

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 Protocoles de préparation de bibliothèques

Deux 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'ARN total disponible, comme indiqué dans le tableau ci-dessous.

#	Intitulé de la prestation	Kit utilisé par la plateforme	Quantité d'ARN total		Type d'ARN étudiés	Directionnel
			Minimale	Optimale		
1	Small RNA-seq / standard quantity (Truseq)	Truseq SmallRNA Sample Prep (Illumina)	1 µg	2 µg	Tous les petits ARN avec 5'P et 3'OH (la taille désirée peut être choisie par le porteur de projet)	Oui
2	Small RNA-seq / low input (NEXTFLEX)	NEXTFLEX Small RNA-Seq Kit (Bio Scientific)	10 ng	1 µg	Tous les petits ARN avec 5'P et 3'OH (la taille désirée peut être choisie par le porteur de projet)	Oui

1.3 Options de séquençage

La profondeur de séquençage, qui dépend de la finalité du projet, est à discuter avec la plateforme au moment de la soumission du projet. Nous recommandons un séquençage en Single-Read d'une longueur de 50 paires de bases. À noter que les bibliothèques small RNA-seq ont une faible diversité en bases (ATGC) ce qui nécessite l'ajout d'une bibliothèque équilibrée (PhiX Illumina) à hauteur de 10%. Le porteur de projet devra donc demander 10% de séquences en plus du nombre de séquences souhaité.

2 Services proposés

- La vérification des échantillons :
 - Quantification et vérification des ARN totaux par fluorimétrie (Qubit ou Varioskan) et électrophorèse capillaire (Bioanalyzer, Agilent).
- La préparation des bibliothèques :

- Préparation des librairies 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 librairies par électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI).
3. Le séquençage avec le séquenceur NextSeq 2000 d'Illumina :
- Séquençage simple 1x50 bases (Single-Read).
4. L'analyse primaire des données :
- Démultiplexage et création des fichiers FASTQ.
 - 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.
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	En fonction du protocole de préparation de librairies (cf. 1.2).
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 ≥ 7 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)	≥ quantité minimale requise (dépend du protocole, cf. 1.2)
Qualité (électrophorèse capillaire)	Ratio 28S/18S ≥ 1,6 et/ou RIN ≥ 7.
2. Préparation des librairies	
Profil de la librairie (électrophorèse capillaire)	Taille des pic(s) > 140 bp. (dépend de la taille des petits ARNs choisis par le porteur de projet)

Pureté de la librairie (électrophorèse capillaire)	Présence minoritaire de dimères d'adaptateur. – bande à 120-130 pb pour le protocole standard. – bande à 130-140 pb pour le protocole low input.
3. Séquençage et analyse primaire des données	
Nombre total de reads (incluant le PhiX)	≥ 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

5 Livraison des résultats par la plateforme

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 les chaînes de caractères MD5 associées 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. Les données seront supprimées du serveur 6 mois après leur mise à disposition.

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 différenciellement 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.