

FICHE PRODUIT : RNA-SEQ

Le RNA-Seq permet d'étudier qualitativement et quantitativement le transcriptome grâce au séquençage à haut débit.

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'ARN total disponible ainsi que des types d'ARN étudiés, 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 ^a
			Minimale	Optimale		
1 ^b	Stranded mRNAseq/ standard quantity (Truseq)	TruSeq Stranded mRNA Prep Illumina	200 ng	1µg	ARN polyA+ de taille > 100b	Oui
2	Stranded mRNA-seq/standard quantity (Ligation)	Illumina Stranded mRNA Prep, Ligation	25 ng	1µg	ARN polyA+ de taille > 100b	Oui
3	mRNAseq/ low input (Smarter)	SMART-Seq v4 UltraLow Input RNA kit (Clontech) + Nextera XT DNA sample preparation Kit (Illumina)	100 cellules	10 ng	ARN polyA+ de taille > 100b	Non
4 ^c	3' mRNA-seq (Lexogen)	QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina (FWD) (Lexogen)	1 ng	500 ng	ARN polyA+, extrémité 3' uniquement	Oui
5 ^d	3' mRNA-seq/ single cell	Chromium Next GEM Single Cell 3' Reagent Kits	1 cellule	1 cellule	ARN polyA+, extrémité 3' uniquement	Non
6 ^{b,c,e}	Stranded Total RNA-seq Ribo-zero/standard quantity (Truseq)	Truseq Stranded Total RNA Sample Prep Illumina	100 ng	1 µg	Tous les ARN de taille > 100b	Oui
7 ^{c,e}	Stranded Total RNA-seq Ribozero Plus (Ligation)/Standard quantity	Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus	1 ng	1 µg	Tous les ARN de taille > 100b	Oui
8 ^d	Small RNA-seq	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

^a Les protocoles directionnels conservent l'information du sens de transcription : les lectures résultantes sont dans le sens inverse comparé au sens de transcription pour tous les protocoles sauf pour les protocoles 4 et 8 pour lesquels les lectures résultantes sont dans le même sens que la transcription. Au contraire, les protocoles non directionnels ne conservent pas l'information du sens de transcription.

^b Ces protocoles (#1 et #6) peuvent être utilisés sur demande spécifique, pour continuer d'anciens projets initiés avec ces protocoles pour éviter un effet batch. Pour tous les nouveaux projets, nous recommandons l'utilisation des versions plus récentes de ces deux protocoles (#2 et #7, respectivement).

^c Ces protocoles (#4, #6 et #7) sont adaptés pour étudier des ARNs dégradés extraits à partir de tissus fixés au formol et inclus en paraffine (FFPE). A noter que la plateforme ne peut s'engager sur la qualité finale des résultats de séquençage à partir de ce type d'échantillons.

^d Nous encourageons les porteurs de projets intéressés à lire la fiche produit dédiée à cette application.

^e Pour les protocoles « Total RNA-seq » (#6 et #7), l'efficacité de la ribo-déplétion est extrêmement variable d'un échantillon à l'autre. Ainsi, la plateforme ne peut pas s'engager sur la proportion de reads correspondants à des ARNr dans les résultats finaux.

Nous recommandons de choisir le même protocole au sein d'un projet. Ainsi, si vous souhaitez comparer vos données à des jeux de données préalablement obtenus, nous vous recommandons de choisir, si possible, le même protocole.

1.2 Options de séquençage

Le tableau ci-dessous fournit des recommandations de longueur de séquençage en fonction des objectifs du projet. Ce tableau ne liste qu'un sous-ensemble des questions possibles pouvant être étudiées par RNA-seq. Nous encourageons les porteurs de projet à nous contacter afin d'obtenir plus d'informations sur ces différentes options de séquençage si nécessaire.

Finalité du projet		Recommandations		
But de l'expérience	Types d'ARN étudiés	Protocole de préparation de librairie*	Type de séquençage	Taille de lecture
Quantification de l'expression de gènes annotés	polyA+	Stranded mRNAseq/standard quantity (Ligation)	Single-read	50 pb
	tous	Stranded Total RNA-seq Ribozero Plus (Ligation)/Standard quantity	Single-read	50 pb
Analyse d'évènements d'épissage alternatif avec un génome de référence, identification de nouveaux transcrits ou assemblage de transcriptome <i>de novo</i>	polyA+	Stranded mRNAseq/standard quantity (Ligation)	Paired-end	100 pb
	tous	Stranded Total RNA-seq Ribozero Plus (Ligation)/Standard quantity	Paired-end	100 pb

* En considérant que la quantité de matériel de départ est suffisante. Si ce n'est pas le cas se référer au tableau précédent pour des protocoles alternatifs.

Ces recommandations concernent un nouveau projet. Si vous souhaitez comparer vos données avec d'autres jeux de données de RNA-seq précédemment obtenus, nous vous recommandons de conserver le même protocole.

La **profondeur de séquençage** dépend de la finalité du projet, de la nature des échantillons et du protocole de préparation de librairies. Ainsi, avec les protocoles de préparation de librairies permettant l'étude de tous

les ARN (protocoles #6 et #7), davantage de molécules d'ARN différentes seront séquencées par rapport à ce qui est réalisé avec des protocoles ciblant les ARN polyadénylés. Ainsi, davantage de lectures seront nécessaires pour obtenir la même couverture sur les ARN polyadénylés. Pour des tissus de mammifères, lorsque la finalité du projet est la quantification de l'expression des gènes annotés, nous recommandons de séquencer 10 millions de séquences par échantillon avec un protocole 3' mRNA-seq (# 4), 30 millions de séquences par échantillon avec un protocole mRNA-seq pleine-taille (#1 à #3) et 50 millions de séquences par échantillon avec un protocole Total RNA-seq (#6 et #7). Une plus grande profondeur de séquençage est nécessaire pour les expériences où la sensibilité de détection est importante, par exemple pour la découverte de nouveaux transcrits ou pour quantifier de manière précise différentes isoformes.

1.3 Plan d'expérience

Il est particulièrement important d'inclure des **réplicats** dans votre planification expérimentale (cf. Hansen et al., Nature Biotechnology 29:572-573, 2011). Nous vous recommandons de définir un plan expérimental aléatoire et équilibré, ainsi que d'essayer de réduire au maximum les effets 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.

2 Services proposés

1. L'aide au montage du projet avec un biologiste et un bioinformaticien de la plateforme
 - Aide à la planification expérimentale
 - Rappel des exigences sur les échantillons de départ
2. La vérification des échantillons :
 - Quantification et vérification des ARN totaux par fluorimétrie (Qubit ou Varioskan) et électrophorèse capillaire (Bioanalyzer, Agilent), lorsque la quantité de matériel de départ le permet.
3. La préparation de bibliothèques :
 - 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 flow cell.
 - Quantification et vérification de la qualité des bibliothèques par électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI).
4. Le séquençage avec le séquenceur NextSeq 2000 d'Illumina :
 - Simple ou pairé avec des tailles de lectures selon les options spécifiées sur le LIMS (<http://ngs-lims.igbmc.fr>) pour chaque projet.
5. 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.
6. L'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'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.

Caractéristiques des échantillons à fournir	
Quantité	En fonction du protocole de préparation des librairies.
Volume minimal	10 µl.
Qualité*	DO260/DO280 ≥ 1,8 Absence de signe de dégradation sur un gel d'agarose <i>ou</i> 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 noms des échantillons doivent être clairement identifiés sur les tubes et dans le LIMS de la plateforme.

*Pour des ARNs extraits à partir de tissus FFPE, nous recommandons d'utiliser des échantillons avec un DV200 ≥ 30% (portion des fragments d'ARN > 200 nucléotides) sur un profil du Bioanalyzer d'Agilent.

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)
Qualité (électrophorèse capillaire)	Ratio 28S/18S ≥ 1,6 et/ou RIN ≥ 7
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 120-130pb).
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 en single-read et nombre de reads ÷ 2 en paired-end

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 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 séquences 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 la livraison de ses données, le porteur du 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 sur un génome de référence en prenant en compte les lectures localisées sur des jonctions d'épissage (à cheval sur plusieurs exons).
- La quantification de l'expression des gènes en utilisant des annotations connues.
- La normalisation, l'analyse exploratoire des données et les analyses statistiques afin de mettre en évidence les gènes significativement différentiellement exprimés entre plusieurs conditions.
- Des analyses fonctionnelles.
- L'analyse d'évènements d'épissage alternatif.

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.