2018-6-20

Guidelines for sequencing SCC libraries

All libraries made using SCC supplied hydrogel after 1-1-2017 are V3 libraries

Upon receiving your libraries:

When your libraries are delivered you will be given the library concentration as we determined using Qubit as well as the library index used for each sample. You will also be given the library size as determined by Bioanalyzer by and a Bioanalyzer trace showing a representative example of your libraries. You will NOT receive traces of all your libraries.

Each of your libraries will also 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 likely have primer dimer contamination, which can negatively impact sequencing. 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 doing this).
>
> This cleanup step can be done by the user, the SCC can help, or it is likely the sequencing core could perform this as part of QC.

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

V3 Sequencing

The V3 design aim was to get rid of the need for custom sequencing primers, include more diversity in certain parts of the barcode for higher quality sequencing, and save cycles by not sequencing the constant regions of the cell barcode.

V3 libraries have been successfully run on MiSeq, HiSeq and NextSeq instruments. While HiSeq usually gives the best data, this comes at a higher cost per base. These libraries are most often run on NextSeqs and the specific parameters for setting up a NextSeq run are given below:

PhiX is not necessary in sequencing.

For V3 sequencing on a NextSeq 75 cycle kit the reads are as follows:

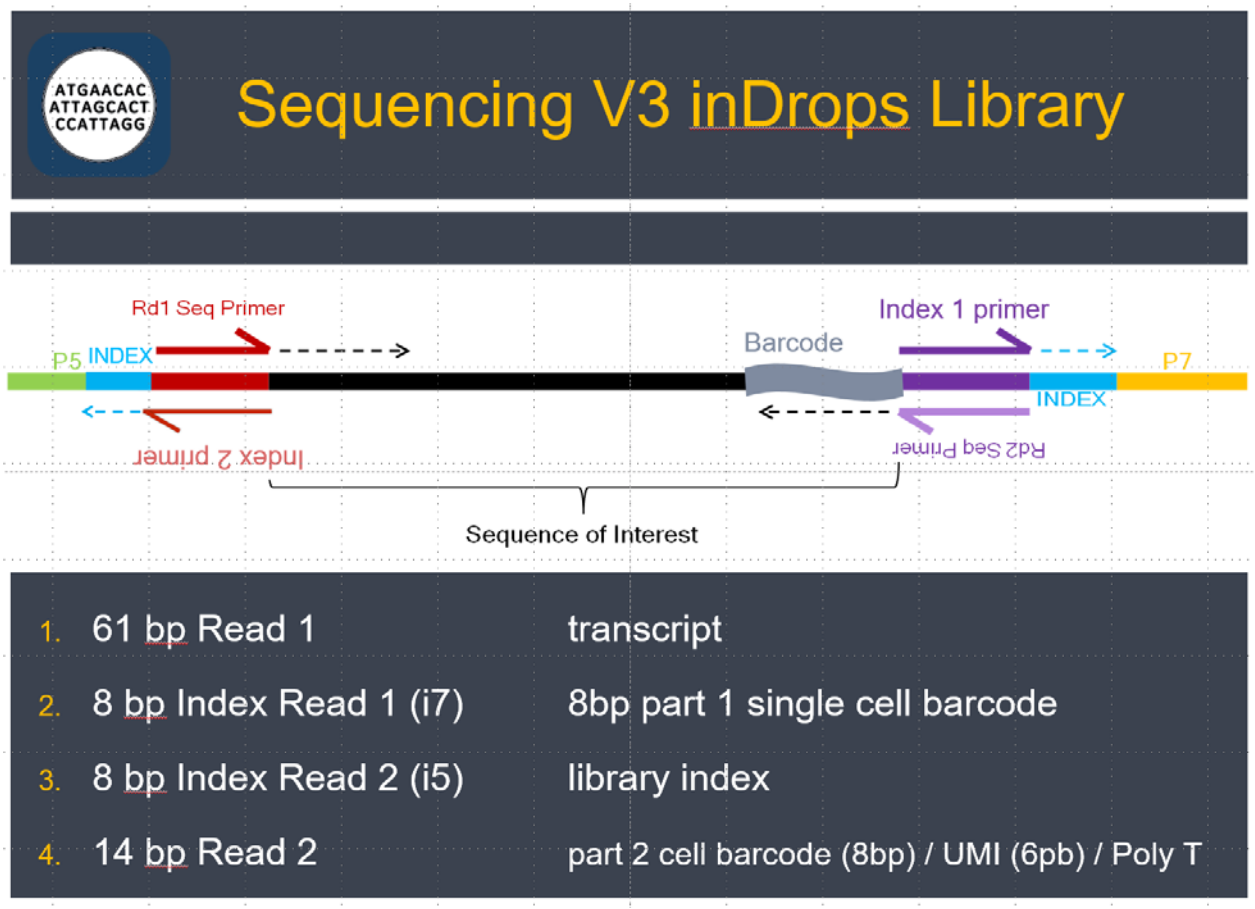
Read 1 – reads into the mRNA
Read 2 – barcode/UMI to polyA tail
Index 1 – part of barcode
Index 2 – library index

NextSeq 75 cycle kit run of V3 libraries

please give your core the following set up parameters:

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

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 8pb sequences that are random combinations of 384 8bp sequences. The first of these is sequenced in the 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 should just enter an arbitrary sequence for i5 and i7 (Index 1 and 2). Don't worry about demultiplexing the sample and generating fastq files. This demultiplexing will be done afterwards bioinformatically.

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.

These are the sequences of the sample indices used on V3 libraries:

V3 Index 1 sequences

idx1	CTCTCTAT
idx2	TATCCTCT
idx 3	GTAAGGAG
idx 4	ACTGCATA
idx 5	AAGGAGTA
idx 6	CTAAGCCT
idx 7	CGTCTAAT
idx 8	TCTCTCCG
idx 9	TCGACTAG
idx 10	TTCTAGCT
idx 11	CCTAGAGT
idx 12	GCGTAAGA
idx 13	CTATTAAG
idx 14	AAGGCTAT
idx 15	GAGCCTTA
idx 16	TTATGCGA
idx 17	GGAGGTAA
idx 18	CATAACTG
idx 19	AGTAAAGG
idx 20	TAATCGTC
idx 21	TCCGTCTC
idx 22	AGCTTTCT
idx 23	AAGAGCGT
idx 24	AGAATGCG

InDrop Designs: V2 vs V3

There are two InDrop designs used by the core. As of September 2016 libraries will be made with the new V3 design. Please be sure you know which version of the hydrogel was used on your sample as this makes VERY IMPORTANT DIFFERENCES IN SEQUENCING.

Libraries are prepared with custom primers that includes a 6-base (V2) or 8-base (V3) long library index. Read 1 adapter is on the 5'/cDNA end (reads gene), and Read 2 adapter is on the 3' barcoded end (reads BC/UMI).

General requirements for all Illumina sequencers:

- Give your core facility the average library size (which will be given to you with your final libraries) and make sure the core facility performs library quantification qPCR before run. It is fine to pool your libraries based on the concentrations given to you from the SCC with your library, but the final pool should be qPCR quantified before running on the sequencer. This will result in optimal cluster generation.

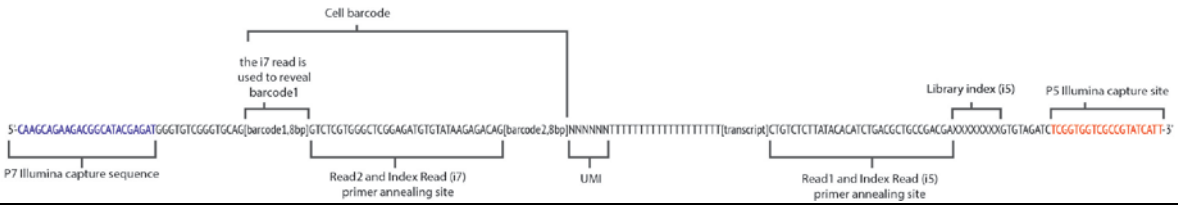
- **V2 ONLY -** *Provide the sequencing core with custom primers (for Read 1, Index Read, and Read 2), or check to see if they already have them in stock. Check with the facility to see how much they need (the DFCI core requests at least 6 µL of 100 µM primer per run). See next page for primer sequences and index sequences.*

  Do NOT mix these custom primers into the standard Illumina primer mix. This will cause sequencing Failure for InDrop libraries. V2 custom primers must be kept separate.

- **V3 libraries** require that fastq files be delivered for ALL reads – including index reads and undetermined reads. If core facility cannot give you the fastq files for all 4 reads then ask for the bcl files.

v2

v3



V2 Sequencing

Our suggested method for sequencing is the Illumina NextSeq. Users typically pool 10,000 to 30,000 cells on one NextSeq run.

Note that the Dana Farber Molecular Biology Core is well versed with InDrops libraries and has all custom primers required for V2 sequencing. (https://mbcf.dfci.harvard.edu/?page_id=4)

> Do NOT mix these custom primers into the standard Illumina primer mix. This will cause sequencing Failure for InDrop libraries. V2 custom primers must be kept separate.

For NextSeq run:

- Use high-yield 75 cycle kit (which comes with 92 cycles)
- 36 cycles on read 1
- 6 cycles on index read
- Remaining 50 cycles on read 2

For diagnostic MiSeq run:

- Use v3 150 cycle kit (all kits comes with +15 cycles for index reads)
- Up to 103 cycles on read 1
- 6 cycles on index read
- 55 cycles on read 2

For HiSeq 2000 run:

- Samples must be multiplexed for this machine, so you are restricted to standard read lengths (be sure to include an index read)
- Use v3 reagents (v4 reagents may be incompatible with our library prep)
- Minimum required reads are Index, 37 cycles R1 and 50 cycles R2. Read 1 can be longer if desired as this end reads into the gene

For HiSeq 2500 run:

- Same as MiSeq.

Sequencing primers (should be HPLC-purified):

Read_1_seq: GGCATTCCTGCTGAACCGCTCTTCCGATCT
Index_i7_seq: AGATCGGAAGAGCGTCGTGTAGGGAAAGAG
Read_2_seq: CTCTTTCCCTACACGACGCTCTTCCGATCT

> Do NOT mix these custom primers into the standard Illumina primer mix. This will cause sequencing Failure for InDrop libraries. V2 custom primers must be kept separate.

These are the sequences of the sample indices used on V2 libraries:

V2 Index sequences
idx1: ATCACG
idx2: CGATGT
idx3: TTAGGC

idx4: TGACCA
idx5: ACAGTG
idx6: GCCAAT
idx7: CAGATC
idx8: ACTTGA
idx9: GATCAG
idx10: TAGCTT
idx11: GGCTAC
idx12: CTTGTA
idx13: AGTCAA
idx14: AGTTCC
idx15: ATGTCA
idx16: CCGTCC
idx17: GTAGAG
idx18: GTCCGC
idx19: GTGAAA
idx20: GTGGCC
idx21: GTTTCG
idx22: CGTACG
idx23: GAGTGG
idx24: GGTAGC