Customer Release Notes

BCL Convert v4.4.6

INTRODUCTION

These Release Notes detail the latest release of BCL Convert v4.4.6, 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.

CHANGES

BCL convert fixes some regressions introduced in earlier versions, related to very long run times with very high sample counts and network file systems. Introduces flexible no-lane-splitting to output concatenated FASTQ files over a subset of lanes.

- Bulk conversion on AWS with EBS or Lustre output was extremely slow (over 4 days per lane), this reduces it back to normal 3 hours per lane.
- Fix a segfault on bcl.bgzf with NoLaneSplitting + bcl-validate-sample-sheet-only=true
- Accept 'CmdExtension' as a valid key in BCLConvert_Settings section
- Fix very slow run time on NovaX instrument, for very high sample counts. Eliminated thread-reduction at 10K files for all run types
- Fix extreme performance regression on v4.4.4 for demuxmap generation with fully masked indexes.
- Flexible no-lane-splitting is now built into bcl-convert.
    - No longer is lane column required to be absent, can have different sampleIDs across lanes.
    - Every sample must have the same file output
    - NOTE: Ora is not supported
- Fix for Report Rebalancing stats reporting identical i5s for all samples.
- Ignore several settings in sample sheet when doing sample sheet validation and/or bcl-convert (no error)
    - KeepFastq
    - GenerateFastqcCoverageStats
    - FastqcDownsampling

KNOWN ISSUES

Known issues of v4.4.4

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.

New issues discovered on v4.4.6

Summary	Resolution/Workaround
BCL->Ora one-step mode does not error out when NoLaneSplitting is enabled and the input sample sheet does not contain lanes.	None. NoLaneSplitting not supported for ORA output
Fastq ORA compression error when bcl-only-matched-reads=true are used.	Use the "--force" option
BCL run time is very long on proteomics prep samples.	Do not use this version v4.4.6 for BCL on proteomics prep samples.

Differences with bcl2fastq

- 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	1127184	Cobus De Beer	Initial release