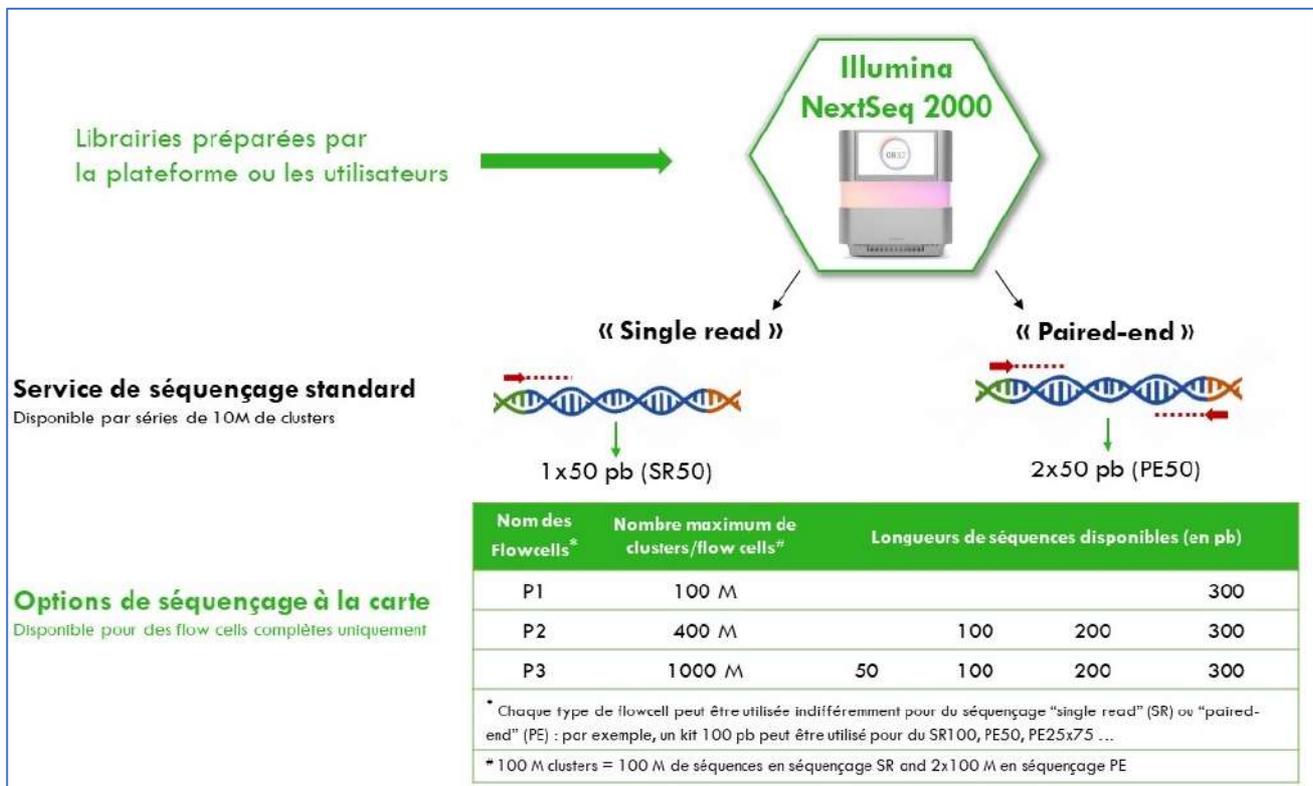


# FICHE PRODUIT : SÉQUENÇAGE DES LIBRAIRIES DES UTILISATEURS

La plateforme propose une prestation de séquençage de bibliothèques préparées par les utilisateurs. Le coût total de ce service comprend la validation des bibliothèques prêtes à séquençer d'une part et le séquençage d'autre part. Si le porteur de projet fournit un seul tube de bibliothèques par flow cell de séquençage en assurant lui-même le multiplexage de ses échantillons, la validation du pool à séquençer est offerte.

## 1 Options de séquençage disponibles



## 2 Services proposés

- La vérification des bibliothèques préparées par l'utilisateur :
  - Quantification et vérification de la qualité des bibliothèques par électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI).
  - Purification éventuelle des bibliothèques avant séquençage pour satisfaire nos critères de qualité (voir ci-dessous)
- Le séquençage avec le séquenceur NextSeq 2000 d'Illumina :
  - Simple ou paillé avec des tailles de lectures selon les options spécifiées sur le LIMS (<http://ngs-lims.igbmc.fr>) pour chaque projet.

- Veuillez noter que pour toutes les options de séquençage à la carte, nous ne séquençons les librairies fournies par les utilisateurs que sur des flow cell entières. De même, pour les longueurs de séquençage utilisées en routine sur la plateforme (1x50b et 2x50b), nous ne séquençons les librairies fournies par les utilisateurs par séries de 10M de séquences uniquement si la taille et les indexes des librairies fournies sont compatibles avec celles préparées par la plateforme. Dans le cas contraire, les librairies seront séquençées sur des flow cells entières.

3. 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 pour l'analyse primaire et les résultats obtenus.

4. L'analyse ultérieure des données (optionnelle, voir section 5 pour plus d'information)

### 3 Informations à fournir par les utilisateurs

Afin de vérifier leur compatibilité avec la technologie NextSeq 2000 d'Illumina et pour optimiser les conditions de séquençage, il est important de transmettre à la plateforme le détail de la construction des librairies générées. Le tableau ci-dessous liste les documents à fournir avec les informations recherchées et leur finalité.

Documents à fournir	Informations recherchées	Finalité
Protocole de préparation des librairies (référence du kit, publication...)	Schéma de la construction des librairies (indexés simples ou doubles, barcodes, UMI...)  Méthode d'ajout des adaptateurs (PCR ou ligation)  Diversité en ATGC des 25 premières bases de la librairie ( <b>Ref 2</b> )  Nécessité d'utiliser un primer de séquençage custom (à fournir : $\geq 20\mu\text{l}$ à $100\mu\text{M}$ dans de l'eau MilliQ dans des tubes LowBind)	Déterminer la compatibilité avec la technologie Illumina et définir les conditions de séquençage  Limiter les problèmes d'index hopping liés à un reliquat d'adaptateurs dans les librairies quand ils sont ajoutés par PCR ( <b>Ref 1</b> ) Déterminer la nécessité de multiplexer les librairies fournies avec la librairie contrôle PhiX d'Illumina ou avec une autre librairie équilibrée pour optimiser le nombre de clusters sur la piste  Séquencer les librairies avec des adaptateurs custom
Listing des échantillons et des indexes utilisés (à renseigner sur le LIMS de la plateforme ou sur un fichier Excel : voir modèle page 4)	Nom, concentration et volume des échantillons ( <b>si les échantillons sont soumis en un pool, le détail de chaque échantillon du pool est à fournir</b> )	Démultiplexer les échantillons sur les données brutes

Documents à fournir	Informations recherchées	Finalité
	Nom, séquence et fournisseur des indexes utilisés.	Plusieurs sets d'indexes commerciaux sont déjà pré-enregistrés dans notre LIMS
Information sur l'utilisation de spikes lors de la génération des librairies (par exemple chromatine de drosophile dans les librairies ChIPseq)	Nature et proportion	Interpréter les résultats d'analyse des contaminants dans le FastQscreen
Si disponible, profils Bioanalyzer (Agilent) ou Fragment Analyzer (AATI)	Taille de l'insert (= taille de la librairie - taille des adaptateurs)  Présence de dimères d'adaptateurs ou de reliquats de primers  Quantité totale des librairies	Déterminer la proportion de fragments « séquençables » (fragments <600 pb sur le NextSeq 2000)  Si un séquençage en paired-end (PE) est demandé, vérifier la taille minimale des librairies pour des séquences non chevauchantes (insert >320 bp pour un séquençage 2x100 PE et insert >220 bp pour un séquençage 2x50 PE)  Optimiser la génération des clusters sur la flow cell et limiter le nombre de séquences non informatives  Quantité optimale : 20µl à 5nM Quantité minimale : 10µl à 3nM (suffisant pour 2 séquençages)

## 4 Contrôles qualité réalisés par la plateforme

Les librairies sont contrôlées selon des critères qualité inhérents à la technologie de séquençage (voir tableau ci-dessous). Les résultats sont envoyés par e-mail au porteur de projet et/ou accessibles sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>).

Vérification des librairies	
Profil des librairies (électrophorèse capillaire)	Taille moyenne comprise entre 200 et 600 pb.
Pureté des librairies (électrophorèse capillaire)	Présence minoritaire de reliquats de primers et de dimères d'adaptateur (bande à 120-130 pb), si applicable.
Quantité totale des librairies	>10µl à 3nM

Après validation des librairies, la plateforme s'engage à utiliser la technologie de séquençage Illumina selon les recommandations du fournisseur. N'ayant aucun contrôle sur la génération des librairies, la plateforme déclinera toute responsabilité sur la qualité des résultats de séquençage finaux. Nous vérifions cependant la qualité des données générées selon nos critères habituels suivants.

Vérification des données de séquençage	
Nombre attendu de clusters* par projet (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 <a href="http://ngs-lims.igbmc.fr">http://ngs-lims.igbmc.fr</a> , pour chaque projet)
Scores de qualité attendus (Phred score)>30	≥ 85% des bases

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

A noter que pour le séquençage des librairies avec une faible diversité en bases nécessitant un multiplexage avec la librairie PhiX Illumina, le nombre total attendu de séquences inclut les séquences s'alignant sur le PhiX.

## 5 Livraison des résultats

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 six mois après leur mise à disposition.**

## 6 Analyse ultérieure (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. Le type d'analyses réalisées dépend de la nature des librairies séquencées. 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.

## 7 Références

- (1) Effects of Index Misassignment on Multiplexing and Downstream Analysis. White paper Illumina. Pub. No. 770-2017-004-D.
- (2) Low-Plex Pooling Guidelines for Enrichment Protocols. Technical note Illumina. Pub. No. 770-2013-060, 23 September 2015.

\* Modèle de fichier Excel pour la description des librairies

Sample name	Concentration (ng/μl)	Quantification method	Volume (μl)	Experimental condition	Remarks	Index 1 (i7) Supplier	Index 1 (i7) Code	Sequence Index 1 (i7)	Index 2 (i5) Supplier 2	Index 2 (i5) Code 2	Sequence Index 2 (i5)
WT1	10	Qubit	10	Contrôle		Illumina:Nextera	N704	TCCTGAGC	Illumina:Nextera	S502	CTCTCTAT