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: EPIGENOMICS

ChIP-seq combines Chromatin ImmunoPrecipitation with high throughput sequencing. It allows identification of DNA-protein interaction sites involved in gene expression regulation through transcription factors or post-translational modifications of histones. We propose standard ChIP-seq or Cut & Run[1] library preparation protocols. Moreover, to analyze nucleotide modifications (methylation, hydroxy-methylation,…), we propose protocols suitable for MEDIP-seq[2].

1 Available library preparation and sequencing options

1.1 Library preparation methods

Several library preparation protocols are currently available on the platform:

#	Application	Kit used by the platform	Quantity of chromatin	
			Minimal	Optimal
1	ChIP-seq	Diagenode MicroPlex	2 ng	10 ng
2	Cut & Run	Diagenode MicroPlex	2 ng	10 ng
3	MeDIP-seq	Takara DNA SMART ChIP-Seq	2 ng	20 ng

1.2 Sequencing options

Libraries are sequenced using the Illumina NextSeq 2000 technology. Both single-read and paired-end reads can be obtained with a size of 50 bp. Table below provides advices regarding read length depending on the objectives of your experiment. The sequencing depth depends on the purpose of the project, the nature of the samples and the library preparation protocol. The project manager is encouraged to contact us for more information regarding these different options if needed.

[1] Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. 6:e21856. doi:10.7554/eLife.21856

[2] Mohn F, Weber M, Schübeler D, Roloff T-C. Methylated DNA Immunoprecipitation (MeDIP). In: Tost J, ed. *DNA Methylation: Methods and Protocols*. Methods in Molecular Biology. Humana Press; 2009:55-64. doi:10.1007/978-1-59745-522-0_5

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

Application	Recommendations			
	Sequencing type	Read length	Number of million reads per sample	PhiX*
ChIP-seq - standard	Single-end	50 bp	≥30	-
ChIP-seq - non-standard (analysis of histone marks such as H3K27me3, interest in repetitive regions, …)	Paired-end	50 bp	≥30	-
Cut & Run	Paired-end	50 bp	≥10	-
MeDIP*	Single-end	50 bp	≥30	25%

* MeDIP libraries have a 9 nucleotides long adapter at the beginning of the sequences. This unbalanced nucleotide composition occurs during identification of cluster positions at the beginning of sequencing and leads to a drop of sequence quality and quantity on NextSeq 2000 sequencers. To correct for nucleotide compositions, Illumina recommends addition of a Spike-in library (Illumina PhiX control)[3].

1.3 Experimental design

It is recommended[4] to include at least 2 biological replicates in your experimental design (the number may vary depending on variability between replicates). It is also important to plan negative controls for each experimental condition in order to be able to extract specific enrichment from noise (see table below). We also encourage to try to reduce batch effects during sample preparation. Project managers who need advices to define the most appropriate experimental design according to their biological questions are encouraged to contact us before starting their experiments.

Application	Recommended negative control (for each condition)
ChIP-seq	• A portion of DNA sample removed prior to IP (Input DNA) • DNA chipped with a non-specific antibody such as IgG • DNA chipped with the same antibody as in the condition of interest but in a condition where no target protein is present • DNA obtained from IP without antibodies (mock)
Cut and Run	• DNA chipped with a non-specific antibody such as IgG
MeDIP	

2 Services provided

1. Sample checking:

- Quantity and quality check using a fluorometer (Qubit or Varioskan) and a capillary electrophoresis machine (Bioanalyzer from Agilent or Fragment Analyzer from AATI), only when the quantity of starting material is not limited

[3] https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/hiseq-phix-control-v3-technical-note.pdf

[4] 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

2. Library preparation:
- 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 run
- Libraries quantification and quality control by capillary electrophoresis (Bioanalyzer from Agilent or Fragment Analyzer from AATI)

3. Sequencing using the Illumina NextSeq 2000 technology:
- Single-read or paired-end sequencing with read lengths according to options specified on the LIMS (http://ngs-lims.igbmc.fr) for each project.

4. Primary data analysis:
- Demultiplexing and generation of FASTQ files
- Sequences 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 5 for more information)

3 Sample preparation (done by the project manager)

The project manager prepares the chipped DNA. Quality of results is closely related to initial sample quality. The project manager should therefore try to avoid any contamination or degradation.

Sample preparation recommendations	
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 DNA that should be provided to the platform	
Quantity	Depends on the library preparation protocol chosen by the project manager
Minimal volume	10 µl
DNA fragment size	Sample can be accompanied by a gel photograph or a capillary electrophoresis profile (ex. Bioanalyzer Agilent). The corresponding file can be uploaded onto the platform LIMS (http://ngs-lims.igbmc.fr) in the corresponding project. The mean DNA fragment size should be below 500 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, EDTA, salts, solvents, etc.). **An additional purification step using AMPure XP or SPRIselect (Beckman Coulter) beads is necessary.** If not realized by the project manager, this purification will automatically be realized by the platform **before quantification of starting material except for Cut & Run projects to preserves small DNA fragments**
Shipping conditions	In solution, in water on dry ice. Use “*low binding*” tubes to limit loss of DNA due to tube adsorption.

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

	Samples are to be registered in the LIMS, then, **the unique ID generated for each sample must be clearly indicated on its 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 after beads purification (Fluorimetry)	≥ minimal required quantity (depending on the library preparation protocol)
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 (including phiX)	≥ 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 in single-end and number of reads ÷ 2 in paired-end

5 Results delivery

For each sample, the following results are available:

- Raw sequencing data (nucleotide sequences in FASTQ format).
- A report (PDF file) presenting the number of raw reads, the percentage of bases with a Phred quality score greater than 30 and the size of raw files (FASTQ) to download;
- A text file providing the MD5 string of each FASTQ file to download. 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.

For ChIP-seq:

- Data aligned onto a reference genome, if available. File formats: BAM, BED, BIGWIG.

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. For instance, for ChIP-seq, the following analyses can be performed:

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

- 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 (using public databases of motifs such as JASPAR) or discovery of new motifs
- Clustering analysis: comparison and clustering of enrichment profiles between several samples at genomics locations or interest (e.g., peaks, TSS).

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.