Table of Contents

Library Construction and Sequencing

FAQ	Answer
What kind of samples can I can submit for sequencing?	• Tissues for Nucleic Acid extraction: - FFPE blocks or section - OCT sections - Fresh frozen sections - Cells • RNA: - Total RNA - FFPE derived RNA - Total Nucleic Acid - DNase-I treated RNA • DNA: - Genomic DNA - FFPE derived DNA - Total Nucleic Acid - Chromatin Immunoprecipitated (ChIP'd) samples - cDNA can also be submitted, but please consult with us in advance - Amplicons • Constructed Libraries- Please refer to the requirements document for more details. • 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.

What library construction strategies are provided?	The following library construction strategies are provided: • FFPE-Genome • PCR-Free Genome • Low input DNA • Bisulphite • ChIP • mRNA using strand specific protocol • Ribodepleted strand specific RNA • miRNA • TCR/BCR • Exome and custom capture • 16S/ITS
How should I isolate my RNA/DNA or immunoprecipitate my chromatin?	We can advise on extraction 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 is the cost of library construction and sequencing?	The costs associated with library production vary depending on starting material. Please contact the Project Manager for a quote or Statement of Work (SOW).
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 quantified upon receipt of samples in advance of library construction. RNA samples are quality and quantity checked in advance of library construction. You will be contacted if your sample falls below the required total mass or is degraded and not suitable for library preparation. 2) If library construction fails, the collaborator will 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 by the GSC 3) For samples requiring multiple lanes of sequencing, 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. Sequencing run quality metrics are reviewed by the lab to ensure high quality sequence is produced. Runs failed due to instrumentation/technical issues will be repeated at no cost to collaborator. The Quality Control team reviews the content of the sequence data to ensure several quality metrics are met. Failures are investigated to determine the root cause and are reported to the customer to discuss possible solutions and options.

How many lanes should I run and how do I determine sequencing coverage?	Sequencing requirements will vary between researchers and between samples. The number of sequencing lanes required depends on the experimental design and your sample. Important variables include: sample quality, sample quantity, genome size, and the availability of a reference genome for comparison and the goal of the project. The following link provided by Illumina may be useful: https://www.illumina.com/documents/products/technotes/technote_coverage_calculation.pdf
Are there any specifications for genomic samples or libraries submitted for sequencing on the HiSeq X?	HiSeq X sequencing is available for whole genome samples (human and other) to an average depth of coverage of 15X or greater. Sequencing of bisulphite or phasing libraries to an average depth of coverage of 15X or greater, is also permitted but unsupported. Illumina does not provide any assurances or guarantees that the performance of the HiSeq X instrument will match published specifications when used for unsupported applications.
Are there any specifications for samples or libraries submitted for sequencing on the NextSeq 500?	The NextSeq 500 has a minor restriction on index sequences that can be used when barcoding libraries. To detect a cluster during template generation, there must be at least 1 base other than G in the first 5 cycles.

Sample Submission

FAQ	**Answer**
When and how do I get a sample submission form?	Sample submission forms are e-mailed to the collaborator once a signed copy of the SOW has been received by the GSC. For certain sample types, an online submission system is available.
What type of containers can I use to submit my samples?	< 24 DNA/RNA samples, submit in 1.5 mL Eppendorf tubes (screw top tubes are not accepted)
	≥ 24 DNA/RNA samples, submit in an Axygen 96-FS-C plate (can be provided by GSC upon request)
	Tissue/cells are accepted only in tube format or matrix tubes
How should I arrange my samples in a 96 well plate?	It is required that samples are sorted by columns in a plate. Wells E12, F12, G12 and H12 must be left empty for internal controls
Why is a plate limited to 92 samples when there are 96 wells in a plate?	Wells E12, F12, G12 and H12 are to be left empty as they are used for internal controls.

Why do I have to provide a tube label when I am providing a sample ID already?	Every sample submitted to the GSC goes through an initial quality check. The label on the physical tube is checked against the tube label provided in the sample submission form.
What if I don't have the specified plate available?	Pre-barcoded Axygen 96-FS-C plate(s) can be provided to the collaborator once the SOW is signed. If you require a plate, please contact GSC_Submissions@bcgsc.ca. Shipping contact information and courier account number are provided by the collaborator to ship a plate. Plate(s) can also be picked-up at the GSC in person.
When and where can I submit my samples?	The sample submission form must be reviewed and approved by GSC personnel prior to submitting samples to the GSC. Regular hours for sample drop-off and plate pick-up: Monday – Friday Times: 9:30-11:30am and 1:30-3:30pm Location: Suite 100-570 West 7th Avenue To enter the building, dial #100 and the receptionist will let you in. The reception is on the ground floor (past the elevators and on the left). Go through to reception and ask the receptionist to call or page anyone from the Biospecimen Core group. We'll come down to reception to meet you.
How do I submit samples via a courier?	Once the sample submission form is approved, samples must be shipped on dry ice and should be addressed to: Dr. Andrew Mungall Biospecimen Core, Room 508 Genome Sciences Centre BC Cancer Suite 100 - 570 West 7th Avenue Vancouver, BC V5Z 4S6 Canada email: amungall@bcgsc.ca Tel: 604-707-5900 ext 3251 When samples have been shipped, we ask that you please email sampleshipments@bcgsc.ca to notify us of your shipment and the associated tracking number, so we can monitor the progress during transit. Please ensure that there is sufficient dry ice for a couple of days. We recommend shipping Monday to Wednesday, as we cannot accept packages on weekends.

Why is spike-in added to genomic DNA samples? Will this affect my sequencing results?	The GSC has extended our QC processes, to enhance our sample identity tracking. We will be adding a small amount of a plasmid (1ng/µg) to each genomic DNA sample upon receipt by the GSC. This plasmid contains a unique insert and allows us to track sample identity and cross contamination throughout the pipeline. The resulting sequence data will contain reads resulting from both the vector (PCR4-Topo) and the insert at a level of 1,000-100,000 reads per lane (spike in reads will not be aligned). Any returned material would also contain this plasmid. This has been extensively tested in both our clinical pipeline as well as several research projects and we are now rolling this process out to all genomic DNA samples received at the GSC. Please advise GSC if you do not wish this tracking spike-in to be added to your sample.

Data and Analysis

FAQ	Answer
What kind of bioinformatics analysis do you provide?	Please see http://www.bcgsc.ca/services/bioinformatic-services for a full list of the standard analyses we provide. If you have a custom analysis in mind that is not listed, please contact us directly.
How much pass filter data am I guaranteed?	We do not have minimum data guarantees, as the data yield depends too much on the sample supplied. We use our internal QC standards to ensure that the best possible data is generated for each sample.
What kind of bioinformatics QC do you do on my samples?	We employ a wide range of quality control metrics in our bioinformatics QC pipeline: - Assessment of technical contaminants such as adapters and sequencing reagents as well as biological contaminants such as bacteria or host species in xenograft samples. - We also look at library type specific metrics such as insert size and duplicate rates for whole genome libraries, ribosomal and mitochondrial content for RNAseq libraries, and capture efficiency for exome libraries. - If multiple samples are submitted from the same patient, we check for possible sample swaps by comparing the samples at positions of common single nucleotide polymorphism. - Selected QC warnings are given if your libraries fall below our standard thresholds. A full list can be found at http://www.bcgsc.ca/services/alert-text-table.
In what format do I receive my sequencing data?	Your data is provided in both fastq and BAM formats by default. Alignment is included in the sequencing price for human samples. For more information regarding the BAM file format, please see https://samtools.github.io/hts-specs/SAMv1.pdf
What software is used for alignment?	Alignment is performed using the Burrows-Wheeler Aligner (BWA) program. Novoalign is used for bisulphite sequence data. Additional alignment, with specific client specified parameters or other aligners may be available upon request at an additional cost. Please contact us for more information.
What reference genome is used for the alignment?	Our current default human genome reference version is hg38, although we support hg19. Please contact us directly for our default genome reference version for any other species.

What if I want my data aligned to a different reference? What if there is no reference for my data?	You can specify any valid reference version for us to use in your alignment when you submit your samples. If we do not have the reference installed in house there will be a cost recovery for installing your reference. If there is no existing public reference for your data, you can provide a custom fasta file, as long as the fasta file can be indexed and is compatible with our aligner. If no reference is provided or the custom reference is not correctly formatted, your BAM file will simply contain all unaligned reads in BAM format.
What reads are included in the BAM file?	All of the raw data is included in the final BAM file, with reads failing the vendor quality checks flagged to allow the user to remove them if desired. Data sequenced on the HiseqX will not contain quality failed reads as the instrument does not output them. Unaligned reads are also included in the BAM file.
Are pooled libraries automatically split by index?	Yes, data from pooled libraries will be supplied to you after splitting by index. Indices are sequenced on a separate read so your data will not contain any indices.
Do you trim the adaptors from my sequence data?	We do not trim adapter sequences from our fastq or BAM files. Generally aligners are able to handle adapter sequence at the end of reads by soft-clipping. The exceptions are bisulphite sequencing reads which are hardclipped in the alignment stage, and miRNA sequencing data for which we do trim adapters due to the short length of the reads and the need for higher sequence specificity in our miRNA profiling pipeline (http://www.bcgsc.ca/platform/bioinfo/software/mirna-profiling). BAM files for both of these library types will not contain adapter sequence.
How do I access my data?	All collaborators will receive an email informing them that their data is available for download from our SFTP site. 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. Data will be automatically deleted from the download site after two weeks. If you are unable to download your data within two weeks, please contact us to re-upload your data. If at that time your data are still available for upload, there may be an additional cost for the re-posting. By default the notification email will be sent to the principal investigator listed on the sample submission and submitter of the samples. Additional email recipients can be specified during submission (Cont. on page 7)

	Once the notification email has been sent, a separate email with login and password details for the SFTP site will be sent to the PI. To protect the privacy of your data, subsequent amendments to the recipient list and creation of additional SFTP accounts will require approval from the PI. If you do not have a SFTP client on your computer, you will need to download and install one before you are able to access your data. Please visit our webpage for a list of some recommended clients that can be downloaded for free, along with links to installation instructions: http://www.bcgsc.ca/services/solseq/sequencing-data-access-using-sftp
How do I match my sample names with the sample names on my data files?	With every data upload, we provide a gsc_library.summary file which can be found in the SFTP folder containing your data. This file provides a mapping between our internal library names and the sample names which you provide on your sample submission form. If you have any problems with your data, please contact data_support@bcgsc.ca
Do you have suggested tools for viewing my BAM files?	Some popular toolsets for working with and viewing sequence data are: IGV (https://www.broadinstitute.org/igv/) Picard (http://broadinstitute.github.io/picard/) SAMTools (http://samtools.sourceforge.net/) For any questions related to the analysis of your data or questions about particular analysis software the SEQanswers (http://seqanswers.com), Biostars (https://www.biostars.org/) and Canadian Bioinformatics Helpdesk (https://bioinformatics.computecanada.ca/) forums are useful resources.
What is your data retention policy?	Sequence data is stored for a minimum of 45 days, and may be deleted after that time without notice.
I need to make my data publicly available for a publication. Can the GSC host my sequencing data or submit my data for me?	As much as the GSC would like to help our collaborators get their data published, we do not have the ability to host data publicly here. However, we have experience submitting to most public repositories including SRA, dbGAP, cgHub and EGA and are happy to answer questions you may have. We are also able to provide submission support on a cost recovery basis.

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 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 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 enable them to advance their careers. 3. The BC Cancer 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