

Illumina 1.8 FASTQ Format

Institute for Integrative Genome Biology
UC Riverside

January 23, 2012

1 CASAVA 1.8 FASTQ

Each entry in a FASTQ file consists of four lines:

- Sequence identifier
- Sequence
- Quality score identifier line (consisting of a +)
- Quality score

An example of a valid entry is as follows:

@HWI-ST279:211:COBFTACXX:3:1101:3469:2181 1:N:0:ACTTGA
GAACTATGCCTGATCAGGTTGAAGTCAGGGGAAACCCTGATGGAGGACCGA
+
CCCCFFFFHHHHHJJJJIIJJJHJJJJJJJJJJJJJJJJJJJJJJ

Each sequence identifier, the line that precedes the sequence and describes it, needs to be in the following format:

@<instrument>:<run number>:<flowcell ID>:<lane>:<tile>:<x-pos>:<y-pos> <read>:<is filtered>:<control number>:<index sequence>

The elements are described below.

	Element	Requirments	Description
1	@	@	Each sequence identifier line starts with @
2	<instrument>	Characters	Instrument ID
3	<run number>	Numerical	Run number on instrument
4	<flowcell ID>	Character	
5	<lane>	Numerical	Lane number
6	<tile>	Numerical	Tile number
7	<x_pos>	Nnumerical	X coordinate of cluster
8	<y_pos>	Numerical	Y coordinate of cluster
9	<read>	Numerical	Read number. 1 can be single read or read 2 of paired-end
10	<is filtered>	Y or N	Y is the read is filtered, N otherwise
11	<control>	Numerical	0 when none of contol bits are on, otherwise it is an even number
12	<index sequence>	ACTG	Index sequence

Table 1: FASTQ Elements