

FICHE PRODUIT : RE-SEQUENÇAGE DE REGIONS CIBLEES

Le re-séquençage de régions ciblées combine des techniques d'enrichissement par capture d'hybrides (SureSelect d'Agilent, etc.) et de séquençage haut débit. Il permet d'analyser des régions génomiques d'intérêt (exome, gènes cibles, etc.) afin d'y détecter des variations génétiques par rapport à un génome de référence.

1 Services proposés

1. La vérification des échantillons :
 - Quantification par fluorimétrie (Qubit ou Varioskan).
 - Control qualité par électrophorèse sur un gel d'agarose.
2. La préparation des librairies :
 - Fragmentation de l'ADN par sonication (Covaris).
 - Vérification de la qualité de l'ADN fragmenté par électrophorèse capillaire (Bioanalyzer, Agilent).
 - 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).
 - Capture des fragments d'ADN couvrant les régions d'intérêt du génome.
3. Le séquençage avec le séquenceur HiSeq 4000 d'Illumina :
 - Chargement des librairies sur la flow cell et génération des clusters avec la Cbot (Illumina).
 - Séquençage double 2x100 bases (paired-end).
4. L'analyse primaire des données :
 - Démultiplexage et création des fichiers FASTQ.
 - Suppression des dimères d'adaptateur.
 - Contrôle de la qualité du séquençage.
 - Détection d'éventuelles contaminations.
 - Création d'un rapport synthétisant les méthodes utilisées par ce pipeline et des résultats obtenus (un rapport détaillé pour chaque échantillon et un rapport global plus synthétique pour chaque projet).
5. Analyse ultérieure des données (optionnelle, détaillée dans le paragraphe 6)

2 Design de la capture

La plateforme donne accès à des designs de capture pour l'exome entier ou pour des captures à la carte. Dans le cas de projet de capture à la carte, le porteur de projet peut réaliser lui-même le design avec l'aide du fournisseur de kit de capture 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 BED contenant les coordonnées des régions capturées devra être fourni à la plateforme. Le fichier correspondant pourra être mis à disposition de la plateforme sur le LIMS (<http://ngs-lims.igbmc.fr>) dans la partie « *Share documents* ».

3 Préparation des échantillons par le porteur de projet

Le porteur de projet réalise la préparation de l'ADN génomique (ADNg) et le fournit pleine taille ou fragmenté.

Caractéristiques de l'ADN à fournir	
Quantité de matériel de départ (ADNg)	3 µg quantifiés avec un fluorimètre.
Concentration	≥ 100 ng/µl.
Volume minimal	20 µl.
Conditions d'envoi	En solution dans de l'eau. Les noms des échantillons doivent être clairement identifiés sur les tubes et dans le LIMS de la plateforme.

4 Contrôles qualité réalisés par la plateforme

Les contrôles qualité listés ci-dessous sont effectués par la plateforme. Les résultats sont envoyés par e-mail au porteur de projet à l'issue de chacune des étapes. Les contrôles qualité des étapes 1 et 2 sont également accessibles sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>).

1. Vérification des échantillons	
Quantité (fluorimètre)	≥ 3 µg.
Qualité	ADN génomique pleine taille : absence de signe de dégradation sur un gel d'agarose. ADN fragmenté : taille moyenne ≤ 500 pb. Les images de l'ADN avant fragmentation sont demandées. Celles-ci peuvent être téléchargées sur le LIMS dans la rubrique « <i>share document</i> ».
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 (i.e. nombre de reads en Single-read et nombre de reads ÷ 2 en Paired-end)	≥ 250 x L millions, où L est le nombre total de lignes demandées par le porteur de projet*
Scores de qualité (Phred score) > 30	≥ 75% des bases pour des lectures de 100 b.

*Pour un projet de re-séquençage d'exome chez l'Homme (taille de capture ≈ 50 Mb) où nous multiplexons 8 échantillons par piste, nous obtenons une couverture nucléotidique moyenne de 60X dans les régions d'intérêt. Pour des tailles de capture plus petites, le nombre total de lectures par échantillon est reconstruit pour obtenir une couverture nucléotidique moyenne de 60X dans les régions d'intérêt. Les échantillons sont multiplexés en conséquence.

5 Livraison des résultats

Pour chaque échantillon, les résultats suivants sont mis à la disposition du porteur de projet :

- Les données brutes de séquençage (séquences nucléotidiques au format FASTQ ayant passé le filtre qualité et ne correspondant pas à des dimères d'adaptateur).
- Un rapport présentant les contrôles qualité de séquençage par échantillon ainsi que les statistiques d'alignement (fichier PDF).

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 et la taille de tous les fichiers bruts (FASTQ) à télécharger.
- Un fichier texte fournissant les chaînes de caractères MD5 associées à chaque fichier FASTQ à télécharger. Le porteur de projet peut utiliser ces informations pour vérifier l'intégrité des fichiers après leur téléchargement.

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 que le porteur de projet est responsable de la sauvegarde et de l'archivage de ses données. La plateforme ne s'engage à les conserver que pour une durée limitée de six 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.
- 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 : les variants détectés sont annotés 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.