

USER GUIDE



CONSTRUCTED LIBRARIES



Bringing Genomics to Life.

ABOUT US

The GSC's CAP, DAP and ISO 27001 certified technology platform is a high-throughput, large-scale DNA and RNA sequencing and analysis facility that has been designed to maximize analytical capacity, diversity, efficiency, scalability and flexibility. Our state-of-the-art clinical, sequencing, bioinformatics and proteomics platforms are ready to be put to use for your research or clinic.

We partner with researchers, guide experimental design, execute high quality processing of complex and valuable biological samples and provide extensive bioinformatics analyses with the aim of making genomics research accessible to our partners and collaborators within the scientific community.

Please feel free to [contact us](#) if you have any questions about the services we provide.



www.bcgsc.ca | info@bcgsc.ca | [@BCCancer_GSC](https://twitter.com/BCCancer_GSC)

CONTENT

ABOUT THIS GUIDE	4
.....	
DNA LIBRARY CONSTRUCTION	5
.....	
RNA LIBRARY CONSTRUCTION	6
.....	
INDEXED LIBRARIES	7
.....	
LIBRARY QUANTITY AND QUALITY	9
.....	
LIBRARY TROUBLESHOOTING	10
.....	
SUBMISSION REQUIREMENTS	12
.....	
ONLINE SUBMISSION GUIDE	14
.....	
SAMPLE SUBMISSION BY COURIER	23
.....	
SAMPLE DROP-OFF	24
.....	
APPENDIX A	25
.....	
APPENDIX B	26
.....	

| THIS GUIDE

This user guide is for the submission of sequencing libraries prepared by your laboratory. If you are submitting DNA or RNA samples for library construction at the GSC please refer to our [DNA or RNA sequencing user guides](#).

The GSC offers sequencing for labs that wish to perform their own library construction. Commercial sources of oligonucleotides and reagents for successful library construction are widely available for purchase. Please note the following when ordering oligonucleotides:

- The adapters must be phosphorylated.
- The adapter can be synthesized with a special linkage between the 3' terminal T and the preceding C. This is a phosphorothioate linkage which renders this overhanging T more nuclease resistant (after annealing the top and bottom adapter oligonucleotides). This provides nuclease resistance for this base, diminishing the probability of adapter dimers.
- Please consult with the constructed library requirements on [page 13](#) before submitting your samples.

1 DNA LIBRARY CONSTRUCTION

Libraries compatible with GSC pipelines can be constructed using Illumina's TruSeq and Nextera DNA sample preparation kits. The sequence for the TruSeq Universal adapter is as follows:

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

Libraries constructed using alternate kits must be compatible with the GSC's standard sequencing primers (please refer to Table 1 on [page 8](#)). GSC compatible primer sequences and index sequences (if applicable) must be provided to the GSC prior to sample submission. We encourage you to perform the research and assess your project's compatibility with our DNA pipeline.

Constructed libraries submitted for HiSeq X sequencing must meet Illumina's minimum coverage requirement of 15X or greater. Sequencing of human 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.

2

RNA LIBRARY CONSTRUCTION

For mRNA-seq libraries, the standard TruSeq Illumina kit is compatible with our pipeline. Commercially available products, with more specialized applications for other types of RNA libraries, may also be compatible with our pipeline. Please [contact us](#) for more information.

microRNA libraries constructed using external adapters may not be compatible with our sequencing pipeline. The GSC has only constructed microRNA libraries using our in-house microRNA adapters; therefore, external adapters and library construction protocol(s) are untested and submission is at the collaborator's risk.

We encourage you to perform the research and assess your project's compatibility with our RNA pipeline.

3 INDEXED LIBRARIES

Index sequences must be provided, and should have a balanced mix of bases at each position to ensure no focusing issues arise during the index read.

Data from libraries submitted as pools will be split using the provided index sequences. Splitting by index is performed with a one-base-mismatch tolerance (i.e. index reads with one mismatch from the expected index will still be counted towards that index). Therefore, please ensure that no two indices in the same pool are different by less than two bases, as any pair of indices that differ by only one base (e.g. AGTCCA and ATTCCA) will be considered ambiguous in the splitting process and reads with either index will be lost from the split bam files. If indices that were not provided are detected in the sequence data with >1% of the total number of reads, these indices will also be split out and disseminated with your data.

It is important to ensure that when you submit pooled libraries the indices in each pool contain an equal representation of each base in each position (e.g. do not have all your indices in one pool start with 'A'). The HiSeq instruments do not focus well when all clusters have the same base in the same position. Please refer to Nextera codes in [Appendix A](#).

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 one base other than G in the first five cycles. Index sequences and protocols used for library construction must be confirmed to be compatible with the GSC's sequencing pipeline prior to sample submission.

TABLE 1. GSC COMPATIBLE ADAPTER & PRIMER SEQUENCES

TYPE	NAME	SEQUENCE
Adapter 5'	Standard GSC	CAAGCAGAAGACGGCATACGAGAT NNNNNN CGGTCTC GGCATTCTGCTGAAC CGCTCTTCCGATCT
Adapter 3'	Standard GSC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCT
Seq read 1 primer	Standard GSC r1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Seq read 2 (index) primer	GSC index	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG
Seq read 3 primer	Standard GSC r3	CGGTCTCGGCATTCTGCTGAACCGCTCTTCCGATCT
Adapter 5'	TruSeq	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC NNNN NN ATCTCGTATGCCGTCTTCTGCTTG
Adapter 3'	TruSeq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTTCCGATCT
Seq read 1 primer	TruSeq r1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Seq read 2 primer	TruSeq Index	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Seq read 3 primer	TruSeq	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

4 LIBRARY QUANTITY & QUALITY

Once constructed libraries have been approved and submitted, the GSC will assess the quantity of the libraries. If the libraries do not pass the GSC quantity check or if there are apparent issues with the libraries, we will contact you. However, despite passing quantity checks, libraries may not result in satisfactory data. The GSC is not responsible for the quality of submitted constructed libraries or for the generation of data from such libraries. If an additional level of quality check is needed using the Agilent Bioanalyzer, we can perform the QC at an extra cost.

It is expected that libraries have been purified using a suitable PCR clean-up kit, have an A260/280 ratio of > 1.8 and an A260/230 ratio of > 1.2 . These requirements are mandatory.

The Agilent Bioanalyzer can be used to provide visual examination of the constructed libraries. The “perfect” library profile (Figure 1) shows a single peak of the expected size. Common additional forms include primer or adapter dimers and broader bands of higher molecular weight (MW) than the expected peak (Figure 2).

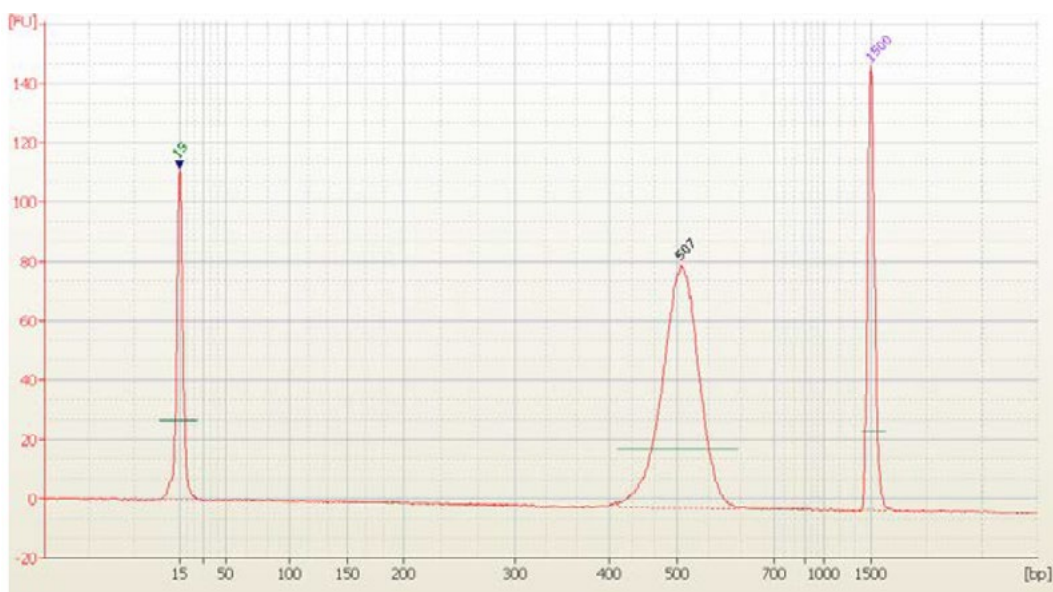


FIGURE 1. Agilent Bioanalyzer profile of a constructed library.

5

LIBRARY TROUBLESHOOTING

Primer dimers can be minimized by size-selection (e.g. using magnetic beads or gel excision) but may not pose a problem unless they completely dominate the reaction. Useful data has been obtained from libraries despite the presence of 50% primer dimers.

Adapter dimers can be a problem because they sequence efficiently. As a result, whatever the proportion of adapter dimers present in your library, at least the same or more of the proportion of reads will be seen in your final data files. Because adapter dimers are very efficient at generating clusters on the flow cell, usually, a higher proportion of reads will be seen in the final data files. Adapter dimers can be minimized by adjusting the adapter:insert ratios during library construction and exercising care in gel extraction or other size selection steps.

Larger MW fractions are typically broader shaped forms when visualized on the Agilent Bioanalyzer and are probably a result of excess amplification during the final PCR step. While some amounts of larger MW fractions are tolerable, the library should be re-amplified from the gel extracted material if they are too prominent.

The proper size of the inserts will depend on the application and the type of sequencing performed. As a general guideline, libraries submitted for sequencing should have inserts no longer than 600 bp.

The GSC cannot guarantee that constructed libraries will be acceptable for sequencing even after samples have been submitted, and we cannot guarantee the quality of sequencing data from these libraries.

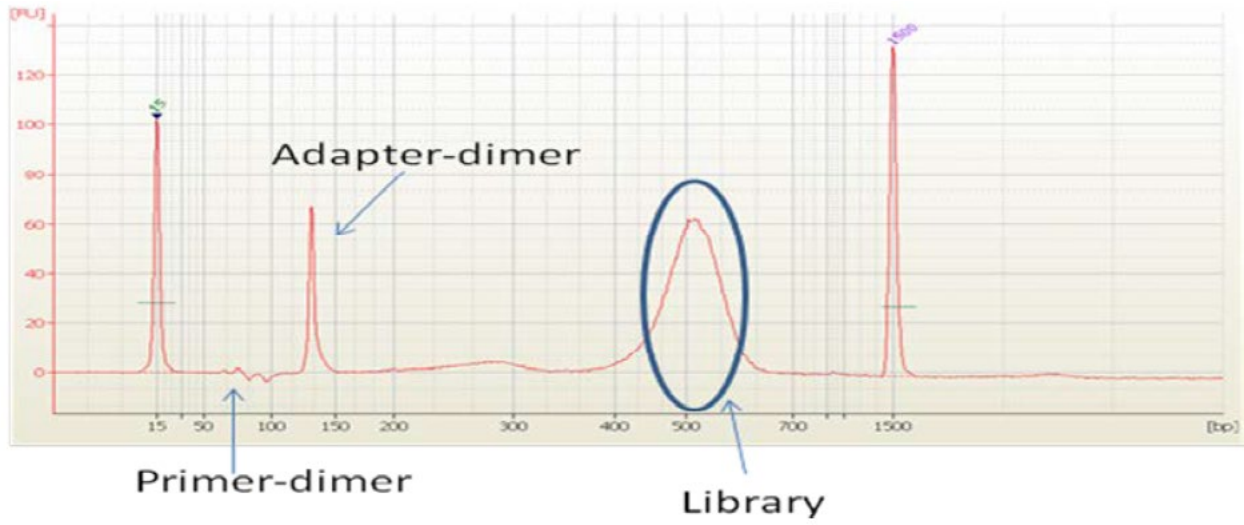


FIGURE 2. Agilent Bioanalyzer profile showing primer dimers and adapter dimers.

6

SUBMISSION REQUIREMENTS

Please refer to [Table 2](#) for minimum submission volumes and concentrations. For maximum yields we strongly recommend that libraries are quantified using qPCR (or Qubit). Nanodrop is not recommended because readings are not accurate at the very dilute concentrations utilized in next generation sequencing protocols. Additionally, spectrophotometric methods can overestimate library quantities by including unadapted, or incorrectly adapted, products.

Average library size should be determined with the Agilent Bioanalyzer 2100 High-Sensitivity DNA kit.

The quantity and concentration of your library is critical. It is not uncommon to require more than the minimal amounts given below in order to achieve full sequence yield. If your sample concentrations/volumes fall just below the minimum required amount, please let us know as we may still be able to run your samples if you do not require maximum sequence yields.

Please submit samples in 1.5 mL tubes.

TABLE 2. LIBRARY SUBMISSION REQUIREMENTS

	MiSeq		NextSeq		Rapid Mode		High Output			
Read length (bp)	Any		Any		Any		75 or 125		150	
Total library size (bp)	~350	~550	~350	~550	~350	~550	~350	~550	~350	~550
Min. conc. (nM)	1.3	2.4	1.1	1.1	1.4	2.4	2.0	2.4	1.5	2.5
Max. conc. (nM)	32	60	50	50	50	90	50	90	80	130
# runs at min. conc.	Minimum volume (ul)									
1	21	21	4	4	21	21	21	21	7	7
2	40	40	6	6	40	40	21	21	12	12
3	59	59	8	8	59	59	21	21	17	17
4	78	78	10	10	78	78	21	21	22	22
5	97	97	12	12	97	97	21	21	27	27
6	116	116	14	14	116	116	21	21	32	32
7	135	135	16	16	135	135	21	21	37	37
8	154	154	18	18	154	154	21	21	42	42
9	173	173	20	20	173	173	21	21	47	47
10	192	192	22	22	192	192	21	21	52	52
11	211	211	24	24	211	211	21	21	57	57
12	230	230	26	26	230	230	21	21	62	62

7 ONLINE SUBMISSION GUIDE

Our guide to online submissions will provide helpful information that will assist you in navigating our online submission website. Please contact us at

GSC_Submissions@bcgsc.ca at any time if you have any questions.

***The online submission form must be completed and approved prior to submitting your samples to the GSC.**

Please note libraries constructed using 10x Genomics kits are not currently supported by our online submission system.

OVERVIEW

Please read this entire section prior to starting your submission. This section will describe all information that is required to successfully fill out your submission form. We recommend ensuring all required information is gathered prior to starting your submission.

To start a new submission, click on the **Dashboard** or **Submissions** tab on the left side panel, and then the “Start New Submission” button. Please note that an active statement of work (SOW) is required to start a new submission and each SOW requires a new submission.

To continue working on an existing submission, click **Submissions** on the left side panel. This is the central location from which you can view all of your current and past submissions, as well as your submission status.

Questionnaire

There are three tabs to complete in this section, corresponding to three categories of information required. Please note that all the fields in the three tabs are mandatory. You need to complete each field in order to proceed to the next page.

1. SOW Selection
2. Sample Details
3. Work Request

Sample Information

There are five mandatory tabs to complete to provide information about your samples. If you are submitting xenograft samples there will be an additional tab to complete.

We recommend that you complete all the fields relevant to your samples. Some fields are mandatory (denoted by *) and other fields are optional (but recommended for tracking purposes).

1. Pool Information
2. Library Information
3. Work Request Assignment
4. Comments and Extra Information
5. Submission Review

QUESTIONNAIRE

1. SOW Selection

- Submission Name: create a submission name that is complex enough to easily and uniquely identify each submission. For example, “Doe Lab ABC cell-line genomes.”
- Statement of Work (SOW): select a SOW from the drop-down menu provided. If your SOW is not listed, please contact SOW@bcgsc.ca. If a SOW has a revision associated, please select the parent SOW. Once selected, options for revisions will appear in a separate drop-down list. Please select the appropriate revision for the sample submission.
- Once the SOW is selected, PI and Dissemination Recipients(s) will be displayed on the right. To edit the dissemination recipients list, please return to the SOW homepage in the SOWs section. Only the PI or a designated Dissemination Editor can edit this list.

2. Sample Details

- Sample Type: only one sample type is allowed per submission.
- Sample Number: enter the number of pooled samples (# of tubes being submitted). Please submit the samples in 1.5 mL tubes.
- Taxonomy: you may enter multiple taxonomy IDs per submission. You may search by typing NCBI taxonomy ID (e.g. 9606) or name (e.g. human). Only selected taxonomy IDs will be available later in your submission. Note that by default, both BAM and fastq files are provided.
- Additional Sample Information: select “Xenograft” if applicable.
- Unused sample: indicate what you want us to do with any unused samples upon completion of your project. If you select “destroy unused sample,” we will discard your

samples six months following data dissemination.

3. Work Request

- Please refer to your SOW to obtain the work request information that will be required in this section.
- If the samples you are submitting have different library sequencing goals, you will need to click “Add Additional Work Request” to be able to enter different goals for each sample.
- Once the sample information is completed and before the form is submitted, you will assign one of the work requests to each sample. At this point, you can also edit or add work request details.
- You will also be able to enter additional comments regarding your work request, should extra or clarifying information be required.

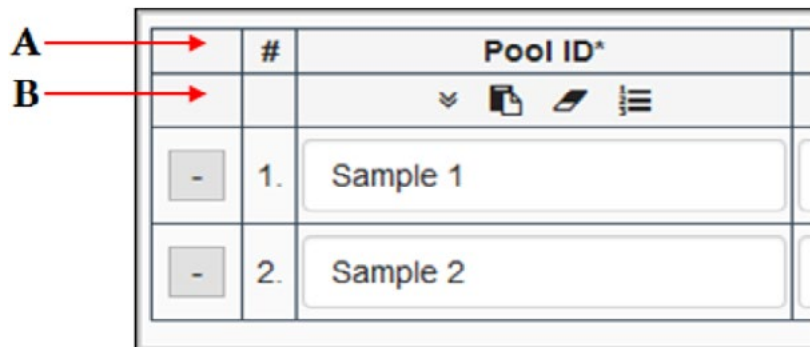
SAMPLE INFORMATION

Overview

There are at least five tabs to complete in this section. Note that mandatory fields are denoted by *. Certain field selections will grey out mandatory fields. In these situations field entry is not required.

For additional information on any given field, click the column heading (Figure 3, Arrow A).

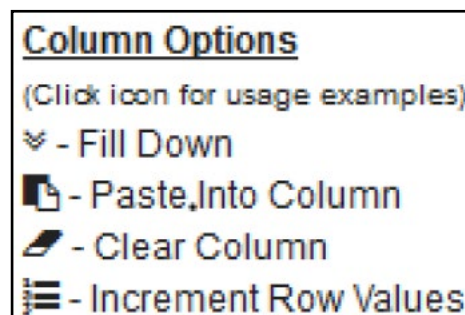
Under each column heading are tools to help you complete each field more conveniently (Figure 3, Arrow B). Please refer to the tool tips on the top right corner to find out how each tool can help (Figure 4).



The screenshot shows a table with three columns. The first column contains minus signs (-). The second column contains numbers 1 and 2. The third column contains text 'Sample 1' and 'Sample 2'. The header row is shaded and contains the text '# Pool ID*'. Below the header, there are four icons: a downward arrow, a document icon, an eraser icon, and a list icon. Red arrows labeled 'A' and 'B' point to the header and the icons respectively.

	#	Pool ID*
-	1.	Sample 1
-	2.	Sample 2

FIGURE 3




The screenshot shows a tooltip titled 'Column Options' with the instruction '(Click icon for usage examples)'. It lists four options with corresponding icons: a downward arrow for 'Fill Down', a document icon for 'Paste Into Column', an eraser icon for 'Clear Column', and a list icon for 'Increment Row Values'.

Column Options	
(Click icon for usage examples)	
▼	- Fill Down
📄	- Paste Into Column
🧼	- Clear Column
☰	- Increment Row Values

FIGURE 4

For each tab, you need to complete all the accessible mandatory fields before you can proceed to the next page. When you click the “Next” button, the webpage will validate each field to make sure all the cells meet our requirements. If there are any errors, cells will display red or orange error message(s). Please edit all the errors before proceeding to the next page. Please note that fields with warnings will be highlighted yellow for your review. It is not necessary to resolve warnings before proceeding.

Your draft submission will be saved when you click “Next”. Note that a submission can be saved at any step using the “Save as Draft” button at the bottom of the page.

You may add or remove a sample by clicking the  icon located on the top left corner, next to your submission name. Note that if you decrease the sample number by one, the last row will be deleted. If you decrease the sample number by two, the last two rows will be deleted.

1. Pool Information

- Pool ID*
- Tube Label*
- Mixed Taxa?*
- Taxonomy For Alignment*
- Mixed Anatomic Site?*
- Anatomic Site*: this is a type-ahead search field. Please enter three or more characters to search and select the best option. If there are no appropriate matching selections, you may enter free text.
- Tissue Disease Status*: this field pertains to the tissue, not the patient. Please select from the drop-down menu provided.
- Mixed Pathology?*

-
- Pathology*: this is a type-ahead search field. Please enter three or more characters to search and select the best option.

2. Library Information

- Library Type*
- Library Construction Method*
- Size Range (bp)*
- Average Size (bp)*
- Pool Type*
- Index Read Type*
- Index Sequence*: please provide the expected index sequence(s). For dual index samples separate the first index (i7) and the second index (i5) with a "-". For example ACGTCATC-TCAGCTAG.

Please see [Appendix A](#) for further information regarding i5 index sequence entry.

Please see [Appendix B](#) for single cell whole genome index sequence entry.


3. Xenotransplant Recipient Information (for xenograft samples only)

- Xenograft*
- Recipient Taxonomy*
- Recipient Strain

4. Work Request Assignment

Prior to preparing your samples, review the sample requirements on [page 13](#) to ensure that your samples meet the requirements for volume and concentration.

- Work Request*
- # of lanes*
- Concentration (nM)*
- Volume (µL)*
- Quantification Method*

Assign a work request to each sample. You may edit the work request by clicking  located on the top right corner of the work request summary box. The required volume and concentration are validated based on the sequencing requested. If you are submitting less than the minimum amounts, you will be asked to accept the risks of failure.

5. Comments and Extra Information (Optional)

Use this field to provide any extra information or details regarding your submission or sample.

6. Submission Review

AFTER SAMPLE SUBMISSION

The submission form will be reviewed for completeness and accuracy within two business days. You will then receive a notification of approval or a request for edits.

Once your submission has been approved, you will receive an email with detailed instructions for shipping your samples.

You can retrieve the submission info at any time from the submissions page under the **Accepted** tab.

If you have any questions during this process please contact the GSC submissions team at GSC_Submissions@bcgsc.ca.

8

SAMPLE SUBMISSION BY 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 Canada
V5Z 4S6

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.



SAMPLE DROP-OFF

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: 9:30-11:30 am and 1:30-3:30 pm

Location:

Suite 100-570 West 7th Avenue, Vancouver, BC V5Z 4S6

To enter the building, dial #100 on the intercom 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.

A APPENDIX

ENTERING ILLUMINA NEXTERA CODES ON SAMPLE SUBMISSION FORMS

Documentation provided by Illumina for Nextera sample preparation includes instructions to enter i5 bases differently on sample sheets when requesting NextSeq or HiSeq X sequencing rather than GA/HiSeq/MiSeq sequencing. The GSC's automated processes require the GA/HiSeq/MiSeq i5 base sequence and will be automatically reverse complement if NextSeq or HiSeq X sequencing is requested. Please only provide the GA/HiSeq/MiSeq adapter bases and **DO NOT** enter as recommended by Illumina.

i5 bases in adapter	i5 index name	i5 bases for entry on sample sheet	
		GA/HiSeq/MiSeq	NextSeq
TAGATCGC	[N/S/E]501	TAGATCGC	GCGATCTA
CTCTCTAT	[N/S/E]502	CTCTCTAT	ATAGAGAG
TATCCTCT	[N/S/E]503	TATCCTCT	AGAGGATA
AGAGTAGA	[N/S/E]504	AGAGTAGA	TCTACTCT
GTAAGGAG	[N/S/E]505	GTAAGGAG	CTCCTTAC
ACTGCATA	[N/S/E]506	ACTGCATA	TATGCAGT
AAGGAGTA	[N/S/E]507	AAGGAGTA	TACTCCTT

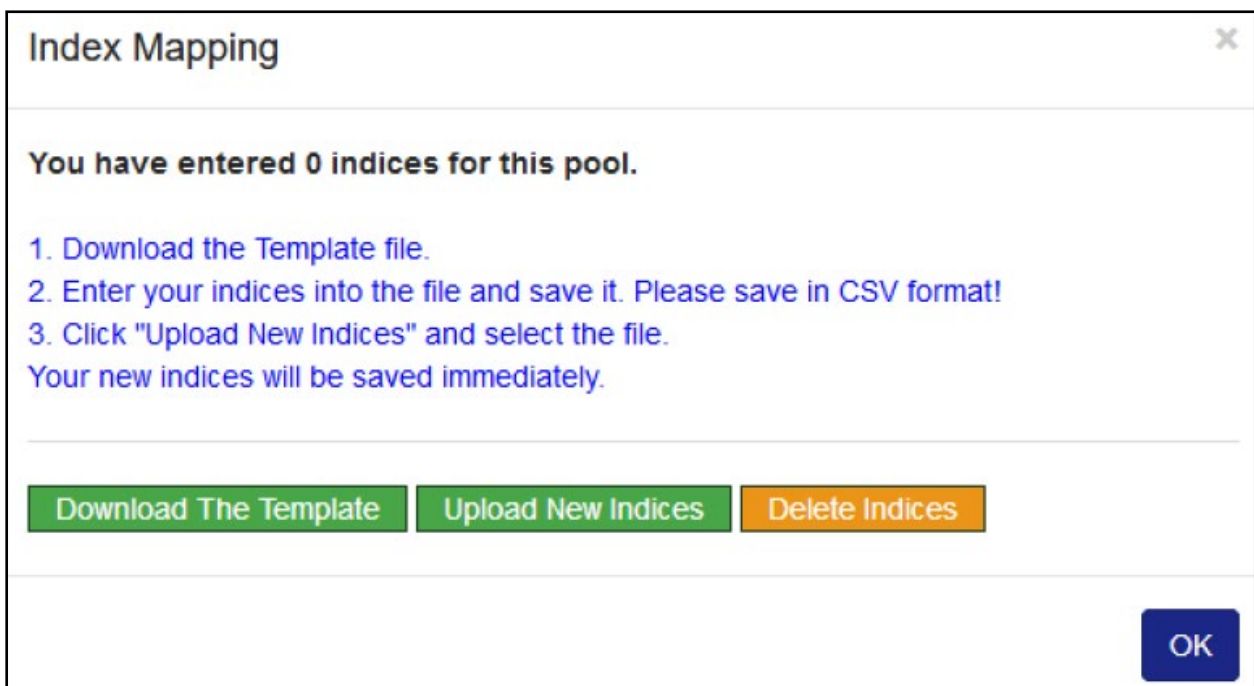
Correct entry of i5 bases
for BCGSC Sample
Submission Forms for all
sequencing requests

Oligonucleotide sequences © 2007-2013 Illumina, Inc. All rights reserved.

B APPENDIX

To facilitate the entry of the large number of index sequences for single cell whole genome samples, an index entry template can be downloaded.

Click in the sequence index entry box. A message will pop up with directions and the template for download:



Follow the directions for index sequence entry making sure to save the file in CSV format. The Sub-Library ID is not mandatory.



Technology platform

Echelon Innovation Centre
570 West 7th Avenue
Vancouver, B.C. V5Z 4S6
Phone: 604-707-5900



Research department

BC Cancer Research Centre
675 West 10th Avenue
Vancouver, B.C. V5Z 1L3
Phone: 604-675-8000