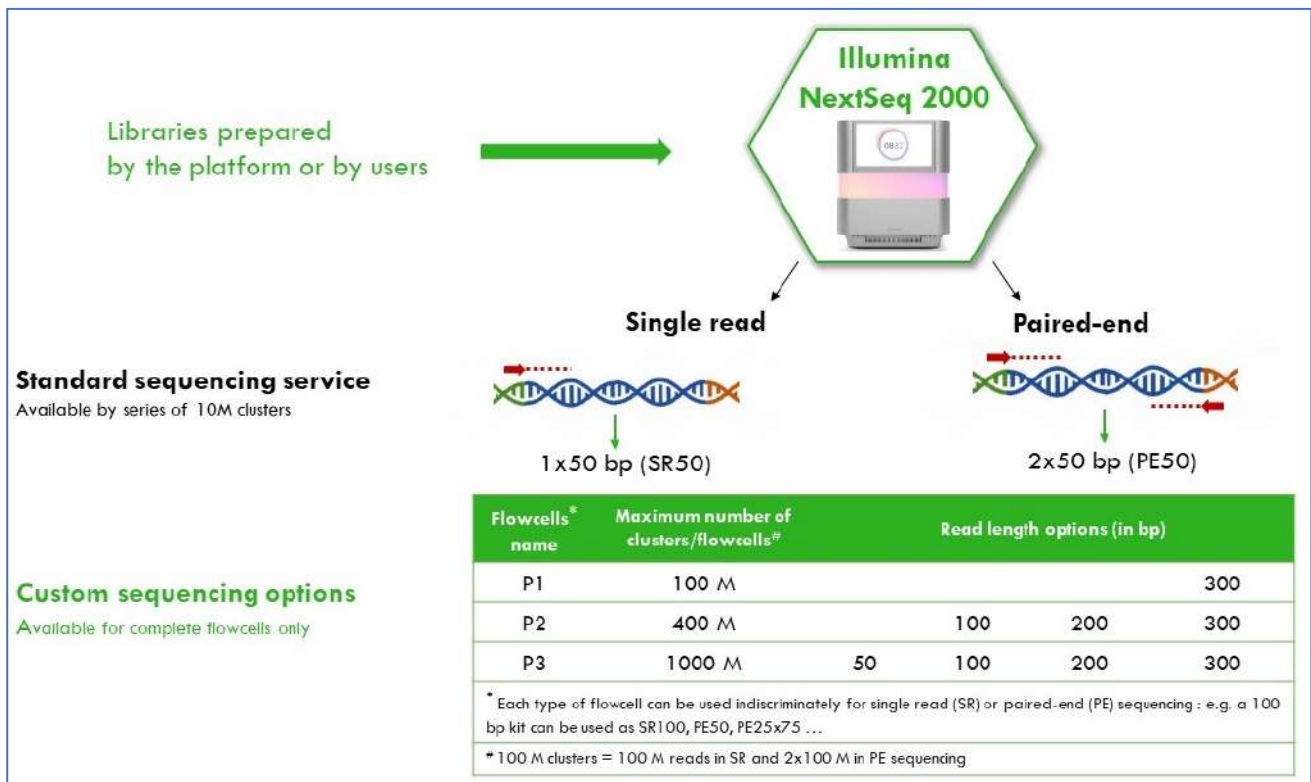


# PRODUCT SHEET: SEQUENCING OF USER-PREPARED LIBRARIES

The platform offers to sequence libraries prepared by project managers. The total cost of this service includes validation of ready-to-load libraries on one hand and sequencing on the other hand. If the project manager provides only one tube of libraries per sequencing flow cell while ensuring the multiplexing of his samples himself, the validation of the pool to be sequenced is offered.

## 1 Sequencing options available



## 2 Services provided

- Library checking:
  - Library quantification and quality control by capillary electrophoresis (Bioanalyzer from Agilent or Fragment Analyzer from AATI).
  - Possible purification of libraries before sequencing to meet our quality criteria (see below)
- Sequencing using Illumina NextSeq 2000 technology:
  - Single-read or paired-end sequencing with read lengths according to options specified on the LIMS (<http://ngs-lims.igbmc.fr>) for each project.
  - Please note that for all a la carte sequencing options, we only sequence user-provided libraries on entire flow cells. Similarly, for the sequencing lengths routinely used on the platform (1x50b and**

**2x50b), we only sequence the libraries provided by users in series of 10M reads if the size and indexes of the libraries provided are compatible with those prepared by the platform. Otherwise, the libraries will be sequenced on entire flow cells.**

3. Primary data analysis:
  - Demultiplexing and generation of FASTQ files.
  - Sequencing quality check.
  - Detection of potential contaminations.
  - Generation of a report summarizing the methods used in the pipeline as well as the results obtained.
4. Downstream data analysis (optional, see section 5 for more information)

### 3 Information to be provided by project managers

In order to verify their compatibility with Illumina's NextSeq 2000 technology and to optimize the sequencing conditions, it is important to transmit to the platform the details of the construction of the generated libraries. The table below lists the documents to provide with the information sought and their purpose.

Documents to provide	Information sought	Purpose
Library preparation protocol (kit reference, publication ...)	<p>Map of the library construction (single or double indexes, bar-codes, UMI ...)</p> <p>Method used for adding adapters (PCR or ligation)</p> <p>ATGC diversity of the first 25 bases of the library (<b>Ref 2</b>)</p> <p>Need to use a custom sequencing primer (to provide: <math>\geq 20\mu\text{l}</math> to <math>100\mu\text{M}</math> in MilliQ water in LowBind tubes)</p>	<p>Determine compatibility with Illumina technology and define sequencing conditions</p> <p>Limit index hopping problems related to remaining adapters in libraries when they are added by PCR (<b>Ref 1</b>)</p> <p>Determine the need to multiplex the libraries provided with Illumina's PhiX control library or with another balanced library to optimize the number of clusters on the lane</p> <p>Sequence libraries with custom adapters</p>
Listing of samples and indexes used (to be filled in on the LIMS of the platform or on an Excel file: see template page 4)	<p>Name, concentration and volume of the samples (<b>if samples are submitted as a pool, the details of each sample in the pool must be provided</b>)</p> <p>Name, sequence and provider of the indexes used.</p>	<p>Demultiplex the samples on the raw data</p> <p>Several sets of commercial indexes are already pre-registered in our LIMS</p>

Documents to provide	Information sought	Purpose
Information on the use of spikes when generating libraries (for example drosophila chromatin in ChIPseq libraries)	Nature and proportion	Interpret contaminant analysis results in FastQscreen
If available, Bioanalyzer (Agilent) or Fragment Analyzer (AATI) profiles	Size of the insert (= size of the library - size of the adapters)  Presence of adapter dimers or remaining primers  Total quantity of libraries	Determine the proportion of "sequenceable" fragments (fragments <600 bp on the NextSeq 2000)  If paired-end (PE) sequencing is requested, check the minimum library size for non-overlapping sequences (insert> 320 bp for 2x100 PE sequencing and insert> 220 bp for 2x50 PE sequencing)  Optimize the generation of clusters on the flow cell and limit the number of non-informative sequences  Optimal amount : 20µl at 5nM Minimal amount : 10µl at 3nM (sufficient for 2 sequencing)

## 4 Quality controls performed by the platform

Libraries are checked according to quality criteria dependent on the sequencing technology in use on the platform (see table below). Quality control results are sent by mail to the project manager and/or available through the platform's LIMS (<http://ngs-lims.igbmc.fr>).

Library checking	
Library profile (capillary electrophoresis)	Average size ranging from 200 to 600 bp.
Library purity (capillary electrophoresis)	Limited presence of remaining primers and adapter dimers (120-130 pb band), if applicable
Total quantity of library	>10 µl at 3nM

After library validation, the platform commits to use the Illumina sequencing technology following Illumina's recommendations. Having no control over the generation of libraries, the platform will not be responsible for the quality of the final sequencing results. However, we check the quality of the data generated according to our usual criteria.

Sequencing data checking	
Expected number of clusters* per project (including PhiX)	≥ Total number of clusters specified in the “Requested services” section from the submission form (pdf file that can be downloaded from the LIMS <a href="http://ngs-lims.igbmc.fr">http://ngs-lims.igbmc.fr</a> , in the “Document” tab for each project)
Expected quality scores (Phred score)>30	≥ 85% of bases

\* Number of reads in single-read and number of reads ÷ 2 in paired-end

It is important to note that for the sequencing of libraries with low diversity in bases requiring multiplexing with the PhiX Illumina library, the total number of expected reads includes the sequences aligning to PhiX.

## 5 Results delivery

For each sample, raw sequencing data are provided (nucleotide sequences in FASTQ format. The files contain reads passing quality filters).

In addition to these sample files, two files are provided for each project:

- A project report (in PDF format) containing the number of raw reads, the percentage of bases with a Phred quality score over 30, various information on data quality and the size of each FASTQ sequence file to be downloaded.
- A text file providing the MD5 strings of each FASTQ file. The project manager is responsible for downloading his files, checking their integrity from MD5 strings and storing them. Data will be removed from the server six months after their delivery.

The project manager is informed of the availability of the data by email once the sequencing process is done. The generated data can be retrieved using a login and a password on the platform FTP server.

**According to the “GenomEast Platform terms and conditions of business”, the project manager is responsible for his data to be saved and archived on its own. Following their transfer to the project manager, the Platform guarantees the conservation of raw data only for a limited period of six months.**

## 6 Downstream analysis (optional)

Data analysis is not part of the standard service but can be done in collaboration between the project manager and the platform. The type of analysis performed will depend upon the nature of the sequenced libraries. We recommend the project managers who would like to collaborate with the platform for data analysis to contact the platform before starting their experiment so that we can define the analyses that best fit to their needs.

## 7 References

- (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.

\* Excel template for library description

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