CHARITÉ
UNIVERSITÄTSMEDIZIN BERLIN

Bioinformatics Workshop

bcl2fastq

2.8.2023

Therese Muzeniek

NGS output (.bcl) to pipeline input (.fastq)

.bcl format

- 001.bcl
- 002.bcl
- 003.bcl
- 004.bcl
- 005.bcl
- 006.bcl
- 007.bcl
- 008.bcl
- 009.bcl

bcl2fastq

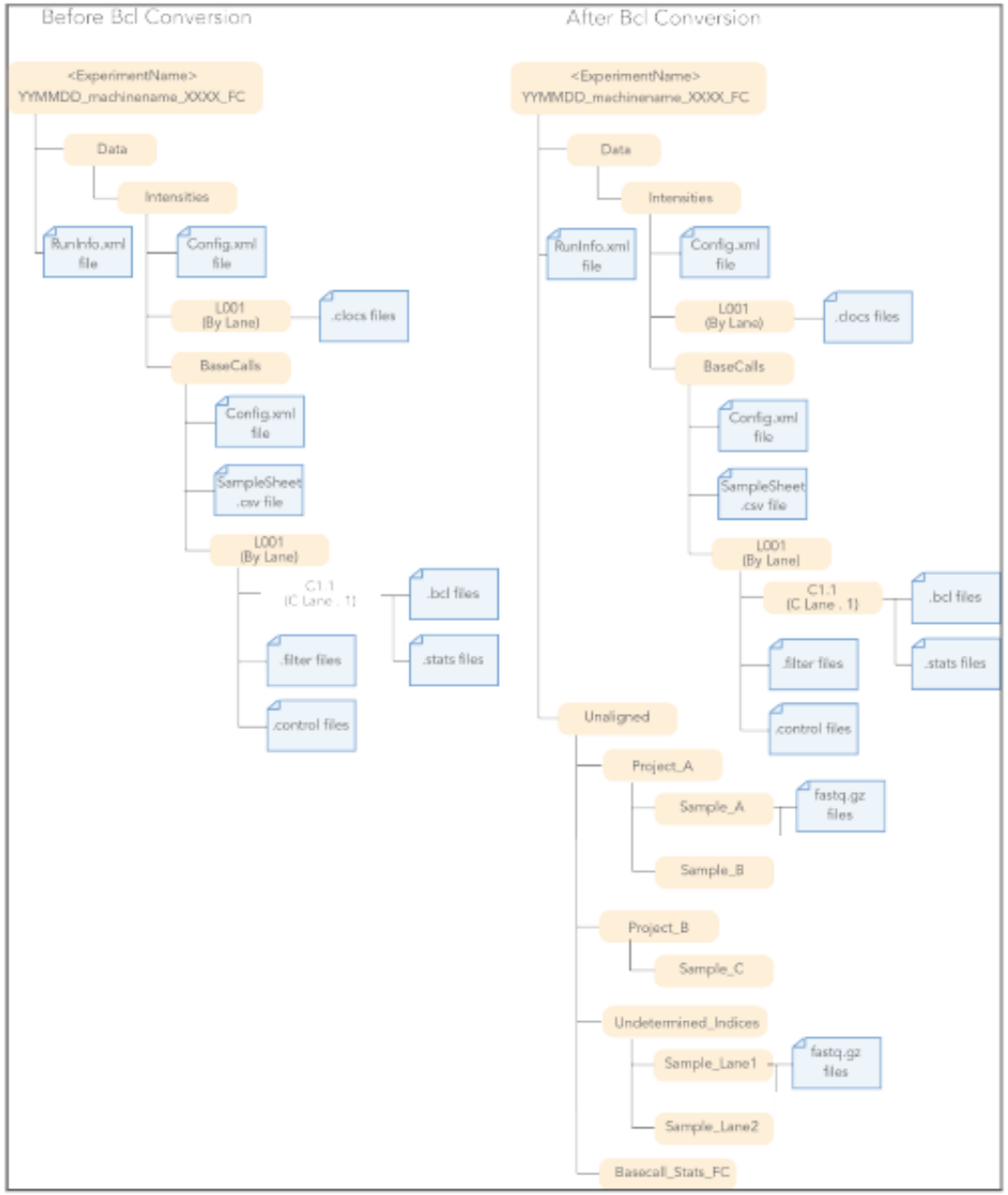
.FASTQ format

bcl2fastq

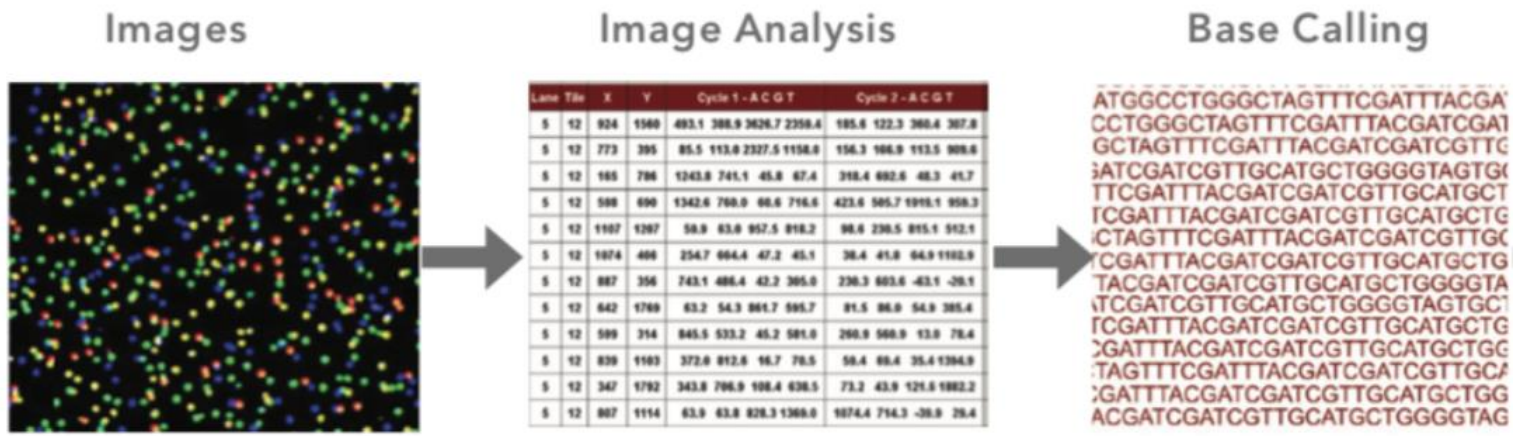
Illumina NGS run

Folder structure

Figure 1 Typical Run Folder Structure after Bcl Conversion and Demultiplexing



bcl2fastq



- .bcl files: How many files do we expect from a 300 cycle kit, paired end reads?

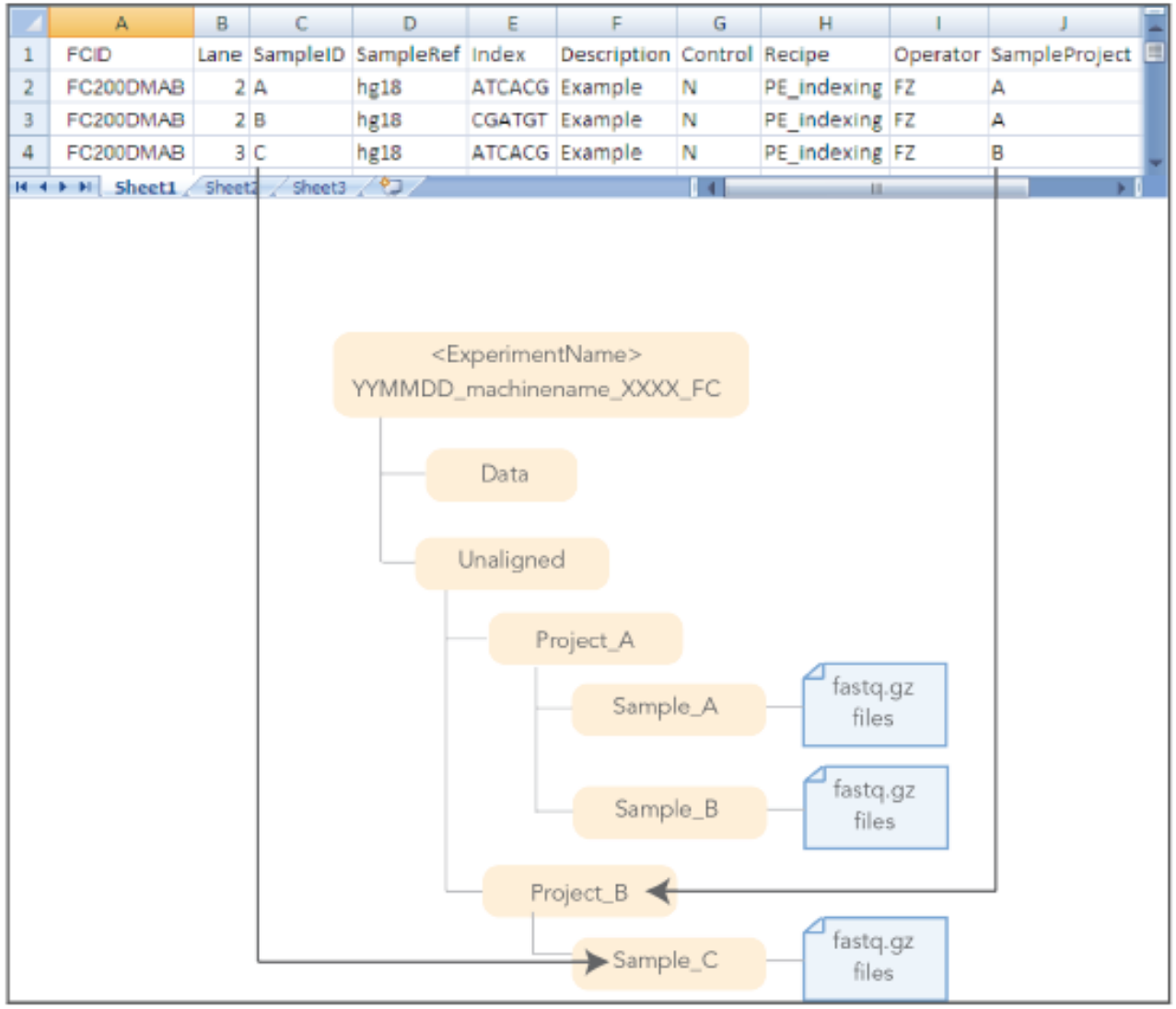
  (318 files: 2 x 151 + 2 x 8 for the index sequencing)

- the bcl2fastq program needs a samples sheet (.csv) with sample information for demultiplexing

bcl2fastq

Sample sheet

Figure 2 Relation between Sample Sheet and Directory Structure



bcl2fastq

Input:

- Sample sheet in .csv format
  - → Information to identify samples by their adapters
  - → Sorting of reads by adapter to different folders
  - → Incorrectly formatted data will cause and error
- .bcl files in the run folder

Output:

- FASTQ files, one folder per sample
- For PE: R1 and R2 in separate folders
  - → Important for the next rule (fastp)