

Customer Release Notes

BCL Convert v4.3.6

INTRODUCTION

These Release Notes detail the latest release of BCL Convert, including known issues.

BCL Convert converts per cycle binary data output by Illumina sequencers containing basecall files and quality scores to per read FASTQ files.

NEW FEATURES

- Support for `-bcl-enable-tile-metrics` command line options, which is true by default and when false the software will output only the header and no content for the following files:
 - `Demultiplex_Tile_Stats.csv`
 - `Quality_Tile_Stats.csv`
- Supporting longer index read length from 27 total bases to 27 bases per index
- Include the following new columns in the `fastq_list.csv` file when present in the sample sheet:
 - `RGID, RGSM, RGPU, RGPL, RGLB, RGCN, RGPM`
 - A custom column of that starts with the substring “RG” and has 4 characters
- Support for `OverrideCycles` order independent format
 - Order Dependent: the order of the mask elements must match that of the `RunInfo.xml` (example: `U7N1Y143;I8;I8;U7N1Y143`)
 - Order Independent: order of the mask elements is flexible
 - Each read must be specified in the read description using “R” for genomic or “I” for index followed by the read number
 - The read description is “:” separated from the read mask (example: `R1:U7N1Y143;I1:I8;I2:I8;R2:U7N1Y143`)
 - Order independent and order dependent specifications cannot be mixed within the same sample sheet
- Library Rebalancing Stats report output according to the “`LibraryInputVolume`” setting in the sample sheet to support library and pooling QC on the iSeq 100 system
 - The “`LibraryInputVolume`” setting must be a real number
 - When the setting is specified, the “`LibraryRebalancing_Stats.csv`” metrics file will be output with the following columns:
 - Lane
 - SampleID
 - Index
 - Index2
 - # Reads
 - % Reads
 - Rebalancing Factor
 - Rebalancing Input Volume

RESOLVED ISSUES

- Performance improvements for processing runs with large number of samples per lane
- Run time statistics are output continuously during BCL conversion

KNOWN ISSUES

- If a directory is specified as input to '--sample-sheet', BCL Convert will hang at the beginning of a run while trying to copy that path as a file to <outdir>/Reports/SampleSheet.csv
- BCL Convert does not validate when "Logs" or "Reports" is provided for a Sample_Project, and the software will be unable to create the subdirectories if these string are provided.
- BCL Convert does not support the --first-tile-only option being specified for SP flow cells, but the new **-tiles** option can be used as a substitute.
- Does not error when no tile list exists in the RunInfo.xml file and -tile or -exclude-tiles is specified in the command line
- When an index collision exists in a lane that has been excluded via the ExcludeTiles setting, the software will still error as it is a sample sheet validation error
- BCL Convert has differences in legacy stats (*all of which are justified*)
 - Formatting differences
 - Json library to parse the data will work
 - AdapterTrimming.txt
 - Different number of cycles listed (bcl2fastq2 lists too many cycles)
 - bcl2fastq2 enumerates too many cycles for some reads, particularly when one read is shorter than the other, and sets the "PercentageOfBases" value of the extra cycles to value 0. The extra cycles are found at the beginning of the enumerated cycles. For example, when Read1 is 88 cycles and Read2 is 76 cycles, bcl2fastq2 will enumerate cycles 0-88 for Read1 and 0-88 for Read2. On the other hand, bcl-convert will enumerate cycles 0-88 for Read1 and 0-76 for Read2. The cycles 0-11 for bcl2fastq2 will have value 0.
 - Different values for "PercentageOfBases" (bcl2fastq2 values are based on incorrect total number of cycles)
 - "PercentageOfBases" value differs between bcl2fastq2 and bcl-convert, particularly when OverrideCycles is used. This is because bcl2fastq2 counts the OverrideCycles-trimmed bases in the total bases, while bcl-convert omits those bases from the total count. For the case in which bcl2fastq2 lists 98 cycles for a 94 cycle read, the bcl2fastq2 "PercentageOfBases" value is equal to BCL-Convert value * (94 / 98), which is inaccurate.
 - Rounding differences (bcl2fastq2 is less accurate)
 - ConversionStats.xml
 - Difference in "Raw" cluster metrics (bcl2fastq2 uses statistical best guess, is not deterministic)
 - Even though "Pf" cluster metrics are identical between bcl2fastq2 and bcl-convert for all samples, the "Raw" cluster metrics differ specifically for 'Sample name="Undetermined"' and 'Sample name="all"' (but are identical for all other samples).
 - Top Unknown Barcodes differ in sequence and value (bcl2fastq2 uses statistical best guess, is not deterministic)
 - Difference in cluster count for "Raw" stats when CreateFastqForIndexReads is enabled
 - Having CreateFastqForIndexReads on causes bcl2fastq2 to sometimes double the intended value, even for single-index inputs. This is likely a bug in bcl2fastq2.

- DemuxSummaryF1L1.txt and DemuxSummaryF1L2.txt
 - Different "Most Popular Unknown Index Sequences" sequences and counts (bcl2fastq2 uses statistical best guess, is not deterministic)
 - Different rounding of decimals (bcl2fastq2 is less accurate)
- **FastqSummaryF#L#.txt**
 - Difference in count for "NumberOfReadsRaw" stats when CreateFastqForIndexReads is enabled
 - Having CreateFastqForIndexReads on causes bcl2fastq2 to sometimes double the intended value, even for single-index inputs. This is a bug in bcl2fastq2.
- Stats.json
 - bcl2fastq2 has a space before the colon ':' but bcl-convert does not
 - bcl2fastq2 example: '<key>: <value>' OR '<key> : <value>'
 - bcl-convert example: '<key>: <value>'
 - Top Unknown Barcodes differ in sequence and value (bcl2fastq2 uses statistical best guess, is not deterministic)
- Html reports
 - Differences in Top Unknown Barcodes sequences and counts (bcl2fastq2 uses statistical best guess) (*same as ConversionStats.xml*)
 - Difference in values for "Clusters (Raw)" and "% PF Metrics" when CreateFastqForIndexReads is enabled
 - Having CreateFastqForIndexReads on causes bcl2fastq2 to sometimes double the intended value, even for single-index inputs. This is a bug in bcl2fastq2.
- IndexMetricsOut.bin
 - Bcl convert will use the value for Sample_Project from the Sample Sheet in the IndexMetrisOut.bin file even when the use of the Sample_Project column is disabled via the command line (no differences to bcl2fastq2 as it is a new feature to disable the column)
 - When Sample_Project is disabled, bcl-convert is expected to behave exactly as if the Sample_Project column did not exist in the sample sheet. In this case, the value for Sample_Project should be the following string: "default". However, bcl-convert uses the value of Sample_Project from the sample sheet.
 - Note that the Sample_Name column behaves correctly - bcl-convert correctly uses Sample_ID in IndexMetricsOut.bin instead of Sample_Name when the Sample_Name column is disabled.

RELEASE HISTORY

| Revision | Workflow | Originator | Description of Change |
|----------|----------|--------------|-----------------------|
| 00 | 1107252 | Daniel Tracy | Initial release |