UB Next-Generation Sequencing and Expression Analysis Core
Illumina HiSeq 2500 Sequencing Services Document

Researchers interested in illumina sequencing services are encouraged to contact the UB Next-Generation Sequencing and Expression Analysis Core (UB Next-Gen Core) to discuss potential projects. We can be reached by email at **cbi-ubnextgencore@buffalo.edu** or by phone at (716) 881-7514. During the consultation period the UB Next-Gen Core will address any researcher questions and discuss sample preparation, number of samples, multiplexing options, type of sequencing run, number of flow cell lanes, and data analysis. When the experimental details are agreed upon, the core can provide the researcher with a custom cost estimate for the project if desired.

Experimental Design

Common questions during the experimental design process include: how many reads will be generated; what is the average length of the reads; how many samples can be multiplexed; and how much coverage can be expected. Because of the flexibility of the illumina HiSeq 2500 the answers can differ depending on several variables, such as the run type and multiplexing strategy. To help better answer these questions, the UB Next-Gen Core has provided information about the different run options available and associated outputs. Keep in mind that the numbers given are approximations, and are meant only to assist the researcher during the experimental design process. The UB Next-Gen Core makes every effort to maximize data quality and quantity, but we **do not** guarantee the output that the researcher will receive from a run.

Run Options:

The illumina HiSeq 2500 can be run in two modes: high-output and rapid. The high-output mode is suitable for larger projects that have more samples or where high depth of coverage is necessary. These flow cells have 8 lanes and produce a greater number of reads per lane with run times varying from 3 – 14 days. Read lengths for the illumina HiSeq 2500 are determined by the number of sequencing cycles. In high-output mode, researchers can choose between 50-cycle or 100-cycle runs, which correspond to read lengths of 50bp or 100bp. Alternatively, the rapid mode provides faster results for fewer samples and can produce reads of up to 150bp. These flow cells have 2 lanes that produce fewer reads per lane with run times of 7 – 40 hours. In rapid mode, researchers can choose between 50-cycle, 100-cycle, or 150-cycle runs. Both modes offer the option to perform the run as single read (unidirectional) or paired-end (bidirectional).

Table 1. Approximate number of reads and bases to expect per flow cell lane depending on run type for the illumina HiSeq 2500.

Run Type	Reads Per Lane	Bases Per Lane
High-Output - 50-Cycle Single Read	150,000,000 - 200,000,000	7,500,000,000 - 10,000,000,000
High-Output - 100-Cycle Single Read	150,000,000 - 200,000,000	15,000,000,000 - 20,000,000,000
High-Output - 50-Cycle Paired End	300,000,000 - 400,000,000	15,000,000,000 - 20,000,000,000
High-Output - 100-Cycle Paired End	300,000,000 - 400,000,000	30,000,000,000 - 40,000,000,000
Rapid - 50-Cycle Single Read	110,000,000 - 150,000,000	5,500,000,000 - 7,500,000,000
Rapid - 100-Cycle Single Read	110,000,000 - 150,000,000	11,000,000,000 - 15,000,000,000
Rapid - 150-Cycle Single Read	110,000,000 - 150,000,000	16,500,000,000 - 22,500,000,000
Rapid - 50-Cycle Paired End	220,000,000 - 300,000,000	11,000,000,000 - 15,000,000,000
Rapid - 100-Cycle Paired End	220,000,000 - 300,000,000	22,000,000,000 - 30,000,000,000
Rapid - 150-Cycle Paired End	220,000,000 - 300,000,000	33,000,000,000 - 45,000,000,000

Project Timeline:

Another important concern that many researchers have during the experimental design phase is when the data will be available. The UB Next-Gen Core is willing to provide researchers with an estimated project completion date, but due to the potential for problems beyond the core's control, such as reagent backorders or instrumentation failures, we **do not** make any guarantees regarding turnaround time. One of the easiest ways a researcher can help ensure a quick turnaround is to provide high quality samples or libraries. This will avoid the delay that occurs while waiting to receive replacements for samples or libraries that do not meet the core's QC standards. Other factors that can significantly influence turnaround time are the type of run and number of lanes. If a researcher is requesting sequencing for less than a full flow cell (<8 lanes for high-output or <2 lanes for rapid), then the core must defer the run until enough samples are received from other researchers to fill the remaining lanes. This is not a problem for the core's most popular run, high-output 50-cycle single read, as we receive samples for this type of run on a consistent basis, making the wait time to fill a flow cell shorter. Data is usually received from high-output 50-cycle single read runs in less than 4-6 weeks. However, if a less common run type is requested, such as high-output 100-cycle paired end, the waiting time could be several months until we receive enough samples to fill all 8 flow cell lanes. With this in mind, there are still a few options to decrease wait time in these situations. The fastest way would be for the researcher to purchase all of the lanes on the flow cell, therefore guarantying that the project would be completed as quickly as possible (generally in less than 4-6 weeks). Another strategy would be to use the instrument in rapid mode, which has flow cells with only 2 lanes. The cost to purchase a full rapid flow cell would be much less than purchasing a full high-output flow cell. The wait to fill a rapid flow cell is also shorter than for a high-output flow cell, since only 2 lanes would need to be filled instead of 8 lanes.

Libraries

The UB Next-Gen Core provides library preparation services and also accepts libraries generated by the researcher for sequencing. Below you will find descriptions and sample requirements for each library type.

Researcher Prepared Libraries – The UB Next-Gen Core will accept any type of prepared library as long as it is compatible with our illumina HiSeq 2500 instrument. It is the responsibility of the researcher to confirm that the adaptors used to prep the library will work with the illumina chemistry. All libraries submitted for sequencing will be assessed internally by the UB Next-Gen Core for quality and quantity. The researcher will be charged a small fee for this service. **The core requires a minimum of 10nM of library and a volume ≥ 15µL.**

DNA and Mate Pair Libraries – DNA sequencing involves creating a library of random fragments representing the entire sample. It is useful for applications such as *de novo* sequencing, resequencing, SNP discovery, and identification of CNVs. The UB Next-Gen Core uses the illumina TruSeq DNA Library Preparation Kit, which provides adaptors for multiplexing up to 24 samples in a single lane. For mate pair sequencing, the DNA is fragmented into 2-5kb segments that are end-repaired with biotin labeled dNTPs. The labeled fragments are circularized and fragmented again into 400-600bp pieces. Fragments with the biotin labels are enriched, end-repaired, and ligated with adapters used for downstream processes. The final library consists of fragments made up of two DNA segments that were originally separated by 2-5kb. Mate pair sequencing is used for the purposes of *de novo* sequencing, genome finishing, and structural variant detection. The UB Next-Gen Core typically uses the illumina Nextera Mate Pair Library Preparation Kit, which allows for multiplexing of up to 24 samples in a single lane.

DNA quality and quantity is critical to the success of the library preparation and sequencing run. Researchers must provide the UB Next-Gen Core with high quality DNA that is free of contamination. The core recommends using any of the available Qiagen kits for DNA extraction. Researchers may choose to use alternative extraction methods, but samples must be column purified and RNase treated before submission. The DNA 260:280 ratio should be ≥1.8 and an agarose gel should be run to check for sample degradation and confirm that the fragment size is ≥50kb. **For quantity, the UB Next-Gen Core requires a minimum of 1µg of DNA in a volume ≤50µL for DNA library preparation. The core requires a minimum of 4µg of DNA in a volume ≤76µL for mate pair library preparation.** However, providing more DNA than the minimum amount necessary for the library preparation is advised so that the core will have enough sample for internal QC evaluations and for cases where protocols need to be repeated. Researchers may quantify samples using a NanoDrop, spectrophotometer, or fluorometer, but it should be noted that the UB Next-Gen Core will use the results from our internal Quant-iT

PicoGreen Assay as the final determinate of DNA concentration. Absorbance-based quantification methods often overestimate DNA quantity as they can be skewed by proteins, free nucleotides, and other contaminants.

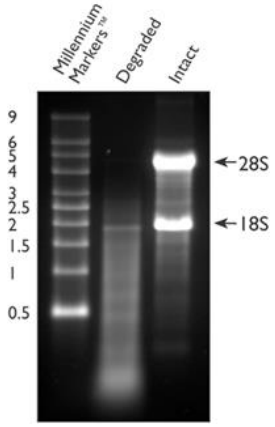
If it is determined that a DNA sample does not meet quality and quantity standards, the UB Next-Gen Core will request a replacement sample from the researcher. The researcher is still responsible for all costs associated with QC assessment for that sample. If a researcher is informed that a sample does not meet the core's QC requirements and requests the sample to be processed anyway, the researcher absolves the UB Next-Gen Core of any liability and is financially responsible for all services regardless of data outcome.

RNA and Small RNA Libraries – RNA sequencing can provide information on differential expression of genes, including gene alleles and differently spliced transcripts; non-coding RNAs; post-transcriptional mutations or editing; and gene fusions. The UB Next-Gen Core uses the illumina TruSeq RNA Library Preparation Kit, which allows for multiplexing of up to 24 samples per lane. This kit starts with total RNA and uses a poly-A enrichment method to select for coding RNAs. However, the protocol is flexible and can be adapted for researchers wanting to include non-coding RNAs or researchers studying organisms without poly-A RNA. In these cases the researcher will submit ribosomal reduced RNA, which will allow the core to create the RNA library without using the poly-A enrichment steps. Small RNA sequencing allows researchers to study expression patterns, isoforms, and discover novel small RNAs. The UB Next-Gen Core uses the illumina TruSeq Small RNA Library Preparation Kit which allows for multiplexing of up to 48 samples per lane. The core has slightly modified the protocol to allow for size fractionation of small RNA using the BluePippin from Sage Science, which delivers a more accurate and consistent size selection than the traditional gel electrophoresis suggested in the original protocol.

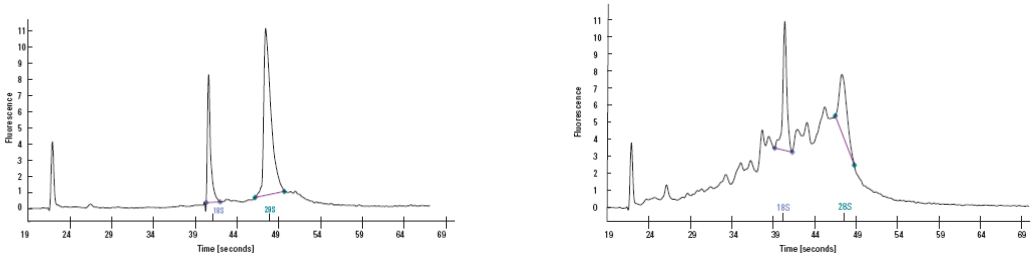
RNA quality and quantity is critical to the success of the library preparation and sequencing run. Researchers must provide the UB Next-Gen Core with high quality RNA that is intact and free of contamination. The core recommends using any of the available Qiagen kits for RNA extraction. Researchers may choose to use alternative extraction methods, but samples must be column purified and DNase treated before submission. The RNA 260:280 ratio should be ≥1.8 and an agarose gel or Agilent Bioanalyzer Chip should be run to confirm RNA quality (see examples below). **The UB Next-Gen Core requires between of 0.1µg - 4µg of total RNA in a volume ≤50µL for RNA library preparation. The small RNA library preparation requires a minimum of 1µg of total RNA in a volume ≤5µL.** However, the core recommends that the researcher provides more than the minimum amount of RNA necessary for the library preparation so that there will be enough sample for internal QC evaluations and for situations where repeating the protocol might be necessary. Researchers may quantify samples using NanoDrop, spectrophotometer, or fluorometer, but it should be noted that the UB Next-Gen Core will use the results from our internal Quant-iT RiboGreen Assay and Agilent Bioanalyzer Chip as the final determinate of RNA concentration and quality. Absorbance-based quantification methods often overestimate RNA quantity as they can be skewed by proteins, free nucleotides, and other contaminants.

If it is determined that an RNA sample does not meet quality and quantity standards, the UB Next-Gen Core will request a replacement sample from the researcher. The researcher is still responsible for all costs associated with QC assessment for that sample. If a researcher is informed that a sample does not meet the core's QC requirements and requests the sample to be processed anyway, the researcher absolves the UB Next-Gen Core of any liability and is financially responsible for all services regardless of data outcome.

Agarose Gel: The image should exhibit the appropriate ribosomal subunit bands (eukaryotic: 18S and 28S, prokaryotic: 16S and 23S). These bands should be sharp and clear and the large subunit band should be about twice as intense as the small subunit band. The UB Next-Gen Core does not accept degraded RNA that has a smeared appearance and lacks sharp, clear ribosomal bands.



Example: Intact vs. Degraded RNA
2µg of degraded total RNA and intact total RNA were run beside Ambion's RNA Millennium Markers on a 1.5% denaturing agarose gel. The 18S and 28S ribosomal RNA bands are clearly visible in the intact RNA sample. The degraded RNA appears a low molecular weight smear. Image from: **http://www.ambion.com**



A **B**

Example: Agilent 2100 Bioanalyzer Chip Electropherograms – Intact vs. Degraded RNA
Figure A is an example of high-quality total RNA. There are two well-defined peaks corresponding to the 18S and 28S ribosomal subunits and the ratio between the 28S and 18S peaks is approximately 2:1. Figure B is an example of partially degraded RNA. The 2.1 ratio between the ribosomal peaks is absent and there is a high presence of degraded products. Image from **http://www.agilent.com/chem/labonachip**

ChIP-Seq Libraries – ChIP-Seq is a combination of chromatin-immunoprecipitation and DNA sequencing. This type of experiment is useful to highlight the interactions between DNA, RNA, and proteins on a genome-wide scale. The UB Next-Gen Core requires that the researcher provide the ChIP DNA, which the core will then make into a ChIP library using the illumina TruSeq ChIP Library Prep Kit. The ChIP library protocol works similarly to the DNA library prep, where the sample fragments are flanked with unique adapters that are used during downstream enrichment, hybridization, amplification, and sequencing.

The quality and quantity of the ChIP DNA is critical to the success of the library preparation and sequencing run. The UB Next-Gen Core encourages researchers to validate all ChIP samples with qPCR, with both positive and negative controls, prior to submission. Because the quantity of immunoprecipitated DNA is typically very small, it is important that the researcher make every effort to avoid contamination of the sample, as it can contribute significantly to the final amplified ChIP-Seq library. Researchers should not use salmon sperm, calf thymus, or any other DNA-based carriers at any step during the immunoprecipitation procedure. **The UB Next-Gen Core requires between 4ng – 10ng of ChIP DNA in a volume ≤ 50µL. The average fragment size of the ChIP DNA should be 300bp – 600bp.** However, the core recommends that the researcher provides more than the minimum amount of ChIP DNA necessary for the library preparation so that there will be enough sample for internal QC evaluations and for situations where repeating the protocol might be necessary. Researchers may quantify samples using NanoDrop, spectrophotometer, or fluorometer, but it should be noted that the UB Next-Gen Core will use the results from our internal Quant-iT PicoGreen Assay and Agilent Bioanalyzer Chip as the final determinate of concentration and quality.

If it is determined that a ChIP DNA sample does not meet quality and quantity standards, the UB Next-Gen Core will request a replacement sample from the researcher. The researcher is still responsible for all costs associated with QC assessment for that sample. If a researcher is informed that a sample does not meet the core's QC requirements and requests the sample to be processed anyway, the researcher absolves the UB Next-Gen Core of any liability and is financially responsible for all services regardless of data outcome.

Sample Submission

The UB Next-Gen Core requires that researchers complete the **Illumina HiSeq 2500 Sample Submission Form** and include it with all samples being submitted to the core.

Sample Drop-Off – The UB Next-Gen Core is open for sample drop-off Monday-Friday 10:00am – 4:00pm. Please call ahead or schedule an appointment to ensure that a technician will be available to accept your samples, as we are often busy working on other projects. Samples should be brought to the UB Next-Gen Core (B3-123) at the New York State Center of Excellence in Bioinformatics and Life Sciences Building (COE) at 701 Ellicott Street in Buffalo, NY 14203. Upon arrival to the COE there will be a receptionist located near the entrance that will be able to let you into the building. Please let the receptionist know that you are dropping off samples to the UB Next-Gen Core as they will be able to contact core personnel and direct you where to go next.

Sample Shipments – The UB Next-Gen Core accepts sample shipments Monday – Friday to the address below. Please call or email the core ahead of time to let personnel know to expect a shipment. Samples should be overnighted on dry ice and packaged carefully to avoid damage to tubes. The core also recommends wrapping tube caps in parafilm to avoid spillage or evaporation.

```
ATTN: UB Next-Generation Sequencing and Expression Analysis Core(B3-123)
State University of New York at Buffalo
New York State Center of Excellence in Bioinformatics and Life Sciences
701 Ellicott Street
Buffalo, NY 14203
```

Data

The UB Next-Gen Core provides demultiplexed data to researchers via sftp as compressed FASTQ files for all projects. However, researchers also have the option to purchase additional bioinformatics analysis from the core.

Data Availability – Researchers will receive an email with instructions for downloading their data from an sftp site. Data will be removed from the site 30 days after the data availability email is sent to the researcher. Once data has been removed from the sftp it is unlikely that the core will be able to recover it and researchers requesting data to be recalled will be charged a fee. The UB Next-Gen Core highly recommends that researchers make backup copies of their data.

Misc. Information

Cost Estimates – The UB Next-Gen Core will provide cost estimates for researcher projects upon request. If the researcher makes any changes to the initial project design the total cost of the project may change and will be reflected in the final invoice.

Financial Responsibility – When a researcher submits a sample submission form they become financially responsible for all labor and reagent costs associated with their project. Researchers choosing to terminate a project before completion must pay for all costs associated with the project up until it was ended. If a researcher is not satisfied with their data, the UB Next-Gen Core will make every effort to resolve the problem. If it is determined that the problem was caused by a technician, instrumentation, or reagent error, the core will perform the experiment again at no additional cost to the researcher. However, if it is concluded that the problem was caused by something on the researcher's end, then the experiment will not be repeated and the researcher will be responsible for all project costs.

Invoices – Most researchers will receive an invoice after a project has been completed. However, the UB Next-Gen Core reserves the right to require payment for services before a project is started if the researcher is external of the University of Buffalo. Invoices are required to be paid within 30 days from the receipt of the invoice

Acknowledgement – Services performed by the UB Next-Gen Core should be acknowledged in publications, scholarly reports, presentations, and posters. Proper acknowledgement provides a visible measure of the impact of the UB Next-Gen Core and is essential for our future funding efforts.