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: RNA-SEQ

RNA-seq provides a snapshot of the transcriptome, thus allowing qualitative and quantitative study of gene expression.

1 Available library preparation and sequencing options

1.1 Library preparation methods

Several 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 and the type of RNAs of interest, as described on the following table.

#	Title of the service	Kit used by the platform	Total RNA quantity		Type of studied RNA	Stranded[a]
			Minimal	Optimal		
1[b]	Stranded mRNAseq/ standard quantity (Truseq)	Truseq Stranded mRNA Prep Illumina	200 ng	1 µg	Only polyA+ RNA of size > 100 b	Yes
2	Stranded mRNA-seq/standard quantity (Ligation)	Illumina Stranded mRNA Prep, Ligation	25 ng	1 µg	Only polyA+ RNA of size > 100 b	Yes
3	mRNAseq/low input (Smarter)	SMART-Seq v4 UltraLow Input RNA kit (Clontech) + Nextera XT DNA sample preparation Kit (Illumina)	100 cells	10 ng	Only polyA+ RNA of size > 100 b	No
4[c]	3' mRNA-seq (Lexogen)	QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina (FWD) (Lexogen)	1 ng	500 ng	Only 3' end of polyA+ RNA	Yes
5[d]	3' mRNA-seq/ single cell	Chromium Next GEM Single Cell 3' Reagent Kits	1 cell	1 cell	Only 3' end of polyA+ RNA	No
6[b,c,e]	Total RNAseq Ribozero/standard quantity (Truseq)	Truseq Stranded Total RNA Sample Prep Illumina	100 ng	1 µg	All RNA of size > 100 b	Yes
7[c,e]	Stranded Total RNA-seq Ribozero Plus (Ligation)/Standard quantity	Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus	1 ng	1 µg	All RNA of size > 100 b	Yes
8[d]	Small RNA-seq	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

[a] 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.

[b] These protocols (#1 and #6) can be used upon request, for old projects started with the same kits to prevent batch effect. For new projects, we recommend using the last version of these protocols (#2 and #7, respectively).

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

[c] *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 can not guarantee the quality of final results on such samples.*
[d] *We encourage project managers interested in this application to read our dedicated product sheet.*
[d] *For total RNA-seq protocols (#6 and #7), the efficiency of ribosomal depletion is highly variable between samples. Thus, the platform can not guarantee the proportion of resulting reads corresponding to rRNA in the final results.*

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 to use the same protocol if possible.

1.2 Sequencing options

Table below provides advices 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 manager is encouraged to contact us for more information regarding these different options if needed.

Goals of the experiment		Recommendations		
Aim of the study	**Type of studied RNA**	**Library preparation***	**Sequencing type**	**Read length**
Expression quantification on annotated genes	polyA+	Stranded mRNAseq/ standard quantity (Ligation)	Single-read	50 bp
	all	Stranded Total RNA-seq Ribozero Plus (Ligation)/Standard quantity	Single-read	50 bp
Alternative splicing analysis guided with a known genome, new transcript identification or *de novo* transcriptome assembly	polyA+	Stranded mRNAseq/ standard quantity (Ligation)	Paired-end	100 bp
	all	Stranded Total RNA-seq Ribozero Plus (Ligation)/Standard quantity	Paired-end	100 bp

**Considering you have sufficient amount of starting material. If not, please refer to the previous table for alternative protocols.*
Recommendations for a new project. If you want to compare with previous RNA-seq data we advise to keep the same library preparation protocol.

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 and #7), 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 to #3), and 50 million reads per sample for a total RNA-seq protocol (#6 and #7). 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.

1.3 Experimental design

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 advices to define the most appropriate experimental design according to their biological questions are encouraged to contact us before starting their experiments.

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 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) 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 from Agilent or Fragment Analyzer from AATI).
4. Sequencing using 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.
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 6 for more information)

3 Sample preparation (done by the project manager)

The project manager prepares total RNA samples. Quality of 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.

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

**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.*

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

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)
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 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 string of each FASTQ file. 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.

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", following data delivery, the project manager is responsible for his data to be saved and archived on its own. Data will be removed from the Platform server six months after their delivery.

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 on a reference genome taking into account reads spanning splice junctions.
- Gene expression quantification using known annotations.

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

- Normalization, exploratory data analyses and statistical analyses in order to highlight significantly differentially expressed genes between different conditions.
- Functional analysis.
- Alternative splicing analysis.

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.