

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

## .bcl format .FASTQ format - 001.bcl - 002.bcl - 003.bcl TGCTAC... - 004.bcl bcl2fastq - 005.bcl - 006.bcl - 007.bcl - 008.bcl - 009.bcl



Illumina NGS run Folder structure

Before Bcl Conversion After Bcl Conversion <ExperimentName> <ExperimentName> YYMMDD\_machinename\_XXXX\_FC YYMMDD\_machinename\_XXXX\_FC Data Data Intensities Intensities Config.xml RunInfo.xml RunInfo.xml Config.xml file file file file L001 (By Lane) L001 (By Lane) .clocs files .clocs files BaseCalls BaseCalls Config.xml Config.xml file file ampleSheet .csv file .csv file (By Lane) L001 (By Lane) (C Lane . 1) C1.1 (C Lane . 1) .bcl files .bcl files stats files. .filter files .stats files filter files Unaligned .control files control files Project\_A fastq.gz files Project\_B Undetermined\_Indices Sample\_Lane1 -Sample\_Lane2 Basecall\_Stats\_FC

Figure 1 Typical Run Folder Structure after Bcl Conversion and Demultiplexing

CHARITÉ



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

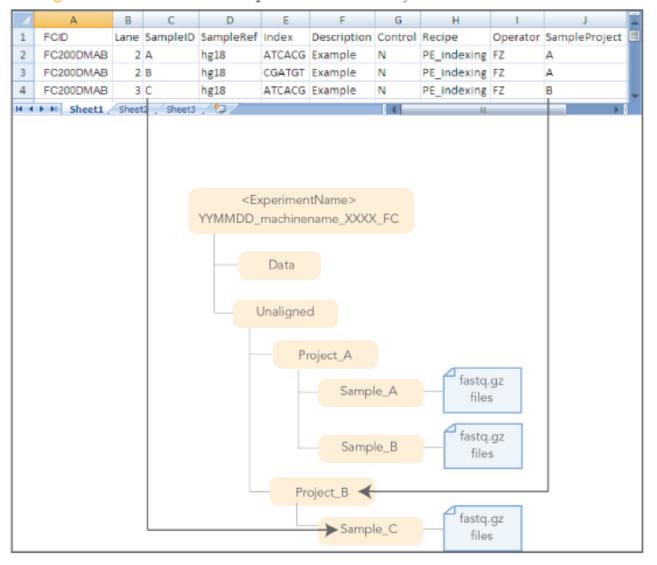
(318 files:  $2 \times 151 + 2 \times 8$  for the index sequencing)

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



Sample sheet

Figure 2 Relation between Sample Sheet and Directory Structure





#### 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)

