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

+33 3 88 65 34 26
thibault@igbmc.fr
keime@igbmc.fr
genomeast.igbmc.fr

PRODUCT SHEET: RNA-SEQ

RNA-seq provides a snapshot of the transcriptome, thus allowing qualitative and quantitative study of gene expression. For this application, we offer different alternatives using either Illumina sequencing technology or MGI technology.

1 Available library preparation protocols

Several different RNA-seq library preparation protocols are currently available on the platform. The choice of the most appropriate protocol mainly depends on the selected sequencing technology, the available amount of total RNA and the type of RNAs of interest.

We recommend choosing the same protocol within a project, i.e., if you want to compare your RNA-seq data with a previously generated RNA-seq dataset we recommend using the same protocol if possible.

1.1 Illumina technology

#	Title of the service[a]	Kit used by the platform	Total RNA quantity		Type of studied RNA	Stranded[b]
			Minimal	Optimal		
1[c]	Library prep stranded mRNA Truseq (Illumina)	Truseq Stranded mRNA Prep (Illumina)	200 ng	1 µg	Only polyA+ RNA of size > 100 b	Yes
2	Library prep stranded mRNA Ligation (Illumina)	Stranded mRNA Prep, Ligation (Illumina)	25 ng	1 µg	Only polyA+ RNA of size > 100 b	Yes
3	Library prep mRNA ultralow Smarter (Takara)	SMART-Seq v4 UltraLow Input RNA kit (Takara) + Nextera XT DNA sample preparation Kit (Illumina)	100 pg	10 ng	Only polyA+ RNA of size > 100 b	No
4[d]	Library prep 3'mRNA (Lexogen)	QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina (FWD) (Lexogen)	1 ng	500 ng	Only 3' end of polyA+ RNA	Yes
5[e]	Library prep 3'mRNA Single cell (10x Genomics)	Chromium Next GEM Single Cell 3' Reagent Kits (10X Genomics)	1 cell	1 cell	Only 3' end of polyA+ RNA	No
6[c, d, f]	Library prep stranded total RNA Ribozero Truseq (Illumina)	Truseq Stranded Total RNA Sample Prep (Illumina)	100 ng	1 µg	All RNA of size > 100 b	Yes
7[d, f]	Library prep stranded total RNA Ribozero Plus Ligation (Illumina)	Stranded Total RNA Prep Ligation with Ribo-Zero Plus (Illumina)	1 ng	1 µg	All RNA of size > 100 b	Yes
8[e]	Library prep small RNA Truseq (Illumina)	Truseq SmallRNA Sample Prep (Illumina)	1 µg	2 µg	All small RNAs with 5'P and 3'OH	Yes

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

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

+33 3 88 65 34 26
thibault@igbmc.fr
keime@igbmc.fr
genomeast.igbmc.fr

[b] Stranded or directional protocols keep the information of the transcribed strand. The resulting reads are in reverse strand as compared to the transcribed one for all protocols except for protocols #4 and 8 for which the resulting reads are in the same strand as the transcribed one.
[c] These protocols (#1 and #6) can be used upon request, for old projects started with the same kits. For new projects, we recommend using the last version of these protocols (#2 and #7, respectively).
[d] These protocols (#4, #6 and #7) are suitable to study degraded RNA extracted from formalin-fixed, paraffin-embedded (FFPE) tissues. Be aware that the platform cannot guarantee the quality of results on such samples.
[e] We encourage project managers interested in this application to read our dedicated product sheet.
[f] For total RNA-seq protocols (#6 and #7), the efficiency of ribosomal depletion is variable between samples. Thus, the platform cannot guarantee the proportion of resulting reads corresponding to rRNA in the results.

1.2 MGI technology

#	Title of the service[a]	Kit used by the platform	Total RNA quantity		Type of studied RNA	Stranded [b]
			Minimal	Optimal		
9	Library prep stranded mRNA (MGI)	NEBNext® Poly(A) mRNA Magnetic Isolation Module (NEB) + MGIEasy RNA Directional Library Prep Kit (MGI)	100 ng	1µg	Only polyA+ RNA of size > 100 b	Yes
10	Library prep stranded total RNA (MGI)	RiboPOOL (si-TOOLs Biotech) MGIEasy RNA Directional Library Prep Kit (MGI)	100 ng	1µg	All RNA of size > 100 b	Yes

[a] Titles as listed on the LIMS of the platform (http://ngs-lims.igbmc.fr)
[b] Stranded or directional protocols keep the information of the transcribed strand. The resulting reads are in reverse strand as compared to the transcribed

2 Sequencing options

The table below provides advice regarding read length depending on the objectives of your experiment. Only a subset of all possible questions that can be studied using RNA-seq are listed, therefore the project leader is encouraged to contact us for more information regarding these different options if needed.

Aim of the study	Minimal sequencing recommendation
Expression quantification on annotated genes	Single read 50 pb
Alternative splicing analysis, new transcript identification	Paired-end 100 pb

Sequencing depth depends on the objectives of the experiment, the nature of the samples and the library preparation method. For instance, using library preparation protocols aimed at studying all types of RNAs at the same time (protocols #6, #7 and #10), we expect to sequence a wider variety of RNA molecules compared to the results obtained using polyA+ protocols. Therefore, more reads are needed to achieve the same coverage on polyadenylated RNA. For mammalian tissues, when the goal of the experiment is to quantify the expression of annotated genes, we recommend to sequence 10 million reads per sample with a 3'end mRNA-seq protocol (#4), 30 million reads per sample for a full-length mRNA-seq protocol (#1, #3 and #9), and 50 million reads per sample for a total RNA-seq protocol (#6, #7 and #10). For experiments where the sensitivity of detection is important, i.e. to discover novel transcripts or precisely quantify known transcript isoforms, a higher sequencing depth is needed.

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

+33 3 88 65 34 26
thibault@igbmc.fr
keime@igbmc.fr
genomeast.igbmc.fr

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.

3 Services provided

1. Assistance in setting up the project with a biologist and a bioinformatician from the platform:
   - Help with experimental design,
   - Reminder of the requirements on the starting samples.
2. Sample checking:
   - Quantity and quality check using a fluorometer (Qubit or Varioskan, Thermo Fisher) and a capillary electrophoresis machine (Bioanalyzer, Agilent), only when the quantity of starting material is not limited.
3. Library preparation:
   - Preparation of fragmented cDNA 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 or Fragment Analyzer, Agilent).
4. Sequencing using Illumina NextSeq 2000 or MGI DNBSEQ G400-RS:
   - Single-read or paired-end sequencing with read lengths according to options specified on the LIMS (http://ngs-lims.igbmc.fr) for each project.
5. 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.
6. Downstream data analysis (optional, see section 7 for more information).

4 Sample preparation (done by the project leader)

The project leader prepares total RNA samples. Quality of RNA-seq results is closely related to initial samples quality. The project leader should therefore try to avoid any contamination (Phenol, DEPC, genomic DNA, etc.) or degradation.

Characteristics of total RNA that should be provided to the platform	
Quantity	Depends on the library preparation protocol chosen by the project manager
Minimal volume	10 µl
Quality*	OD260/OD280 ≥ 1.8 No degradation on agarose gel or 28S/18S ≥ 1.6 and/or RIN ≥ 7 on an Agilent Bioanalyzer profile
Shipping conditions	In solution, in water on dry ice. Sample names must be clearly indicated on the tubes as well as in the platform's LIMS

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

+33 3 88 65 34 26
thibault@igbmc.fr
keime@igbmc.fr
genomeast.igbmc.fr

**For RNAs extracted from FFPE tissues, we recommend to use samples with a DV200≥30% (fraction of RNA fragments>200 nucleotides) on an Agilent Bioanalyzer profile.*

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), quality controls perform at step 3 are available in the report provided with the data (see section 6).

5.1 Illumina technology

1. Sample checking	
Quantity (Fluorometry)	≥ minimal required quantity (depending on the library preparation protocol)
Quality (capillary electrophoresis)	Ratio 28S/18S ≥ 1.6 and/or RIN ≥ 7
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 pb 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 in single-read and number of reads ÷ 2 in paired-end*

5.2 MGI technology

1. Sample checking	
Quantity (fluorimetry)	≥ 100 ng
Quality (capillary electrophoresis)	Ratio 28S/18S ≥ 1,6 and/or RIN ≥ 7
2. Individual library preparation	
Double strand cDNA profile after adaptor ligation and amplification (capillary electrophoresis)	Average size ranging from 300 to 400 bp
3. DNA nanoballs (DNB) preparation per sequencing lane	
Quantity of the pool of single strand circularized libraries (fluorimetry)	≥ 80 fmol
Concentration of DNA nanoballs (fluorimetry)	≥ 8 ng/µl
4. Sequencing and primary data analysis	
Total number of reads per lane	≥ 350 million (single read) ≥ 2x350 million (paired-end)
Quality score (Phred score) > 30	≥ 85% of bases

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

+33 3 88 65 34 26
thibault@igbmc.fr
keime@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[1], 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[2].

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 on a reference genome taking into account reads spanning splice junctions.
- Gene expression quantification using known annotations.
- Normalization, exploratory data analyses and statistical analyses to highlight significantly differentially expressed genes between different conditions.
- Functional analysis.
- Alternative splicing analysis.

This list is not exhaustive, and we recommend the project leader 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.

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