thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

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

Seven 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.

1†	mRNAseq/ standard quantity	TruSeq RNA Sample Prep	200 ng	1 g	Only polyA RNA of size 100 b	No
2	Stranded mRNAseq/ standard quantity	Directional mRNA-Seq SamplePrep	200 ng	1 g	Only polyA RNA of size 100 b	Yes
3	mRNAseq/low input (Smarter)	SMART-Seq v4 UltraLow Input RNA kit Nextera XT DNA sample preparation Kit	100 cells	10 ng	Only polyA RNA of size 100 b	No
4	mRNA-seq/ single cell	SMARTer Ultra Low RNA Kit for the Fluidigm C1 System Nextera XT DNA sample preparation Kit	1 cell	1 cell	Only polyA RNA of size 100 b	No
5	Total RNAseq Ribozero/standard quantity	Truseq Stranded Total RNA SamplePrep	100 ng	1 g	All RNA of size 100 b	Yes
6	Total RNAseq/ low input (Ovation)	Ovation RNA-Seq System V2 Ovation SP Ultralow Library systems	500 pg	10 ng	All RNA of size 100 b	No
7	Small RNA-seq	Truseq SmallRNA SamplePrep	1 g	2 g	All small RNAs with 5'P and 3'OH (desired size can be chosen by the project manager)	Yes

** Whether the protocol is directional or not. Stranded protocols keep the information of the transcribed strand, i.e. the resulting reads are in reverse strand as compared to the transcribed one for protocols 2 and 5 and in the same strand as the transcribed one for protocol 7. In contrary, non-stranded protocols do not keep the information of the transcribed strand.*

† This protocol can be used upon request, for old projects started with this protocol. We do not recommend this protocol for new projects, since protocol #2 offers the same advantages and in addition keep the information of the transcribed strand.

‡ We encourage project managers interested in small RNA-seq to read our dedicated product sheet.

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.

thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

Libraries are sequenced using the Illumina HiSeq 4000 technology. Both single-read and paired-end reads can be obtained with a size of 50 or 100 bp. 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.

Expression quantification on annotated genes	polyA	Stranded mRNAseq/ standard quantity	Single-read	50 bp
	all	Total RNA Ribozero/ 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	Paired-end	100 bp
	all	Total RNA Ribozero/ 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.

depends on the objectives of the experiment, the nature of the samples and the library preparation method. For instance, using library preparation protocols used to study all types of RNAs at the same time (protocols #5 and #6), 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, we thus generally recommend multiplexing up to 8 libraries per lane prepared with polyA protocols (protocols #1 to #3) or up to 5 libraries per lane prepared with total RNA protocols (protocols #5 and #6) when the goal of the experiment is to quantify the expression of annotated genes. 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.

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.

thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

1. 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.
2. 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 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 generation of clusters on the Cbot (Illumina).
   - Single-read or paired-end sequencing of 50 or 100 bp according to the read length chosen by the project manager.
4. Primary data analysis:
   - Demultiplexing and generation of FASTQ files.
   - Adapter dimer removal.
   - 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 (one report per sample and one global report for the project).
5. Downstream data analysis (optional, see section 6 for more information)

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.

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 8 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

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).

thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

Quantity (Fluorometry)	minimal required quantity (depending on the library preparation protocol)
Quality (capillary electrophoresis)	Ratio 28S/18S 1.6 and/or RIN 8
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)
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 75 of bases for 100 b reads

For each sample, the following results are available:

- Raw sequencing data (nucleotide sequences in FASTQ format. The files contain reads passing quality filters and do not contain dimer adapter sequences).
- A report with sequences quality controls (in PDF 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 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 can use these MD5 to check the integrity of the files after their download (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.

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 in order to highlight significantly differentially expressed genes between different conditions.
- Functional analysis.
- Transcript prediction.
- Alternative splicing analysis.

thibault@igbmc.fr
jost@igbmc.fr
genomeast.igbmc.fr

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.