du génome entier (WGS, Whole-genome Sequencing) humain, plusieurs régions peuvent être difficiles à cartographier avec seulement des lectures courtes. Le séquençage à lectures longues peut compléter les données de WGS à lectures courtes standard pour aider à traiter ces régions complexes. La technologie Illumina Complete Long Reads utilise un flux de travail de séquençage de nouvelle génération (SNG) standard pour produire des séquences contiguës de lectures longues sur les systèmes de séquençage d'Illumina avec un seul pipeline d'analyse (figure 1)<sup>1-3</sup>. Illumina Complete Long Read Prep with Enrichment, Human ajoute une approche ciblée pour un séquençage à lectures longues plus rentable\*. La chimie d'enrichissement d'Illumina Complete Long Read offre une grande flexibilité avec les cibles et la conception des sondes pour aider à surmonter les difficultés posées par les régions difficiles à cartographier ou fournir des renseignements supplémentaires à partir du séquençage par mise en phase.

## Conception de panels de sondes d'enrichissement pour les lectures longues

Illumina Complete Long Read Prep with Enrichment, Human utilise une stratégie de conception de sonde différente pour capturer des fragments plus longs (~ 7 à 10 kb) que l'approche généralement utilisée pour capturer des fragments courts (~ 200 à 500 pb).

\* Requierit ≥ 30× plus de données de WGS à lectures courtes standard du même échantillon pour l'analyse. Il est possible d'utiliser les fichiers FASTQ d'un échantillon d'analyse antérieur.

Le logiciel DesignStudio<sup>MC</sup> d'Illumina est un outil gratuit et convivial pour la conception de panels de sondes d'enrichissement. L'algorithme DesignStudio prend en compte la teneur en GC, la spécificité du résultat et l'espacement des sondes, c'est-à-dire le nombre de sondes qui se trouveront dans la région cible. L'espacement standard pour les panels d'enrichissement à lectures courtes de 120 mer est une fenêtre de sonde de 250 à 350 pb. Pour la conception de panels d'enrichissement à lectures longues, l'espacement des sondes a été testé à plusieurs longueurs et une fenêtre d'un kilobase s'est avérée optimale pour une capture rentable et à grande efficacité.

L'efficacité de l'enrichissement par hybridation dépend fortement de la spécificité de la sonde. Le pourcentage d'enrichissement sur cible affecte directement le nombre de séquençage nécessaire pour atteindre la profondeur de couverture cible. Une spécificité élevée est plus difficile à atteindre pour les régions répétitives. Cependant, l'utilisation d'une fenêtre de sonde plus grande offre une meilleure flexibilité pour exclure les sondes peu performantes, éviter les régions de répétition (jusqu'à la taille de fenêtre de 1 kb) et maintenir l'efficacité de l'enrichissement avec moins de sondes (figure 2). L'algorithme DesignStudio peut utiliser ces considérations pour recommander le positionnement de la sonde. Les panels tiers doivent utiliser des directives similaires pour une meilleure performance et une meilleure rentabilité. L'espacement des sondes d'enrichissement standard est également entièrement compatible.

### Flexibilité dans la conception des sondes et la stratégie cible

Illumina Complete Long Read Prep with Enrichment, Human offre une grande flexibilité pour choisir et concevoir des panels de sondes personnalisés pour s'adapter aux objectifs de l'étude. Les régions ciblées individuelles peuvent couvrir des bases uniques jusqu'à des centaines de kilobases. La taille totale du panel peut aller de panels



Figure 1 : Élément d'un flux de travail intégré : accédez à des données de WGS à lectures longues ciblées et rentables à l'aide d'un protocole de préparation de bibliothèques optimisé et évolutif, d'une chimie de séquençage Illumina éprouvée et d'une analyse secondaire DRAGEN. Requierit ≥ 30× plus de données du génome entier à lectures courtes standard du même échantillon pour l'analyse. Il est possible d'utiliser les fichiers FASTQ d'un échantillon d'analyse antérieur.

personnalisés de 2,5 Mb à > 95 Mb. Les chercheurs peuvent utiliser des lectures longues ciblées pour améliorer la couverture dans des régions spécifiques connues pour avoir une faible cartographiabilité avec des données à lectures courtes. Les lectures longues peuvent également être ciblées pour couvrir des gènes entiers jusqu'à de longues régions multigéniques pour permettre la mise en phase des variants et la définition des haplotypes.

Plusieurs panels préconçus sont disponibles dans l'outil DesignStudio (tableau 1). Ces panels ciblent des gènes complexes pertinents sur le plan médical (CMRG, Challenging Medically Relevant Genes)<sup>4</sup>, des gènes fréquemment ciblés par des tests pharmacogénétiques (PGx)<sup>5-7</sup>, des gènes sur la liste des résultats secondaires de l'American College of Medical Genetics and Genomics (ACMG) (ACMG SF v3.1)<sup>8</sup> ou la région complète du complexe majeur d'histocompatibilité (CMH)<sup>9</sup>. Illumina Human Comprehensive Panel, qui cible principalement les régions discrètes à faible couverture dans les gènes codant les protéines, est également disponible en tant que panel préconçu ou préfabriqué prêt à être expédié (Illumina, référence n° 20113836)<sup>10,11</sup>. Le logiciel DesignStudio prend en charge la conception de panels personnalisés à partir de fichiers BED<sup>†</sup> ou la modification de préconceptions existantes.

### Profondeur de séquençage recommandée pour les panels de sondes personnalisés

Illumina Complete Long Read Prep with Enrichment, Human fournit des performances hautement uniformes et fiables. Pour les panels préconçus testés, une performance optimale a été obtenue avec environ 1,5 Gb de données de séquence (environ 5 millions de lectures appariées) pour un panel cible de 1 Mb (figure 3). Pour les nouveaux panels dont les performances sont inconnues, il est recommandé d'utiliser 3 Gb de données de séquence (environ 10 millions de lectures appariées) pour un panel cible de 1 Mb, sous réserve d'une réduction avec une optimisation supplémentaire.

### Couverture et mise en phase de haute précision des régions difficiles

Les panels de sondes d'enrichissement à lectures longues axés sur l'amélioration de régions spécifiques à faible couverture, comme Illumina Human Comprehensive Panel et le panel CMRG, améliorent la précision de l'appel des variants dans les régions complexes ciblées (figure 4). L'enrichissement à lectures longues avec le panel CMRG permet également d'améliorer l'exhaustivité de la couverture et la détection des variants dans les régions codant les protéines (figure 5, figure 6).

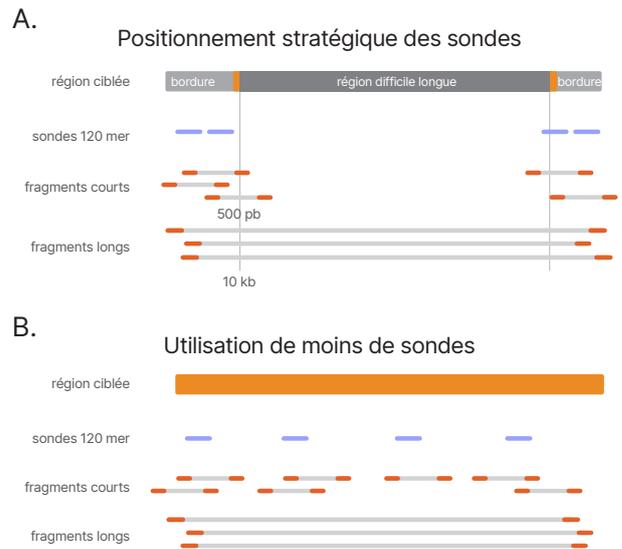


Figure 2 : L'hybridation de fragments longs augmente l'efficacité de l'enrichissement : l'hybridation de fragments longs offre des avantages par rapport à la capture de fragments courts, notamment (A) le positionnement stratégique de sondes en dehors des régions difficiles pour la conception de sondes, comme la teneur en GC extrême, la faible complexité ou les répétitions, et (B) moins de sondes sont nécessaires pour capturer chaque région cible. L'algorithme DesignStudio recherche une teneur en GC optimale dans des sections de 1 kb dans les régions cibles et les régions à la spécificité la plus élevée pour positionner les sondes.

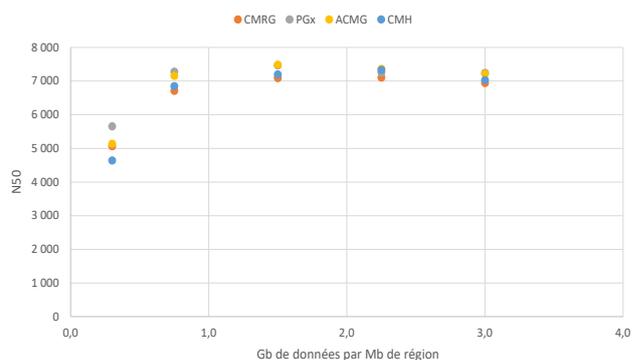


Figure 3 : Exigences de séquençage pour les panels de sondes personnalisés : le titrage des données de séquençage requises pour un N50 maximal montre que 1,5 Gb (~ 5 millions de lectures appariées) par Mb de région cible résout efficacement les régions ciblées pour la génération de données Illumina Complete Long Read.

† BED, format Browser Extensible Data.

Tableau 1 : Panels de sondes d’enrichissement préconçus pour Illumina Complete Long Read Prep with Enrichment, Human

Panels <sup>a</sup>	Panel CMRG	Panel PGx	Panel ACMG	Panel du CMH
Gènes ciblés	391 gènes pertinents sur le plan médical connus pour poser des difficultés difficiles à surmonter avec des lectures courtes <sup>5</sup>	98 gènes fréquemment ciblés par des tests pharmacogénomiques <sup>6-8</sup>	78 gènes uniques de la liste des résultats secondaires de l’ACMG (ACMG SF v3.1) <sup>9</sup>	> 140 gènes dans la région complète du CMH dans l’assemblage GRCh38.p14 <sup>10</sup>
Taille de la région cible <sup>b</sup>	22,5 Mb	8,1 Mb	7 Mb	4,9 Mb
Débit de séquençage par échantillon <sup>c</sup>	~ 67,5 Gb	~ 24,3 Gb	~ 21 Gb	~ 14,7 Gb
Nbre de sondes	~ 22 500	~ 8 200	~ 6 900	~ 5 000
N50 <sup>d</sup>	6,1 kb	7,3 kb	7,3 kb	7,3 kb
Bloc de phase N50 <sup>d,e</sup>	82,8 kb	94,4 kb	94,4 kb	357 kb
Taille moyenne de la région cible <sup>e</sup>	58 kb	83 kb	88 kb	5 000 kb
Uniformité <sup>d,f</sup>	97,9 %	99,0 %	99,5 %	97,8 %
Enrichissement par lectures élargies <sup>d,f</sup>	80,1 %	79,3 %	66,3 %	67,5 %
% de SNV hétérozygotes mis en phase <sup>d</sup>	98,9 %	98,9 %	99,6 %	98,6 %

a. CMRG, gènes complexes pertinents sur le plan médical; PGx, pharmacogénomique; ACMG, American College of Medical Genetics and Genomics; CMH, complexe majeur d’histocompatibilité.  
 b. La taille de la région cible est la somme des longueurs d’emplacement de sonde élargies, fusionnées là où elles se chevauchent.  
 c. Requiert une analyse de séquençage de 2 × 150 pb et 5 à 10 millions de lectures appariées (~ 1,5 à 3 Gb de données) par Mb de région cible, générant environ 30× la couverture finale d’Illumina Complete Long Reads. Les exigences de données par échantillon pour les panels personnalisés ne sont qu’un point de départ recommandé. Les utilisateurs peuvent optimiser les données attribuées en fonction des performances du panel.  
 d. Données générées à l’aide de 50 ng d’ADN génomique HG002 (Coriell, référence n° NA24385). Les performances peuvent varier en fonction de l’entrée d’ADN et de la qualité de l’échantillon.  
 e. Les tailles des blocs de phase sont limitées aux tailles des régions cibles contiguës individuelles.  
 f. Uniformité de la couverture calculée en % > 0,2× la moyenne. Enrichissement par lectures élargies calculé comme 100 \* (lectures élargies alignées sur la cible/nombre total de lectures alignées).

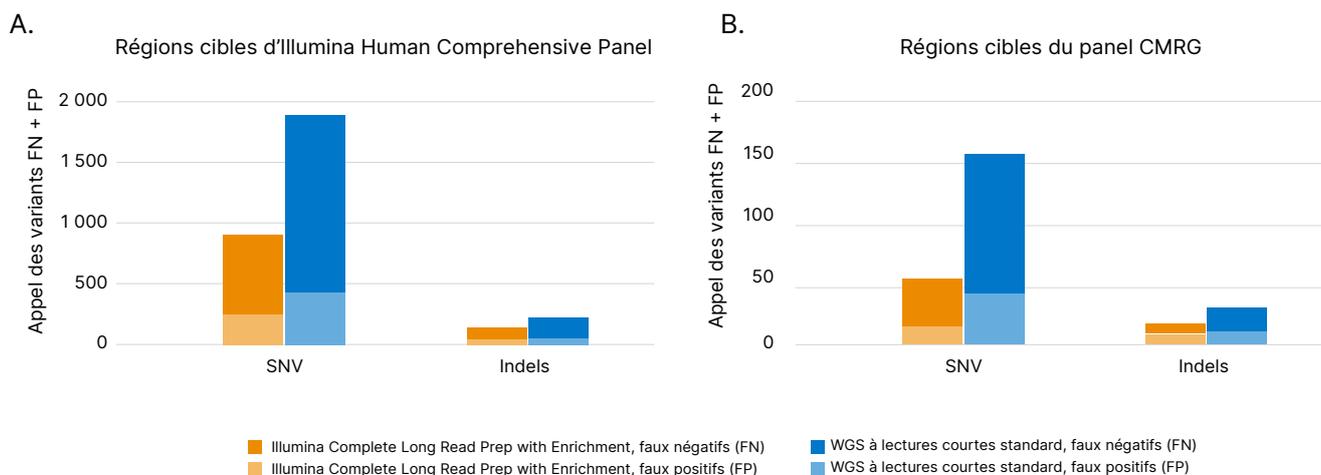


Figure 4 : Lectures longues ciblées pour améliorer la précision de l’appel des variants dans les régions complexes : appels des variants faux négatifs (FN) et faux positifs (FP) pour les variants mononucléotidiques (SNV, Single Nucleotide Variant) et les insertions/suppressions (indels) dans les régions géniques HG002 ciblées par (A) Human Comprehensive Panel ou (B) le panel CMRG, à l’aide d’Illumina Complete Long Read Prep with Enrichment (orange) par rapport au WGS à lectures courtes standard (bleu).



Figure 5 : Les lectures longues ciblées améliorent les régions à faible couverture : les tracés d’Integrative Genomics Viewer (IGV) du séquençage à lectures longues de *HBG1* à l’aide du WGS d’Illumina Complete Long Read Prep, Human (en haut) et Illumina Complete Long Read Prep with Enrichment, Human et du panel CMRG (au milieu) par rapport au WGS à lectures courtes standard (en bas).

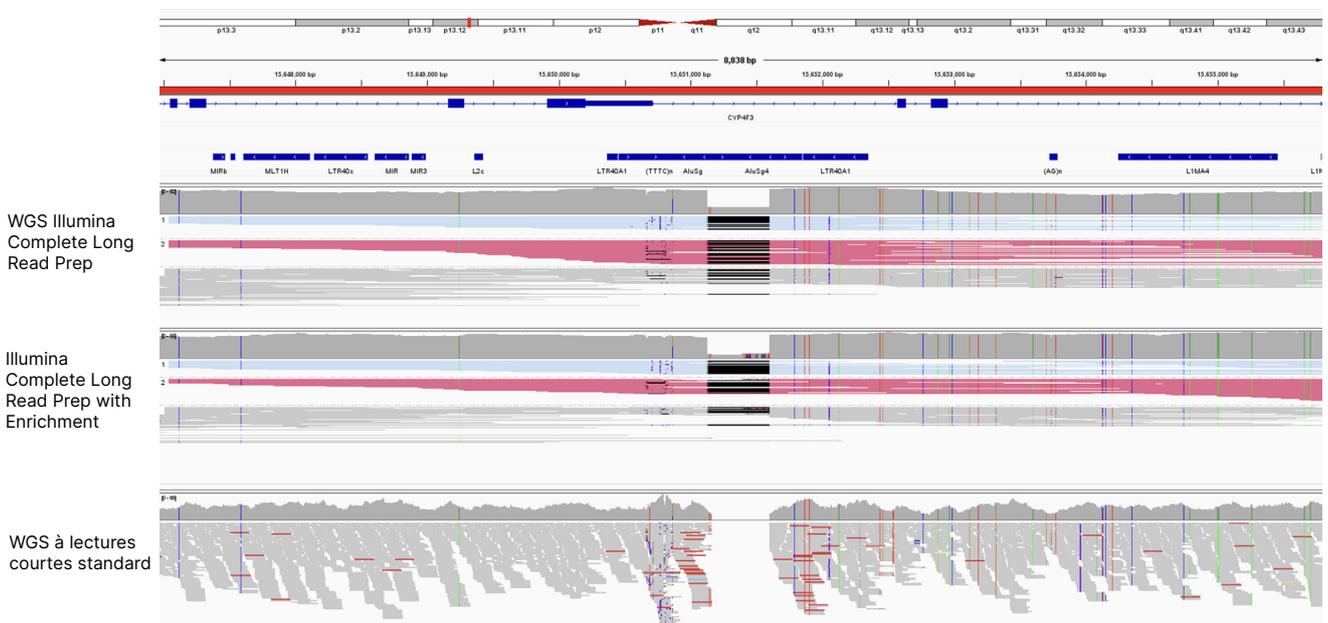


Figure 6 : Résolution claire des limites de suppression avec des lectures longues ciblées : tracés d’IGV à partir du séquençage à lectures longues et de la mise en phase de *CYP4F3* à l’aide du WGS d’Illumina Complete Long Read Prep, Human (en haut) et Illumina Complete Long Read Prep with Enrichment, Human et du panel CMRG (au milieu) par rapport au WGS à lectures courtes standard (en bas). Allèle 1 en bleu, allèle 2 en rose.

### Blocs de phase longs pour définir les haplotypes

Le bloc de phase N50+ de chaque panel est associé à la longueur contiguë des régions cibles (figure 7, tableau 1). Les panels CMRG, PGx et ACMG sont conçus pour cibler les gènes d'intérêt pleine longueur et ont généré un bloc de phase N50 moyen d'environ 80 à 95 kb pour permettre la mise en phase complète des allèles hétérozygotes (figure 8). Le panel du CMH cible une seule région contiguë d'environ 4,9 Mb et a généré un bloc de phase N50 moyen de plus de 350 kb pour permettre la résolution de la région génique pleine longueur (figure 9).

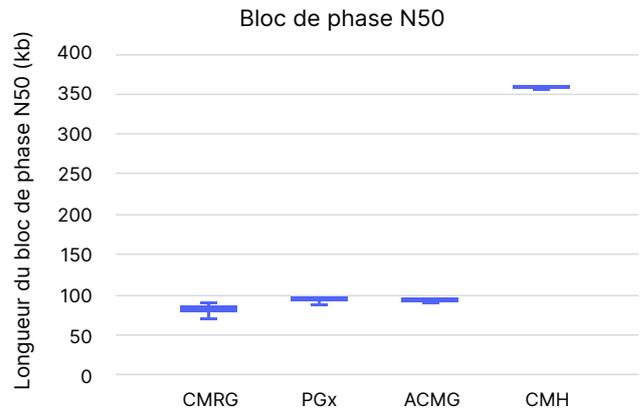


Figure 7 : Le bloc de phase N50 dépend de la longueur des régions cibles contiguës : les panels CMRG, PGx et ACMG ciblent les gènes d'intérêt pleine longueur et ont généré un bloc de phase N50 moyen d'environ 80 à 95 kb. Le panel du CMH cible la région génique du complexe majeur d'histocompatibilité et a généré un bloc de phase N50 moyen de plus de 350 kb. La taille moyenne de la région cible est de 58 kb pour le panel CMRG, de 83 kb pour le panel PGx, de 88 kb pour le panel ACMG et de 5 000 kb pour le panel du CMH.

‡ Le bloc de phase N50 reflète la longueur du bloc le plus court de la séquence contiguë à 50 % de la longueur totale de l'ensemble des régions cibles.

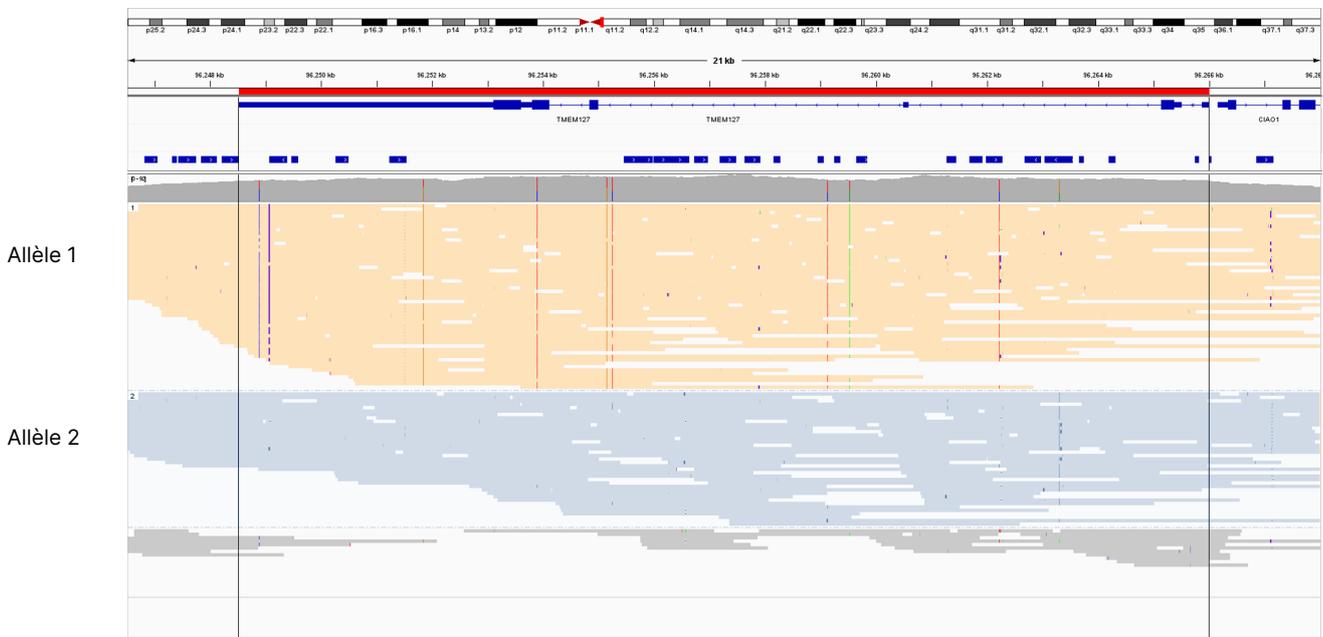


Figure 8 : Les lectures longues ciblées permettent la mise en phase de régions avec des SNV hétérozygotes : les tracés d'IGV du séquençage à lectures longues montrent la mise en phase complète dans un bloc de phase pour *TMEM127*, un gène de 21 kb, à l'aide d'Illumina Complete Long Read Prep with Enrichment, Human et du panel ACMG. Allèle 1 en jaune. Allèle 2 en bleu.

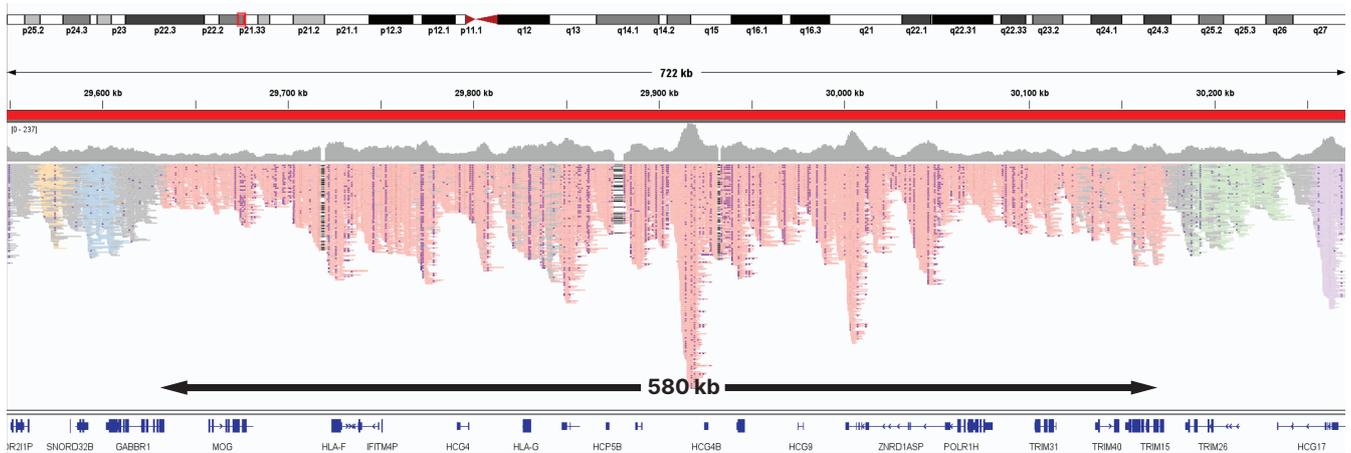


Figure 9 : Les lectures longues ciblées aident à définir les haplotypes dans les gènes polymorphiques : tracés d’IGV du séquençage à lectures longues à l’aide d’Illumina Complete Long Read Prep with Enrichment, Human. Mise en phase sur une région de 722 kb dans le locus du CMH. Une région de 580 kb (rose) est encapsulée dans un bloc de phase.

## Résumé

Illumina Complete Long Read Prep with Enrichment, Human complète le WGS à lectures courtes éprouvé d’Illumina et concentre le séquençage à lectures longues là où il offre la plus grande valeur. Les chercheurs ont une grande flexibilité pour choisir des panels préconçus ou utiliser l’algorithme DesignStudio pour concevoir des panels personnalisés pour un enrichissement ciblé à lectures longues. Les panels de sondes d’enrichissement peuvent être ciblés pour améliorer la couverture ou ajouter des renseignements par la mise en phase des gènes entiers pour un WGS rentable et hautement précis avec une solution de flux de travail complète.

## En savoir plus

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

[Outil de conception de tests DesignStudio](#)

[Technologie de séquençage à lectures longues](#)

## Références

1. Illumina. Illumina Complete Long Read Prep, Human data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf). Publié en 2022. Consulté le 22 septembre 2023.
2. Illumina. Comprehensive whole-genome sequencing with Illumina Complete Long Read Prep, Human technical note. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf](https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf). Publié en 2022. Consulté le 22 septembre 2023.
3. Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. <https://www.illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html>. Publié en 2023. Consulté le 22 septembre 2023.
4. Wagner J, Olson ND, Harris L, et al. Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat Biotechnol.* 2022;40(5):672-680. doi:10.1038/s41587-021-01158-1
5. PharmGKB. VIPs: Very Important Pharmacogenes. [pharmgkb.org/vips](https://www.pharmgkb.org/vips). Consulté le 22 septembre 2023.
6. National Library of Medicine. GTR: Genetic Testing Registry. Precision HealthPGx Panel (25 Genes). [ncbi.nlm.nih.gov/gtr/tests/593428/](https://www.ncbi.nlm.nih.gov/gtr/tests/593428/). Mis à jour le 29 novembre 2022. Consulté le 22 septembre 2023.

7. Pratt VM, Everts RE, Aggarwal P, et al. Characterization of 137 Genomic DNA Reference Materials for 28 Pharmacogenetic Genes: A GeT-RM Collaborative Project. *J Mol Diagn.* 2016;18(1):109-123. doi:10.1016/j.jmoldx.2015.08.005
8. Miller DT, Lee K, Abul-Husn NS, et al. ACMG SF v3.1 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2022;24(7):1407-1414. doi:10.1016/j.gim.2022.04.006
9. Kulski JK, Suzuki S, Shiina T. Human leukocyte antigen super-locus: nexus of genomic supergenes, SNPs, indels, transcripts, and haplotypes. *Hum Genome Var.* 2022;9(1):49. doi:10.1038/s41439-022-00226-5
10. Bekritsky MA, Colombo C, Eberle MA. Identifying genomic regions with high quality single nucleotide variant calling. Publié en 2021. Consulté le 30 août 2023.
11. Illumina. Illumina Human Comprehensive Panel data sheet. [illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf](https://illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf). Publié en 2024. Consulté le 26 janvier 2024.



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

© 2024 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](http://www.illumina.com/company/legal.html).  
M-GL-02189 FRA v1.0