Deep Sequencing Core Guidelines for the Preparation and Submission of Samples

Your sample (i.e. your "library") must be sized, quantified, and proofed. These are described below. You will need an Analysis Form for each sample. Each sample is treated as an individual project for cluster formation, sequencing, and data retrieval, please assist us by providing a ticket for each sample when it is submitted.

Uniform Size: The clusters are formed by seeding a lawn of library fragments onto the flow cell, then attaching and amplifying each sequence by bridge amplification. For an optimal cluster formation library fragments should be close to the same size (within 50bp). DNA libraries should have a median fragment size under 1Kb (500bp ± 50bp is recommended); Chromatin IP libraries are often in the 300bp range, and small RNA libraries are usually under 100bp. For cluster formation, we need to know the median size (not the mean). The size-selection and determination should be done on an agarose gel followed by purification of fragments on a column (see protocol). Please put the size information on the Analysis Form.

Quantification: Optimal seeding of the clusters also depends upon delivering the correct amount of your library to the flow cell, in order to do this we need an accurate quantification of the material in your library. Standard spectrophotometric measurement of material after it is gel-purified (sized) is recommended. The final material should be delivered to the core at 10nM in sterile water or EB (minimum of 25 µl). If you aren't able to do this, please let us know and provide us with the concentration in ng/ml and the median fragment size.

Proofing: This step is critical for two reasons. (1) demonstrates that your library has fragments which can be used to create clusters and (2) indicates whether your library contains fragments with inserts from your target genome. By cloning an aliquot of your library and sequencing 20 clones, you will know if the linkers for attachment are present, you will confirm which sequencing primer should be used for cluster formation, and you will know if your library has any inserts from the genome of interest or if the bulk of the clones contain primer-dimers or other unwanted material. Please remember that only a few unwanted clones here will represent potentially millions of sequences after the analysis. Please note on your Analysis Form what number of clones you sequenced and how many had attachment linkers (if only half had linkers, then we need to adjust the amount of material used to form clusters to compensate for this).

Attachment sequences:

AATGATACGGCGACCACCGA or
CAAGCAGAAGACGGCATACGA

Sequencing Primers, PLEASE indicate which one to use for your sample:

Solexa **DNA** Seq primer
ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT

Solexa **RNA** Seq primer (used for libraries made from RNA Kit adapters)
CGA CAG GTT CAG AGT TCT ACA GTC CGA CGA TC

Data: Each run generates a large number of image files. These are transferred to BINAR (the High Performance Computing Cluster) where the data analysis pipeline reduces the images to sequences. The pipeline can also run an alignment to a reference genome if you request it. At the time of sample submission, please note the following items on the Analysis Form:

BINAR Account: Name of person with Account, and account ID. This person should have a folder in their account, entitled **DeepSEQ** into which the data from the pipeline will be deposited. DeepSEQ users will have workspace (not storage) available to them for ongoing projects on BINAR, this is near-line, high-speed data holding, not archive storage.

Images: If you want to save your images (this is LARGE) please note this on the form. We do not store images more than 2 weeks. Additionally, they may not be stored on BINAR which is used solely for workspace and analysis, not storage. You will need to arrange a storage solution if you wish to save images. Your images will either be made available to you by sneaker drive or deposited for a brief stay in your BINAR account.

GERALD: Gerald is the name of the genome alignment feature of the data pipeline. If you would like to have your sequences aligned with Gerald, please indicate what species and which reference genome (if there is more than one).

IrfanView: IrfanView is free-ware available at http://www.irfanview.com This program is good for opening image files from the Deep Sequencer (as well as any other image format). After your run has completed, some sample images will be made available to you. Irfanview is a good way to look at them. We generally put them on the PMM server, which allows guest login.

Sample Drop-Off: The location for sample drop off is in 2 Biotech, Lab 207. Samples will be logged in, tickets collected, and the library sample tubes put in a designated box in the -20°C. Please see Lynn Kerr or Ellie Kittler for sample login. Please let us know when you are coming over if possible.

We are trying to keep the web site up to date, please check
http://www.umassmed.edu/nemo