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 GENOMIC DNA

The platform proposes sequencing or re-sequencing, either of complete genomes or of targeted regions by capturing regions of interest. We also provide expertise in bioinformatics analyses for genetics variants detection for sequenced and assembled genomes.

1 Available library preparation and sequencing options

1.1 Library preparation methods

Several library preparation protocols are currently available on the platform. The choice of the most appropriate protocol for a project mainly depends on the available amount of DNA and on the experimental design as described on the following table.

#	Title of the service	Kit used by the platform	Genomic DNA quantity		Size of capture
			Minimal	Optimal	
1	Whole genome / standard quantity	TruSeq Nano Illumina	100 ng	1µg	-
2	Whole genome / low input	Microplex Diagenode	2 ng	20 ng	-
3	Human Exome / Twist	Human Comprehensive Exome Twist Bioscience	50 ng	500 ng	36,8 Mb
4	Human Exome / Agilent	Sureselect XT Human Exome Agilent	100 ng	1 µg	35,1 Mb
5	Custom Capture	Several protocols are available. Please, contact us.	100 ng	1 µg	Dependent on capture design

The platform has expertise in the use of several capture solutions such as the ones from Agilent, Illumina, IDT, Roche and Twist Bioscience. For instance, for Human exome, the capture solution currently in used is the Human Comprehensive Exome panel from Twist Bioscience[1]. In addition to standard panels, we can use custom capture designs. For custom kits, the project manager can either work directly with enrichment kit providers or contact the platform for advises and help on capture design. In case the project manager wishes to use a custom design and if a collaboration is set up with the platform for data analysis, a BED[2] file with genomic coordinates of target regions must be provided to the platform. The corresponding file can be uploaded onto the platform LIMS (http://ngs-lims.igbmc.fr) in the corresponding project.

1.2 Sequencing options

Libraries are sequenced using the Illumina NextSeq 2000. Paired-end reads are obtained with a size of 150 bp. The following formula may be used to estimate the number of reads needed to reach targeted coverage:

[1] https://www.twistbioscience.com/products/ngs/fixed-panels/human-comprehensive-exome

[2] https://genome.ucsc.edu/FAQ/FAQformat.html#format1

Number of reads (N):

$$N = \frac{D \times C}{2 \times 150}$$

Where:

- D: Total length (in number of bases) of regions of interest or genome size
- C: Mean coverage desired or recommended (see table 1)

Table 1: Recommended mean of coverage for a minimal and sufficient coverage of the regions of interest.

Project	Recommended mean of coverage
Exome (germline)	≥ 100 X*
Exome (*de novo*)	≥ 200 X
Exome (FFPE)	≥ 150 X
Genome (germline)	≥ 30 X**

*_**Example 1**: to have a minimum mean coverage of 100X using Twist Bioscience's Human Comprehensive Exome kit for a germline variant detection project, we advise to sequence ≥ 10M reads per sample: (36,8 x 10^6 x 100) / (2 x 150)_

Example 2**: to have a minimum mean coverage of 30X for the human genome (3Gb), we recommend to sequence 300M reads per sample: (3 x 10^9 x 30) / (2 x 150)

2 Services provided

1. Sample checking:
   - Quantity and quality check using a fluorometer (Qubit or Varioskan) and a capillary electrophoresis machine (Fragment Analyzer, Agilent), only when the quantity of starting material is not limited
2. Library preparation:
   - Optional: DNA fragmentation depending on sample type and protocol used
   - Preparation of libraries and ligation of indexed sequencing adapters to DNA fragments. Indexes are DNA sequences used to identify each sample. Usage of indexes allows for pooling multiple samples on a single sequencing lane
   - Libraries quantification and quality control by capillary electrophoresis (Bioanalyzer from Agilent or Fragment Analyzer from AATI).
   - Optional: capture of DNA fragments within regions of interest
3. Sequencing using the Illumina NextSeq 2000 technology:
   - Paired-end sequencing of 150 bp
4. Primary data analysis
   - Demultiplexing and generation of FASTQ files
   - Sequence quality check
   - Detection of potential contaminations
   - Generation of a report summarizing the methods used in the primary data analysis pipeline as well as the results obtained
5. Downstream data analysis (optional, see section 6 for more information)

3 Sample preparation (done by the project manager)

The project manager provides the platform with full length or fragmented genomic DNA (gDNA). The success of the experiment is closely linked to the quality of the starting samples. Particular care must be taken to avoid any trace of contamination or degradation in the samples.

Characteristics of DNA that should be provided to the platform	
Quantity	Depends on the library preparation protocol chosen by the project manager
Minimal volume	10 µl
Quality	DNA sample must be depleted of any contaminants that may inhibit enzymatic reaction during the library preparation (proteins, EDTA, salts, solvents, etc.) Optional: Additional purification, using AMPure XP or SPRI-select beads (Beckman Coulter) may be necessary, but is only possible with already fragmented DNA. If not realized by the project manager, this purification will be realized by the platform before quantification of starting material
Shipping conditions	In solution, in water, shipped on cold packs. Samples are to be registered in the LIMS, then, **the unique ID generated for each sample must clearly indicated on the tube**

4 Quality controls

Quality controls listed below are performed by the platform. Quality controls performed at steps 1 and 2 are also available through the platform's LIMS (http://ngs-lims.igbmc.fr).

1. Sample checking	
Quantity (Fluorimetry)	≥ minimal required quantity (depending on the library preparation protocol)
Quality	Full size genomic DNA: no sign of degradation on an electrophoresis profile Fragmented DNA: mean size ≤ 500 bp.
2. Library preparation	
Library profile (capillary electrophoresis)	Average size ranging from 200 to 600 bp
Library purity (capillary electrophoresis)	Limited presence of adapter dimers (120-130 bp band) and primers (30-60 bp band).
3. Sequencing and primary data analysis	
Total number of clusters* per project	≥ total number of clusters specified in the "Requested services" section from the submission form (pdf file that can be downloaded from the LIMS http://ngs-lims.igbmc.fr, in the "Document" tab for each project)
Quality score (Phred score) > 30	≥ 85% of bases

* number of reads ÷ 2 in Paired-end

5 Results delivery

For each sample, raw sequencing data are provided (nucleotide sequences in FASTQ format).

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 string of each FASTQ file. The project manager is responsible for downloading his files, checking their integrity from MD5 strings and storing them. A documentation is available on the following webpage: http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5.

The project manager is informed of the availability of the data by email once the sequencing process is done. This email contains a login and a password to be used to retrieve the generated data on the platform FTP server.

According to the "GenomEast Platform terms and conditions of business", following data delivery, the project manager is responsible for his data to be saved and archived on its own. Data will be removed from the Platform server six months after their delivery.

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 following analyses can be performed:

- Alignment to a reference genome.
- Assessment of the capture efficacy.
- Variant discovery (SNV and short indels).
- Functional annotation regarding genomic features (3'UTR, exon, intron, etc.) and the impact of variants (synonymous, non-synonymous, stop codon, impact on splicing, etc.).
- Annotation with public databases such as dbSNP, 1000 genomes, Hapmap, EVS, etc.
- Variant ranking.

This list is not exhaustive and we recommend the project manager 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 the project manager's needs.