

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. Pour cette application, nous proposons différentes alternatives, en utilisant soit la technologie de séquençage Illumina soit la technologie MGI. 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

Plusieurs protocoles de préparation de bibliothèques différents sont actuellement proposés par la plateforme. Le choix du protocole le plus adapté dépend principalement de la technologie de séquençage, de la quantité d'ADN disponible et du design expérimental du projet.

1.1 Technologie Illumina

#	Intitulé de la Prestation ⁽¹⁾	Kit utilisé par la plateforme	Quantité d'ADN génomique		Taille des régions capturées
			Minimale	Optimale	
1	Library prep DNA Truseq Nano (Illumina)	TruSeq Nano Illumina	100 ng	1 µg	-
2	Library prep ChIP (Diagenode)	Microplex Diagenode	2 ng	20 ng	-
3	Library prep DNA Exome (Twist Bioscience)	Human Comprehensive Exome Twist Bioscience	50 ng	500 ng	36,8 Mb
4	DNA-seq ciblé, à la carte	Plusieurs protocoles possibles. Nous contacter	100 ng	1 µg	Dépend du design

⁽¹⁾ Intitulés tels que référencés sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>)

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, 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 Technologie MGI

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

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

#	Intitulé de la Prestation ⁽¹⁾	Kit utilisé par la Plateforme ⁽²⁾	Quantité d'ADN Génomique		Taille des régions capturées
			Minimale	Optimale	
1	Library prep DNA PCR free (MGI)	MGIEasy FS PCR-free DNA Library Prep	50 ng	1 µg	-
2	Library prep DNA PCR free (MGI)	MGIEasy PCR-free DNA Library Prep <ul style="list-style-type: none"> • ADN fragmenté (300-500bp) • ADN pleine taille (pic ≥ 50Kb) 	80 ng 200 ng	200 ng 1 µg	-

¹ Intitulés tels que référencés sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>)

² L'intégrité des échantillons de départ va déterminer le protocole de préparation de librairie utilisé :

- #1 : ADN génomique pleine taille (bande unique de haut PM sur gel d'agarose) : « MGIEasy FS PCR-Free DNA Library Prep Set User Manual » avec fragmentation enzymatique de l'ADN génomique.
- #2 : ADN génomique montrant des signes de dégradation (présence d'un smear sous la bande de haut PM) : « MGIEasy PCR-Free DNA Lib Prep Set User Manual avec une fragmentation adaptée de l'ADN génomique par sonication au COVARIS.

2 Options de séquençage

Nous recommandons de séquencer en paired-end 2x150 pb.

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)	≥ 100 X*
Exome (de novo)	≥ 200 X
Exome (FFPE)	≥ 150 X
Génome (germline)	≥ 30 X**

***Exemple 1** : pour avoir une couverture moyenne minimale de 100 X en utilisant le kit Human Comprehensive Exome de Twist Bioscience de 36,8Mb, pour un projet de détection de variants germinaux, nous conseillons de séquencer ≥ 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 : $(3 \times 10^9 \times 30) / (2 \times 150)$

3 Services proposés

1. La vérification des échantillons :
 - Quantification de l’ADN génomique par fluorimétrie (Qubit ou Varioskan) et vérification de l’intégrité par électrophorèse capillaire (Fragment Analyzer, Agilent), lorsque la quantité de matériel de départ le permet.
2. La préparation des bibliothèques :
 - Optionnel : Fragmentation de l’ADN selon la nature des échantillons et le protocole utilisé.
 - Préparation des bibliothèques 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 bibliothèques par fluorimétrie (Qubit ou Varioskan) ou électrophorèse capillaire (Bioanalyzer d’Agilent ou Fragment Analyzer d’Agilent).
 - Optionnel : capture des fragments d’ADN couvrant les régions d’intérêt du génome.
3. Le séquençage avec le NextSeq 2000 d’Illumina ou le DNBSEQ G400-RS de MGI :
 - 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).

4 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 générales de l’ADN à fournir	
Quantité	Dépend du protocole de préparation des bibliothèques choisi.
Volume minimal	10 µl.
Qualité	<p>Dans des conditions optimales, l’ADN doit présenter des ratios A260/280>1.8 et A260/230>1.7</p> <p>ADN génomique pleine taille : absence de signe de dégradation sur un profil d’électrophorèse.</p> <p>ADN fragmenté : taille moyenne ≤ 500 pb.</p> <p>L’ADN doit être dépourvu de contaminants qui pourraient inhiber les réactions enzymatiques de préparation des bibliothèques (protéines, EDTA, sels, solvants, etc.).</p> <p>Optionnel : Une purification additionnelle, utilisant les billes AMPure XP ou SPRI-select (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.</p>

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.
--------------------	--

5 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>).

5.1 Technologie Illumina

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

5.2 Technologie MGI

1. Vérification des échantillons	
Quantité (Fluorimétrie)	≥ 2 µg
Qualité (Electrophorèse capillaire et spectrophotométrie)	ADN génomique pleine taille : absence de signe de dégradation sur un profil d'électrophorèse pour une fragmentation enzymatique Ratio A260/280 > 1.8 et A260/230 > 1.7
2. Préparation des librairies individuelles	
Profil des ADNs (Electrophorèse capillaire)	300 bp ≤ Pic ≤ 800 bp
3. Préparation des DNA nanoballs (DNB) par piste de séquençage	
Quantité des librairies circulaires simple brin (fluorimétrie)	≥ 75, 60 et 30 fmol à partir de 200~1000ng, 100~200ng, 50~100ng d'ADN génomique de départ
Concentration des nanoballs d'ADN (fluorimétrie)	≥ 8 ng/µl
4. Séquençage et analyse primaire des données	

Nombre total de reads/piste (i.e. nombre de reads ÷ 2 en Paired-end)	≥ 400 millions
Scores de qualité (Phred score) > 30	≥ 80% des bases

6 Livraison des résultats

Un e-mail sera envoyé au porteur de projet pour l'informer qu'il peut télécharger ses données sur le serveur sftp/https de la plateforme³, en utilisant le login et le mot de passe dédiés à son projet, indiqués sur l'interface web de la plateforme (<https://ngs-lims.igbmc.fr>). Si le porteur de projet a ajouté des collaborateurs pour son projet, ce mail sera également envoyé à ces collaborateurs, qui auront également accès à ces login et mot de passe pour ce projet.

Les fichiers suivants seront mis à disposition :

- Les fichiers de séquences au format FASTQ.
- Un rapport présentant les méthodes utilisées par le pipeline d'analyse primaire de la plateforme et les résultats obtenus.
- Un fichier texte fournissant l'empreinte numérique MD5 associée à chaque fichier FASTQ à télécharger. Le porteur de projet devra utiliser ces empreintes pour vérifier l'intégrité des fichiers après leur téléchargement⁴.

Conformément aux « Conditions Générales de la Plateforme GenomEast », il est rappelé qu'après la livraison de ses données, le porteur du projet est responsable de leur sauvegarde et de leur archivage. L'accès au serveur sftp/https n'est valable que durant six mois à partir de la date de livraison des données.

Les reliquats d'échantillons et de bibliothèques seront détruits 6 mois après livraison des données brutes si non réclamés par le porteur de projet.

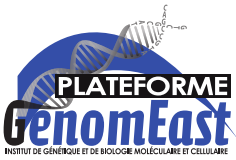
7 Analyse ultérieure des données (optionnelle)

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.

³ Une documentation est disponible sur la page suivante :
<http://genomeast.igbmc.fr/wiki/doku.php?id=help:downloading>

⁴ Une documentation est disponible sur la page suivante :
<http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5>



GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
jost@igbmc.fr
slegras@igbmc.fr
genomeast.igbmc.fr

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.