User Guide

TELL-Seq™ and Standard Illumina Library Run, Mixed Sample Demultiplexing User Guide

Universal Sequencing Technology's TELL-Seq™ libraries and standard Illumina libraries can be pooled in the same sequencing run on an Illumina sequencing system. We recommend loading TELL-Seq libraries and standard Illumina libraries on separate lanes in the same flow cell. This document describes procedures on how to set up the sample sheet for TELL-Seq and standard Illumina (Flex) libraries on separate lanes in the same flow cell for a NovaSeq™ run, and how to demultiplex the different types of libraries after the run.

1. **Sample sheet for preparing a 2-lane NovaSeq sequencing run**

To ensure compatibility with downstream demultiplexing software, each lane should have only one type of library; either a TELL-Seq library or a standard Illumina library. When both types of libraries are sequenced in the same flow cell, the number of flow cycles for both index1 and index2 for all samples in the flow cell must be the same. TELL-Seq libraries utilize an 18-base molecular barcode as the I7 index (i.e., index 1), as well as an 8-base sample barcode as the I5 index (i.e., index 2). Any sequencing runs containing TELL-Seq libraries must be sequenced <u>at least 18 and 8 bases for index 1 and 2</u>, respectively, to properly demultiplex and use TELL-Seq's linked read information. Illumina's Flex libraries used here have a 10-base barcode for both indices.

When TELL-Seq libraries and Flex libraries are sequenced in a run together, set 18 cycles for index 1 and 10 cycles for index 2. When an index length is shorter than the sequencing cycle number, add additional Ns for the missing bases for the index sequences on the sample sheet. TELL-Seq libraries should designate 18 Ns as the index 1 sequence on the sample sheet. The sample sheet file *SampleSheet.csv* shown below should be set up for a pooled library run (e.g., TELL-Seq in lane1 and Flex in lane2).

Lane	Sample_ID	Sample_Name	Sample_Plate	Sample_Well	Index_Plate	Index_Plate_Well	I7_Index_ID	index	I5_Index_ID	index2	Sample_Project	Description
1	TellSeq-01	TellSeq-01						NNNNNNNNNNNNNNNNNN		TGTTCTAGNN		
1	TellSeq-02	TellSeq-02						NNNNNNNNNNNNNNNNNN		TCGATTGANN		
1	TellSeq-03	TellSeq-03						NNNNNNNNNNNNNNNNNN		ACTTAGCANN		
1	TellSeq-04	TellSeq-04						NNNNNNNNNNNNNNNNNN		AAGGTTCANN		
2	Flex-01	Flex-01						ACTCGGCAATNNNNNNNN		TTCAGTTGTC		
2	Flex-01	Flex-01						GTCTAATGGCNNNNNNNN		CCTGACCACT		
2	Flex-03	Flex-03						CCATCTCGCCNNNNNNNN		AACCATAGAA		
2	Flex-04	Flex-04						CTGCGAGCCANNNNNNNN		TGGCCGGATT		

Note: TELL-Seq custom sequencing primers must be spiked into standard sequencing primer wells on the sequencing cartridge for a run with mixed library types.

2. **Sample sheet after sequencing run**

To proceed with sample demultiplexing after the sequencing run, the *SampleSheet.csv* file must be modified to remove Ns **trailing** the known sample barcode sequences in the index 1 and index 2. All samples from the same lane must designate the same length for index 1 and index 2, respectively.

Lane	Sample_ID	Sample_Name	Sample_Plate	Sample_Well	Index_Plate	Index_Plate_Well	I7_Index_ID	index	I5_Index_ID	index2	Sample_Project	Description
1	TellSeq-01	TellSeq-01						NNNNNNNNNNNNNNNNNN		TGTTCTAG		
1	TellSeq-02	TellSeq-02						NNNNNNNNNNNNNNNNNN		TCGATTGA		
1	TellSeq-03	TellSeq-03						NNNNNNNNNNNNNNNNNN		ACTTAGCA		
1	TellSeq-04	TellSeq-04						NNNNNNNNNNNNNNNNNN		AAGGTTCA		
2	Flex-01	Flex-01						ACTCGGCAAT		TTCAGTTGTC		
2	Flex-01	Flex-01						GTCTAATGGC		CCTGACCACT		
2	Flex-03	Flex-03						CCATCTCGCC		AACCATAGAA		
2	Flex-04	Flex-04						CTGCGAGCCA		TGGCCGGATT		

3. Demultiplexing of Flex library samples

To produce a standard set of FASTQ files for each Flex library, Demultiplexing can be performed using the normal `bcl2fastq` workflow as in the following:

bcl2fastq -R [--runfolder-dir] -o [--output-dir] --create-fastq-for-index-reads --sample-sheet /path/to/SampleSheet.csv

4. Demultiplexing of TELL-Seq library samples

The above command will result in a set of FASTQ files that are not sample demultiplexed (i.e., Undetermined) for the TELL-Seq lane.

```
Undetermined_S0_L001_I1_001.fastq.gz

Undetermined_S0_L001_I2_001.fastq.gz

Undetermined_S0_L001_R1_001.fastq.gz

Undetermined_S0_L001_R2_001.fastq.gz
```

- Execute following step to demultiplex TELL-Seq samples:

demult Undetermined_S0_L001_I2_001.fastq.gz Undetermined_S0_L001_R1_001.fastq.gz Undetermined_S0_L001_R2_001.fastq.gz sample_index_list

- Sample Index List

An example `sample_index_list` is shown below,

```
TGAACCTT       T501
TGCTAAGT       T502
CGCTCCGC       T516
```

Each row specifies a sample. The 1st column specifies the actual index sequence used for demultiplexing in the FASTQ files. The 2nd column specifies a custom name the user can give for the sample. The columns are tab delimited.

5. Download Demultiplex Utility `demult` and Examples

Utility package, including example sample sheets, can be downloaded from github repo, https://github.com/universalsequencing/demultiplex.

Appendix

Comparable sequencing quality between runs with both library types and runs with a single library type

TELL-Seq libraries require custom sequencing primers for Illumina sequencing systems. When TELL-Seq libraries are pooled with other Illumina libraries for sequencing in the same run, these custom sequencing primers can be spiked into standard sequencing primer wells so that both types of libraries can be sequenced together. To confirm that the sequencing quality from having these two libraries in the same run is comparable to a run with single library type, we constructed four TELL-Seq human libraries with different sample index (8-base index2) and four Illumina Nextera™ DNA Flex human libraries with different dual indexes (10-base index1 and 10-base index2), and sequenced them using a S1 flow cell on a NovaSeq™ system in the following run types:

1. Flex libraries alone in the S1 flow cell with 150-10-10-150 sequencing cycle
2. TELL-Seq libraries alone in the S1 flow cell using TELL-Seq custom sequencing primers loaded into custom primer wells with 150-18-8-150 sequencing cycle
3. Flex libraries and TELL-Seq libraries on separate lanes in the same S1 flow cell using TELL-Seq custom sequencing primers spiked into standard primer wells with 150-18-10-150 sequencing cycle

The QC metrics for the Flex alone versus Flex with TELL-Seq and TELL-Seq alone versus TELL-Seq with Flex are comparable (see Table 1).

Table 1. Sequencing run QC metrics among three run types

NovaSeq Sequencing run QC metrics		Run Type 1 Illumina Flex libraries alone in a flow cell	Run Type 3 Illumina Flex libraries and TELL-Seq libraries on separate lanes in the same flow cell		Run Type 2 TELL-Seq Libraries alone in a flow cell
			Flex library lane	TELL-Seq library lane	
Cluster PF (%)		82.6	82.5	81.8	89.0
%>=Q30	**Read1**	91.1	90.9	92.7	94.2
	Read2	95.1	95.6	91.6	93.6
	Read3	94.9	94.1	94.0	94.6
	Read4	89.0	88.0	85.8	87.4

Furthermore, Type 1 and Type 2 sequencing runs were demultiplexed following standard Illumina library and TELL-Seq library demultiplexing pipeline, respectively, while Type 3 sequencing runs, in which two different types of libraries were sequenced in the same flow cell but on different lanes, were demultiplexed based on the instruction provided in this guide. Demultiplexed libraries were mapped against GRCh38 reference and mapping rates were summarized in Table 2. There is no significant difference for the same type of libraries between run types.

Table 2. Sequencing run QC metrics among three run types

Mapping Rate	Run Type 1 Illumina Flex libraries alone in a flow cell	Run Type 3 Illumina Flex libraries and TELL-Seq libraries on separate lanes in the same flow cell		Run Type 2 TELL-Seq Libraries alone in a flow cell
		Flex library lane	TELL-Seq library lane	
Library #1	99.63	99.57	98.08%	98.34%
Library #2	99.65	99.60	97.86%	98.17%
Library #3	99.76	99.74	98.20%	98.45%
Library #4	99.80	99.77	98.10%	98.39%

In summary, TELL-Seq libraries and custom sequencing primers are compatible with standard Illumina libraries and sequencing primers and will not adversely affect the sequencing performance when these two types of libraries are sequenced in the same sequencing run.

UNIVERSAL
SEQUENCING
innovation for all

Revision History

Document #	Version	DCR Reference and Comment
100022-USG	1.0	DCR-210082 Initial Release