Sequencing Lanes with Single Index Tag Samples

Introduction

Illumina's TruSeq™ sample preparation kits enable multiplexing, which lowers the cost of sequencing per sample and increases sample throughput. Multiplexing incorporates an indexing adapter tag, allowing pooling of multiple samples within a single flow cell lane. Illumina sequencing control software processes these tags in an index read.

In flow cell lanes containing a single index tag, all of the clusters in that lane appear in a single channel rather than being distributed between four different channels. This can occasionally cause failures in analysis of the index tag in that lane. This Technical Note provides guidance for sequencing lanes with single samples using CASAVA 1.7.

All Lanes Containing a Single Sample

Since there is only one tag, the index sequence (and thus the sample identity) is known, making sequencing the tag unnecessary. In a run where every lane has only one sample, a multiplexing recipe would not be used and the index tag would not be sequenced.

Runs Containing Multiple and Single Samples

In a run containing both multiple and single samples, the lanes with single samples may have a high proportion of reads that do not align to the index tag and are placed in the unknown folder after demultiplexing. The data from these reads are not affected, but they might not be used for alignments and downstream analysis.

Since sequencing the index tags from lanes with only one sample is not necessary, the lanes are treated as non-multiplexed and omitted from the sample sheet used for demultiplexing. When the run is demultiplexed during CASAVA analysis, these lanes that do not appear in the sample sheet will be transferred to the unknown folder. Data that are transferred to the unknown folder can then be aligned in the `config.template.txt` file used by CASAVA 1.7.

Example of a Mixed Run

In this example, the sequencing run contains the following samples:

Lane	Sample	Index
1	One human library	2
2	Two human libraries	4 and 7
3	Two human libraries	6 and 12
4	PhiX	No index
5	One human and one *Arabidopsis* library	5 and 6
6	Two *Arabidopsis* libraries	1 and 2
7	Three *Arabidopsis* libraries	7, 8, and 9
8	One *Arabidopsis* library	11

Here, a PhiX control was run by itself in lane 4 for matrix and basecalling, which is necessary because the *Arabidopsis* genome has a high AT percentage. Lanes 1, 4, and 8 all contain single samples and are omitted from the sample sheet (Figure 1). Note that lanes 1 and 8 have an index tag, and lane 4 does not, but they will all end up in the unknown folder and the `config.template.txt` will call these using SAMPLE unknown and lane-specific parameters in the `config.template.txt` file.

Figure 1: Sample Sheet for Example Sequencing Run

	A	B	C	D	E	F	G	H	I	J
1	FCID	Lane	SampleID	SampleRe	Index	Descriptic	Control	Recipe	Operator	
2	FC612PV	2	hu4	human	TGACCA	desc	N	R1	a.scientist	
3	FC612PV	2	hu7	human	CAGATC	desc	N	R1	a.scientist	
4	FC612PV	3	hu6	human	GCCAAT	desc	N	R1	a.scientist	
5	FC612PV	3	hu12	human	CTTGTA	desc	N	R1	a.scientist	
6	FC612PV	5	hu5	human	ACAGTG	desc	N	R1	a.scientist	
7	FC612PV	5	ara6	arabid	GCCAAT	desc	N	R1	a.scientist	
8	FC612PV	6	ara1	arabid	ATCACG	desc	N	R1	a.scientist	
9	FC612PV	6	ara2	arabid	CGATGT	desc	N	R1	a.scientist	
10	FC612PV	7	ara7	arabid	CAGATC	desc	N	R1	a.scientist	
11	FC612PV	7	ara8	arabid	ACTTGA	desc	N	R1	a.scientist	
12	FC612PV	7	ara9	arabid	GATCAG	desc	N	R1	a.scientist	
13										

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

This sample sheet is saved as a comma-separated value (CSV) file. The `config.template.txt` for this run would appear:

```
# Global parameters for all lanes and samples

ANALYSIS eland_pair

USE_BASES Y*n,y*n

ELAND_SET_SIZE 40


# Alignment targets for the samples described
in the sample sheet:

REFERENCE human ELAND_GENOME /data/genomes/
human

REFERENCE arabid ELAND_GENOME /data/genomes/
arabidopsis


# Alignment targets for lanes that were not
in the sample sheet:

SAMPLE unknown 1:ELAND_GENOME /data/genomes/
human

SAMPLE unknown 4:ELAND_GENOME /data/genomes/
phi

SAMPLE unknown 8:ELAND_GENOME /data/genomes/
arabidopsis


# Do not align reads from the indexed lanes
that don't match an expected index

SAMPLE unknown 23567:ANALYSIS none
```

The sample sheet and the `config.template.txt` is used with `demultiplex.pl` from CASAVA 1.7. For details on this process, refer to the CASAVA 1.7 User Guide.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, HiScan, TruSeq, and Eco are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub. No. 770-2010-019 Current as of 8 December 2010

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA