USER GUIDE

CONSTRUCTED LIBRARIES

BC CANCER RESEARCH
Provincial Health Services Authority

CANADA'S MICHAEL SMITH
GENOME SCIENCES CENTRE

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

CONTENT

ABOUT THIS GUIDE

This user guide is for the submission of sequencing libraries prepared by your laboratory. If you are submitting a different sample type, please refer to the other user guides on **our website**.

The GSC offers sequencing for labs that wish to perform their own library construction.

Library construction kits should be compatible with standard sequencing primers (i.e. Illumina TruSeq or Nextera primers). If custom sequencing primers or a custom sequencing recipe are required, please contact us prior to submission.

1 LIBRARY CONSTRUCTION

Constructed libraries need to be compatible with standard Illumina sequencing primers (i.e. TruSeq or Nextera) and must be constructed with dual indexed adapters to allow for pooling with other samples. Illumina adapter and primer sequences can be found on the Illumina website. The use of UDIs (Unique Dual Indexes) is strongly recommended to reduce the effect of index hopping.

Libraries constructed using non-TruSeq or Nextera chemistry/adapters must be assessed for compatibility with sequencing requirements.

If your construction method results in libraries with biased base distribution (i.e. an imbalance in the distribution of the 4 bases at any cycle(s)) common in some library types such as amplicon and certain 10x Genomics libraries, please let us know via the comments section. Please also confirm if you have any guidance on the appropriate percentage of PhiX genome spike-in to overcome this bias in sequencing. Illumina sequencing instruments can be adversely affected by bias in libraries, potentially affecting base quality and read yield.

Please contact us if you are using a custom library construction prior to submission. 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 platform.

2 INDEXED LIBRARIES

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

Data from libraries 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.

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 instruments do not focus well when all clusters have the same base in the same position.

At the GSC, we ask that all indices be provided to us in the forward format, regardless of the sequencing platform. We will automatically reverse complement the i5 indices as needed depending on the sequencer workflow (forward or reverse). Please refer to **Appendix A** for details regarding index sequence entry.

Certain sequencing instruments (those with 2-colour chemistry) have 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.

De-multiplexing at the GSC is optimized for libraries with dual index reads read separately from the insert reads. We strongly recommend against inline indices that are read as part of the insert as this can confound our ability to accurately split reads in complex library pools having potentially large numbers of unassigned reads. If your libraries are constructed using inline indices please contact us prior to submission.

3 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 fragment analysis (e.g. 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).
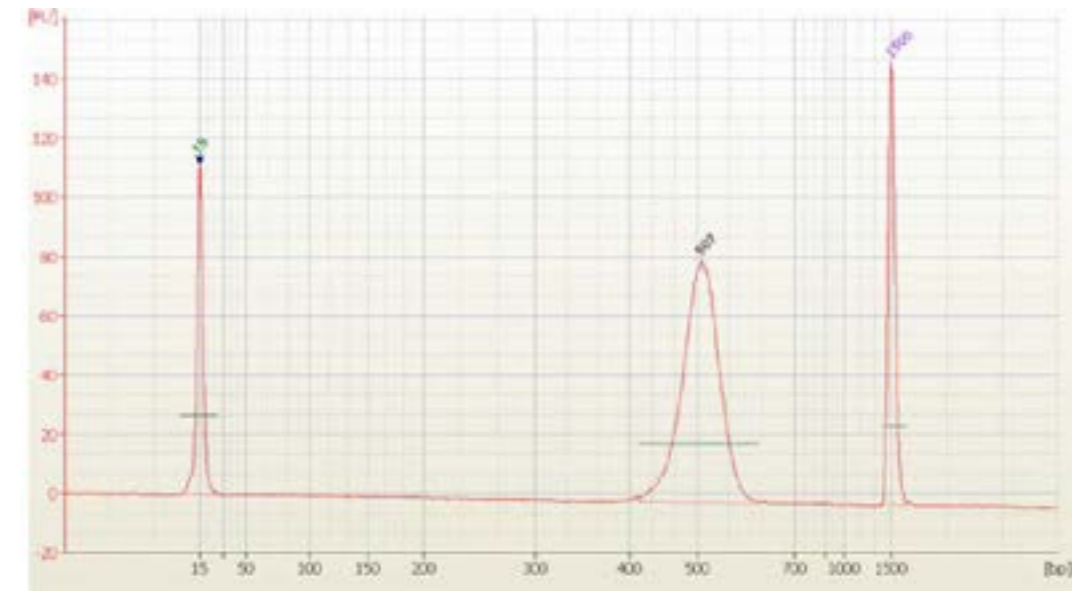

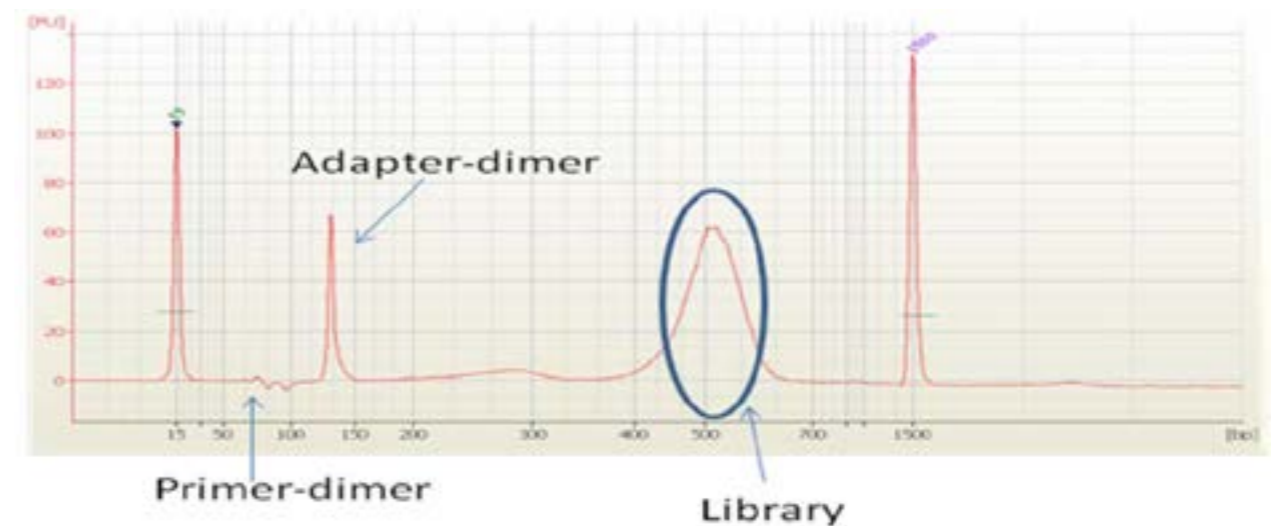**FIGURE 1.** Agilent Bioanalyzer profile of a constructed library.

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

4 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 short-read 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.

5 SUBMISSION REQUIREMENTS

For NovaSeq X sequencing, we request a minimum of 15 µL of 4 nM library resuspended in EB buffer or equivalent. This includes a library volume required for MiSeq index QC before we pool with other in-house libraries. More library may be required for a large number of reads (>600 million). For a full NovaSeq X 25B lane (6 billion reads), 25 µL of 4 nM library is required. For a full NovaSeq X 25B flow cell (8 lanes, 48 billion reads), 130 µL of 4 nM library is required.

For NextSeq 2000 sequencing, we request a minimum of 21 µL of 2.5 nM library resuspended in EB buffer of equivalent.

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 unadaptered, or incorrectly adaptered, 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. 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.

Samples must be submitted in 1.5 mL tubes.

6 ONLINE SAMPLE SUBMISSION

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

Our "Online Submssions Webpage" can be found **here**. Please use Firefox or Google Chrome browsers.

This guide to online submissions will provide helpful information that will assist you in navigating our online submission website.

Please write to **GSC_Submissions@bcgsc.ca** at any time if you have any questions.

GETTING STARTED

1. You can access the GSC Sample Submission Interface by entering the link below in Firefox or Google Chrome: **https://www.bcgsc.ca/samplesubmissions**

2. Enter your username and password, and then click Login:

Username

Password

Forgot Password?

Log In

3. To start a new submission, click Dashboard or Submissions on the left side panel:

- Dashboard
- Submissions
- $ SOWs
- Sample Tracker
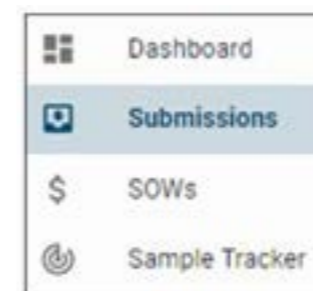
- Click the ▦ (Dashboard) icon on the left side panel, and then click Start New Submission under Quick Links, or

- Click the (Submissions) icon on the left side panel, and then click CREATE NEW SUBMISSION located on the top right corner.

HOW TO CREATE A NEW SAMPLE SUBMISSION

1. Submission Name: This is an optional field to help you identify your submission. Create a submission name that is complex enough to easily and uniquely identify each submission.

2. Type of Submission: Select from the drop-down menu. Only one sample type is allowed per submission.

3. Statement of Work: Select from the drop-down menu. Once selected, PI, Dissemination Recipient(s), and Scope of Work will be displayed. To make any changes to this information, please contact **SOW@bcgsc.ca**.

4. Sample Number: Enter the number of constructed library pools (number of tubes) being submitted. Samples must be submitted in 1.5 mL Eppendorf tubes.

5. Taxonomy: You may enter multiple Taxonomy IDs per submission. Search for the taxa of your samples by typing NCBI Taxonomy ID (e.g., 9606) or name (e.g., human). Only selected Taxonomy IDs will be available later in the "SAMPLE INFORMATION" section under the Taxonomy field.

6. Sample Options: Select Xenograft if applicable. If selected, there will be additional fields to complete.

7. Prepopulate the form by uploading an Excel file: Advanced users (i.e., those who often use the GSC Sample Submission Interface) have an option to prepopulate the sample information form by uploading an Excel file. This is not recommended for first time submitters.

- Click the checkbox next to Prepopulate the form by uploading an Excel file:

- Click "DOWNLOAD TEMPLATE".

- Enter your sample information directly into the Excel file that was generated:

  » You can copy and paste entries from a pre-existing Excel file that contains your sample information.

  » If a field has a drop-down menu, you must select from the available options.

  » Some fields may require validation after the sample information form has been populated.

  » All fields do not need to be completed in order to use this feature to prepopulate the sample information form.

  » Any missing or invalid fields can be corrected within the actual sample information form.

  » Do not use the Prepopulate form by uploading an Excel file tool to enter Index Information. Index Information should be entered directly into the actual sample information form.

- Once completed, save the Excel file on your computer.

- Go back to the "Create a New Submission" section and click "UPLOAD EXCEL FILE".

- A dialogue window will open that allows you to navigate to the location on your computer where you saved the Excel file.

- The "SAMPLE INFORMATION" section will be auto-populated based on the entries in the Excel file.

8. 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.

9. Click "START DRAFT" on the bottom right corner.

SAMPLE INFORMATION FORM

Overview

1. Mandatory fields are denoted by *****. Certain field selections will grey out mandatory fields. In these situations, field entry is not required. For example, if Tissue Disease Status is "Normal," Pathology will be greyed out, so you do not need to provide this information.

2. For additional information on any given field, click the ⓘ icon next to the column heading:

3. There are tools on the top right corner to help you complete each field more conveniently:

- **Paste Range:** Paste Range can be helpful when you have a pre-existing Excel file that you would like to copy and paste line items from. For the fields without drop-down menu, you can copy and paste directly into the form from your Excel file without using this tool.

- **Fill Down**: Fill Down can be helpful when you have several consecutive rows that will contain the same value.

- **Clear Column:** Clear Column can be helpful when you would like to reset values in the entire column.

4. Validator will give an orange warning message for any entries that we want you to double-check. If the information looks correct, you can proceed as is. You will receive a red error message when the entry is missing or incorrect. These entries must be corrected before you can submit your sample information form:

Sample Information

1. Pool ID: This is a unique identifier referring to the actual sample being submitted. You will receive your data back linked to Pool ID.

Pool ID Requirements:

- Between five and 80 characters with no symbols and no spaces (hyphen and underscore are allowed);

- Sufficiently complex to help distinguish your samples from other samples; and

- If submitting two or more identical samples, indicate that they are identical by using the same Sample ID.

2. Tube Label: Unless otherwise instructed, ensure that your samples are submitted in 1.5 mL Eppendorf tubes. Enter Tube Label as it is written on the tube. This can be the same as the Sample ID.

Tube Label requirements:

- Less than 13 characters with no symbols (hyphen and underscore are allowed)

- All labels must be unique to differentiate between tubes.

3. Mixed Taxa?: Select from the drop-down menu. Select "Yes" if your pool contains more than one taxon.

4. Taxonomy for Alignment: Auto-filled from the SET UP section or select from the drop-down menu if more than one taxonomy is available.

5. Mixed Anatomic Site?: Select from the drop-down menu. Select "Yes" if your pool contains multiple anatomic sites.

6. Anatomic Site: Enter the site from which the sample you are submitting was derived (e.g., peripheral blood, liver). This is a type ahead input field. Begin typing the anatomic site and a list of suggestions will appear for selection. If necessary, you may enter free text.

7. Tissue Disease Status: Select from the drop-down menu.

8. Mixed Pathology?: Select from the drop-down menu. Select "Yes" if your pool is composed of samples with more than one associated pathology.

9. Pathology: This is a type ahead input field. Begin typing the disease name. A list of suggestions will appear for selection. The list is generated from http://disease-ontology.org. If there are no appropriate matching selections, you may enter free text. **Do not use abbreviations**. An orange warning message will display, but you can still proceed.

10. Library Type: Select from the drop-down menu what type of constructed libraries you are submitting. For Amplicon and Specific Capture, please add PhiX recommendations in the COMMENTS section.

11. Library Construction Method: Enter the library construction method.

12. Size Range (bp): Enter the total size range including adapters (e.g., 250-350).

13. Average Size (bp): Enter the total average size including adapters.

14. Pool Type: Select from the drop-down menu:

- Illumina Compatible if using a library construction kit resulting in libraries which can be sequenced with standard Illumina sequencing primers.

- Custom if using a library construction kit that requires any custom sequencing primers. Additional information such as sequencing primer sequences can be provided later in the Comments section.

15. Index Read Type: Select from the drop-down menu the expected index primer sequence reads. Your libraries must be constructed with dual indexed adapters to allow for pooling with other samples.

16. Contains UMI Adapter?: For ChIP and Single Cell Whole Genome libraries, indicate if the pool contains a UMI Adapter Sequence.

17. UMI Adapter Details: For ChIP and Single Cell Whole Genome libraries, provide details about the adapter, such as the length of the UMI itself, length of the UMI adapter that needs to be trimmed off, and the location of the UMI (beginning of read 1, index read 2 of the dual index, etc).

18. Index Information: Provide the expected index sequence(s). For dual index samples, separate the first index (i7) and the second index (i5) with a "-" (e.g, ACGTCATC-TCAGCTAG). Please see **Appendix A** - We ask that i5 indices be entered in the forward direction, regardless of which platform will be used for sequencing.

In order to enter Index Information:

a. Double-click on the blue "0 sequences" button:

b. Index Sequences for Sample window will pop up. You can copy and paste directly into this window:

c. Once Indices and Sub-Library ID (optional) have been entered click ADD INDICES.

d. Make sure the blue button displays the correct number index sequences in your pool.

19. Chemistry Version: For Chromium libraries, select from the drop-down menu. There are 3 selections available: V1, V2, and V3. V1 can be selected for V1.1 and accordingly for other versions.

20. Chromium Index Names: For Chromium libraries, enter Chromium Index Names (such as SI-GA-A1) and number of cells for each Sub-Library (optional). **Do not enter the Chromium Index Sequences.**

In order to enter Chromium Index Names:

a. Double-click on the blue "0 index names" button:

b. Chromium Index Names for Sample will pop-up. You can copy and paste directly into this window:

c. Once Index Name and Number of Single Cells have been entered, click "ADD INDEX NAMES".

d. Make sure the blue button displays the correct number of index names in your pool.

21. Concentration (nM): Concentration of nucleic acid that you are submitting in nM – **Do not enter the unit**. Prior to preparing your samples, review to ensure that your samples meet the requirements of volume, concentration, and amount for submission.

22. Volume (uL): Volume of nucleic acid that you are submitting in uL – **Do not enter the unit.** Prior to preparing your samples, review to ensure that your samples meet the requirements of volume, concentration and amount for submission.

23. Quantification Method: Select the method used to quantify your samples from the drop-down menu. Note that qPCR and Qubit are the preferred quantification methods.

WORK REQUESTS & WORK REQUEST ASSIGNMENT

1. Please refer to your SOW to obtain the work request information that will be required in this section.

2. Select Platform, Read Type, and Millions of Reads from the drop-down menu for each sample.

3. If all the samples have the same work request, use the FILL DOWN tool on the top right corner.

COMMENTS

Use the "COMMENTS" section to provide extra information or details regarding your submission of samples such as PhiX recommendations for Amplicon and Specific Capture libraries.

SUBMITTING THE SAMPLE INFORMATION FORM

Once you have reviewed all the fields and are satisfied with the sample information form, click "SUBMIT FOR REVIEW" on the top right corner.

1. If any of the samples do not meet the minimum recommended amount, click on the "I accept the risks" checkbox in order to proceed.

2. If there are any errors that need to be addressed, you will see the following message. Click "VIEW" and you will be taken directly to the cell whose error must be fixed before submitting.
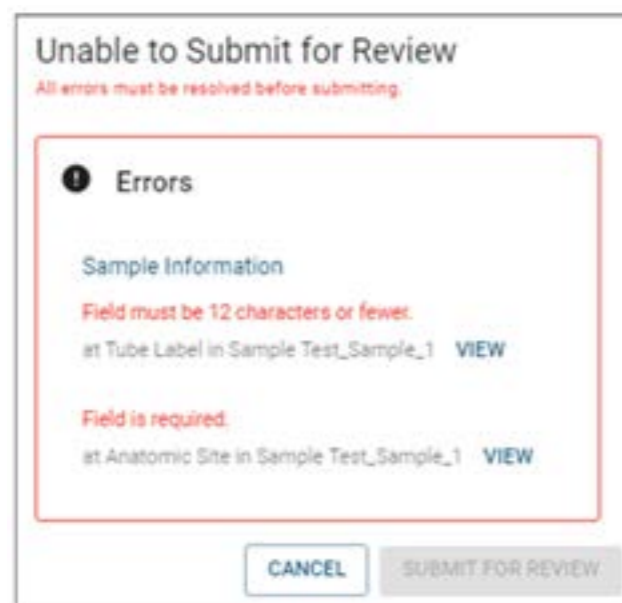
3. Once your sample information form has been successfully submitted, you will receive "Sample Information Form Received for Review" email notification.

AFTER SUBMITTING SAMPLE INFORMATION FORM FOR REVIEW

1. Your sample information form will be reviewed for completeness and accuracy. You will then receive "Edits Requested" or "Submission Approved" email notification.

For "Edits Requested":

- In order to resubmit the form, **you must confirm or amend** all of the fields which have been highlighted yellow or orange by the reviewer. As you address each field in the form, its colour will change to green or blue. **When there are no more yellow or orange fields left, you can resubmit.**

- Some fields may also contain a comment made by the reviewer, in which case those fields will be marked with a small red triangle. You can hover your mouse over those fields to see the reviewer comments.

- You also have the option of making edits to any field, including those which have not been highlighted by the reviewer. Any such fields will become highlighted purple, and are subject to final review by the GSC Submission team.

- Here is a legend of the highlighting colours that you may encounter while making edits:

  » **Yellow**: Needs Attention; the reviewer believes an edit may be required.

  » **Orange**: Changed by Reviewer; the reviewer has already changed the value of this field, and would like you to confirm (or re-edit). Hover over the field with your mouse to see the change that has been made.

  » **Blue**: Amended by Submitter; to be reviewed by the GSC Submission team.

  » **Green**: Confirmed by Submitter; subject to final review by the GSC Submission team.

  » **Purple**: Additional Edit by Submitter; subject to final review by the GSC Submission team.

- You should use the following three modes ("Edit", Highlight" and "Read-only" when addressing reviewer comments and changes:
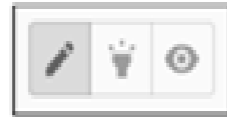
» **Edit**: This is the default mode. When in 'Edit' mode, you are able to alter any field to change its value. Whenever you have made an edit to a field, it will turn blue or purple.

» **Highlight**: When you switch to 'Highlight' mode, clicking on a field should turn it green. You may highlight a field green to indicate that you agree with an edit made by the reviewer, or that you believe the highlighted field already has the correct value. (Clicking on a field multiple times while in highlight mode will toggle between 'unconfirmed' and 'confirmed' state.)

» **Read-only**: All error and warning messages will be hidden.

2. Once your submission has been approved, you will receive a "Submission Approved" email notification with detailed instructions for dropping off or shipping your samples.

7 SAMPLE SUBMISSION BY COURIER

Once the sample submission form is approved, samples should be shipped on dry ice and 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 (*If shipping via Purolator, use V5Z 1B3)

email: amungall@bcgsc.ca

Tel: 604-707-5900 ext 675411

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.

If shipping on dry ice, 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.

8   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

Please email us at GSC_Submissions@bcgsc.ca when you are on your way to the GSC. Once you get to our building, please call our office directly at 604-707-5900 ext 675411. We will meet you at the lobby to take delivery of the samples.

A   APPENDIX A

ENTERING ILLUMINA NEXTERA CODES ON SAMPLE SUBMISSION FORMS

Documentation provided by Illumina and other manufacturers for dual indexed library construction includes instructions to enter i5 bases differently on sample sheets depending on the sequencer to be used. **The GSC's automated processes require the forward workflow i5 base sequence and will be automatically reverse complemented if an applicable sequencing instrument is used.** Please only provide the NovaSeq 6000 v1.0/MiSeq/NovaSeq X bases.

IDT for Illumina UD Indexes

These unique dual (UD) index adapters are arranged in the plate to enforce the recommended pairing strategy. The index adapters are 10 bases long, instead of the typical eight bases. The IDT for Illumina UD Indexes include IDT for Illumina–DNA/RNA UD Indexes, IDT for Illumina–PCR UD Indexes, and IDT for Illumina–Nextera DNA UD Indexes.

Index 1 (i7) Adapters

CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG

Index 2 (i5) Adapters

AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC

▼ Plate A/Set 1 Index Adapters

Index Name	i7 Bases in Adapter	i7 Bases for Sample Sheet	i5 Bases in Adapter	i5 Bases for Sample Sheet NovaSeq 6000 with v1.0 reagent kits, NovaSeq X Series, MiniSeq with Rapid reagents, MiSeq, HiSeq 2000/2500, NextSeq 1000/2000 (Sample Sheet v2)	i5 Bases for Sample Sheet iSeq, NovaSeq 6000 with v1.5 reagent kits, MiniSeq, NextSeq 500/550, HiSeq 3000/4000/X, NextSeq 1000/2000 (Sample Sheet v1)
UDP0001	CGCTCAGTTC	GAACTGAGCG	TCGTGGAGCG	TCGTGGAGCG	CGCTCCACGA
UDP0002	TATCTGACCT	AGGTCAGATA	CTACAAGATA	CTACAAGATA	TATCTTGTAG
UDP0003	ATATGAGACG	CGTCTCATAT	TATAGTAGCT	TATAGTAGCT	AGCTACTATA
UDP0004	CTTATGGAAT	ATTCCATAAG	TGCCTGGTGG	TGCCTGGTGG	CCACCAGGCA
UDP0005	TAATCTCGTC	GACGAGATTA	ACATTATCCT	ACATTATCCT	AGGATAATGT

(B) APPENDIX B

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

Double-click on the blue "0 sequences" button. A message will pop up with directions and the template for download:

Index Sequences for Sample

> (i) Index sequences must be separated by a hyphen (e.g. ACGTACGT-TGCATGCA). Each index sequence must be at least 6 bases long. The combined index sequence length must not exceed 36 bases.

You have entered 0 indices for this pool.

1. Download the template.

2. Enter your indices into the file and save it. Please save in CSV format!

3. Click 'Upload New Indices' and select the file.

4. Click 'Add Indices'

[DOWNLOAD TEMPLATE] [CLEAR INDICES] [UPLOAD NEW INDICES]

[CANCEL] [ADD INDICES]

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

BC CANCER RESEARCH
Provincial Health Services Authority

CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE