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

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

PRODUCT SHEET: SEQUENCING OF GENOMIC DNA

The platform proposes sequencing or re-sequencing services, either of complete genomes or of targeted regions by capturing regions of interest. For this application, we offer different alternatives using either Illumina sequencing technology or MGI technology. We also provide expertise in bioinformatics analyses for genetics variants detection for sequenced and assembled genomes.

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 sequencing technology, the available amount of DNA and on the experimental design.

1.1 Illumina technology

#	Title of the service [1]	Kit used by the platform	Genomic DNA quantity		Size of capture
			Minimal	Optimal	
1	Library prep DNA Truseq Nano (Illumina)	TruSeq Nano Illumina	100 ng	1 µg	-
2	Library prep ChIP (Diagenode)	Microplex Diagenode	2 ng	20 ng	-
3	Library prep DNA Exome (Twist Bioscience)	Human Comprehensive Exome Twist Bioscience	50 ng	500 ng	36,8 Mb
4	Custom targeted DNA-seq	Several protocols are available. Please, contact us.	100 ng	1 µg	Dependent on capture design

[1] Titles as listed in the LIMS of the platform (http://ngs-lims.igbmc.fr)

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] https://www.twistbioscience.com/products/ngs/fixed-panels/human-comprehensive-exome
[2] https://genome.ucsc.edu/FAQ/FAQformat.html#format1

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

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

1.2 MGI technology

#	Title of the service [1]	Kit used by the Platform [2]	Genomic DNA quantity		Size of capture
			Minimal	Optimal	
1	Library prep DNA PCR free (MGI)	MGIEasy FS PCR-free DNA Library Prep	50 ng	1 µg	-
2	Library prep DNA PCR free (MGI)	MGIEasy PCR-free DNA Library Prep • Fragmented DNA (300-500bp) • Full length DNA (peak ≥ 50Kb)	 80 ng 200 ng	 200 ng 1 µg	-

[1] Titles as listed on the LIMS of the platform (http://ngs-lims.igbmc.fr)

[2] Library protocol to use depending on the starting material purity:

- #1: Full size genomic DNA (high-MW single band on agarose gel): « MGIEasy FS PCR-Free DNA Library Prep Set User Manual » with enzymatic fragmentation of genomic DNA.
- #2: Genomic DNA showing signs of degradation (presence of a smear under the high-MW band): « MGIEasy PCR-Free DNA Lib Prep Set User Manual with sonication of genomic DNA using COVARIS.

2 Sequencing options

We advise sequencing in paired-end mode with reads of 150 bp.
The following formula may be used to estimate the number of reads needed to reach targeted coverage:

Number of reads (N):

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

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)*

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

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

3 Services provided

1. Sample checking:
   - Quantity check using a fluorometer (Qubit or Varioskan) and quality check using 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 using a fluorometer (Qubit or Varioscan) or by capillary electrophoresis (Bioanalyzer from Agilent or Fragment Analyzer from Agilent),
   - Optional: capture of DNA fragments within regions of interest.
3. Sequencing using Illumina NextSeq 2000 or MGI DNBSEQ G400-RS:
   - Paired-end 2x150 bp sequencing.
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)

4 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	In optimal conditions, DNA shall have the following ratios A260/280>1.8 and A260/230>1.7 Full size genomic DNA: no sign of degradation on an electrophoresis profile Fragmented DNA: mean size ≤ 500 bp. 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**

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

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

5 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).

5.1 Illumina technology

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

5.2 MGI technology

1. Sample checking	
Quantity (Fluorimetry)	≥ 2 µg
Quality	Full size genomic DNA: no sign of degradation on an electrophoresis profile, for enzymatic fragmentation Ratios A260/280>1.8 and A260/230>1.7
2. Individual library preparation	
DNA profil (capillary electrophoresis)	300 bp ≤ Peak ≥ 800 bp
3. DNA nanoballs (DNB) preparation per sequencing lane	
Quantity of the pool of single strand circularized libraries (fluorimetry)	≥ 75, 60 and 30 fmol from 200~1000ng, 100~200ng, 50~100ng of starting genomic DNA
Concentration of DNA nanoballs (fluorimetry)	≥ 8 ng/µl
4. Sequencing and primary data analysis	
Total number of clusters* per lane	≥ 400 millions
Quality score (Phred score) > 30	≥ 80% of bases

* number of reads ÷ 2 in Paired-end

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

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

6 Results delivery

An e-mail will be sent to the project leader informing him that he can download his data from the platform's sftp/https server[3], using the login and password dedicated to his project, indicated on the platform's web interface (https://ngs-lims.igbmc.fr). If the project leader has added collaborators to his project, this e-mail will also be sent to these collaborators, who will also have access to the login and password for this project.

The following files will be made available:

- Raw sequencing data in FASTQ format.
- A report describing the methods used by the platform primary analysis pipeline and the results obtained.
- A text file providing the MD5 string of each FASTQ file to download. The project leader must use these MD5 strings to verify the integrity of the files after they have been downloaded[4].

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. Access to the sftp/https server is only valid for six months from the date of data delivery.

Leftover samples and libraries will be destroyed 6 months after delivery of the raw data if not claimed by the project leader.

7 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.

[3] A documentation is available on the following webpage:
http://genomeast.igbmc.fr/wiki/doku.php?id=help:downloading
[4] A documentation is available on the following webpage:
http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5