

## FICHE PRODUIT : RNA-SEQ

Le RNA-seq permet d'étudier qualitativement et quantitativement le transcriptome grâce au séquençage à haut débit. Pour cette application, nous proposons différentes alternatives, en utilisant soit la technologie de séquençage Illumina soit la technologie MGI.

### 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 sélectionnée, de la quantité d'ARN total disponible ainsi que des types d'ARN étudiés.

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.

#### 1.1 Technologie Illumina

| #                  | Intitulé de la prestation <sup>a</sup>                            | Kit utilisé par la plateforme  | Quantité d'ARN total |           | Type d'ARN étudiés                   | Directionnel <sup>b</sup> |
|--------------------|---|--|----------------------|-----------|--------------------------------------|---------------------------|
|                    |   |  | Mini-male            | Opti-male |                                      |                           |
| 1 <sup>c</sup>     | Library prep stranded mRNA Truseq (Illumina)                      | TruSeq Stranded mRNA Prep (Illumina)   | 200 ng               | 1 µg      | ARN polyA+ de taille > 100b          | Oui                       |
| 2                  | Library prep stranded mRNA Ligation (Illumina)                    | Stranded mRNA Prep, Ligation (Illumina)  | 25 ng                | 1 µg      | ARN polyA+ de taille > 100b          | Oui                       |
| 3                  | Library prep mRNA ultralow Smarter (Takara)                       | SMART-Seq v4 UltraLow Input RNA kit (Takara) + Nextera XT DNA sample preparation Kit Illumina) | 100 pg               | 10 ng     | ARN polyA+ de taille > 100b          | Non                       |
| 4 <sup>d</sup>     | Library prep 3'mRNA (Lexogen)                                     | QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina (FWD) (Lexogen)                             | 1 ng                 | 500 ng    | ARN polyA+, extrémité 3' uniquement  | Oui                       |
| 5 <sup>e</sup>     | Library prep 3'mRNA Single cell (10x Genomics)                    | Chromium Next GEM Single Cell 3' Reagent Kits (10X Genomics)                                   | 1 cellule            | 1 cellule | ARN polyA+, extrémité 3' uniquement  | Non                       |
| 6 <sup>c,d,f</sup> | Library prep stranded total RNA Ribozero Truseq (Illumina)        | Truseq Stranded Total RNA Sample Prep (Illumina)   | 100 ng               | 1 µg      | Tous les ARN de taille > 100b        | Oui                       |
| 7 <sup>d,f</sup>   | Library prep stranded total RNA Ribozero Plus Ligation (Illumina) | Stranded Total RNA Prep Ligation with Ribo-Zero Plus (Illumina)                                | 1 ng                 | 1 µg      | Tous les ARN de taille > 100b        | Oui                       |
| 8 <sup>e</sup>     | Library prep small RNA Truseq (Illumina)                          | Truseq SmallRNA Sample Prep (Illumina)   | 1 µg                 | 2 µg      | Tous les petits ARN avec 5'P et 3'OH | Oui                       |

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

<sup>b</sup> 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.

<sup>c</sup> Ces protocoles (#1 et #6) peuvent être utilisés sur demande spécifique, pour continuer d'anciens projets initiés avec ces protocoles. Pour tous les nouveaux projets, nous recommandons l'utilisation des versions plus récentes de ces deux protocoles (#2 et #7, respectivement).

<sup>d</sup> 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.

<sup>e</sup> Nous encourageons les porteurs de projets intéressés à lire la fiche produit dédiée à cette application.

<sup>f</sup> Pour les protocoles « Total RNA-seq » (#6 et #7), l'efficacité de la ribo-déplétion est 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.

## 1.2 Technologie MGI

| #  | Intitulé de la prestation <sup>a</sup> | Kit utilisé par la plateforme  | Quantité d'ARN total |          | Type d'ARN étudiés            | Directionnel <sup>b</sup> |
|----|--|--|----------------------|----------|-------------------------------|---------------------------|
|    |  |  | Minimale             | Optimale |                               |                           |
| 9  | Library prep stranded mRNA (MGI)       | NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB) + MGIEasy RNA Directional Library Prep Kit (MGI) | 100 ng               | 1µg      | ARN polyA+ de taille > 100b   | Oui                       |
| 10 | Library prep stranded total RNA (MGI)  | RiboPOOL (si-TOOLS Biotech) MGIEasy RNA Directional Library Prep Kit (MGI)                             | 100 ng               | 1µg      | Tous les ARN de taille > 100b | Oui                       |

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

<sup>b</sup> Les lectures résultantes sont dans le sens inverse comparé au sens de transcription.

## 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  | Recommandation minimale de séquençage |
|---|---------------------------------------|
| Quantification de l'expression de gènes annotés                                   | Single read 50 pb                     |
| Analyse d'évènements d'épissage alternatif, identification de nouveaux transcrits | Paired-end 100 pb                     |

La **profondeur de séquençage** dépend de la finalité du projet, de la nature des échantillons et du protocole de préparation de bibliothèques. Ainsi, avec les protocoles de préparation de bibliothèques permettant l'étude de tous les ARN (protocoles #6, #7 et #10), 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 #9) et 50 millions de séquences par échantillon avec un protocole Total RNA-seq (#6, #7 et #10). 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.

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.

### 3 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, Thermo Fisher) et électrophorèse capillaire (Bioanalyzer d'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 ou Fragment Analyzer, Agilent).
4. Le séquençage avec le NextSeq 2000 d'Illumina ou le DNBSEQ G400-RS de MGI :
  - Simple ou paillé 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 7 pour plus d'informations).

### 4 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 bibliothèques.   |
| Volume minimal                              | 10 µl.   |
| Qualité*                                    | DO260/DO280 ≥ 1,8<br>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. |

| Caractéristiques des échantillons à fournir |   |
|---|---|
| Conditions d'envoi                          | En solution dans de l'eau sur de la carboglace.<br>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 $\geq$ 30% (portion des fragments d'ARN $>$ 200 nucléotides) sur un profil du Bioanalyzer d'Agilent.

## 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>), les contrôles qualités de l'étape 3 sont disponibles dans le rapport fourni lors de la livraison des données (voir paragraphe 6).

### 5.1 Technologie Illumina

| 1. Vérification des échantillons                   |  |
|--|--|
| Quantité (Fluorimétrie)                            | $\geq$ quantité minimale requise (dépend du protocole)   |
| Qualité (électrophorèse capillaire)                | Ratio 28S/18S $\geq$ 1,6 et/ou RIN $\geq$ 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               | $\geq$ Nombre total de clusters indiqué dans le paragraphe « Requested services » de la fiche de soumission (fichier pdf téléchargeable dans l'onglet « Document » du LIMS <a href="http://ngs-lims.igbmc.fr">http://ngs-lims.igbmc.fr</a> , pour chaque projet) |
| Scores de qualité (Phred score) $>$ 30             | $\geq$ 85% des bases   |

\* Nombre de reads en single-read et nombre de reads  $\div$  2 en paired-end

### 5.2 Technologie MGI

| 1. Vérification des échantillons  |   |
|---|---|
| Quantité (Fluorimétrie)   | $\geq$ 100 ng                               |
| Qualité (électrophorèse capillaire)   | Ratio 28S/18S $\geq$ 1,6 et/ou RIN $\geq$ 7 |
| 2. Préparation des librairies individuelles   |   |
| Profil des ADNc double brin après ligation des adaptateurs et amplification (électrophorèse capillaire) | Taille moyenne comprise entre 300 et 400 pb |
| 3. Préparation des DNA nanoballs (DNB) par piste de séquençage  |   |
| Quantité des pools de librairies circulaires simple brin (fluorimétrie)                                 | $\geq$ 80 fmol                              |
| Concentration des nanoballs d'ADN (fluorimétrie)  | $\geq$ 8 ng/ $\mu$ l                        |
| 4. Séquençage et analyse primaire des données   |   |

|                                      |   |
|--------------------------------------|---|
| Nombre total de séquences par piste  | ≥ 350 millions (single read)<br>≥ 2x350 millions (paired-end) |
| Scores de qualité (Phred score) > 30 | ≥ 85% des bases   |

## 6 Livraison des résultats par la plateforme

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 plateforme<sup>1</sup>, 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 téléchargement<sup>2</sup>.

**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 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.

<sup>1</sup> Une documentation est disponible sur la page suivante :  
<http://genomeast.igbmc.fr/wiki/doku.php?id=help:downloading>

<sup>2</sup> Une documentation est disponible sur la page suivante :  
<http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5>