Requirements for Illumina Sequencing: Constructed Library Submission

Genome Sciences Centre, BC Cancer

January 2018

Table of Contents

Submission of Constructed Libraries

GSC offers sequencing for labs that wish to perform their own library construction. Oligonucleotides and reagents required for library construction can be purchased from Illumina and other companies. Commercial sources of oligonucleotides have been successfully used for library construction. 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 in Appendix A on page 9 before submitting your samples.

Library Construction Method for DNA

- 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 4). 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.

Library Construction Method for RNA

- For mRNA-seq libraries the standard TruSeq Illumina kit is compatible with our pipeline, as well as our in house chemistries. 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.***

Indexed Libraries

- 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. Index sequences and protocols used for library construction must be confirmed to be compatible with the GSC's sequencing pipeline prior to sample submission.
- 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 by 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 2 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 HiSeq instruments do not focus well when all clusters have the same base in the same position. Please refer to Nextera codes on sample submission on page 5.

Table 1: GSC Compatible Adaptor & Primer Sequences

Type	Name	Sequence
Adapter 5’	Standard GSC	CAAGCAGAAGACGGCATACGAGAT**NNNNNN**CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
Adapter 3’	Standard GSC	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Seq read 1 primer	Standard GSC r1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Seq read 2 (index) primer	GSC Index	GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCG
Seq read 3 primer	Standard GSC r3	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
Adapter 5'	TruSeq	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC**NNNNNN**ATCTCGTATGCCGTCTTCTGCTTG
Adapter 3’	TruSeq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
Seq read 1 primer	TruSeq r1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Seq read 2 (index) primer	TruSeq Index	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Seq read 3 primer	TruSeq r3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Adapter 5'	Direct Seq	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC**NNNNNN**ATCTCGTATGCCGTCTTCTGCTTGC**TG**
Adapter 3’	DirectSeq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT**GAC**
Seq read 1 primer	Direct Seq r1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTC**TG**
Seq read 2 (index) primer	TruSeq Index	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Seq read 3 primer	Direct Seq r3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**GAC**

BC Genome Sciences Centre guidelines to 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 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 (figure 1 below).

		i5 bases for entry on sample sheet	
i5 bases in adapter	i5 index name	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

Figure 1: Correct entry of i5 bases for BCGSC Sample Submission Forms for all sequencing requests.

Library Quantity and 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 quality QC 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 2), shows a single peak of the expected size. Common additional forms include primer dimers (Figure 3), adapter dimers (Figure 3), and broader bands of higher molecular weight (MW) than the expected peak.

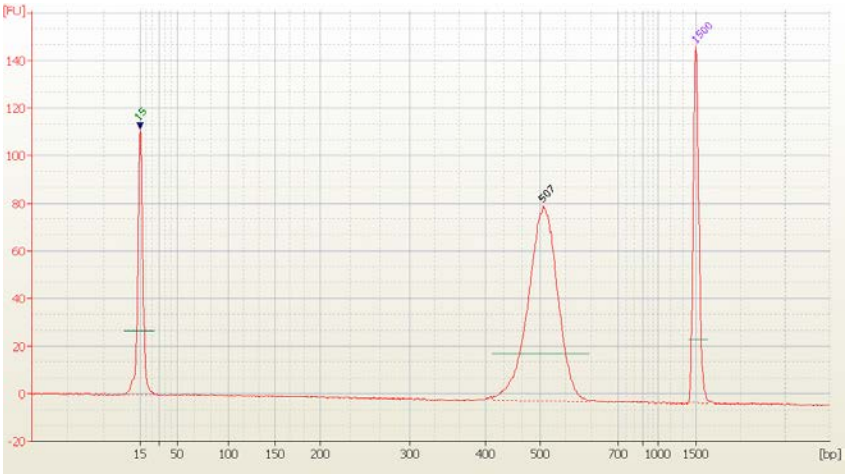

Figure 2: Agilent Bioanalyzer profile of a constructed library

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 more hump 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 600bp.
- The GSC cannot guarantee that constructed libraries will be accepted for sequencing even after samples have been submitted and, we cannot guarantee the quality of sequencing data from these libraries.

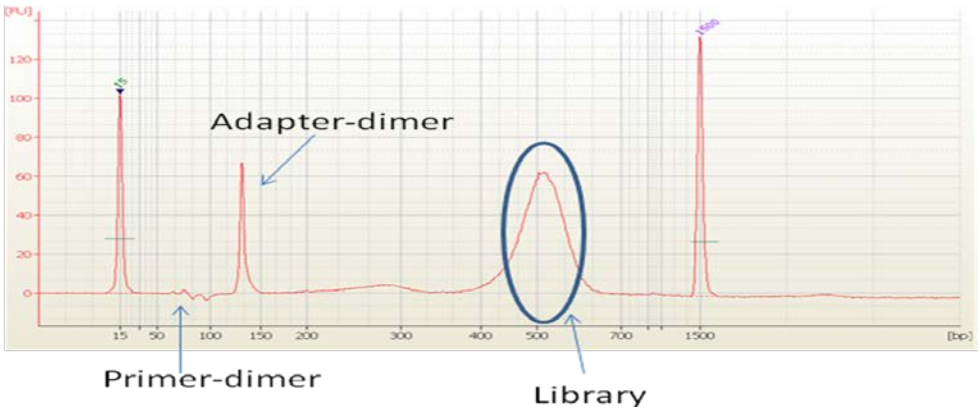

Figure 3: Agilent Bioanalyzer profile showing primer dimers and adapter dimers

Submission requirements: Volumes and concentration

- Please refer to "Appendix A" 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 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. It is not uncommon to require 5 times the minimal amounts given above in order to achieve full sequence output capacity. If your sample concentrations 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.

References & Acknowledgement Policy

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:

- 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:
    - In peer-reviewed publications incorporate the following sentence into the Acknowledgements section of the article: “The authors wish to acknowledge the Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, Canada for [activity]”.
    - Or alternatively, the GSC can be cited in the Materials and Methods section. A suggested sentence for inclusion is: “[Activity] was performed by the Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, Canada”.
- 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.

The Michael Smith 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.

Appendix A

	MiSeq		NextSeq		HiSeq 2500 Rapid		HiSeq2500 High Output		HiSeq X	
Run Size	1 run = 1 lane		1 run = 4 lanes		1 run = 2 lanes		1 run = 1 lane		1 run = 1 lane	
	total library size ~350	total library size ~550	total library size ~350	total library size ~550	total library size ~350	total library size ~550	total library size ~350	total library size ~550	total library size ~350	total library size ~550
Minimum concentration (nM)	1.3	2.4	1.1	1.1	1.4	2.4	2	2.4	1.5	2.5
Maximum concentration (nM)	32	60	50	50	50	90	50	90	80	130
# runs at minimum concentration	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	Minimum Volume (ul)	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