Customer Release Notes

BCL Convert v4.4.4

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

- Updates for greater flexibility and efficiency.
  - New SampleSheet settings
    - OverrideReads: Redefine read structure by overriding RunInfo (read cycle boundaries, genomic/index status)
    - Index Orientation: Specify index2 orientation as forward or reverse
    - Custom columns support:
      - Use custom_* fields for per-readgroup customer data
      - v2 sample sheets ignore these fields without error – helpful for sample management
  - New output files
    - New Demultiplex_Detailed_Stats.csv file provides cycle & transition error details for mismatch tolerant demultiplexing (miscall metrics)
    - Extended columns to per-tile Demultiplex_Tile_Metrics.csv file, to include derived stats identical to those in the aggregate Demultiplex_Stats.csv file
  - Generate undetermined read stats when enabling --bcl-only-matched-reads
    - Even though Undetermined FASTQs are not produced, the stats will still be produced.
  - Memory usage optimization
    - New settings reduce memory usage for large sample batches

Resolved Issues

- Fixed an issue where BCL seemingly omits tiles with zero reads from Reports files: Demultiplex_Tile_Stats.csv, Quality_Tile_Metrics.csv, and IndexMetricsOut.bin
- Fixed an issue where UMI sequences in header could be different from UMI sequences in reads
- Fixed an issue where BCL output very large cluster counts in legacy stats
- Fixed an issue where BCL LibraryRebalancing_Stats.csv %Reads column zeroed out when no Lane column used
- Fixed BCL outputs legacy stats with invalid characters
- Fixed a BCL convert discordance in indexmetricsout.bin
- Fixed an issue where BCL->Ora with FastqCompressionFormat set to dragen-interleaved in the sample sheet and samples with mask settings will revert to outputting non-interleaved Ora files but will still produce a fastq_list.csv with interleaved Ora file names in the Read1File column (and blank Read2File column)
- Fixed create-fastq-with-index-reads command line option not supporting true/false

KNOWN ISSUES

Summary	Resolution/Workaround
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	Specify the sample sheet file.
BCL does not detect when LibraryInputVolume setting is blank	Blank/empty value is the same as not providing the setting, which is the same as the default setting of being disabled.
BCL conversion appends FASTQ files when using "—force". FASTQ output may get concatenated if user uses the same output directory twice for BCL.	Do not run BCL conversion multiple times using the same output folder
Sample sheet validation can pass, but demultiplexing fails	Barcode collision detection works slightly differently for ss validation and real runs. Some errors are not caught at ss validation step.

- 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 (bcl2fastq 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	CN 1122438	Cobus De Beer	Initial release