UB Next-Gen Sequencing and Expression Analysis Core
Roche/454 GS FLX Titanium Sequencing Services Document

GS FLX TITANIUM SEQUENCING SERVICES - PROJECT WORKFLOW

1) Experimental Design and Consultation

Researchers interested in GS FLX Sequencing Services are required to contact the UB Next-Gen Sequencing and Expression Analysis Core to discuss potential sequencing projects prior to sending samples. We can be reached by email at **cbi-ubnextgencore@buffalo.edu** or by phone at (716) 881-7514. During experimental design consultation the researcher and the UB Next-Gen Sequencing and Expression Analysis Core will discuss sample preparation, type of sequencing run, number of samples, and multiplexing options for the sequencing run. Once an experimental design is agreed upon, we can provide a cost estimate for the project. The UB Next-Gen Sequencing and Expression Analysis Core requires researchers to read the **GS FLX Titanium Sequencing Services Document** and complete the **Sample Submission Form** prior to sending samples.

Read Length and Number of Reads

A very common question during the experimental design process is how many reads can be expected from a GS FLX Titanium Sequencing Run and what is the average length of the reads. This is a difficult question to give an absolute answer to because the number of reads and the length of the reads can vary between runs and samples. The average read length of a GS FLX Titanium Sequencing Run is between 300-500bp. However, the read length can vary, especially if sequencing amplicons of different sizes. The number of reads from a run is highly dependent on the number of regions on the sequencing plate. For example, a 2 Region GS FLX Titanium Shotgun Sequencing Run typically gets between 300,000-500,000 reads per region, making the total number of reads for the run 600,000-1,000,000. A 16 Region GS FLX Titanium Shotgun Sequencing Run typically gets between 20,000-40,000 reads per region, making the total number of reads for the run 320,000-640,000. As the number of regions on a sequencing plate increases the number of reads decreases. This is due to the dividing gasket taking up more space on the sequencing plate. With the release of the GS FLX Titanium Rapid Library MID Adaptors and the ability to design amplicon fusion primers with MID tags, researchers can choose to bioinformatically separate their shotgun and amplicon samples using these 'barcoding' tags instead of using physical separation on the plate. This reduces the number of reads that are lost due to physical separation by the dividing gasket. Table 1 shows the approximate number of reads that can be expected from a typical GS FLX Titanium Sequencing Run. The numbers in Table 1 are approximations and are meant only to assist you during your experimental design process. The UB Next-Gen Sequencing and Expression Analysis Core makes every effort to try to maximize read length and number of reads from each run, but we **do not** make any guarantees about read length or the number of reads you will receive from your run.

Table 1. Approximate number of reads to expect per run depending on the number of regions on the sequencing plate.

# of Regions	# of Reads per Region	Total # of Reads per Run
2	300,00 - 500,000	600,000 - 1,000,000
4	100,000 - 200,000	400,000 - 800,000
8	60,000 - 100,000	480,000 - 800,000
16	20,000 - 40,000	320,000 - 640,000

2) DNA and GS FLX Titanium Library Preparation

Shotgun and Paired End – The quality and quantity of the DNA sample is critical to the success of the sequencing run. Any contamination in the starting material will be directly reflected in the data. The UB Next-Gen Sequencing and Expression Analysis Core provides genomic DNA preparation for bacterial samples only. It is the researcher's responsibility to ensure that the bacterial pellets provided for genomic DNA extraction are pure and not contaminated with other strains. The researcher must send pellets as described in our **Bacterial Pellet Preparation Protocol**. Researchers using DNA from other organisms must provide their own high quality DNA for sequencing. The UB Next-Gen Sequencing and Expression Analysis Core recommends using the Qiagen Genomic Buffers Set in conjunction with the Qiagen Genomic-tip columns for DNA preparation. If an alternative extraction method is used it is very important that the sample is RNase treated and free of organics and particulate matter. It is the researcher's responsibility to make certain that the DNA contains only the target DNA and is free of contaminating DNA. The DNA provided by the researcher must meet the criteria in Table 2 before library preparation can proceed.

Table 2. DNA quality and quantity criteria for GS FLX Titanium Shotgun and Paired End Library Preparation.

DNA Quality and Quantity Criteria	Shotgun	3kb Paired End	8kb Paired End	20kb Paired End
DNA Structure	Double-Stranded	Double-Stranded	Double-Stranded	Double-Stranded
OD260/280	≥ 1.8	≥ 1.8	≥ 1.8	≥ 1.8
Fragment Size	≥ 1.5 kb	≥ 1.5 kb	≥ 1.5 kb	≥ 1.5 kb
Minimum Amount of Starting Material	500 ng	5 µg	15 µg	30 µg
Maximum Volume	100µL	200µL	150µL	150µL

Providing more sample than the minimum amount necessary for the library preparation is advised just in case something needs to be repeated. DNA samples should be run on an agarose gel and the image must be provided to the UB Next-Gen Sequencing and Expression Analysis Core. Samples should be quantified using a NanoDrop, spectrophotometer, or fluorometer. However, we use the quantifications from our Invitrogen Quant-iT PicoGreen Assay as the final determinate of sample concentration. This is because most researchers determine the concentration of their DNA based on absorbance at 260nm on a spectrophotomer or NanoDrop, which may overestimate concentration. Absorbance-based quantification methods can be skewed by proteins, free nucleotides, and other contaminants. The Quant-iT PicoGreen Assay is highly sensitive and is not affected by the presence of contaminants. If it is determined that a sample does not meet our sample quality and quantity standards, the UB Next-Gen Sequencing and Expression Analysis Core will request a new sample from the researcher. The researcher is still responsible for all costs associated with DNA quality and quantity assessment for any sample that does not meet standards and must be resubmitted. If a sample does not meet our requirements and the researcher still chooses to have the sample processed, the researcher absolves the UB Next-Gen Sequencing and Expression Analysis Core of any liability and is financially responsible for all services regardless of the outcome of the data.

Amplicon – GS FLX Titanium Amplicons are sequenced in both directions. The researcher is responsible for providing samples that have already been amplified with fusion primers. Amplicon fusion primers have a directional GS FLX Titanium Fusion Primer Sequence at the 5' end of the primer and the template specific sequence at the 3' end of the primer. In addition, the researcher may add an optional MID sequence between the fusion primer sequence and the template specific sequence (Figure 1). The MID will allow multiplexing of samples that can be separated during data analysis. Researchers may choose to combine the use of MID tags along with physical separation of samples on the sequencing plate (2, 4, 8, or 16 regions) to maximize the number of individual samples on a sequencing run.

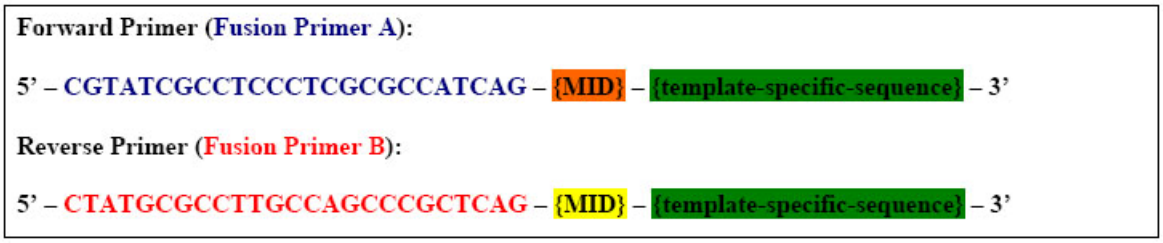

Figure 1. GS FLX Titanium Fusion Primer Sequence with optional MIDs.

Because experimental design and sample preparation for GS FLX Amplicon can be complex, it is imperative that the researcher consults with the UB Next-Gen Sequencing and Expression Analysis Core BEFORE ordering amplicon fusion primers. As with all GS FLX Sequencing applications, the quality of the starting material is critical to the success of the sequencing run. PCR optimization and individual sample testing may be necessary with the amplicon fusion primers. The addition of the fusion adaptors and MIDs may result in problems with primer duplex and hairpin formation, making optimization important. The amplicon fusion primers must be HPLC purified and a high fidelity polymerase should be used for amplification. The UB Next-Gen Sequencing and Expression Analysis Core recommends using the Roche FastStart High Fidelity PCR System or Invitrogen Platinum Taq DNA Polymerase High Fidelity. Reactions should be set up in a volume ≥ 50µL. PCR products should be run on an agarose gel and the image must be provided to the UB Next-Gen Sequencing and Expression Analysis Core. The exact amplicon length (base pairs) should also be provided as this information is used in calculations that determine the quantity of sample to add to the emulsion PCR (emPCR) reaction. Samples should be quantified using a NanoDrop, spectrophotometer, or fluorometer. Products for amplicon sequencing should NOT be purified or cleaned prior to sending to the UB Next-Gen Sequencing and Expression Analysis Core. The GS FLX Amplicon Library Preparation Protocol includes steps to remove excess primers and nucleotides prior to emPCR, and purification of the amplicons by the researcher before these steps can cause problems.

SAMPLE SUBMISSION REQUIREMENTS

A **GS FLX Sequencing Services Sample Submission Form** must be completed by the researcher and is required to include the following information for each project:

1) Lab Contact Information – Please provide the email address and phone number of the preferred contacts for the project. This information will be used to communicate with the researcher/technician if we have any questions or problems with the samples.

2) Project Information – The researcher must provide a brief description of the experimental design. The following information should include:

a) Type and Number of Runs (Shotgun, Paired End, Amplicon)

b) Number of Samples

c) Multiplexing Plans – Do you plan to use MIDs or separate your samples in 2, 4, 8, or 16 regions?

3) Sample Names – Verify that each sample has a unique name and that the sample tubes submitted have the correct sample identification clearly written on each tube.

4) Sample Volumes – Accurately measure each sample volume with a pipet.

5) DNA Extraction Information (Shotgun and Paired End Only) – Confirm that you have used the recommended Qiagen Genomic Buffers Set in conjunction with the Qiagen Genomic Tip Protocol or indicate that you would like the UB Next-Gen Sequencing and Expression Analysis Core to extract DNA from your bacterial pellets (must be prepared as described in our **Bacterial Pellet Preparation Protocol**). If you used an alternative method for DNA extraction, please provide a copy of the protocol.

6) Amplification Conditions and Fusion Primers (Amplicon Only) – Provide a description of your PCR amplification conditions and list all of your primers. Indicate which amplicons were amplified with which primers, exact amplicon size (base pairs), and which samples should be pooled together.

7) Sample Concentrations (ng/uL) – Measure each sample concentration using a NanoDrop, spectrophotometer, or fluorometer. Accurate measurements are crucial for success of the labeling procedure.

8) 260/280 Ratios (Shotgun and Paired End Only) – This ratio indicates the quality of the DNA.

9) Agarose Gel Electrophoresis Image – Provide the gel image with the sample wells and ladder sizes clearly labeled. The image will be used to confirm sample quality and size.

3) GS FLX Titanium emPCR Titration

The UB Next-Gen Sequencing and Expression Analysis Core uses emulsion PCR (emPCR) titration to determine the optimal conditions for a library that will provide the best possible data from the sequencing run. During emPCR, the library is bound to "capture beads" and amplified in an oil emulsion. The emulsion ensures functional clonality by physically separating the "capture beads" in "microreactors." The objective is to have a single library fragment bound to each "capture bead" and one bead per "microreactor." The reason that a titration is necessary is because each library behaves differently during emPCR. The titration is a functional quantitation to determine the correct amount of library to add to the emPCR. The experiment is performed by setting up small scale emPCR reactions with different amounts of library and choosing the condition which results in the appropriate number of library bound beads. Using this condition during the sequencing emPCR will maximize the number of high quality reads from the sequencing run. The UB Next-Gen Sequencing and Expression Analysis Core recommends that each library be titrated prior to sequencing. However, if a researcher has several libraries from similar samples it may be possible to only titrate one library and extrapolate that information to the other libraries to reduce project costs. If a researcher chooses not to do an emPCR titration of their library, they accept full financial responsibility regardless of the number of reads received from the sequencing run.

4) GS FLX Titanium emPCR and Sequencing Run

After a library has passed all quality control specifications and has been titrated, it is ready to be sequenced. The sequencing process starts by setting up an emPCR using the optimal conditions determined from the emPCR titration. After the library bound beads are amplified, they go through a process called enrichment, which removes the oil waste from the beads and selects for beads with amplified library fragments. The number of beads are counted using a Beckman Coulter Z1 Particle Counter and loaded onto the picotiter plate (PTP). The PTP is loaded onto the GS FLX Instrument, where the pyrosequencing process takes place. Nucleotides are flowed across the PTP in a set order during the sequencing run and when a nucleotide complementary to the library strand is flowed into a well, the polymerase adds the nucleotide. This nucleotide incorporation results in a reaction that generates a light signal. A CCD camera on the instrument takes an image of each nucleotide flow and is able to record the signal generated from hundreds of thousands of library containing beads in parallel. The GS FLX Titanium Sequencing Run has 200 nucleotide flow-cycles, which takes about 10 hours to complete. After the run is finished the CCD camera images must be processed by the GS FLX Titanium Software (more details on this process can be found at **www.454.com**).

The UB Next-Gen Sequencing and Expression Analysis Core offers GS FLX Titanium XLR 70X75 Shotgun, Paired End, and Amplicon Sequencing Runs. Please see the **"Read Length and Number of Reads"** section of this document for information about the number of reads and average read length you can expect from a run. The UB Next-Gen Sequencing and Expression Analysis Core will make every effort to maximize the number of reads and read length for each library, but we **do not** make any guarantees about read length or the number of reads you will receive from your run.

5) Data Processing and Assembly

The GS FLX Titanium Sequencing Run raw data is in the form of a series of CCD images. In simple terms, the GS FLX Image and Signal Processing Software is necessary to extract the data from the CCD images by creating flowgrams for each well that contains a bead with amplified library. The software does this by tracking the bead containing wells on an X,Y axis over the series of CCD images and uses the signal intensity for each nucleotide flow to generate the flowgram for each well. Each flowgram represents the sequence read from one amplified library bead and can be thought of as the Roche/454 version of the electropherogram. The software creates a flowgram/read for each well that contains a library bead, which typically results in over 1,000,000 flowgrams/reads per run. The data is stored in standard flowgram format (SFF) files for downstream analysis. The software also generates a set of metrics and FASTA files, which can be used to evaluate the sequencing run (more details on the GS Processing Software can be found at **www.454.com**).

The UB Next-Gen Sequencing and Expression Analysis Core provides the initial signal and image processing for each GS FLX Titanium Sequencing Run. When a researcher's data is available we will use an ftp site to send the data. The researcher will be given information allowing access to their data from a password protected ftp site. The UB Next-Gen Sequencing and Expression Analysis Core allows researchers to access both raw and processed data. We highly recommend that the researcher downloads the data immediately from the ftp site and makes a backup. Please keep in mind that if you wish to keep both the raw and processed data you will need more than 60 gigs of storage space. However, most researchers choose to keep only the processed data, which includes .sff, FASTA, and metrics files. The UB Next-Gen Sequencing and Expression Analysis Core can also provide assembly data using the GS De Novo Assemler Software. At this time we do not offer full data analysis services, but we are in the process of setting up a pipeline. Please contact us by email at **cbi-ubnextgencore@buffalo.edu** or by phone at (716) 881-7514 to discuss our most current data analysis services.

6) Service Requests and Misc. Information

Service Requests – The first step to requesting GS FLX Titanium Sequencing Services from UB Next-Gen Sequencing and Expression Analysis Core is to contact us by email at **cbi-ubnextgencore@buffalo.edu** or by phone at (716) 881-7514 to discuss the sequencing project. After the initial contact, the researcher must read the **GS FLX Titanium Sequencing Services Document** and complete the **Sample Submission Form** prior to sending samples. Information about how to prepare samples for sequencing can be found in the **"DNA and GS FLX Titanium Library Preparation"** section of this document.

Sample Drop-Off – The UB Next-Generation Sequencing and Expression Analysis Core Facility is open for sample drop-off Monday-Friday 9:00am-3:00pm. Please call ahead or schedule an appointment to make sure that a technician will be available to accept your samples, as we are often busy working on other projects. We will not accept samples unless we have already received the **Sample Submission Form** or if the form accompanies the samples. Samples should be brought to the University at Buffalo Next-Generation Sequencing and Expression Analysis Core Facility (B3-123) at the New York State Center of Excellence in Bioinformatics and Life Sciences Building (COE) located at 701 Ellicott Street in Buffalo, NY 14203. There is a receptionist at the entrance of the COE who will be able to let you in the building and contact us to let us know that there is a researcher dropping off samples. Alternatively, samples can be shipped overnight on dry ice to the below address, but please email or call the core to let us know to expect the sample shipment.

```
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
```

Project Scheduling – Projects are completed in the order in which samples are received. The UB Next-Gen Sequencing and Expression Analysis Core will try to accommodate your deadline if we are provided with sufficient lead time. However, we can not make any guarantees as to when your data will be available. There are a number of factors that influence project completion; such as the number of projects already in the queue, reagent shipping, and instrumentation/software issues. The UB Next-Gen Sequencing and Expression Analysis Core does not keep a stock of GS FLX Sequencing Reagents, so items must be custom ordered for each project. Orders usually take at about two weeks to receive from the vendor, but can be significantly delayed if an item is on backorder. Please contact the core if you would like to discuss your project completion timeline.

Reagent Ordering – The UB Next-Gen Sequencing and Expression Analysis Core orders reagents for a project after we receive the samples and the **GS FLX Titanium Sample Submission Form**. However, the UB Next-Generation Sequencing and Expression Analysis Core Facility is willing to pre-order sequencing reagents before samples are received in an effort to streamline project completion, if a researcher completes the **GS FLX Titanium Reagents Pre-Order Form**. This form will authorize the core to order all consumables necessary for the researcher's sequencing project and also confirms that the researcher is financially responsible for all items ordered for their project even if the researcher does not end up using the reagents.

Data Availability – When a researcher's data is available we will use an ftp site to send the data. The researcher will be given information allowing access to their data from a password protected ftp site. Please access and download your data as soon as possible. It is highly recommended that the researcher makes a backup copy of their data. Data will be removed from the ftp site 30 days after the data availability email is sent to the researcher. There will be a $100.00 data recall charge if a researcher requests data to be re-uploaded to the ftp site after the data has already been removed. The UB Next-Gen Sequencing and Expression Analysis Core does not guarantee that data will be able to be recalled after the 30 day data availability period.

Project Costs – The UB Next-Gen Sequencing and Expression Analysis Core is happy to provide a cost estimate for a researcher's project after the initial discussion of the project. Our cost estimates include library preparation, emPCR titration, and the sequencing run. Keep in mind that the cost estimate provided to the researcher is just an estimate. If the researcher makes any changes to the original project plan (number of samples, type of run, etc.) the final cost of the project will reflect these changes.

Researcher Financial Responsibility – Once a researcher requests an order they are financially responsible for all reagents and shipping costs for items ordered on behalf of their project. Even if the researcher does not use all of reagents requested for their project they are still responsible to pay for them, as we are unable to return unused items to the vendors. Also, if it is determined that at any point throughout the process a sample does not meet quality or quantity control measures the researcher is still responsible for any reagents and labor charges up to the point that the sample failed to pass quality/quantity control standards.

Invoices for GS FLX Titanium Services – The UB Next-Gen Sequencing and Expression Analysis Core will send the researcher an invoice for the final cost of the project. Invoices are required to be paid within one month from the receipt of the invoice. University at Buffalo researchers that will be paying from Research Foundation Accounts can pay their invoices with an Interdepartmental Invoice (IDI). UB researchers paying from State Accounts or other types of accounts should let the core know which type of account they will be paying from so that we can send you the appropriate type of invoice. Non-UB researchers will receive a standard UB Next-Gen Sequencing and Expression Analysis Core Invoice and will be expected to pay with a check.