

FICHE PRODUIT : SÉQUENÇAGE D'ADN GENOMIQUE

La plateforme propose le séquençage ou le re-séquençage, soit de génomes complets, soit de régions ciblées par capture de régions d'intérêt. La plateforme possède une expertise dans l'analyse bioinformatique pour la détection de variations génétiques par rapport à un génome de référence.

1 Protocoles de préparation de bibliothèques et options de séquençage proposés

1.1 Protocoles de préparation de bibliothèques

Plusieurs protocoles de préparation de bibliothèques différents sont actuellement proposés par la plateforme. Le choix du protocole le plus adapté dépend principalement de la quantité d'ADN disponible et du design expérimental du projet, comme indiqué dans le tableau ci-dessous.

#	Intitulé de la prestation	Kit utilisé par la plateforme	Quantité d'ADN génomique		Taille de la capture
			Mini-male	Opti-male	
1	Whole genome / standard quantity	TruSeq Nano Illumina	100 ng	1 µg	-
2	Whole genome / low input	Microplex Diagenode	2 ng	20 ng	-
3	Human Exome / Twist	Human Comprehensive Exome Twist Bioscience	50 ng	500 ng	36,8 Mb
4	Human Exome / Agilent	Sureselect XT Human Exome Agilent	100 ng	1 µg	35,1 Mb
5	Custom Capture	Plusieurs protocoles possibles. Nous contacter	100 ng	1 µg	Dépend du design

La plateforme possède une expertise dans l'utilisation de solutions de capture ciblées comme celles d'Agilent, Illumina, IDT, Roche et Twist Bioscience. Par exemple, pour l'exome humain, la prestation proposée actuellement en routine utilise le panel Human Comprehensive Exome de chez Twist Bioscience¹. En plus des panels existants, nous proposons des solutions de capture à la carte. Dans ce cas, le porteur de projet peut réaliser lui-même le design des oligonucléotides de capture avec l'aide du fournisseur de kit de capture de son choix ou contacter la plateforme pour obtenir des conseils et de l'aide dans cette démarche. Dans le cas d'un projet de capture à la carte et si de l'aide a été demandée pour réaliser les analyses ultérieures, un fichier BED² contenant les coordonnées des régions capturées devra être mis à disposition de la plateforme sur le LIMS (<http://ngs-lims.igbmc.fr>) dans le projet correspondant.

1.2 Options de séquençage

Les bibliothèques seront séquençées à l'aide de la technologie NextSeq 2000 d'Illumina en paired-end 2x150 pb.

¹ <https://www.twistbioscience.com/products/ngs/fixe-panels/human-comprehensive-exome>

² <https://genome.ucsc.edu/FAQ/FAQformat.html#format1>

Pour estimer le nombre de reads nécessaires à atteindre une couverture minimale cible : le calcul peut être réalisé de la façon suivante :

Nombre de reads (N) :

$$N = \frac{D \times C}{2 \times 150}$$

Où :

- D : Taille totale (en nombre de bases) des régions d'intérêt ou du génome
- C : Couverture moyenne souhaitée ou recommandée (voir tableau 1)

Tableau 1: Couverture moyenne recommandée pour une couverture minimale suffisante des régions d'intérêts.

Projet	Couverture moyenne recommandée
Exome (germline)	$\geq 100 X^*$
Exome (<i>de novo</i>)	$\geq 200 X$
Exome (FFPE)	$\geq 150 X$
Génome (germline)	$\geq 30 X^{**}$

***Exemple 1 :** pour avoir une couverture moyenne minimale de 100 X en utilisant le kit Human Comprehensive Exome de Twist Bioscience, pour un projet de détection de variants germinaux, nous conseillons de séquencer $\geq 10M$ de reads par échantillon : $(36,8 \times 10^6 \times 100) / (2 \times 150)$

****Exemple 2 :** pour avoir une couverture moyenne minimale de 30 X pour le génome humain (3Gb), nous recommandons de séquencer 300M de reads par échantillon : $(3 \times 10^9 \times 30) / (2 \times 150)$

2 Services proposés

1. La vérification des échantillons :

- Quantification et vérification de l'ADN génomique par fluorimétrie (Qubit ou Varioskan) et électrophorèse capillaire (Fragment Analyzer, Agilent), lorsque la quantité de matériel de départ le permet.

2. La préparation des librairies :

- Optionnel : Fragmentation de l'ADN selon la nature des échantillons et le protocole utilisé.
- Préparation des librairies et ligation d'adaptateurs de séquençage possédant un index. Les index sont des séquences d'ADN utilisées pour identifier chaque échantillon. L'utilisation d'index permet de séquencer plusieurs échantillons sur une même piste de flow cell.
- Quantification et vérification de la qualité des librairies par électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI).
- Optionnel : capture des fragments d'ADN couvrant les régions d'intérêt du génome.

3. Le séquençage avec le séquenceur NextSeq 2000 d'Illumina :

- Paired avec une taille de lectures de 150 pb (paired-end 2x150).

4. L'analyse primaire des données :

- Démultiplexage et création des fichiers FASTQ.
- Contrôle de la qualité des séquences.
- Détection d'éventuelles contaminations.
- Création d'un rapport synthétisant les méthodes utilisées par le pipeline d'analyse primaire et les résultats obtenus.

5. Analyse ultérieure des données (optionnelle, voir le paragraphe 6 pour plus d'informations).

3 Préparation des échantillons par le porteur de projet

Le porteur de projet réalise la préparation des échantillons d'ADN génomique (ADNg) pleine taille ou fragmenté. Le succès final de l'expérience est étroitement lié à la qualité des échantillons de départ. Un soin tout particulier doit être pris pour éviter toute trace de contamination ou de dégradation dans les échantillons.

Caractéristiques de l'ADN à fournir	
Quantité	En fonction du protocole de préparation des librairies choisi.
Volume minimal	10 µl.
Qualité	L'ADN doit être dépourvu de contaminants qui pourraient inhiber les réactions enzymatiques de préparation des librairies (protéines, EDTA, sels, solvants, etc.). Optionnel : Une purification supplémentaire, utilisant les billes AMPure XP ou SPRIselect (Beckman Coulter) peut être nécessaire, mais n'est possible qu'avec de l'ADN déjà fragmenté. Si non effectuée par le porteur de projet, cette purification sera réalisée par la plateforme avant quantification des échantillons de départ.
Conditions d'envoi	En solution dans de l'eau, expédié avec un bloc froid. Les échantillons doivent être enregistrés dans le LIMS puis l'identifiant unique généré pour chaque échantillon par le LIMS doit être clairement indiqué sur son tube.

4 Contrôles qualité réalisés par la plateforme

Les contrôles qualité listés ci-dessous sont effectués par la plateforme. Les contrôles qualités des étapes 1 et 2 sont accessibles sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>).

1. Vérification des échantillons	
Quantité (Fluorimétrie)	≥ quantité minimale requise (dépend du protocole)
Qualité	ADN génomique pleine taille : absence de signe de dégradation sur un profil d'électrophorèse. ADN fragmenté : taille moyenne ≤ 500 pb.
2. Préparation des librairies	
Profil de la librairie (électrophorèse capillaire)	Taille moyenne comprise entre 200 et 600 pb.
Pureté de la librairie (électrophorèse capillaire)	Présence minoritaire de dimères d'adaptateur (bande à 120-130 pb) et de primers (bandes à 30-60 pb).
3. Séquençage et analyse primaire des données	
Nombre total de clusters* par projet	≥ nombre total de clusters indiqué dans le paragraphe « Requested services » de la fiche de soumission (fichier pdf téléchargeable dans l'onglet « Document » du LIMS http://ngs-lims.igbmc.fr , pour chaque projet)
Scores de qualité (Phred score) > 30	≥ 85% des bases

* nombre de reads ÷ 2 en Paired-end

5 Livraison des résultats

Pour chaque échantillon, les données brutes de séquençage (séquences nucléotidiques au format FASTQ) sont mises à la disposition du porteur du projet.

Deux fichiers additionnels sont fournis par projet :

- Un rapport (fichier PDF) présentant le nombre de lectures brutes, le pourcentage de bases ayant un score de qualité Phred supérieur à 30, diverses informations sur la qualité des données et la taille de tous les fichiers bruts (FASTQ) à télécharger.
- Un fichier texte fournissant la chaîne de caractère MD5 associée à chaque fichier FASTQ à télécharger. Le porteur de projet est responsable du téléchargement de ses fichiers, de la vérification de leur intégrité à partir des chaînes de caractères MD5 et de leur stockage. Une documentation est disponible à l'adresse suivante : <http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5>.

Un e-mail de livraison des données informe le porteur de projet qu'il peut télécharger ses données de séquence en utilisant un login et un mot de passe sur le serveur FTP de la plateforme.

Conformément aux « Conditions Générales de la Plateforme GenomEast », il est rappelé qu'après livraison de ses données, le porteur de projet est responsable de leur sauvegarde et de leur archivage. Les données seront supprimées du serveur de la plateforme 6 mois après leur mise à disposition.

6 Analyse ultérieure des données (optionnelle)

L'analyse ultérieure des données n'est pas prise en charge dans la prestation standard. Néanmoins, elle peut être réalisée sous la forme d'une collaboration avec des membres de la plateforme. Elle peut comprendre :

- L'alignement des séquences sur le génome de référence.
- L'évaluation de l'efficacité de la capture.
- La recherche de variants : variation ponctuelle (Single Nucleotide Variant, SNV), petites insertions, délétions, variants structuraux, CNVs.
- L'annotation fonctionnelle des variants par rapport aux annotations du génome (p. ex. 3'UTR, exon, etc.) de même que leur impact (synonyme, non-synonyme, codon stop, variant d'épissage, etc.).
- L'annotation des variants avec des banques de données publiques comme dbSNP, 1000 genomes, Hapmap, EVS, etc.
- Priorisation des variants.

La liste ci-dessus n'est pas exhaustive et nous recommandons aux porteurs de projet qui souhaiteraient initier une collaboration avec la plateforme de nous contacter avant de commencer leur projet pour que nous puissions les aider à définir les analyses qui pourraient répondre au mieux aux questions biologiques posées.