



CEITEC

Central European Institute of Technology
BRNO | CZECH REPUBLIC

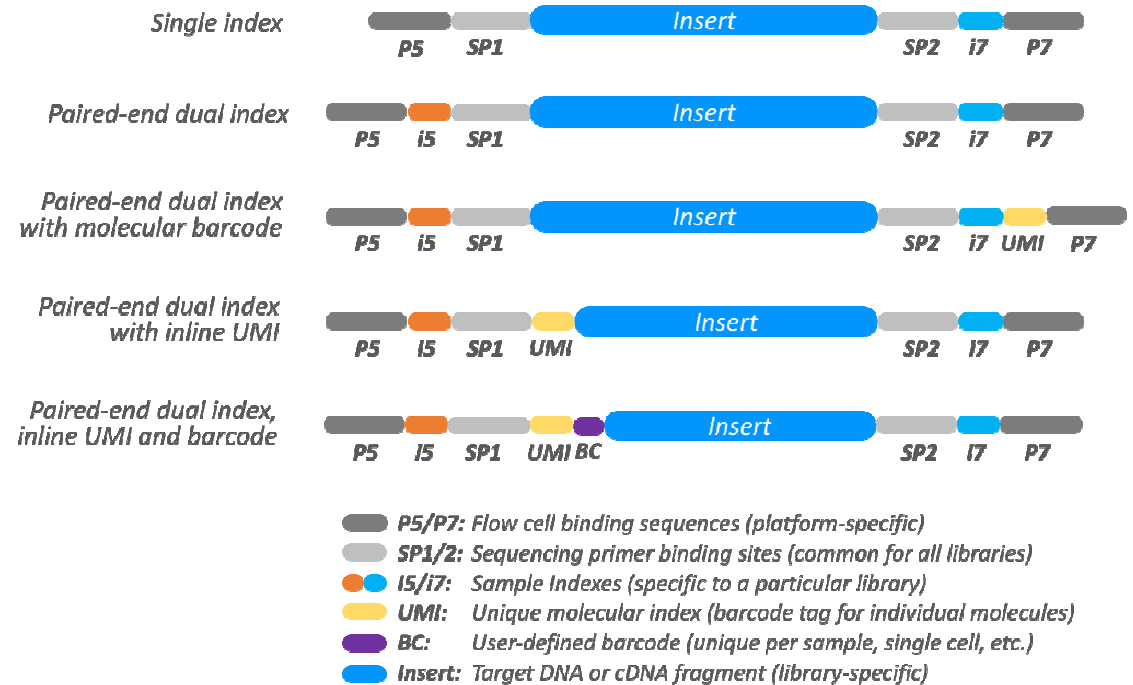
Modern Genomic Technologies (LF:DSMGT01)

Lecture 2 : NGS libraries and basic data quality control

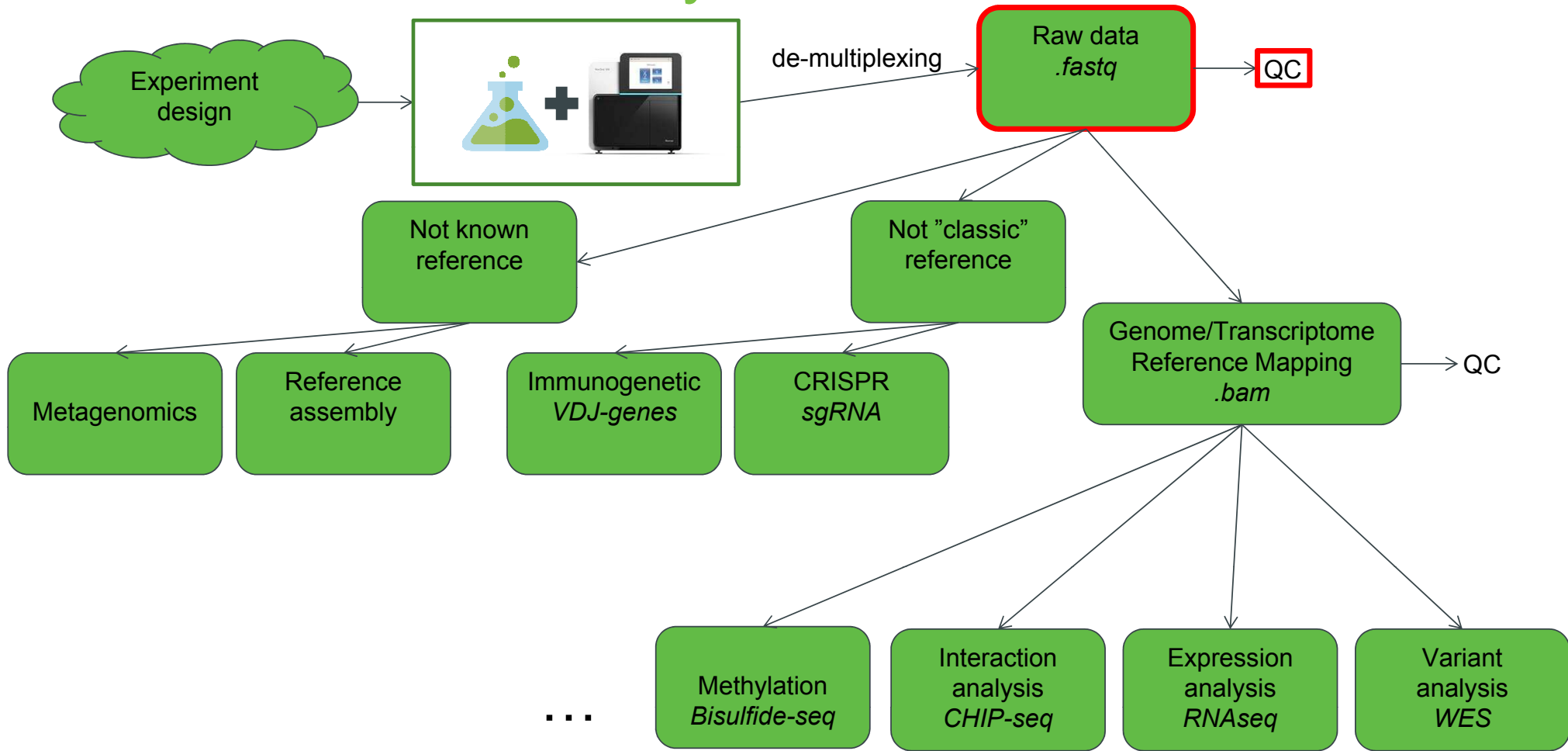
Vojta Bystry
vojtech.bystry@ceitec.muni.cz

De-multiplexing

- Bcl2fastq tool
 - Needs sample sheet with indexes
 - Number of barcode mismatches
 - Check undetermined



NGS data analysis



Fastq format - quality

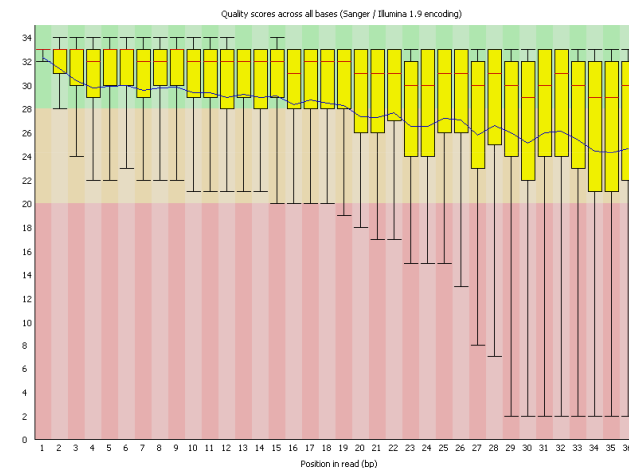
- **Fastq - q stands for quality – coded phred score**

CFFFFEFFFGCEECECFGGGAFF87@E:++6C<+3:,8,33,,:,,,::,,:,,

$$Q = -10 \cdot \log_{10} P$$

Quality	Error probability
5	31%
10	10%
20	1%
30	0.1%

- **Very good for early problem detection**
- **Reasonable for trimming and read filtering**
 - RNA seq - above phred score 5



FastQC Report

Summary

Return to [start page](#)

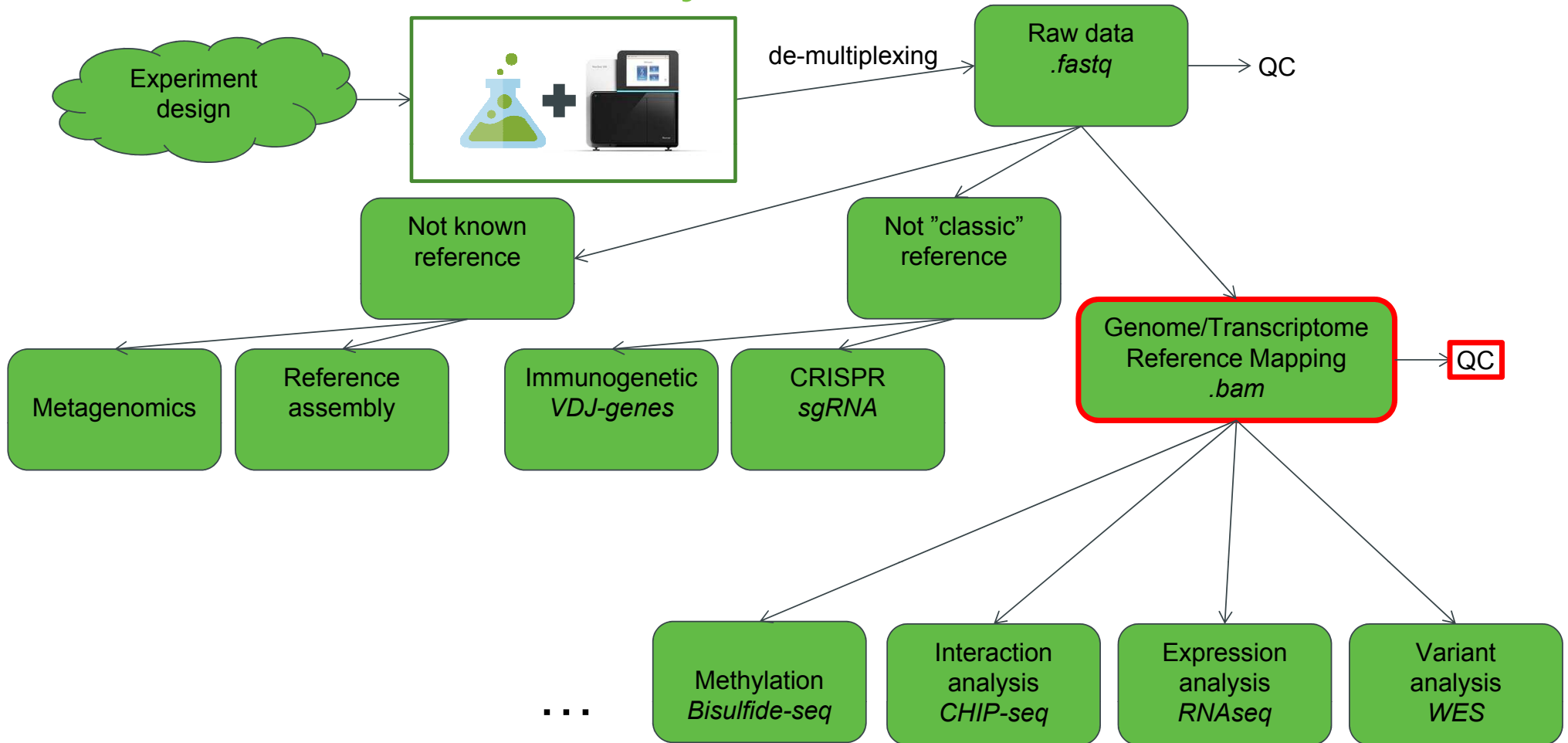
- ✔ [Basic Statistics](#)
- ✔ [Per base sequence quality](#)
- ✔ [Per tile sequence quality](#)
- ✔ [Per sequence quality scores](#)
- ! [Per base sequence content](#)
- ! [Per sequence GC content](#)
- ✔ [Per base N content](#)
- ✔ [Sequence Length Distribution](#)
- ✔ [Sequence Duplication Levels](#)
- ✔ [Overrepresented sequences](#)
- ✔ [Adapter Content](#)

✔ Basic Statistics

Measure	Value
Filename	MU_a_ytHl_R1.fastq.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	252819865
Sequences flagged as poor quality	0
Sequence length	161
%GC	40



NGS data analysis



 CEITEC @CEITEC_Brno

Thank you for your attention!

 CEITEC

www.ceitec.eu

Vojta Bystry
vojtech.bystry@ceitec.muni.cz