

Next Generation Sequencing FAQ

Is your company certified as a service provider for Illumina sequencing?

Yes, we are a certified Illumina service provider.

What kind of analysis software do you use for sequence analysis?

The Illumina sequencers come with the integrated software that can perform the regular bioinformatics analysis functions. In addition, we have extensive experience with other open access software such as Galaxy, Adnuril, BioBike, etc. In other words, we are willing to perform additional bioinformatics analysis with different software and tools to satisfy a variety of demands from our customers.

What kind of QC methods do you adopt for the customer's sample(s)?

We will perform QC on your Total RNA samples prior to sequencing them, using the Agilent Bioanalyzer (industry standard) to determine the RNA Integrity Number (RIN). If the RIN is lower than 8, we consider these samples as having failed QC. If the samples do not pass QC, we will ask you to send us new samples. If we do not receive new samples, thereby cancelling the sequencing order, there will be a QC fee of \$50/sample. This fee will be waived if you choose to continue with the sequencing order. A strict positive and negative control will be applied to this QC process.

The library QC will also be performed using the Agilent Bioanalyzer to determine library size and purity. Also, prior to loading the libraries on the sequencer, we perform qPCR quantification. The cost for this is included in the sequencing service.

For the QC of the final data from our sequencing service, the integrated software will generally do the job and provide adequate QC information, however, we can also provide additional QC data such as FASTQC upon request.

The data we output will pass our Q30 filter, which means that the error rate in base calls is less than 1 in 1000, or 0.1%. We will get a percentage at the end of the run which states the percentage of bases that have a Q score > 30, and this percentage is usually $\geq 85\%$.

What is the lead time for your WGS, WES, and RNA-Seq services?

Our sequencing lead times depend on workload, but are typically 2-4 weeks for sequencing data + 2-3 weeks for standard analysis. Please inquire for lead times for more detailed analysis.

Can you recommend any preferred RNA/DNA isolation kits to provide the best quality samples for these services?

For miRNA sequencing, you may isolate your RNA sample with a mirVana™ miRNA Isolation Kit, mirVana™ PARIS™ Kit or PureLink® miRNA Isolation kit to enable adequate enrichment of your sample. If preferred, it should be okay to use Qiagen's kit, as long as the RNA Integrity Number (RIN) of the sample is above 8.0 to make sure there is no significant degradation. (Samples with a RIN lower than 8.0 may contain smaller degraded RNA fragments that can be sequenced in addition to the miRNAs). We prefer receiving Total RNA sample however so we can perform the bioanalyzer QC to check the RNA quality before beginning library construction. There will be no extra-charges if you submit Total RNA instead of isolated miRNA.

For miRNA sequencing we require 200ng-2ug of total RNA in 10ul of nuclease free water, as quantified with a fluorometric method. Lower amounts might result in inefficient ligation and low yield. If you are supplying purified miRNA, please submit a minimum of 50-100 ng of purified small RNA in 10ul is required. Purified small RNAs must be in nuclease free water or 10 mM Tris-HCl, pH 8.5.

For other services, there are generally no preferred DNA/RNA isolation kits as long as minimum requirements for QC are met.

Do I need any special software to view the data report files?

You will first need to unzip the fastq.gz files with winrar to view with vim, nano, gedit, Notepad plus plus. Nextgen workbench generally can view the data immediately without unzipping.

The RNASeq_sample_cufflinks_output.tar.gz will also need to be unzipped to see the text files by vim, nano, gedit, or Notepad plus plus. The unzipped data can also be imported as text into the data tab of Excel 2007. Other versions may differ.

To analyze further, Excel can sort, filter, find, match, etc and has many functions for conditional coloring, graphing, etc.

For the data, why are there no "ALT", "POS" and "ID" in your CSV file?

CSV files are meant to be used in conjunction with the raw fastq/bam files by certain commercial bioinformatics software. Many have a proprietary algorithms that takes the ALT POS and ID from the Fastq and indexes them for its own use. They are not meant to be used alone.

Would you let me know the description for the following: Genotype Sample_0, Coverage Sample_0, Coverage By Allele Sample_0, Variant Quality Sample_0, and Genotype Quality Sample_0?

- Genotype (Sample Name) = nucleotide.
- Coverage (Sample Name) = will fail if there are insufficient reads.
- Coverage by allele (Sample Name) = will fail on one strand (+/-) read if only one side of the pair was sequenced.
- Variant Quality (Sample Name) = confidence of variant called.
- Genotype Quality (Sample Name) = confidence of genotype sequenced.

Concerning the column of "CHROM", would you let me know the rules of naming?

type of accession number | actual number | type of sequence | actual curated version used

Does your NGS service including RNA isolation?

We do not provide RNA isolation service by default. However we can provide this service. For more information, please refer to Cat#IR19001.

How many biological replicates do I need for each condition?

To increase confidence and reduce experimental error it is suggested that you submit at least 3 replicates per sample. Note that this is to serve as a guideline only and the final number of replicates and samples is to be determined by the end user based on their final experimental conditions.