GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

PRODUCT SHEET: SEQUENCING OF USER-PREPARED LIBRARIES

The platform offers to sequence libraries prepared by project managers.

1 Services provided

1. Library checking:
   - Library quantification and quality control by capillary electrophoresis (Bioanalyzer from Agilent or Fragment Analyzer from AATI).
2. Sequencing using Illumina HiSeq 4000 technology:
   - Loading of libraries in a flow cell and cluster generation on the Cbot (Illumina).
   - Single-end or paired-end sequencing of 50 or 100 bp according to the read length chosen by the project manager. We can realize other read lengths upon demand but only for complete flow cells (8 lanes).
   - Libraries submitted by the project manager will be sequenced on entire lane(s) of flow cell (e.g. libraries will not be multiplexed with libraries prepared by the platform or by other users). We run into a sequencing process as soon as the 8 channels of the flow cell are fulfilled with samples to be sequenced with the same read length.
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)

2 Information to be provided by project managers

In order to verify their compatibility with Illumina's HiSeq 4000 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 ...)	Map of the library construction (single or double indexes, barcodes, UMI ...) Method used for adapters (PCR or ligation) ATGC diversity of the first 25 bases of the library (**Ref 2**)	Determine compatibility with Illumina technology and define sequencing conditions Limit index hopping problems related to remaining adapters in libraries when they are added by PCR (**Ref 1**)

GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

Documents to provide	Information sought	Purpose
	Need to use a custom sequencing primer (to provide: ≥20µl to 100 µM in MilliQ water in LowBind tubes)	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 Sequence libraries with custom adapters
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)	Name, concentration and volume of the samples (**if samples are submitted as a pool, the details of each sample in the pool must be provided**) Name, sequence and provider of the indexes used.	Demultiplex the samples on the raw data Several sets of commercial indexes are already pre-registered in our LIMS
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)	Determine the proportion of "sequenceable" fragments (fragments <600 bp on the HiSeq 4000) 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)
	Presence of adapter dimers or remaining primers	Optimize the generation of clusters on the flow cell and limit the number of non-informative sequences
	Total quantity of libraries	Optimal amount : 20µl at 5nM Minimal amount : 10µl at 3nM (sufficient for 2 sequencing)

GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

3 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 reads per lane	>250 million in single-end >500 million in paired-end
Expected quality scores (Phred score)>30	≥ 85% of bases for sequences of 50 bases ≥ 75% of bases for sequences of 100 bases

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 reads expected per lane includes the sequences aligning to PhIX.

4 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.

Instead of or in addition to all files cited above, the project manager may ask to the platform to provide non demultiplexed files in BCL (binary base call) format. In that case, the following files are provided to the project manager:

- An archive in tar.gz format containing BCL, xml and RTA*.txt files from the sequencing run folder, preserving the directory tree and the name of each folder, for the lanes where samples from this project have been sequenced.

GenomEast Platform
CERBM GIE – IGBMC
1 rue Laurent Fries
67404 Illkirch Cedex
France

+33 3 88 65 34 26
thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

- A SampleSheet.csv file (csv file containing information relative to sequenced samples, required to demultiplex BCL files using Cell Ranger mkfastq or Illumina bcl2fastq).
- A text file providing the MD5 strings of each file. The project manager can use these MD5 to check the integrity of the files after their download.

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.

5 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.

6 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*	*Controle*			*Illumina:Nextera*	*N704*	*TCCTGAGC*	*Illumina:Nextera*	*S502*	*CTCTCTAT*