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: CHIP-SEQ

ChIP-seq combines Chromatin ImmunoPrecipitation (ChIP) with high throughput sequencing to identify binding sites of DNA-associated proteins.

1 Available library preparation and sequencing options

1.1 Library preparation methods

One ChIP-seq library preparation protocol is currently available on the platform.

#	Kit used by the platform	Quantity of chromatin	
		Minimal	Optimal
1	Diagenode MicroPlex	2 ng	10 ng

1.2 Sequencing options

Libraries are sequenced using the Illumina HiSeq 4000 technology. Both single-read and paired-end reads can be obtained with a size of 50bp. Table below provides advices regarding read length depending on the objectives of your experiment. Project managers are encouraged to contact us for more information regarding these different options if needed.

Project	Recommendations	
	Sequencing type	Read length
General case:	Single-end	50 bp
Particular case: - Analysis of certain histone marks (ex: H3K27me3) - Analysis of proteins known to bind in repetitive elements	Paired-end	50 bp

For mammalian genomes, we recommend multiplexing up to 8 libraries per lane. Depending on the protein of interest, a higher coverage may be needed. This would lead to multiplex fewer samples on a single flow cell lane.

1.3 Experimental design

Experimental design recommendation	
Negative control	It is **strongly advised to use a negative control** to improve the detection of regions enriched by the ChIP. For example: • Wild type DNA Chipped with the same antibody as the one used for the mutant • A portion of DNA sample removed prior to IP (Input 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

	• DNA obtained from IP with an antibody not known to be involved in DNA binding or chromatin modification such as IgG • DNA obtained from IP without antibodies (mock)
Biological replicates	It is recommended[1] providing samples from at least 2 independent experiments (this number is highly dependent on variability between replicates). For any questions on experimental design, please contact the platform before starting your experiments.

2 Services provided

1. Sample checking:
   - Quantity checking using a fluorimeter (Qubit)
2. Library preparation:
   - 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)
3. Sequencing using Illumina HiSeq 4000 technology:
   - Loading of libraries on a flow cell and cluster generation on the Cbot (Illumina)
   - Single-read or paired-end sequencing of 50 bases as requested
4. Primary data analysis:
   - Demultiplexing and generation of FASTQ files
   - Sequences quality check
   - Detection of potential contaminations
   - Mapping of reads onto the reference genome, if available
   - Generation of a report summarizing the methods used in the primary analysis pipeline as well as the results obtained.
5. Downstream data analysis (optional, see section 5 for more information)

3 Sample preparation (done by the project manager)

The project manager performs the ChIP. It is worth noting that during this step, DNA amplification (ex: WGA, etc.) is not recommended.

Sample prep recommendation	
IP blocking agent	Avoid DNA (salmon sperm) as blocking agent. It is better to use yeast tRNA, etc.
IP enrichment validation	By doing qPCR on known targets

Characteristics of Chipped DNA

[1] Landt, Stephen G. et al. "ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia." *Genome Research* 22.9 (2012): 1813–1831. *PMC*. Web. 13 Mar. 2017.

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

Starting material (Input or ChIP)	Depends on the library preparation protocol chosen by the project manager (measured using a fluorometer, e.g. Qubit)
DNA fragment size	Sample **must** be accompanied by a gel photograph. The corresponding file can be uploaded in the “Share document” part of the platform’s LIMS (http://ngs-lims.igbmc.fr). The mean DNA fragment size should be below 500 pb (Smear in a range of 100 to 700 bp). The platform recommends performing DNA shearing using sonication (ex: Covaris).
Quality	DNA sample must be depleted of any contaminants that may inhibit enzymatic reaction during the library preparation (proteins, salts, solvents, etc.). **An additional purification step, using Agencourt AMPure XP (Beckman Coulter), is necessary**. If not realized by the project manager, this purification will automatically be realized by the platform before quantification of starting material.
Shipping condition	In solution, in water and in “low binding” tubes to limit loss of DNA due to tube adsorption. 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 by the platform. Quality controls performed at steps 1 and 2 are available through the platform’s LIMS (http://ngs-lims.igbmc.fr). In case of problem the project manager will be contacted by e-mail.

1. Sample checking	
Quantity after **AMPure** beads purification (fluorometer)	≥ 2 ng
2. Library prep	
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 analysis of data	
Total number of clusters (i.e. number of reads in Single-read and number of reads ÷ 2 in Paired-end)	≥ 250 x L millions, where L represents the total number of lanes requested by the user
Quality score (Phred Score)>30	≥ 85% of bases for 50 b reads

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).
- Data aligned onto a reference genome, if available (BAM and BED file formats).

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

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. Data will be removed from the server six months after their delivery. 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", 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 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 following analyses can be performed:

- Peak calling: localization of binding sites of the DNA-associated protein of interest.
- Peak annotation: detected peaks are annotated with respect to genomic features (nearest genes or transcripts, exons, introns, etc.).
- Motif search or de novo motif discovery: detection of known motifs (from public databases such as JASPAR) or discovery of new motifs.
- Clustering analysis: comparison of enrichment profiles between several samples at given genomic locations.

This list is not exhaustive and we recommend the project manager that 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.