

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

La plateforme GenomEast propose une prestation de 3' mRNA-seq sur cellules uniques fraîches en combinant la technologie microfluidique de 10X Genomics et le séquençage haut débit d'Illumina. La plateforme est aujourd'hui équipée de la dernière génération de Chromium X de 10X Genomics, ce qui nous permet de bénéficier de leur dernière chimie « Chromium GEM-X Single Cell Gene Expression v4 ».

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 système Chromium X permet, à partir d'une suspension cellulaire, d'encapsuler individuellement jusqu'à 20 000 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 d'une séquence de capture spécifique (Capture Seq 1) pour l'analyse simultanée d'analytes additionnels.

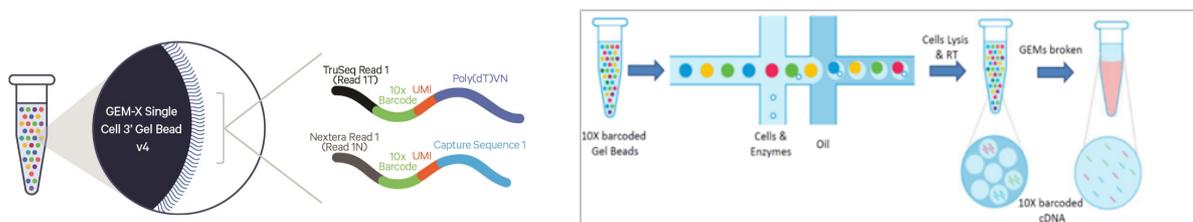


Fig.1 : Principe de la technologie (source : 10X Genomics)

Après formation des nano gouttelettes, les billes d'hydrogel sont dissoutes et les cellules sont lysées par chauffage. Les ARNm alors libérés sont retenus sur les oligonucléotides de capture 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, et la librairie de séquençage est finalisée sur l'ensemble des ADNc portant un code-barres par addition des adaptateurs de séquençage et amplification par PCR.

Sur la plateforme, les librairies finales (**Fig.2**) sont séquençées sur un NextSeq 2000 en lecture par paires. Le read 1 contenant le code-barres de la cellule et l'UMI est utilisé pour démultiplexer les cellules et quantifier les transcrits. Le read 2 quant à lui est utilisé pour identifier les transcrits.

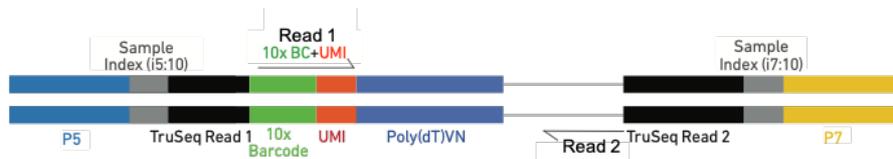


Fig.2 : Structure des bibliothèques 3'mRNA-seq finales (source : 10X Genomics)

2 Planification des expériences

Lorsqu'il soumet sa demande sur l'interface web de la plateforme (<http://ngs-lims.igbmc.fr/>), 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 X, 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 préférable de déposer ses cellules sur la plateforme du lundi au jeudi au plus tard à 16 h.

Le taux de capture sur le Chromium X est estimé à environ 70% (ie 70% des gouttelettes contiennent une cellule). Le taux de multipléts quant à lui dépend du nombre total de cellules chargées sur le système. 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 multipléts toléré et du nombre final de cellules souhaité.

Multiplet Rate (%)	# of Cells Loaded	# of Cells Recovered
~0.20%	~725	~500
~0.40%	~1,450	~1,000
~0.80%	~2,900	~2,000
~1.20%	~4,350	~3,000
~1.60%	~5,800	~4,000
~2.00%	~7,250	~5,000
~2.40%	~8,700	~6,000
~2.80%	~10,150	~7,000
~3.20%	~11,600	~8,000
~3.60%	~13,050	~9,000
~4.00%	~14,500	~10,000
~4.40%	~15,950	~11,000
~4.80%	~17,400	~12,000
~5.20%	~18,850	~13,000
~5.60%	~20,300	~14,000
~6.00%	~21,750	~15,000
~6.40%	~23,200	~16,000
~6.80%	~24,650	~17,000
~7.20%	~26,100	~18,000
~7.60%	~27,550	~19,000
~8.00%	~29,000	~20,000

Tableau 1 : Taux de multipléts 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 :
 - Comptage des cellules et test de viabilité à l'aide d'un compteur de cellules à double fluorescence.
2. Encapsulation des cellules sur le Chromium X et préparation des bibliothèques :
 - Préparation des bibliothèques à l'aide du « Chromium GEM-X Single Cell 3' Kit v4 » selon les protocoles en vigueur chez 10X Genomics;
 - Quantification et vérification de la qualité des bibliothèques finales par électrophorèse capillaire (Bioanalyzer 2100 d'Agilent).
3. Le séquençage avec le séquenceur NextSeq 2000 d'Illumina :
 - Séquençage paillé de 28 bases pour le read 1 et 85 bases pour le read 2 ;
4. L'analyse primaire des données :
 - Démultiplexage et création des fichiers FASTQ ;
 - Contrôle de la qualité des séquences ;
 - 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 X doit présenter les caractéristiques suivantes :

- Tampon de suspension : PBS Ca²⁺, Mg²⁺ free + 0.04 % BSA ;
- Taux de viabilité : >70% ;
- Taille des cellules : 5 à 30 µm.

La concentration optimale de la suspension de départ dépend du taux de capture ciblé :

Optimal Input Cell Concentration	Cell Recovery Target
700-1,200 cells/ µl	500-10,000 cells
1,300-1,600 cells/ µl	10,000-20,000 cells

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 X. 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://www.10xgenomics.com/support/single-cell-gene-expression/documentation/steps/sample-prep>
- <https://www.worthington-biochem.com/tools-resources/tissue-dissociation-guide>

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 à leur pourcentage de viabilité. Il est important de réduire au maximum le délai entre la préparation de la suspension et son chargement sur le Chromium X. 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 X.

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 X. 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	
Concentration de la suspension cellulaire de départ (Hématimètre)	700 à 1 000 cellules/ μ l pour une capture de 500 à 10 000 cellules 1 300 à 1 600 cellules/ μ l pour une capture de 10 000 à 20 000 cellules
Volume total de cellules	$\geq 25 \mu$ l
Viabilité des cellules (Double fluorescence)	$\geq 70\%$
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 (incluant le PhiX)	\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 http://ngs-lims.igbmc.fr , 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

6 Livraison des résultats

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 plateforme¹, 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 téléchargement².

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

² Une documentation est disponible sur la page suivante :

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.

En outre, les reliquats 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 des lectures (reads 2), le comptage des UMI, le filtrage des code-barres cellulaires, la génération d'une matrice gènes x code-barres ;
- 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.

8 Prestations additionnelles disponibles

8.1 Analyse multiomique « CITE-seq »

Le « CITE-seq » (Cellular Indexing of Transcriptomes and Epitopes by Sequencing) est une technique qui permet l'analyse simultanée du transcriptome et de l'expression de protéines de surface à l'échelle de la cellule unique sur des milliers de cellules en parallèle.

Cette méthode repose sur l'utilisation d'anticorps conjugués à des oligonucléotides pour convertir la détection de protéines de surface en une lecture quantitative et séquençable (**Fig.3**). Les oligonucléotides sont composés d'une amorce PCR suivi d'un code-barre unique à chaque anticorps et d'une séquence de capture complémentaire à l'une des séquences de capture présentes sur les billes 10X Genomics (**Fig.1**). Les oligonucléotides agissent comme des transcrits synthétiques qui seront capturés dans les nano-gouttelettes simultanément aux transcrits cellulaires et qui, au sein d'une même cellule, seront indexés avec le même code-barre « 10X BC » que l'ensemble des transcrits cellulaires de cette cellule. La société BioLegend propose à son catalogue deux types d'anticorps compatibles avec le protocole Sc 3'mRNA-seq de 10X Genomics : les anticorps TotalSeq™-A utilisant un poly(dT) comme séquence de capture et les anticorps TotalSeq™-B qui utilisent la « Capture Sequence 1 ».

La plateforme GenomEast se propose de réaliser cette prestation chez l'homme ou la souris en utilisant les anticorps TotalSeq™-A selon les recommandations de BioLegend (<https://www.biolegend.com/en-us/totalseq/single-cell-rna>) ou bien TotalSeq™-B selon un protocole supporté par 10X Genomics. Dans le cadre de cette prestation, l'achat des anticorps TotalSeq™ et l'immunomarquage des suspensions cellulaires est à la charge du porteur de projet. Ce dernier devra fournir ses échantillons de cellules marquées à une concentration finale compatible avec un chargement sur le système Chromium X tel que décrit **page 3**.

Lors de la soumission de sa demande, le porteur de projet précisera le nombre et les références exactes des anticorps utilisés afin que la plateforme puisse sélectionner le protocole de préparation de librairie adéquat et, si demandé, pour le démultiplexage final des échantillons.

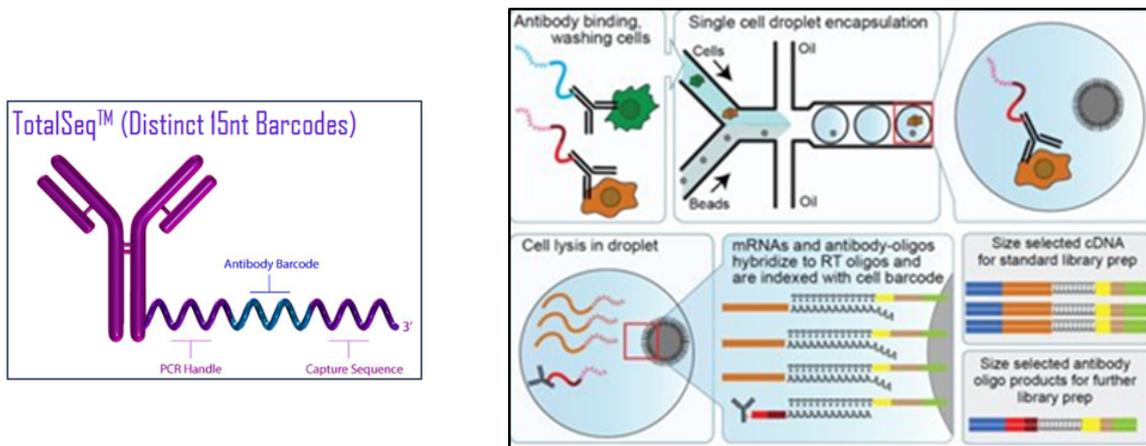


Fig.3 : Structure des anticorps TotalSeq™ (source BioLegend) et Analyse CITE-seq (source : <https://cite-seq.com/>)

8.2 Multiplexage d'échantillons sur chip avec la solution « GEM-X Universal 3' Gene Expression v4 4-plex »

La Société 10X Genomics propose aujourd'hui une solution de multiplexage sur cellules fraîches en 4-plex avec leur dernière chimie GEM-X. Le multiplexage se déroule directement sur le chip de chargement dans le système Chromium et ne nécessite aucun marquage préalable des suspensions de départ. Chaque suspension est encapsulée individuellement sur le chip avec un set de billes de gel spécifique (Fig.4). En sortie de chip, l'ensemble des gouttelettes issues de 4 suspensions sont regroupées pour générer une seule librairie de séquençage. Les cellules individuelles seront démultiplexées et réassignées à l'échantillon de départ informatiquement après analyse des données de séquençage. Le nombre total de cellules chargées sur le chip ainsi que le taux de capture pour les 4 suspensions réunies demeurent identiques aux valeurs indiquées dans le **tableau 1**. Ainsi, si en simplex il est possible de capturer au maximum 20 000 cellules pour un échantillon, en 4-Plex, il sera possible de capturer jusqu'à 4 x 5000 cellules. Pour plus de détails, veuillez contacter la plateforme.

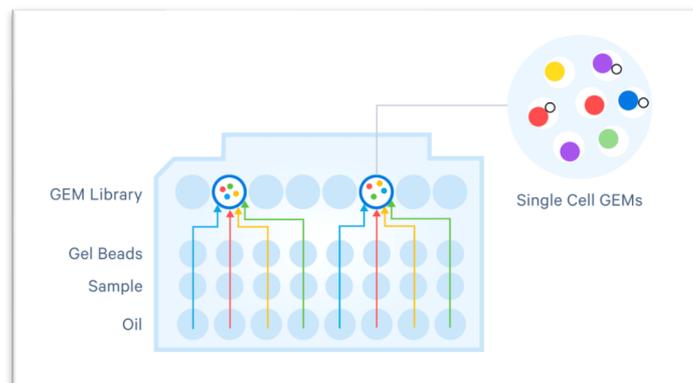


Fig.4 : Multiplexage d'échantillons sur chip (source : 10X Genomics)