Technical document for library preparation and sequencing

The general process of sequencing service at BMKGENE consists of six parts, as illustrated below in Chart 1: sample shipment, sample quality assessment, library preparation, sequencing, sequencing data quality control, and data release. During this process, the client would need to provide permission twice to proceed with the project, respectively after the sample quality assessment and after the data quality control. Only if the confirmation of proceeding is received at BMKGENE, will the project move forward to the next stage. A fast-track process without clients' confirmation could be issued as well if it is requested by the client.

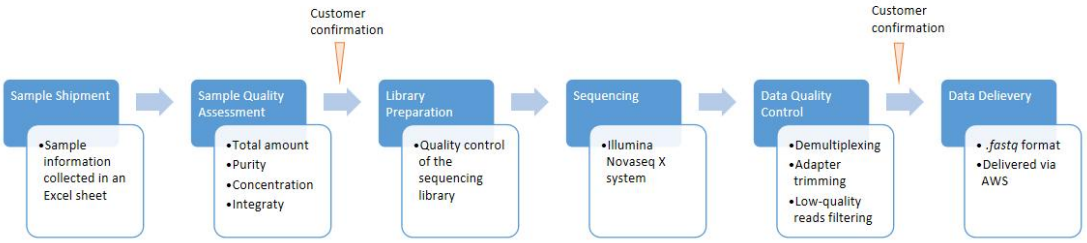

Chart 1. General process of sequencing service at BMKGENE

The detailed information on reagents and methods used in sample quality assessment, library preparation,and sequencing for the ten required sequencing services is described in the rest of part of this document.

Please be informed that, if more flexibility and variability are required in the project, we highly encourage clients to discuss with us before declining the possibility of proceeding. BMKGENE has developed a grading system for handling low-quality samples based on the sequencing experience in the past fifteen years. This system gathers successful and failed experiences in samples with low quality and low concentration, samples with severe pollution, and samples with degraded nucleic acid molecules. With the ongoing gaining of experience, we constantly adjust our sample requirements, aiming to widen the sequencing possibilities for challenging samples and projects.

2. RNA sequencing with Illumina PE150 mode

1) Sample quality assessment

a) Sample quality control

Upon receiving the samples, sample quality will be assessed. Below are the systems used to assess sample qualities.

Measurement	System	Kit
Concentration	Qubit 4.0	Qubit TM dsDNA HS Assay Kit
Molecule integrity	Agarose gel electrophoresis	-
Purity/Impurity (if request)	LabChip GX Touch HT Nucleic Acid Analyser	RNA Assay Reagent Kit (Revvity)

b) Sample requirements

- Microbial samples

Measurement	Requirement
Qubit Concentration	≥ 50 ng/μl
Total Amount	≥ 1 μg
Volume	≥ 20 μl
OD260/280	1.8 - 2.0
OD260/230	1.0 - 2.5
RIN	> 6.5
Agarose gel electrophoresis	The main band is clear. No or limited degradation.

- Human/animal/plant/fungal samples

Measurement	Requirement
Qubit Concentration	≥ 10 ng/μl
Total Amount	≥ 200 ng
RIN	For plants: RIN≥ 4.0; For animals: RIN≥ 4.5; 5.0≥ 28S/18S≥ 1.0;
Volume	≥ 20 μl
Nanodrop OD260/OD280	1.7-2.5
Nanodrop OD260/OD230	0.5-2.5

c) Sample quality reports will be sent to our customers via email for their ***confirmation***.

Note: Clients can proceed with risky samples which do not reach the sample requirements listed above. Our lab technicians can give a suggestion based on their experience on whether it is worth proceeding with the risky samples. Clients can then make decisions.

2) Library preparation

a) rRNA depletion library preparation

- The rRNA depeletion in the bacterial sample will be performed with the TransNGS® Ribo-Cap rRNA Depletion Kit (Bacteria) (TransGen Biotech), by following the official protocol. Here listed the four major steps in the workflow.

  i. rRNA probe hybridization

ii. RNase H digestion

iii. DNase I digestion

iv. RNA magnetic bead-based RNA purification

- For rRNA removal in animal-origin samples, the TransNGS® Ribo-Cap rRNA Depletion Kit (Animal) (TransGen Biotech) can also be chosen to remove host rRNA. The official manual can be found here (https://www.transgenbiotech.com/download/pdf/KD101_2025-08-07.pdf).
- For rRNA removal in plant-origin samples, the TransNGS® Ribo-Cap rRNA Depletion Kit (Plant) will be applied.
- For fungal rRNA removal, the three TransNGS kits listed above will be used.
- **Library preparation** will be performed with the Hieff NGS Ultima Dual-mode mRNA Library Prep Kit for Illumina (Yeasen) by following the official protocol. Here listed the major steps in the workflow.

  i. RNA fragmentation

     - The fragmentation method is enzymatic by default. By request, sonic fragmentation could be applied as well.

  ii. Synthesis of the 1st strand cDNA

  iii. Synthesis of the 2nd strand cDNA with the stranded method

  iv. Purification of the double-stranded DNA

  v. Adapter ligation

  vi. Purification and size selection for the ligated reaction

  vii. Library amplification with PCR

  viii. PCR product purification

b) polyA selected library preparation

- mRNA enrichment with polyA selection will be performed with the VAHTS mRNA Capture Beads (Vazyme) by following the official protocol (https://www.vazymeglobal.com/product-center/rna-library-preparation-module/vahts-mrna-capture-beads-2-0).
- **Library preparation** will be performed with the VAHTS Universal V8 RNA-seq Library Prep Kit for Illumina. The detailed protocol can be found here (https://www.vazymeglobal.com/product-center/rna-library-preparation-kit/vahts-universal-v8-rna-seq-library-prep-kit-for-illumina). Here listed the major steps in the workflow.

  i. mRNA fragmentation.

     - The fragmentation method is enzymatic by default. If request, sonic fragmentation could be applied as well.

  ii. Synthesis of the 1st strand cDNA.

  iii. Synthesis of the 2nd strand cDNA, End Repair, and dA-Tailing.

  iv. Adapter ligation.

  v. Purification and size selection for the ligated reaction.

  vi. Library amplification with PCR.

  vii. PCR product purification.

3) Library quality control

a) The Qubit 4.0 fluorescence quantifier will be used for preliminary DNA quantification. The qualified DNA concentration should be above 1 ng/μl.

b) The Qsep400 high-throughput analysis system will be used to evaluate the inserted fragments of the library, and after that, qPCR will be used to determine the effective concentration of the library. The major peak of the inserts should be around 350 - 550 bp, and the effective concentration of the library should be above 2 nM.

4) Sequencing

- Qualified libraries will be pooled and sequenced with other samples in a 25B flow cell on the Illumina NovaSeq X system with the PE150 mode.

5) Data quality control

For all the samples matching our sample quality standards, BMKGENE guarantees high-quality sequencing data to meet the agreed data volume in the contract. Here, the high quality refers to the clean data in which more than 85% of the bases are above Q30.

The resulting raw reads will be trimmed by quality and the adapter sequences, which will result in the clean reads for delivery. The detailed process is listed below.

a) Raw sequencing data quality control.
   i. Assessing total data volume, read length, and read numbers.
   ii. Counting the percentage of reads with a Phred quality score above Q20 and Q30.
   iii. Adapter trimming and demultiplexing.
   iv. Filtering out reads that have more than 50% bases with a Q score below 10.

6) Bioinformatic workflow

A brief overview of our bioinformatic service for the bacterial RNA data is listed below.

i. Quality control.
ii. Alignment with the reference genome.
iii. Gene expression quantification.
iv. Sample correlation based on the expression profiles.
v. Differential expression analysis.
vi. Function annotation and enrichment on differentially expressed genes.
vii. Differential alternative splicing analysis.
viii. Small RNA prediction and annotation.
ix. Novel gene prediction.
x. Transcript's SNP identification.

7) Data delivery and storage

a) Once data quality control is completed, data quality report will be sent to our customers.

b) Upon receiving the report, our clients would need to ***confirm*** whether they are satisfied with and will accept the sequencing data.

c) Once the confirmation is received by BMKGENE, the project is considered completed, and the data release process starts.

d) By default, the sequencing data will be delivered by AWS (Amazon Web Services) in *.fastq* format. It is also flexible to choose data delivery by hard drive or other methods.

e) Md5 files for each batch of data and reports will be provided together in data release.

f) Sequencing data will be stored free of charge for six months upon receiving the project completion confirmation at BMKGENE. **This length of data storage period**

should be stated in the final contract. Sequencing data will be automatically deleted before the first week of the 7th month if no further notice is received .

g) The AWS link for data downloading will be available for one month after project completion, due to technical restriction. However, upon request the data will be uploaded again to the AWS after the first link expires.

h) The RNA samples will be stored for four months after project completion. **This length of sample storage period should be stated in the final contract.**

i) The data delivery or turnaround time is shown as below, which starts to count when samples are received at BMKGENE. If a shorter TAT is required, please consult with the local sales manager.

Sample Number per Batch	Working days
1-30	30