Microarray and sequencing platform

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: TARGET RESEQUENCING

Target resequencing combines enrichment capture assays with high throughput sequencing. It is used to detect genetic variations in genomic regions of interest (exome, genes involved in particular diseases, etc.).

1 Services provided

1. Sample checking:
   - Quantity check using a fluorometer (Qubit or Varioskan).
   - Quality check on an agarose gel.
2. Library preparation:
   - Fragmentation of DNA by sonication (Covaris).
   - Quality check of fragmented DNA using capillary electrophoresis (Bioanalyzer, Agilent).
   - Preparation of libraries and ligation of indexed sequencing adapters to DNA fragments. Indexes are DNA sequences of ≥ 6 nt long 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).
   - Capture of DNA fragments within regions of interest.
3. Sequencing using Illumina HiSeq 4000 technology:
   - Loading of libraries on a flow cell and cluster generation on the Cbot (Illumina).
   - Paired-end sequencing 2x100 bases.
4. Primary data analysis:
   - Demultiplexing and generation of FASTQ files.
   - Adapter dimer removal.
   - Sequencing quality check.
   - Detection of potential contaminations.
   - Generation of a report summarizing the methods used in the pipeline as well as the results obtained (one report per sample and one global report for the project).
5. Downstream analysis of data (optional, see section 6 for more information).

2 Capture design

The platform provides access to all exon or custom capture kits. For custom kits, the project manager can either work directly with the enrichment kit provider 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 file with genomic coordinates of target regions should be provided to the platform. The corresponding file can be uploaded in the "Share document" part of the platform's LIMS (http://ngs-lims.igbmc.fr).

3 Sample preparation (done by the project manager)

The project manager provides the platform with full length or fragmented genomic DNA.

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

Characteristics of DNA that should be provided to the platform	
Starting material (genomic DNA)	3 µg (measured using a fluorometer).
Concentration	≥ 100 ng/µl.
Minimal volume	20 µl.
Shipping conditions	In solution, in water. Sample names must be clearly indicated on the tubes as well as in the platform's LIMS.

4 Quality controls

Quality controls listed below are performed and corresponding results are sent to the project manager after each of the following steps. 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 (fluorometer)	≥ 3 µg.
Quality	Full length genomic DNA: no degradation on agarose gel. Fragmented DNA: mean size ≤ 500 bp (Smear in a range of 100 to 700 bp). The platform recommends performing DNA shearing using sonication (e.g. Covaris). Sample must be accompanied by a gel photograph. The corresponding file can be uploaded in the "Share document" part of the platform's LIMS.
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).
3. Sequencing and primary data analysis	
Total number of clusters (i.e. number of reads in Single-read and number of reads ÷ 2 in Paired-end sequencing)	≥ 250 x L millions, where L represents the total number of lanes requested by the user *.
Quality score (Phred Score) over 30	≥ 75% of bases for 100 length reads

*For a human exome resequencing project (size of captured regions ≈ 50 Mb) where we multiplex 6 samples per lane, we obtain a mean nucleotide depth ≥ 60X in target regions. For shorter capture design, the total number of reads per sample is re-considered to obtain a mean nucleotide depth ≥ 60X in target regions. Samples are thus multiplexed accordingly.

5 Results delivery

For each sample, the following results are available:

- Raw sequencing data (nucleotide sequences in FASTQ format. The files contain reads passing quality filters without dimer adapter sequences).

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 report with sequence quality controls (in PDF 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 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 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. 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", the project manager is responsible for his data to be saved and archived on its own. Following their transfer to the Beneficiary, the Platform guarantees the conservation of raw data 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 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.