

Questions Frequently Asked about Illumina Sequencing

Table of Contents

Samples, sample submission and sequencing	1
Cost and turnaround time	3
Data and analysis	4
Reference and acknowledgement policy	5

Samples, sample submission and sequencing

What kind of samples can I have sequenced?

You can sequence:

- Genomic DNA
- RNA
- microRNA (small RNA)
- Chromatin Immunoprecipitated (ChIP'd) samples

Sequencing of other starting material may also be available. Please contact us with your specific project needs, to confirm our ability to process your samples.

For transcriptome projects, can I send total RNA or do I need to make cDNAs?

You can send us:

- tissue/cells
- total RNA
- DNase-I treated RNA
- cDNA can also be submitted, but please consult with us in advance.

The costs associated with library production vary depending on starting material. Please contact the Project Manager for a quote or Statement of Work (SOW).

How should I isolate my RNA/DNA or immunoprecipitate my chromatin?

We can advise on production of RNA, isolation of DNA or immunoprecipitation, but cannot provide optimized protocols. All isolation protocols will need to be optimized in your own lab. Some variability is unavoidable when working with biological samples.

What happens if the run/library fails?

The solution will depend on the source of the problem. Several quality checks are included in the process of constructing libraries and sequencing, in an effort to minimize the potential for failure.

- 1) DNA samples are quality checked in advance of library construction. You will be contacted if your sample fails Quality Control.
- 2) If library construction fails, the collaborator would be consulted to find a solution. There is no standard policy, as failure can be attributed to many causes. If the cause is found to be sample related, a replacement sample may be submitted. Costs for library construction that fails due to sample issues will be recovered.
- 3) Poor Sequencing results. A single lane is initially run to assess library quality. If the lane fails any of several quality metrics, the Quality Control team reviews the data to identify the source of the



problem. Concerns about library construction, are reported to the customer to discuss possible solutions and options.

The short answer is that we work with the collaborator at every step to ensure the best possible results are obtained.

How many lanes should I run?

Sequencing requirements will vary between researchers and between samples. The number of lanes of sequencing needed depend on the experimental design and your sample. Important variables include: sample quality, sample quantity, genome size, the availability of a *reference* genome for comparison and the goal for the project.

Cost and turnaround time

How much will my sequencing cost?

Sequencing cost is dependent on:

- sample type (tissue, RNA, DNA, ChIP)
- the number of lanes of sequence requested.

The total cost for sequencing includes the cost for construction of a library suitable for sequencing on an Illumina HiSeq instrument plus the cost for each lane of sequence. To obtain an accurate quote, please contact us with the details of your project.

Why aren't prices posted on this website?

The Genome Sciences Centre operates on a collaborative, cost-recovery basis. This means that costs vary between sample types and will vary with time as new and updated technologies become available. Costs reflect the actual cost of sequence production.

Can I save money by constructing the libraries myself?

Our library construction pipeline has been streamlined to offer maximum value. In general, your costs will be **less** if you submit your sample to the GSC for library construction rather than doing it in-house. We also caution researchers that libraries are not always straightforward to construct, and inexperience may result in libraries that have a higher failure rate. You are welcome to submit pre-constructed libraries for sequencing especially if your lab has experience in this area. Please contact us in advance of constructing your own libraries to ensure that your protocol is compatible with our pipeline.

If I submit multiple samples, can I obtain a discount?

We operate on a cost-recovery basis, which means we cannot offer discounts. We may have options for collaborators who submit >24 samples at one time, as that is more cost efficient than submitting a few at a time. Please contact us for more information.

How long will it take to receive my data?

Samples are placed in the queue in the order in which they are received, so the time will depend on the workload in the pipeline at the time of sample submission. Our [terms of service](#) offer more information about our policies. For more information about the queue or how your sample is progressing within the queue, please contact the Project Manager.

Data and analysis

In what format do I receive my sequencing data?

Your data is provided as a binary alignment file (bam) for all sequenced libraries. Scripts are available for download (such as [Picards SamToFastq](#)) which convert bam formatted files to fastq files for independent off-site alignment.

If the UCSC genome browser supports the reference genome used for the alignment, additional files for visualizing the alignment results in the genome browser are also available on request at no additional charge.

These file types include:

- 'wig' (and/or 'bigwig')
- 'bedgraph' (and/or 'bigbedgraph')
- 'bed' (and/or 'bigbed') formats

Data processing is limited to the file types listed. GSC bioinformaticians can be contracted to perform additional analysis, please contact us for a quote specific to your requirements.

Do you send my files on DVD or email them to me?

The amount of data is too large for email or DVD. Data is available via password protected ftp, and must be downloaded within 2 weeks of posting. A link is provided by email, when your data becomes available.

Do you keep all data (picture files...)?

No, images are not stored. Other data is stored for a minimum of 45 days, and may be deleted after that time without notice.

How much pass filter data am I guaranteed?

We don't have minimum data guarantees, as the data yield depends too much on the sample supplied. We would use our internal QC standards to ensure that the best possible data is generated for each sample.

Does the cost of sequencing include alignments? If so, what software is used?

The price includes a preliminary alignment by the Burrows-Wheeler Aligner (BWA) program with (<http://bio-bwa.sourceforge.net/bwa.shtml>). Additional alignment, with specific client specified parameters using alignment programs such as Maq, Eland, [NovoAlign](#) etc is available upon request and may require additional payment.

Does the alignment provide multiple-matches, none, only 1, or random placement in the case of ambiguity?

The resulting bwa alignment (referred to as a bam file) is a binary alignment/map file which includes all reads. Predefined columns within the file will flag sequence reads which passed or failed chastity filtering, and mark those reads which are unique and have either multiple, or no matches. For more information regarding the bam output format and tools for working with these alignment files, please refer to samtools (<http://samtools.sourceforge.net>).

How many mismatches (and at what quality?) are tolerated?

Most alignment software defaults to 2 base mismatches per alignment, for speed and accuracy. We routinely run our alignment software with this as the default setting.

Did you align the filtered data or all of the raw data?

All of the raw data is aligned. No filtering is applied directly to the reads. Reads are however flagged using various filtering approaches (including quality) allowing reads to be filtered out by the end user.

How do I identify / separate out the barcodes for the samples in each lane of an indexed / multiplexed library?

Adapter Trimming removes the barcodes from the bam files:

<http://www.bcgsc.ca/platform/bioinfo/software/adapter-trimming-for-small-rna-sequencing>

Are the barcodes part of the 3' adaptor used in the miRNA library construction?

Yes, the adapter sequence is at the 3' end of each read.

Could we obtain the original FASTQ files at no charge (or at a minimal additional charge), in addition to the BAM files?

No. But you can use Picards SamToFastq (<http://picard.sourceforge.net/command-line-overview.shtml#SamToFastq>) to create a fastq file from the provided bam file. Even though the fastq files are used as the input to generate the alignments, due to their large size, they are immediately deleted once the bam files have been generated. The bam files contain all of the alignment information including the quality information. SamToFastq enables quality scores to be extracted from the bam file into the precursor constituent fastq files.

How do I access my data?

All collaborators will receive an email informing them that their data is available for download. The email is a receipt, identifying which data has recently been made available in addition to the previously uploaded data sets from the same project. This allows the collaborator to track sequence data as it is generated.

In addition to the data set notification, the user name and password for your project account is provided along with a web link to instructions explaining the download process.

http://www.bcgsc.ca/services/solseq/data_access

What do I do with my data once I have downloaded it?

We strongly encourage you to perform data analysis using a Unix-based operating system. This enables you to run third party and open source software readily available on the internet. For a partial list of various bioinformatics applications please refer to the following link:

<http://seqanswers.com/wiki/Software>

Also, for any questions related to the analysis of your data or questions about particular analysis software the seqanswers forum (<http://seqanswers.com>) is a useful resource.

Recently NCBI has announced that they will no longer be able to receive sequence data for publications. Can the GSC host my sequencing data instead?

As much as the GSC would like to help our collaborators get their data published, we do not have the storage infrastructure available. Instead we would encourage our collaborators to publish their sequence data to the EBI Sequence Read Archive (SRA).

For more information about the EBI SRA please refer to:

http://www.ebi.ac.uk/ena/about/sra_submissions

Reference and acknowledgement policy

Do I have to cite the GSC for the sequencing work performed when I publish my research? How should I cite work performed by the GSC?

We require our collaborators to acknowledge the work performed by the GSC in the following ways depending on the level of collaborative effort between the GSC and the researcher:

1. If the data was generated as a fee for service (cost-recovery collaborative service alone, i.e. when no intellectual contribution has been made). The GSC should be cited using either of the methods below:
 - A. In peer-reviewed publications incorporate the following sentence into the Acknowledgements section of the article: "The authors wish to acknowledge the BC Cancer Agency Genome Sciences Centre, Vancouver, Canada for [activity]".
 - B. Or alternatively, the GSC can be cited in the Materials and Methods section. A suggested sentence for inclusion is: "[Activity] was performed by the BC Cancer Agency Genome Sciences Centre, Vancouver, Canada".
2. Where intellectual contributions have been made by researchers at the GSC, collaborators are required to discuss potential and pending publications based on these contributions with the relevant GSC scientists or staff to identify appropriate co-authorship. This will ensure that our scientists and staff receive the appropriate credit for their work, and enables them to advance their careers.

The BC Cancer Agency Genome Sciences Centre (GSC) tracks contributions to the wider scientific community. This is a means to measure our ongoing support for the activities of our collaborators, as well as to ensure we meet the requirements of both our funding partners and our charter as a non-profit agency.