

FICHE PRODUIT : SMALL RNA-SEQ

L'application de séquençage haut débit small RNA-seq permet une analyse qualitative et quantitative de l'expression des petits ARN tels que les microARN.

1 Plan d'expérience

1.1 Réplicats biologiques

Il est particulièrement important d'inclure des réplicats dans les projets (cf. Hansen et al., Nature Biotechnology 29:572-573, 2011). Nous recommandons de définir un plan expérimental aléatoire et équilibré, pour essayer au maximum de réduire les effets de lots durant la préparation des échantillons. Nous vous encourageons à nous contacter avant de commencer vos expériences pour toute question relative à votre planification expérimentale.

1.2 Multiplexage

Pour un projet de small RNA-seq, nous recommandons habituellement de multiplexer jusqu'à 20 échantillons sur une piste de séquenceur HiSeq 4000 d'Illumina.

2 Services proposés

1. La vérification des échantillons :
 - Quantification et vérification des ARN totaux par fluorimétrie (Qubit ou Varioskan) et électrophorèse capillaire (Bioanalyzer, Agilent).
2. La préparation des bibliothèques :
 - Préparation des bibliothèques et ligation d'adaptateurs de séquençage possédant un index. Les index sont des séquences d'ADN de taille supérieure à 6 nt, 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 électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI).
3. Le séquençage avec le séquenceur HiSeq 4000 d'Illumina :
 - Chargement des bibliothèques sur la flow cell et génération de clusters avec la Cbot (Illumina).
 - Séquençage simple 1x50 bases (Single-Read).
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. L'analyse ultérieure des données (optionnelle, détaillée dans le paragraphe 6).

3 Préparation des échantillons par le porteur de projet

Le porteur de projet réalise la préparation des échantillons d'ARN totaux. Le succès final de l'expérience est étroitement lié à la qualité des échantillons d'ARN totaux de départ. Un soin tout particulier doit être pris pour éviter toute trace de contamination (phénol, DEPC, ADN génomique, etc.) ou de dégradation.

La plateforme souhaite mettre en garde les porteurs de projets sur le fait que certains kits utilisant des colonnes ne conservent pas les petits ARNs. La plateforme recommande donc l'utilisation de kits de purification des ARNs qui permettent de conserver les ARNs de petites tailles (p. ex. Trizol).

Caractéristiques des échantillons à fournir	
Quantité minimale	1 µg (optimale : 2 µg).
Concentration	De 150 à 300 ng/µl.
Volume minimal	10 µl.
Qualité	DO260/DO280 ≥ 1,8. Absence de signe de dégradation sur un gel d'agarose ou ratio 28S/18S ≥ 1,6 et/ou RIN ≥ 8 sur un profil du Bioanalyzer d'Agilent.
Conditions d'envoi	En solution dans de l'eau sur de la carboglace. Les tubes 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 qualités des étapes 1 et 2 sont également accessibles sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>).

1. Vérification des échantillons	
Quantité (Fluorimétrie)	≥ 1µg.
Qualité (électrophorèse capillaire)	Ratio 28S/18S ≥ 1,6 et/ou RIN ≥ 8.
2. Préparation des librairies	
Profil de la librairie (électrophorèse capillaire)	Pic(s) d'environ 140-150 bp.
Pureté de la librairie (électrophorèse capillaire)	Présence minoritaire de dimères d'adaptateur (bande à 120-130 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	≥ 85% des bases pour des lectures de 50 b.

5 Livraison des résultats par la plateforme

Pour chaque échantillon, les résultats suivants sont mis à la disposition du porteur du 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 (au format PDF).

Deux fichiers additionnels sont fournis par projet :

- Un rapport (fichier PDF) présentant le nombre de lectures brutes, le pourcentage de lectures ayant un score de qualité Phred moyen 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 du 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 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, mais elle peut être réalisée sous la forme d'une collaboration avec des membres de la plateforme. Celle-ci peut comprendre :

- Suppression des adaptateurs en 3' de chaque séquence.
- Alignement sur un génome de référence.
- Quantification des miRNA et autres petits ARN connus en utilisant des banques de données publiques (miRBase, Rfam, etc.).
- Normalisation et analyse statistique afin de mettre en évidence les petits ARN significativement différemment exprimés entre plusieurs conditions.
- Prédiction de nouveaux miRNA

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.