

2019-7-31

Guidelines for sequencing SCC V3 libraries

Upon receiving your libraries:

When your libraries are finished you will be given the average library size and concentration as determined by Agilent Tapestation. You will receive traces of all your libraries please send this report to your sequencing core and/or bioinformatics team.

Each of your libraries will have a separate library index. We suggest pooling your samples by the concentration you are given then having the sequencing core do qPCR quantitation of the pooled sample for optimal cluster generation.

IF your libraries show any peaks below 200bp you will receive notification that your libraries have primer dimer contamination, which can negatively impact sequencing. This is particularly critical if your libraries will be sequenced on a NovaSeq. If this is significant we suggest you pool your samples at 2x or greater the concentration you will use for sequencing and do a final 0.8x SPRI cleanup on the sample (note, there will be sample loss on cleanup).

This cleanup step can be done by the user, the SCC, or the sequencing core could perform this prior to QC of the pooled libraries.

Once you make your pool and are ready to sequence see the DETAILED instructions below as sequencing InDrops libraries:

For V3 sequencing please ask the core to deliver fastq files for ALL reads – including index reads. Or you can ask for the BCL files

Cores that have run a lot of inDrops libraries are: Biopolymers Facility (HMS), Bauer Core (Harvard), DFCI Molecular Biology Core Facility (DFCI).

Sequencing

Libraries have been successfully run on MiSeq, HiSeq, NextSeq, and NovaSeq instruments. While HiSeq usually gives the best data, this comes at a higher cost per base. Libraries are most often run on a NextSeq and the specific parameters for setting up a NextSeq run are given below:

PhiX is not necessary in sequencing.

NextSeq 75 cycle kit run

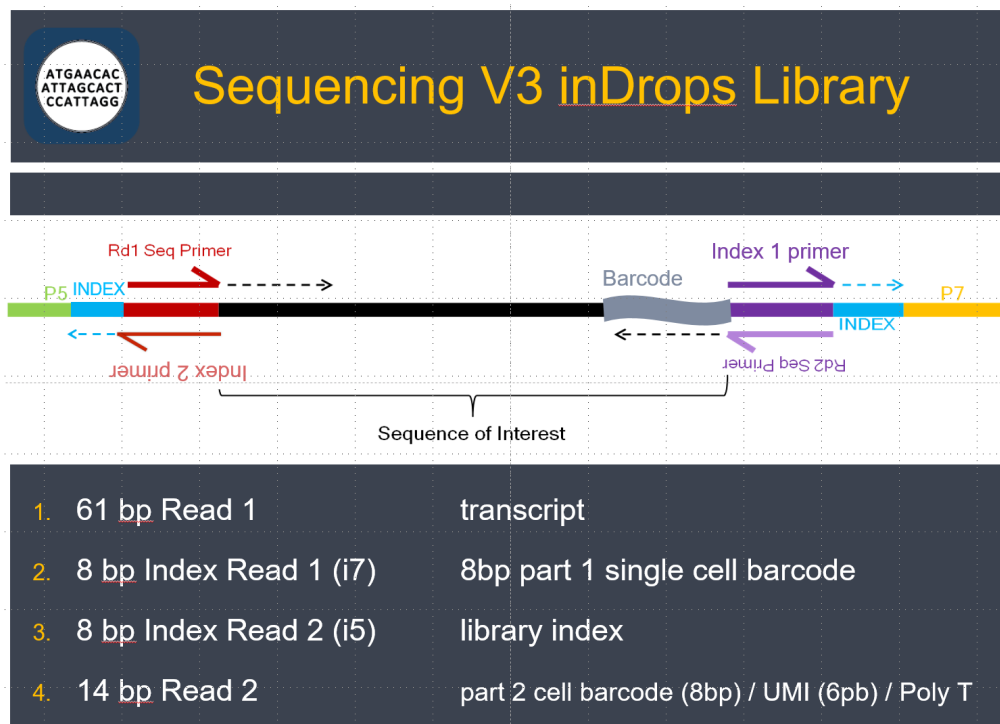
Platform NextSeq
RE-HYB No
Read type Paired-end
Cycles read1 61
Cycles read2 14
Indexing Dual Index
Cycles index 1 8
Cycles index 2 8

NovaSeq 100 cycle kit run

Read type Paired-end
Cycles read1 86
Cycles read2 14
Indexing Dual Index
Cycles index 1 8
Cycles index 2 8

Our current experience with inDrops runs on a NovaSeq indicate that we are having some trouble on the Index reads. Some libraries are only getting 6 bp of the 8bp index read. This is leading to a lower percent of valid reads after filtering for valid barcodes 50-60%. Due to this, the NovaSeq seems most cost effective if sequencing with the full S1 or S2 flow cells.

Specifically, we have found that index 6, 10, and 18 give demultiplexing issues on the NovaSeq.



- The i5 indexes (Index 2) are the library indexes we give to each sample or group of up to 3000 cells.
- The cell barcode is made up of two 8bp sequences that are random combinations of 384 8bp sequences.
 - The first of these is sequenced in the Index 1 (i7) sequencing read in our samples (single-cell barcodes), which are read as Index 1.
 - The second half of the cell barcode is read in the first 8bp of the Read2 followed by the 6bp UMI read.

When setting up your sample sheet for the run you can enter an arbitrary sequence for i5 and i7 (Index 1 and 2). The sample demultiplexing will be done in the first step of your analysis pipeline.

You either need the core to give you the fastq files for all reads – Read 1 and 2 and Index 1 and 2 – as well as reads from the undetermined file. If your core can not give you these files then you need the RAW BCL files for all four sequencing reads. What this means is that all of your data will go to one

“undetermined file” off the instrument because none of the index reads will match what the instrument is expected to see. Pipelines for demultiplexing all of this exist through the Chan Bioinformatics core.

V3 Index 2 (library index) sequences

	Sequence in PCR primer.	
	i5 Bases for Sample Sheet MiSeq, HiSeq	i5 Bases for Sample Sheet MiniSeq, NextSeq, HiSeq
v3 index	2000/2500	3000/4000
idx 1	CTCTCTAT	ATAGAGAG
idx 2	TATCCTCT	AGAGGATA
idx 3	GTAAGGAG	CTCCTTAC
idx 4	ACTGCATA	TATGCAGT
idx 5	AAGGAGTA	TACTCCTT
idx 6	CTAAGCCT	AGGCTTAG
idx 7	CGTCTAAT	ATTAGACG
idx 8	TCTCTCCG	CGGAGAGA
idx 9	TCGACTAG	CTAGTCGA
idx 10	TTCTAGCT	AGCTAGAA
idx 11	CCTAGAGT	ACTCTAGG
idx 12	GCGTAAGA	TCTTACGC
idx 13	CTATTAAG	CTTAATAG
idx 14	AAGGCTAT	ATAGCCTT
idx 15	GAGCCTTA	TAAGGCTC
idx 16	TTATGCGA	TCGCATAA
idx 17	GGAGGTAA	TTACCTCC
idx 18	CATAACTG	CAGTTATG
idx 19	AGTAAAGG	CCTTTACT
idx 20	TAATCGTC	GACGATTA
idx 21	TCCGTCTC	GAGACGGA
idx 22	AGCTTTCT	AGAAAGCT
idx 23	AAGAGCGT	ACGCTCTT
idx 24	AGAATGCG	CGCATTCT
idx 25	TGAGTACC	GGTACTCA
idx 26	GAATATTC	GAATATTC