

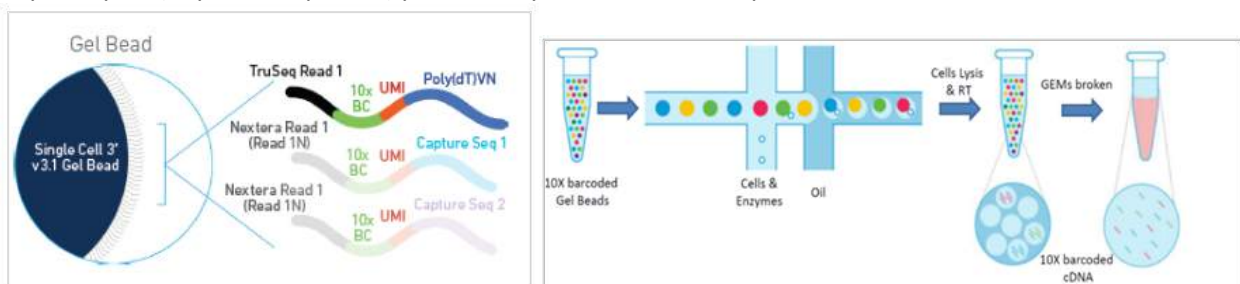
# FICHE PRODUIT : 3' mRNA-SEQ SUR CELLULES UNIQUES

La plateforme GenomEast propose une prestation de 3' mRNA-seq sur cellules uniques en combinant la technologie microfluidique de 10X Genomics et le séquençage haut débit d'Illumina. Le RNA-seq sur cellules uniques (scRNA-seq) vise à analyser le transcriptome des différentes cellules dans un échantillon donné. Il permet ainsi d'étudier l'hétérogénéité cellulaire, de caractériser, sur la base de leur profil d'expression génique, les différentes sous-populations de cellules présentes dans un échantillon.

## 1 La technologie 3'mRNA-seq de 10X Genomics

Le Chromium Controller développé par 10X Genomics est un équipement qui permet, à partir d'une suspension cellulaire, d'encapsuler individuellement de 500 à 10 000 cellules dans des gouttelettes renfermant l'ensemble des réactifs nécessaires à la lyse cellulaire et à la synthèse d'ADNc ainsi qu'une bille recouverte d'oligonucléotides spécifiques (Fig.1). Ces oligonucléotides sont constitués :

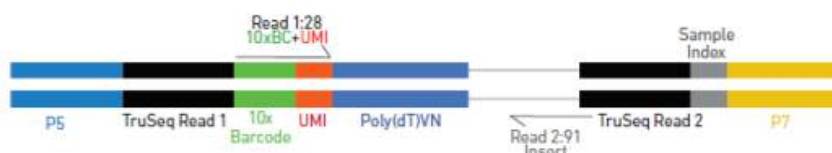
- d'une amorce PCR (Truseq Read 1 ou Nextera Read 1) ;
- d'un code-barres (BC), unique à chaque bille et donc à chaque cellule ;
- d'un Unique Molecular Identifier (UMI), une séquence aléatoire courte unique à chaque oligonucléotide sur la bille et donc à chaque transcrite ou molécule capturé(e) ;
- d'une queue poly (dT) pour la capture des ARN polyA+ des cellules cibles ou bien de séquences de capture spécifiques (Capture Seq 1 et 2) pour l'analyse simultanée de protéines de surface.



**Fig.1 : Principe de la technologie (source : 10X Genomics)**

Après isolement dans les nano gouttelettes, les cellules sont lysées par chauffage et les ARNm ainsi libérés sont retenus sur les billes par hybridation complémentaire grâce à leur queue polyA. Une réaction de transcription inverse permet de convertir les ARNm en ADNc marqués avec les codes-barres et UMI spécifiques. L'émulsion est alors brisée, les billes d'hydrogel sont dissoutes et la librairie de séquençage est finalisée sur l'ensemble des ADNc barre codés par addition des adaptateurs de séquençage et amplification par PCR.

Sur la plateforme, les librairies finales (Fig.2) sont séquençées sur un HiSeq 4000 en lecture par paires 2x100 bases. Le read 1 contenant le code-barres de la cellule et l'UMI est utilisé pour dé-multiplexer les cellules lors de l'analyse primaire. Le read 2 quant à lui est utilisé pour identifier et quantifier les transcrits.



**Fig.2 : Structure des librairies 3'mRNA-seq finales (source : 10X Genomics)**

Seule la préparation de bibliothèques 3’ mRNA-seq est proposée en prestation standard sur notre plateforme. Les applications additionnelles telles que le CITE-seq ou le cell hashing peuvent néanmoins être réalisées sous la forme de collaboration scientifique.

## 2 Planification des expériences

Toute demande de prestation doit être soumise sur l’interface web de la plateforme à l’adresse suivante : <http://ngs-lims.igbmc.fr/> en créant une nouvelle requête après avoir ouvert un compte utilisateur.

Lorsqu’il soumet sa demande, le porteur de projet doit spécifier dans la description du projet : le nombre total d’échantillons (nombre de suspensions cellulaires différentes) à traiter simultanément sur le Chromium, le nombre total de cellules qu’il souhaite charger par puits et le nombre de cellules qu’il espère capturer. Il doit également évaluer le risque biologique lié à ses échantillons de cellules et informer la plateforme des mesures de confinement appropriées (Bio Safety Level BSL1 ou BSL2). **La plateforme ne prendra en charge aucun projet impliquant des échantillons identifiés comme BSL3 ou BSL4.**

Dans la mesure du possible, le porteur de projet doit planifier la date de l’expérience avec la plateforme au minimum 2 semaines avant le run. De plus, afin de pouvoir compléter l’expérience dans les plages horaires de travail habituelles, il est souhaitable de déposer ses cellules sur la plateforme du lundi au jeudi au plus tard à 16 h.

Le taux de capture sur le Chromium est estimé à environ 65% (ie 65% des gouttelettes contiennent une cellule). Le taux de multiplets quant à lui dépend du nombre total de cellules chargées sur le Chromium. Il est conseillé de se référer au tableau 1 ci-dessous pour déterminer le nombre optimal de cellules à charger en fonction du taux de multiplets toléré et du nombre final de cellules souhaité.

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.4%	~800	~500
~0.8%	~1,600	~1,000
~1.6%	~3,200	~2,000
~2.3%	~4,800	~3,000
~3.1%	~6,400	~4,000
~3.9%	~8,000	~5,000
~4.6%	~9,600	~6,000
~5.4%	~11,200	~7,000
~6.1%	~12,800	~8,000
~6.9%	~14,400	~9,000
~7.6%	~16,000	~10,000

**Tab.1 : Taux de multiplets et nombre de cellules capturées en fonction du nombre de cellules chargées (source : 10X Genomics)**

### 3 Services proposés

1. La vérification des échantillons de départ (optionnel mais conseillé) :
  - Comptage des cellules et test de viabilité au bleu trypan.
2. La préparation de bibliothèques :
  - Préparation des bibliothèques selon les recommandations de 10X Genomics à l'aide des « Chromium Single Cell 3' Reagent kits » ;
  - Quantification et vérification de la qualité des bibliothèques finales par électrophorèse capillaire (Fragment Analyzer ou Bioanalyzer 2100 d'Agilent).
3. Le séquençage avec le séquenceur HiSeq 4000 d'Illumina :
  - Chargement des bibliothèques sur la flow cell et génération des clusters avec la Cbot ;
  - Séquençage 2x100 bases ;
  - A noter que les bibliothèques 10X Genomics sont séquencées en présence de 30% de la bibliothèque contrôle PHIX d'Illumina pour parer au manque de diversité des 25 premières bases du read 1.
4. L'analyse primaire des données :
  - Démultiplexage et création des fichiers FASTQ ;
  - Suppression des dimères d'adaptateur ;
  - 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.
5. L'analyse ultérieure des données (optionnelle, voir le paragraphe 7 pour plus d'informations).

### 4 Préparation des échantillons par le porteur de projet

Selon les recommandations de 10X Genomics, **dans des conditions optimales**, la suspension cellulaire chargée sur le Chromium doit présenter les caractéristiques suivantes :

- concentration : 700 à 1000 cellules/ $\mu$ l ;
- tampon de suspension : PBS + 0.04 % BSA ;
- taux de viabilité : >70% ;
- taille des cellules : 5 à 30  $\mu$ m.

Pour augmenter les chances d'atteindre le taux de récupération de cellules désiré, il est impératif d'éviter les agrégats de cellules et d'éliminer au maximum les débris cellulaires et les cellules mortes avant chargement sur le Chromium. Ces facteurs vont en effet impacter négativement le taux de récupération puisque les cellules mortes et les débris seront capturés au même titre que les cellules vivantes.

Différents protocoles de dissociation cellulaire et de préparation de suspensions cellulaires sont disponibles à ces deux adresses :

- <https://support.10xgenomics.com/single-cell-gene-expression/sample-prep>
- <http://www.worthington-biochem.com/tissuedissociation/default.html>

Lorsqu'il vient déposer ses cellules sur la plateforme, le porteur de projet doit fournir les informations suivantes :

- La concentration de ses cellules ;
- Leur % de viabilité ;
- Le volume total de la suspension cellulaire.

## 5 Contrôles qualité réalisés par la plateforme

Le succès final de l'expérience est étroitement lié à la précision du comptage des cellules et de la détermination de leur viabilité. Il est également important de réduire au maximum le délai entre la préparation de la suspension et son chargement sur le Chromium Controller.

Si le pourcentage de viabilité est très inférieur à 70%, en particulier si le nombre total de cellules est faible (<1000 cellules), la plateforme se réserve le droit de stopper l'expérience et d'annuler le run sur le Chromium Controller.

S'il le souhaite, le porteur de projet peut demander à la plateforme d'utiliser ses propres résultats de comptage de cellules pour le chargement sur le Chromium Controller. Dans ces conditions, la plateforme effectuera un comptage de vérification seulement après chargement sur la machine si le matériel de départ le permet.

1. Vérification des suspensions cellulaires de départ (optionnel mais conseillé)	
Nombre total de cellules à charger sur le Chromium Controller (N) (Hématimètre)	$1000 \leq N \leq 10000$ .
Volume total de cellules à charger sur le Chromium Controller (V)	$5 \mu\text{l} \leq V \leq 43.2 \mu\text{l}$ .
Viabilité des cellules (Bleu trypan)	$\geq 70\%$ .
2. Préparation des bibliothèques	
Profil de la bibliothèque (électrophorèse capillaire)	Taille moyenne comprise entre 200 et 600 pb.
Pureté de la bibliothèque (électrophorèse capillaire)	Présence minoritaire de dimères d'adaptateur (bande à 120-130pb).
3. Séquençage et analyse primaire des données	
Nombre total de clusters (i.e. nombre de reads en « single-read » et nombre de reads $\div 2$ en « paired-end »)	$\geq 200 \times L$ millions. où L est le nombre total de lignes demandées par le porteur de projet.
Scores de qualité (Phred score) > 30	$\geq 65\%$ des bases pour des lectures de 100 b.

## 6 Livraison des résultats

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).

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 la chaîne de caractère MD5 associée à chaque fichier FASTQ à télécharger. Nous conseillons au porteur de projet de vérifier l'intégrité des fichiers après leur téléchargement (pour plus d'informations sur le MD5 la page suivante peut être consultée : <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 que le porteur du projet est responsable de la sauvegarde et de l'archivage de ses données. La plateforme s'engage à les conserver que pour une durée limitée de six mois après leur mise à disposition.**

## 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'extraction des code-barres et des UMI des lectures (reads 1), l'alignement sur un génome de référence des lectures (reads 2), le filtrage des code-barres et des UMI, la génération d'une matrice gènes x code-barres, le filtrage des cellules ;
- L'analyse de l'expression des gènes dans l'ensemble des cellules : réduction de dimension, clustering, analyse d'expression différentielle ;
- L'inférence de trajectoire en utilisant ces données.

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.