

# PRODUCT SHEET: SMALL RNA-SEQ

Small RNA-seq allows expression analysis and discovery of new miRNA and other small non-coding RNA, using high-throughput sequencing technology.

## 1 Experimental design

### 1.1 Biological replicates

It is very important to include replicates in your experimental design (cf. Hansen et al., Nature Biotechnology 29:572-573, 2011). A randomized and balanced experimental design is also important. We also encourage project managers to try to reduce batch effects during sample preparation. Project managers who need advice to define the most appropriate experimental design according to their biological questions are encouraged to contact us before starting their experiments.

### 1.2 Library preparation methods

Two different RNA-seq 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 total RNA, as described on the following table.

#	Title of the service	Kit used by the platform	Total RNA quantity		Type of studied RNA	Stranded
			Minimal	Optimal		
1	Small RNA-seq / standard quantity (Truseq)	Truseq SmallRNA Sample Prep (Illumina)	1 µg	2 µg	All small RNAs with 5'P and 3'OH (desired size can be chosen by the project manager)	Yes
2	Small RNA-seq / low input (NEXTFLEX)	NEXTFLEX Small RNA-Seq Kit (Bioo Scientific)	10 ng	1 µg	All small RNAs with 5'P and 3'OH (desired size can be chosen by the project manager)	Yes

### 1.3 Sequencing options

Sequencing depth, which depends on the objectives of the experiment, is to be discussed with the platform at the time of project submission. We recommend 50 bp Single-Read sequencing. Note that the small RNA-seq libraries have a low base diversity (ATGC) which requires the addition of a balanced library (PhiX Illumina) up to 10%. The project manager will therefore have to request 10% more sequences than the desired number of sequences.

## 2 Services provided

### 1. Sample checking:

- Quantity and quality check using a fluorometer (Qubit or Varioskan) and a capillary electrophoresis machine (Bioanalyzer, Agilent).

### 2. Library preparation:

- Preparation of fragmented cDNA libraries and ligation of indexed sequencing adapters to DNA fragments. Indexes are DNA sequences of  $\geq 6$  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 NextSeq 2000 technology:
    - Single-read sequencing 1x50 bases.
  4. Primary data analysis:
    - Demultiplexing and generation of FASTQ files.
    - Sequencing quality check.
    - Detection of potential contaminations.
    - Generation of a report summarizing the methods used in the 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 prepares total RNA samples. Quality of small RNA-Seq results is closely related to initial samples quality. The project manager should therefore try to avoid any contamination (Phenol, DEPC, genomic DNA, etc.) or degradation.

Most RNA extraction kits using columns do not keep small RNAs. We therefore recommend using RNA purification kits retaining RNAs of small size (e.g. Trizol).

Characteristics of total RNA that should be provided to the platform	
Quantity	Depends on the library preparation protocol chosen by the project manager (cf. 1.2)
Minimal volume	10 µl
Quality	- OD260/OD280 ≥ 1.8 or no degradation on agarose gel or 28S/18S ≥ 1.6 and/or RIN ≥ 7 on an Agilent Bioanalyzer profile
Shipping condition	In solution, in water on dry ice. 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 (Fluorometry)	≥ minimal required quantity (depending on the library preparation protocol, cf. 1.2)
Quality (capillary electrophoresis)	28S/18S ≥ 1.6 and/or RIN ≥ 7
2. Library preparation	
Library profile (capillary electrophoresis)	peak(s) > 140 bp (depending on the type of small RNA chosen by the project manager)
Library purity (capillary electrophoresis)	Limited presence of adapter dimers (120-130 bp band). <ul style="list-style-type: none"> <li>• 120-130 bp band for standard protocol</li> <li>• 130-140 bp band for low input protocol</li> </ul>

3. Sequencing and primary data analysis	
Total number of reads (including the PhiX)	≥ Total number of clusters specified in the “Requested services” section from the submission form (pdf file that can be downloaded from the LIMS <a href="http://ngs-lims.igbmc.fr">http://ngs-lims.igbmc.fr</a> , in the “Document” tab for each project)
Mean quality score (Phred Score) > 30	≥ 85% of bases.

## 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 strings 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.

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 guarantee 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:

- 3' adapter trimming.
- Alignment on a reference genome.
- Quantification of known miRNA and other small non-coding RNA using public databases (miRBase, Rfam, etc.).
- Normalization and statistical analysis in order to highlight significantly differentially expressed small non-coding RNA between different conditions.
- Prediction of new miRNA.

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