bcl2fastq2 Software

Release Notes

bcl2fastq2 v2.16.0

May 2015

Revision History

Revision	Date	Description of Change
A	3/1/2015	Initial release of version 2.16.0
B	5/5/2015	Additional known issue reported; added language to clarify which instruments are supported

Introduction

These Release Notes detail the key changes to the bcl2fastq2 converter since the release of bcl2fastq2 v2.15.0. This new version updates and replaces the version listed below.

Software Application	Prior Version	New Version
bcl2fastq2	2.15.0	2.16.0

This is a required software update for customers currently using bcl2fastq2 v2.15.0, as this release includes important features and bug fixes outlined below. The changes outlined here are changes to bcl2fastq2 since the release of 2.15.0.

Version 2.16.0 is used to convert bcl files generated on NextSeq 500 and all HiSeq (3000, 4000 and HiSeqX) systems, as long as they are running RTA version 2.0 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 sequencing running earlier versions of RTA, use bcl2fastq v1.8.4, available on illumina.com.

I.   bcl2fastq2 v2.16.0

NEW FEATURES:

- The default value for --minimum-trimmed-read-length has been updated to 35 from 32 and for --mask-short-adapter-reads has been updated to 22 from 10. See the User Guide for more information on these options.

- A table with a list of the top unknown barcodes has been added to the reported statistics. See the User Guide for more information on how to access this table.

- Adapter trimming logic has been improved to handle indels in the adapter sequence, with the insertion or deletion of a single base now scored as a mismatch. This leads to an overall improvement in the accuracy of adapter trimming and masking.

- Added a new --sample-sheet option to specify the sample sheet location. The default sample sheet location remains unchanged (top level of the run folder).

- Improved thread allocation logic and updated conversion logic to reduce the overall execution time. Users may see speed improvements of up to 4x, depending on compute hardware.

- Improved file compression options:

    - FASTQ files are now compressed in the BGZF format by default. This allows parallel decompression by downstream applications. This can be disabled with the option: --no-bgzf-compression which output the same compression format implemented in the previous version 2.15.

    - Added a new option for the FASTQ compression level: --fastq-compression-level, with 9 being the most compressed level, and 1 the least. This is set to a default value of 4, whereas bcl2fastq2 v2.15.0 used a default value of 6. Testing indicates that level 4 is substantially faster than level 6, but the files produced are less than 10% different in size.

- Improved options for fastq output files:

  - Added a new option to output FASTQ files by Sample Name, without lane splitting: The option is --no-lane-splitting. See the User Guide for more information on how to use this option. When this option is used, the resulting FASTQ files will be named by a combination of Sample Name and sample number.

  - Changes to files names relative to v2.15: A sample number is assigned to each Sample Name and Sample ID combination supplied in the sample sheet, according to the order in which they are listed. The sample number can be used to track information included in the reports files.

  - If the Sample_Name is absent in the sample sheet, the FASTQ files will be named according to the Sample_Id. With the previous version (2.15), if the Sample Name was not provided, the FASTQ file was named Undetermined.

- Improved error handling and robustness for corrupted file situations.

  - Changed the name of the option --ignore-missing-locs to --ignore-missing-positions. The usage and function of the option remain unchanged.

  - Improved robustness of --ignore-missing-positions, --ignore-missing-filters, --ignore-missing-controls and --ignore-missing-bcls to allow successful execution in a wider range of corrupt files scenarios.

DEFECT REPAIRS:

- A fix was made in the error handling of barcode collision situations. Barcodes in the sample sheet are now validated against each other to detect possible collisions. If 2 barcodes can map to the same sample when considering allowed mismatches, bcl2fastq2 exits with an error. Previously, no error was thrown.

- The --tiles option now accepts a comma separated list of tiles, as originally intended.

- If the option --minimum-trimmed-read-length is greater than the smallest read length, it is automatically adjusted to the size of the smallest read length.

- The option --use-bases-mask now accepts masking all of the bases of a read.

- Bcl2fastq2 no longer generates empty FASTQ files for samples that have no data in a given lane.

- Multiple adapter sequences entered in the sample sheet can now be separated by a '+', to offer compatibility with IEM.

- If non-ACTGN characters are included the in barcodes in the sample sheet, bcl2fastq now exits with an error.

- When missing positions files and --ignore-missing-positions is set, the coordinates written in the FASTQ files will now be unique. Previously, in this situation, the coordinates written into the FASTQ files was 0:0 for all clusters.

- Fixed a bug in reports generation that would only occur when installing the rpm in a non-default location.

- Information in the reports files is now broken out by Lane and Sample ID, which is a unique identifier for each sample within each lane. Previously, this information was provided per Sample Name, which was not required to be unique.

- bcl2fastq2 now trims off any whitespace at the end of each barcode in the Sample Sheet.

KNOWN ISSUES:

- The DemultiplexingStats.xml will not include the expected statistics for two mismatches to the specified barcodes, even if the 2 mismatch option is specified in the command line via the --barcode-mismatches option. However, if 2 mismatches are specified, this setting will be properly applied during demultiplexing.

- When using the --use-bases-mask option, the --minimum-trimmed-read-length option must be manually set to be less than or equal to the minimum read length. Otherwise, bcl2fastq 2.16.0 will N-pad the reads to the default –minimum-trimmed-read-length of 35 base pairs.

- In the sample sheet, only adapter-related settings (e.g., Adapter, MaskAdapter) in the [Setting] section of the sample sheet will be recognized and applied by bcl2fastq2 v2.16. Other settings will be ignored.

- 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.

- FASTQ file outputs are named according to the Sample Name provided in the sample sheet (or, if no Sample Name is provided, FASTQ file names use the Sample ID). However, the information in the reports files is presented by Sample ID. If the Sample Names provided are not unique, you will need to refer to the Sample Sheet to match the information included in the reports files to the correct corresponding FASTQ file.

- If a user wishes to apply adapter sequence masking, specific settings may be required in some situations: If a user wishes to mask entire reads in which the number of non-adapter basepairs is below the --mask-short-adapter read setting, the --minimum-trimmed-read-length must be set to the length of the read, and the adapter sequence to be masked must be specified in the Sample Sheet as Adapter rather than MaskAdapter.

- There is a defect in the --write-fastq-reverse-complement option: using this option will correctly output the sequence in reverse compliment orientation, but the quality score string is not reversed as it should be. Users will need to perform post processing of the fastq files output when using this option to reverse the quality strings.