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.
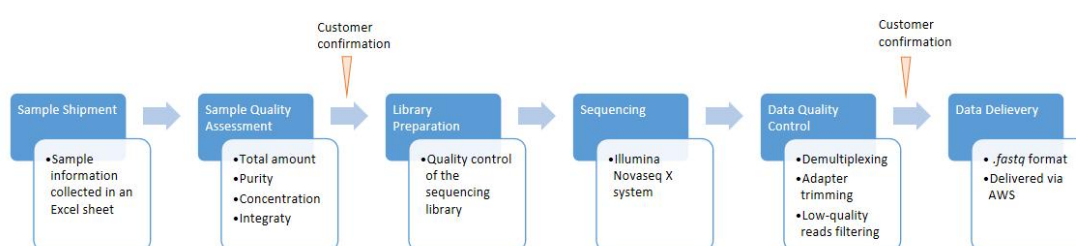

Chart 1. General process of sequencing service at BMKGENE

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.

1. Whole Genome Sequencing with Illumina

1) Sample quality assessment

a) Sample quality control

Upon receiving the samples, sample quality will be assessed. Below are the methods used in assessing sample qualities.

Measurement	System	Kit
Concentration	Qubit 4.0	Qubit ™ dsDNA HS Assay Kit
Molecule integrity	Agarose gel electrophoresis	1% agarose gel
Purity/Impurity	LabChip GX Touch HT Nucleic Acid Analyser	DNA Assay Reagent Kit (Revvity)

b) Sample requirements

- Microbial samples

Measurement	Requirement
Qubit Concentration	≥ 1 ng/µl
Total Amount	≥ 60 ng
Volume	≥ 20 µl
Agarose gel electrophoresis	The main band is clear. No or limited degradation.

- Human/animal/plant samples

Measurement	Requirement
Qubit Concentration	≥ 1 ng/µl
Total Amount	≥ 30 ng
Volume	≥ 20 µl
Agarose gel electrophoresis	The main band is clear. No or limited degradation.

c) Sample quality reports will be sent to our customers via E-Mail for *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, and clients can then make a decision.

2) Library preparation

- After quality control on DNA samples, sequencing libraries will be prepared by using the VAHTS Universal DNA Library Prep Kit for Illumina. The workflow of library preparation follows the official manual of the kit (https://www.vazymeglobal.com/product-center/dna-library-preparation-kit/vahts-universal-dna-library-prep-kit-for-illumina-v4). Major steps of this procedure are listed below.

  i. DNA fragmentation and end repair.

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

  ii. Adapter ligation.

  iii. Purification of the ligated DNA products.

  iv. Fragment size selection in the purified DNA.

  v. PCR enrichment.

      vi.   Purification of PCR products.

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 Qsep-400 bio-fragment analyzer will be used to evaluate the inserted size of the library, and the Qubit 4.0 will be used to estimate the rough concentration of the library. Additionally, qPCR will be used to determine the effective concentration of the library. The major peak of the inserts should be around 430 - 530 bp, and the effective concentration of the library should be above 4 nM.

4) Sequencing

- Qualified libraries will be pooled and sequenced 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 clean data is generated through a data quality control process. 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) 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
1-150	30