

bcl2fastq2 Software Release Notes

bcl2fastq2 2.17.1

June 2015

FOR RESEARCH USE ONLY

© 2015 Illumina, Inc. All rights reserved.

Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPro, CytoChip, DesignStudio, Epicentre, GAIIX, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, ForenSeq, MiSeq, MiSeqDx, MiSeq FGx, NeoPrep, Nextera, NextBio, NextSeq, Powered by Illumina, SeqMonitor, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

Introduction

These Release Notes detail the key changes to the bcl2fastq2 converter since the release of bcl2fastq2 v2.16.0. This new version updates and replaces the version listed below.

Software Application	Prior Version	New Version
bcl2fastq2	2.16.0	2.17.1

This is a required software update for customers currently using bcl2fastq2 v2.16.0, as this release includes important features and bug fixes outlined below. The changes outlined here are changes to bcl2fastq2 since the release of 2.16.0.

Version 2.17.1 is used to convert bcl files to FASTQ, and is compatible with MiSeq, NextSeq 500 and all HiSeq (2000, 2500, 3000, 4000 and HiSeqX) systems running RTA version 1.18.54 and above. Installers and the software User Guide, which includes installation instructions, are available for download from Illumina.com.

For FASTQ conversion of bcl files generated on Illumina GAIIX and HiScan-SQ sequencing instruments, or any other type of sequencer running earlier versions of RTA, use bcl2fastq v1.8.4, available on illumina.com.

I. bcl2fastq2 2.17.1

NEW FEATURES:

- Added support for MiSeq, HiSeq 2000 and HiSeq 2500 for systems running RTA 1.18.54 or above.
- Underscores are now allowed in the SampleId, SampleName, and Project columns in the [Data] section of the sample sheet. These underscores will be propagated into the FASTQ file names and the project folder names.
- Only alpha-numeric, '-' or '_' characters are allowed in the [Data] section of the sample sheet. An exception is thrown if other characters are encountered.
- If multiple samples have the same sample names, the sample number found in the FASTQ file name is appended to the sample name in the HTML report.
- Several new sample sheet settings have been added. See the *User Guide* for details.
- In the statistics file AdapterTrimming.txt, a table with the fraction of reads containing a given number of untrimmed bases has been added.
- A new statistics file has been added called: FastqSummaryF1L#.txt, where # is the lane number. This file contains the number of raw and passed filter reads for each sample number and tile.

- A new statistics file has been added called: DemuxSummaryF1L#.txt, where # is the lane number. This file contains the percentage of each tile that each sample makes up. It also contains the 1000 most common unknown barcode sequences.
- Converting by single index in a dual index run is supported.
- Unique Molecular Identifiers (UMIs) are now supported through the sample sheet. See the User Guide for more details.

OPTIMIZATIONS:

- Improved memory allocation scheme to decrease execution time.
- IndexMetricsOut.bin has been updated to use the number of clusters that passed filter rather than the total number of clusters.
- Minor updates to the HTML report have been made to improve readability.
- Trailing blank spaces in sample sheet settings are now ignored. The uniqueness of the SampleId values in the sample sheet is now enforced.

DEFECT REPAIRS:

- When generating the reverse complement of the bases through the command line parameter: --write-fastq-reverse-complement, the quality scores are now also reversed as well.
- FASTQ files of 0 size are no longer generated for empty samples.
- If the Interop folder is read-only, an exception is no longer thrown.
- The command line option --no-bgzf-compression no longer produces corrupt FASTQ files.
- The regular expressions used in --use-bases-mask now treat the '*' as matching 0 or more cycles.
- A whole read can now be masked with --use-bases-mask.
- The DemultiplexingStats.xml now includes the expected statistics for two mismatches to the specified barcodes if the 2 mismatch option is specified in the command line via the --barcode-mismatches option.
- Barcodes specified in the sample sheet can now be shorter than the number of bases indicated in RunInfo.xml. Index bases after this length will be ignored.

KNOWN ISSUES:

- When 'N' is contained in the sample sheet barcode, bcl2fastq2 matches only to 'N' in the sequence.
- When control files are missing, bcl2fastq2 does not give any warnings or errors with default settings.
- If "--no-bgzf-compression" is used, it is reflected in the command line shown at the start of the log.
- When "default"/"unknown"/"Undetermined"/"all" is used as the sample name or project name, statistics in html report will be generated incorrectly.

- When using the --use-bases-mask option to generate shorter reads, the --minimum-trimmed-read-length option must be manually set to be less than or equal to the shortest resulting unmasked read length. Otherwise, bcl2fastq 2.17.0 will N-pad the reads to the default --minimum-trimmed-read-length of 35 base pairs.
- If a user wishes to mask an entire read following adapter trimming, where the remaining non-adapter basecalls numbers less than the --mask-short-adapter-reads value, then the --minimum-trimmed-read-length value must be set to the length of the read. Also the sequence to be masked must be specified in the SampleSheet as Adapter rather than MaskAdapter.
- The "--mask-short-adapter-reads" option only works when a read is trimmed to below the length specified by the "--minimum-trim-read-length" option (default of 35). No special trimming strategy is implemented to support Nextera mate pair adapters.
- If adapter trimming is performed allowing indels, one extra mismatch will be allowed when the matched adapter sequence is longer than 10bp and falls at the end of the read.
-