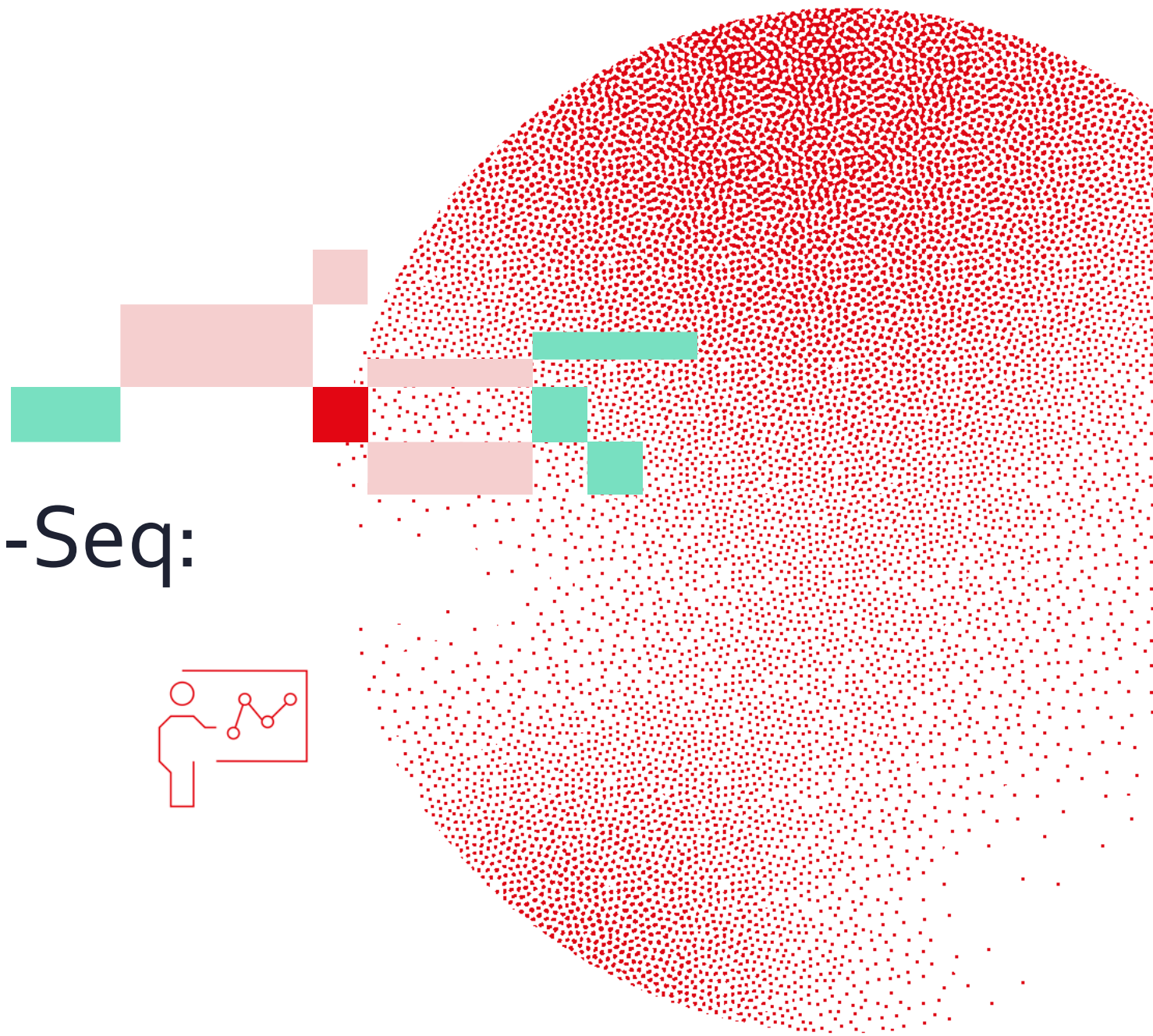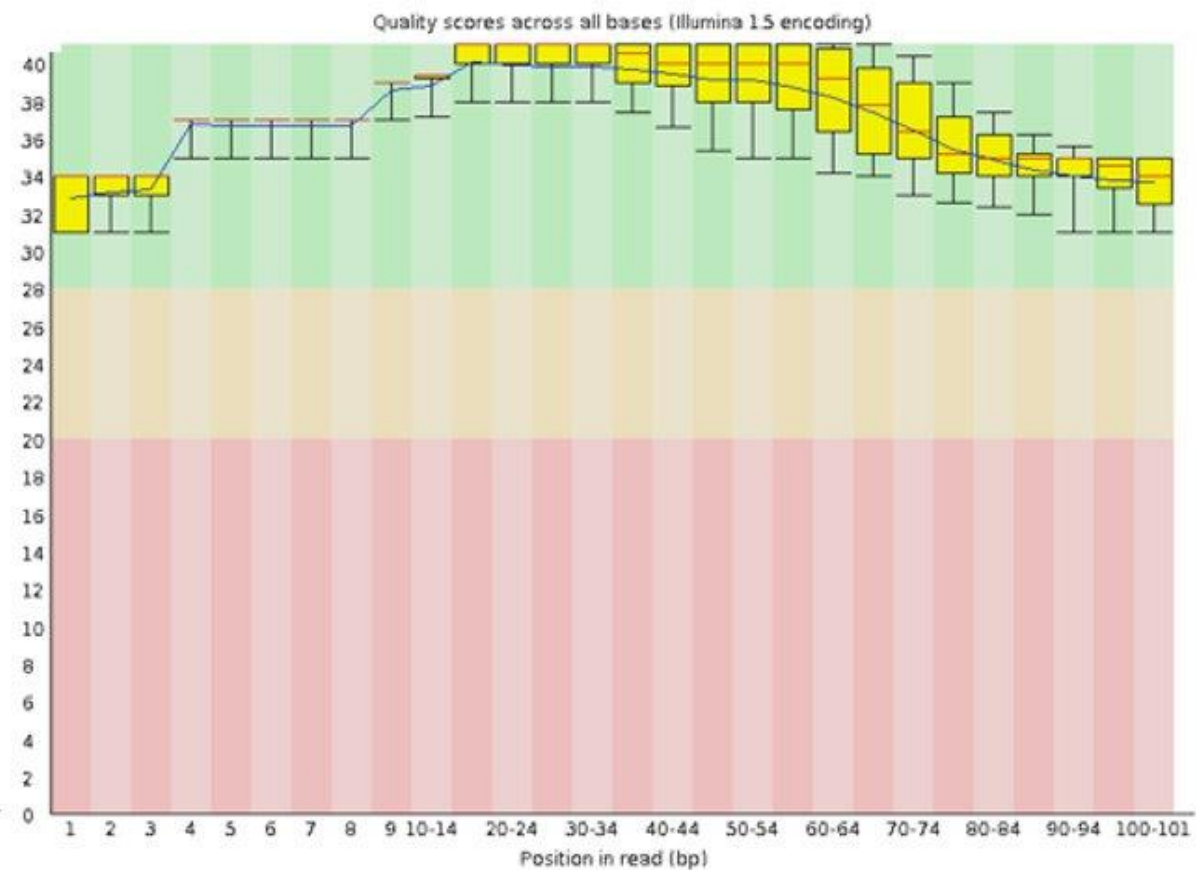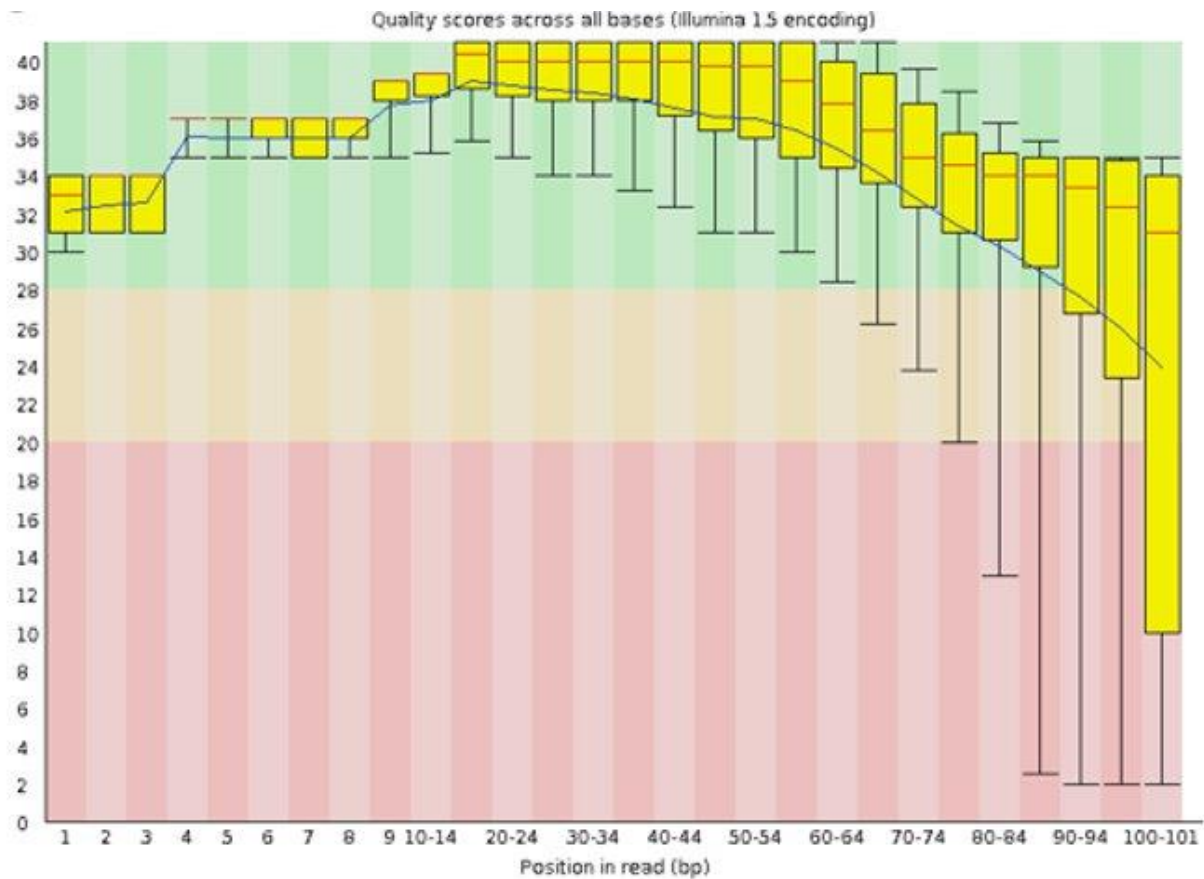# Introduction to RNA-Seq: Sequence trimming
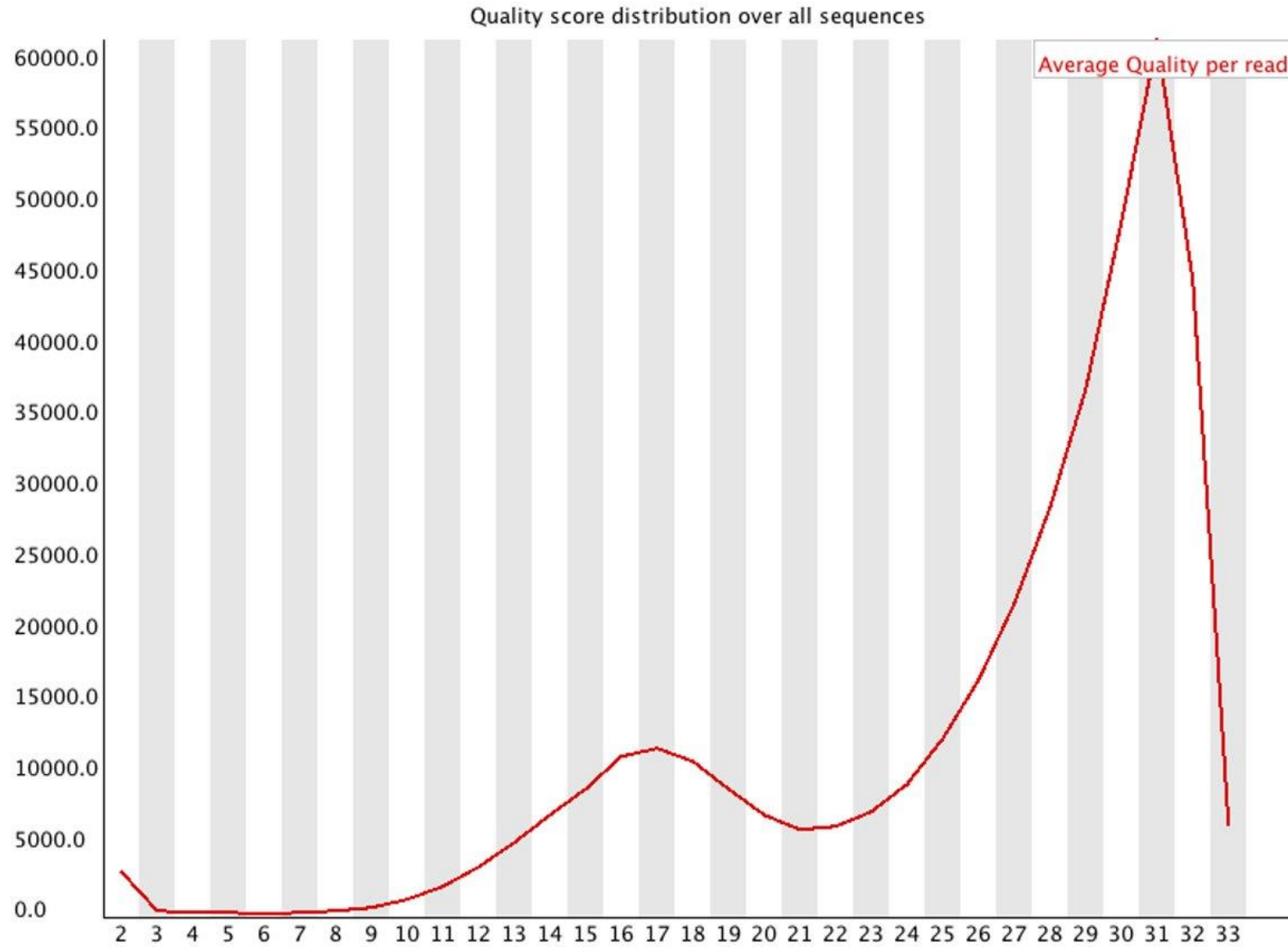
Wandrille Duchemin

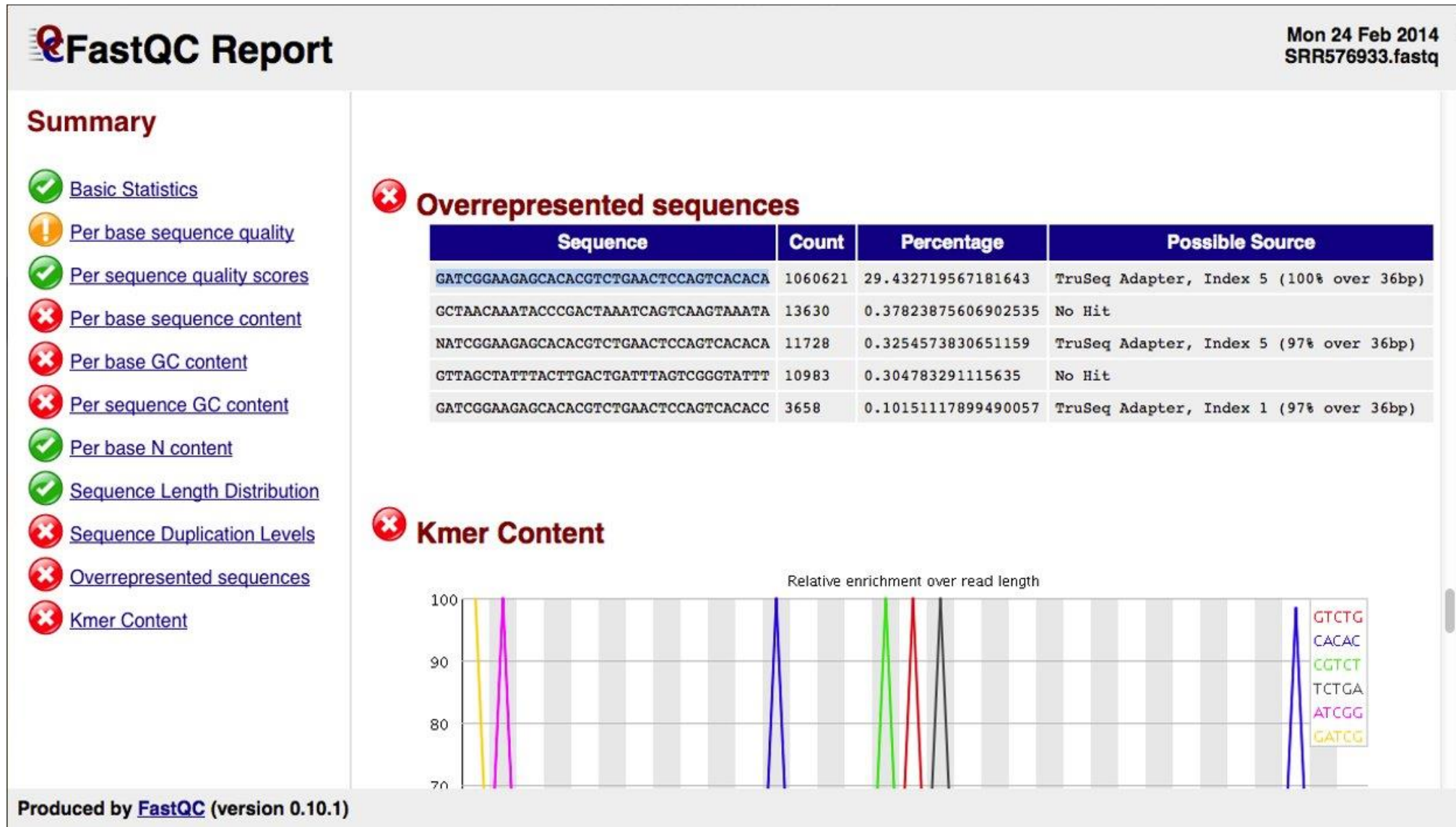# Par base sequence quality

Par sequence quality score

Quality score distribution over all sequences
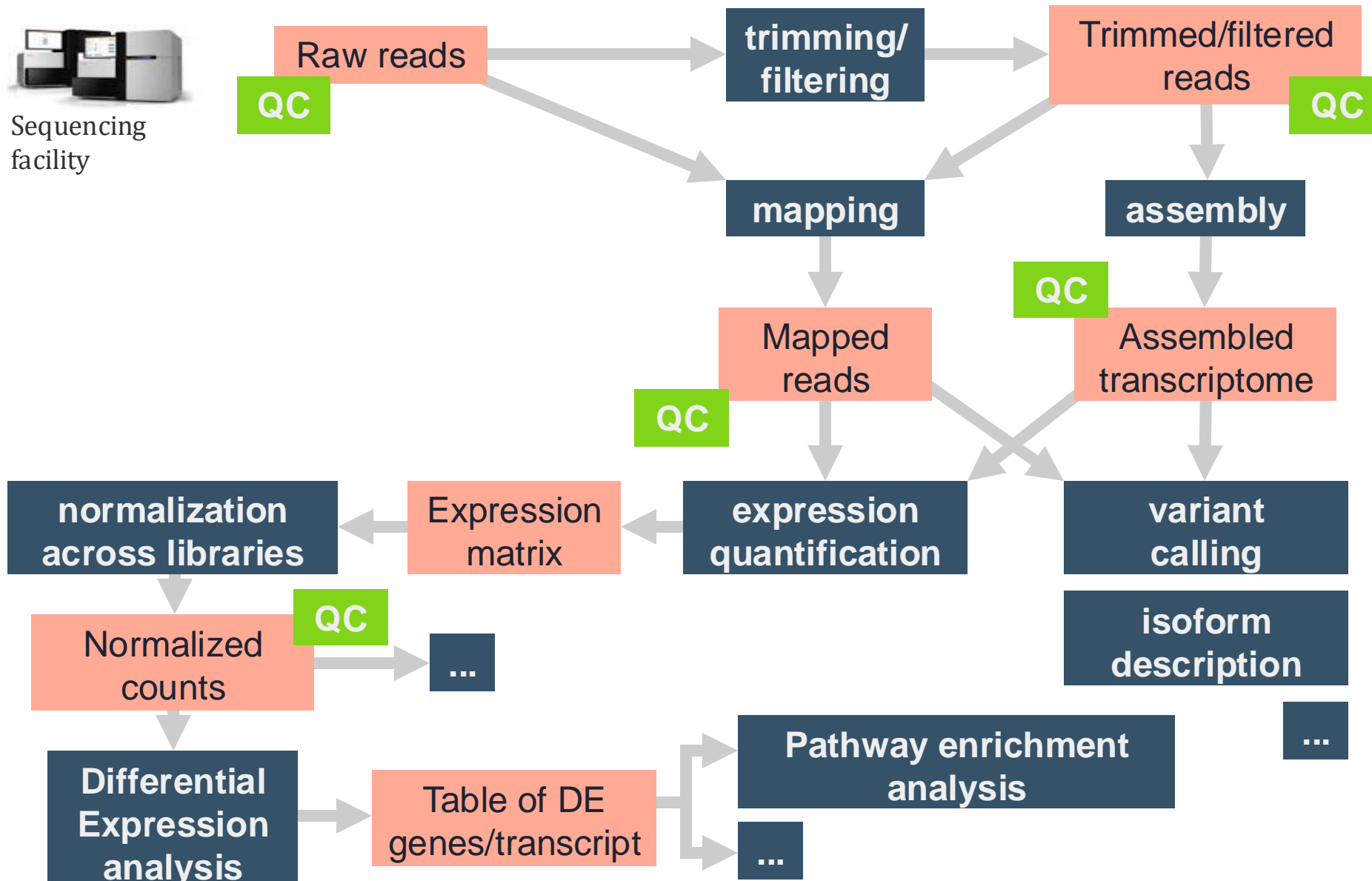
Over-represented sequences

Trimming: doing it or not?

Trimming: doing it or not?

Trimming: doing it or not?

Sequencing facility

Raw reads → trimming/ filtering → Trimmed/filtered reads

assembly

Assembled transcriptome
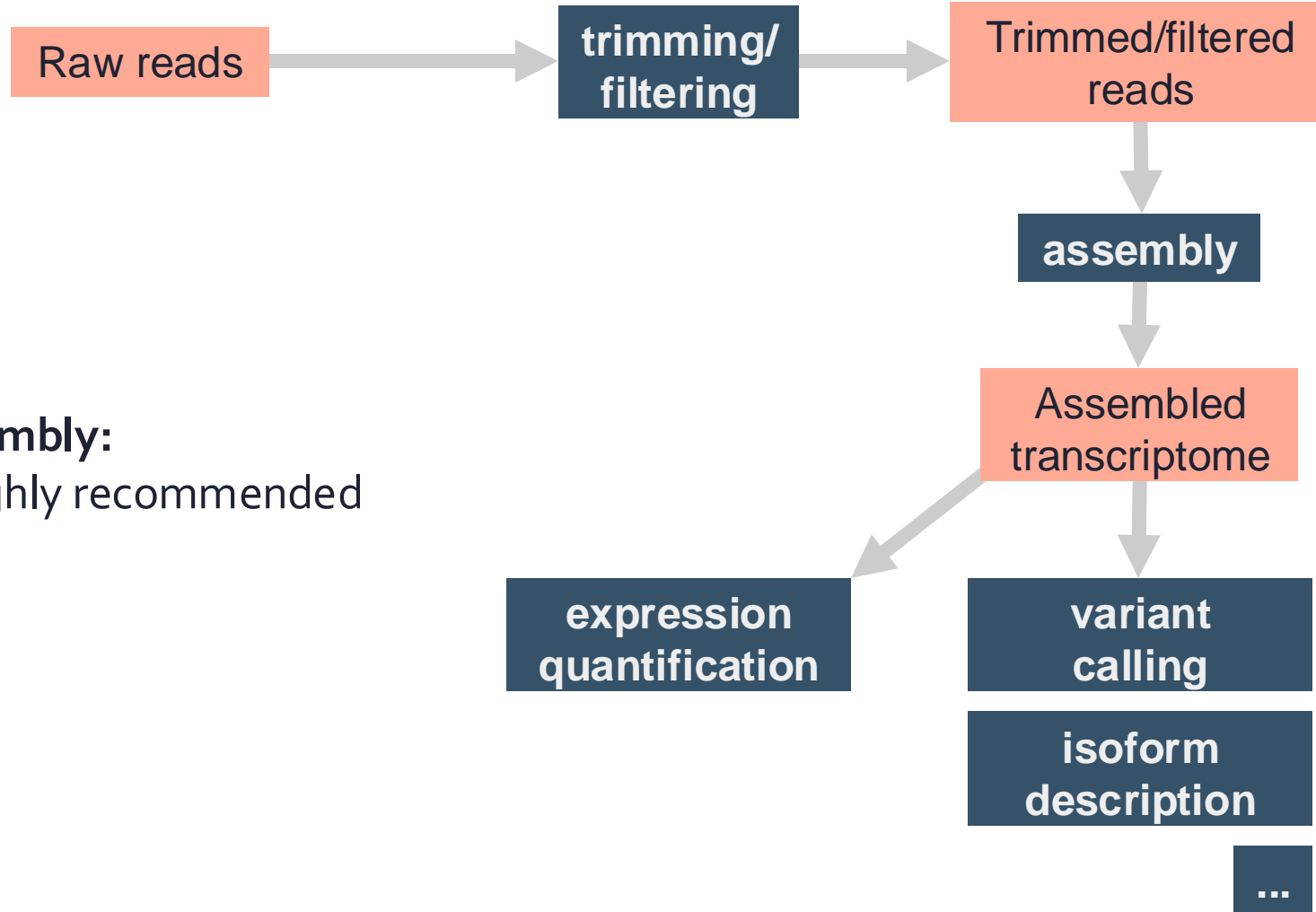
Transcriptome assembly:
Trimming is highly recommended (at least light)

expression quantification

variant calling

isoform description

...

Trimming: doing it or not?

Sequencing facility

Raw reads → trimming/filtering → Trimmed/filtered reads → mapping → Mapped reads → variant calling …

Variant analysis:
Trimming is highly recommended (at least light)

Trimming: doing it or not?

Sequencing facility

Raw reads → trimming/filtering → Trimmed/filtered reads → mapping → Mapped reads → expression quantification
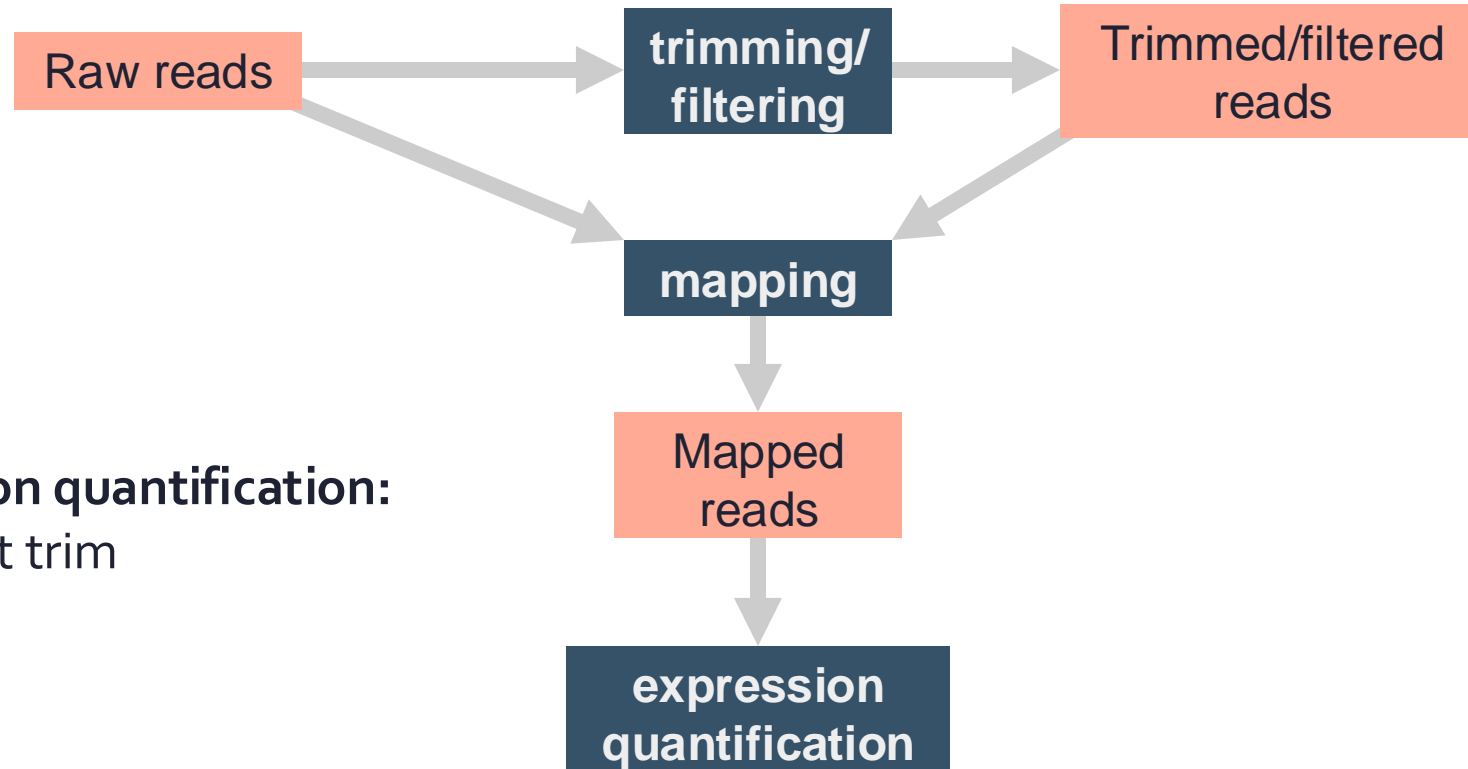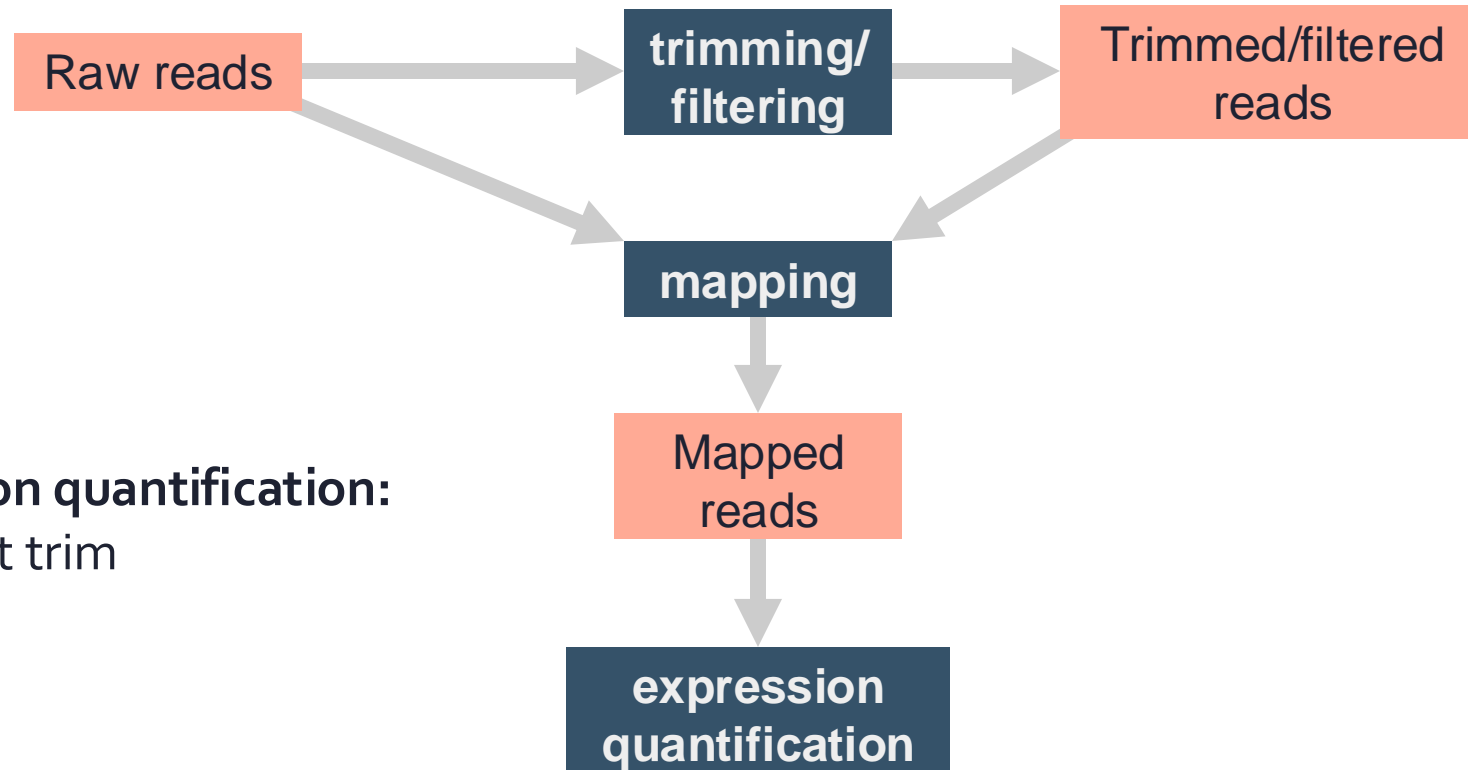
Counting, expression quantification:
No trimming or light trim

Trimming: doing it or not?

Sequencing facility

Raw reads → trimming/filtering → Trimmed/filtered reads

mapping

Mapped reads

expression quantification

Counting, expression quantification:
No trimming or light trim

Aggressive trimming (high quality threshold, low length filtering) can have negative impacts on expression quantification

see: William et al. (2016) Trimming of sequence reads alters RNA-Seq gene expression estimates. BMC bioinfo.

Trimming: doing it or not?

Sequencing
facility

Raw reads → trimming/ filtering → Trimmed/filtered reads

mapping

Counting, expression quantification:
No trimming or light trim

Mapped reads

QC

expression quantification

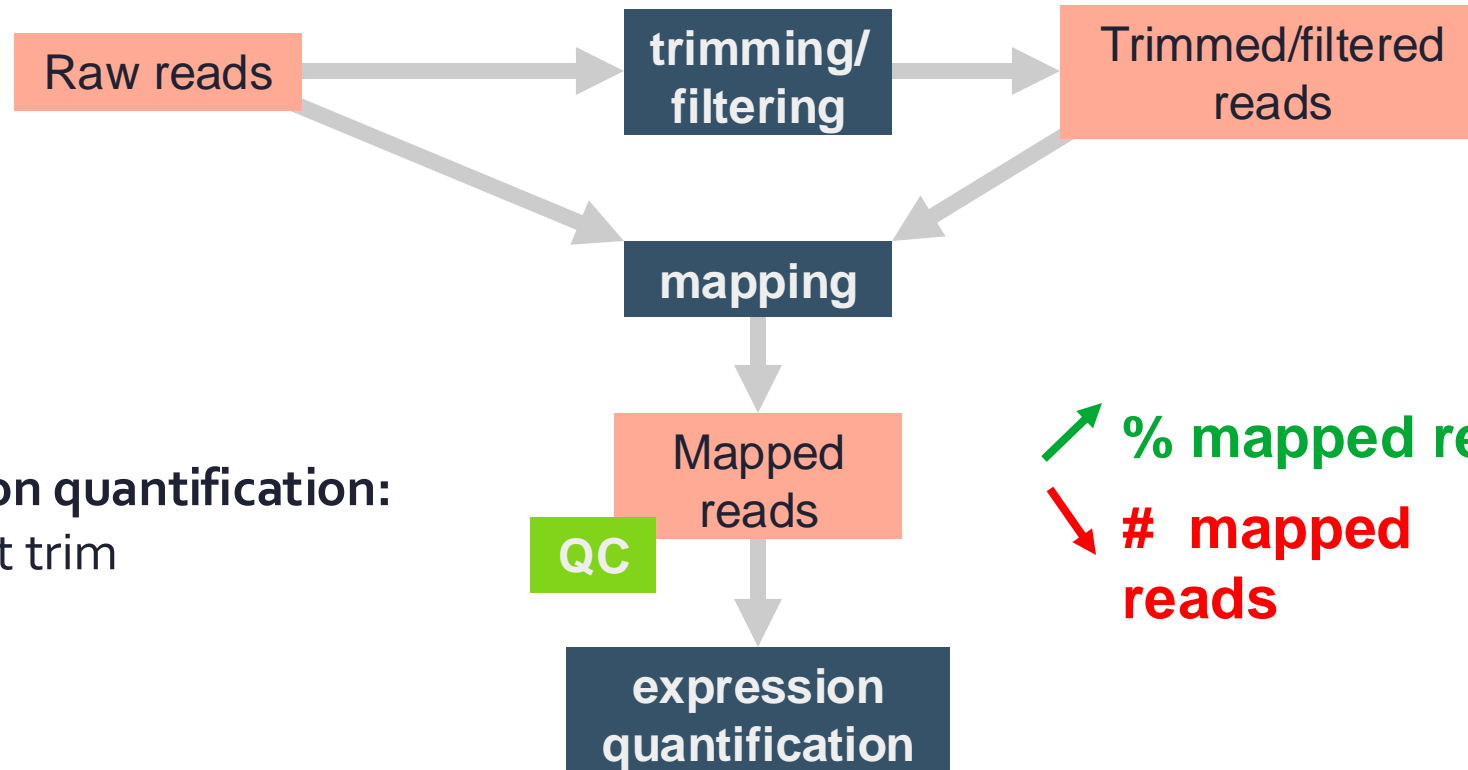↗ **% mapped reads**

↘ **# mapped reads**

Aggressive trimming (high quality threshold, low length filtering) can have negative impacts on expression quantification

see: William et al. (2016) Trimming of sequence reads alters RNA-Seq gene expression estimates. BMC bioinfo.

Trimmomatic

http://www.usadellab.org/cms/?page=trimmomatic

ILLUMINACLIP: Cut adapter and other illumina-specific sequences from the read.

SLIDINGWINDOW: Performs a sliding window trimming approach. It starts scanning at the 5" end and clips the read once the average quality within the window falls below a threshold.

LEADING: Cut bases off the start of a read, if below a threshold quality

TRAILING: Cut bases off the end of a read, if below a threshold quality

CROP: Cut the read to a specified length by removing bases from the end

HEADCROP: Cut the specified number of bases from the start of the read

MINLEN: Drop the read if it is below a specified length

AVGQUAL: Drop the read if the average quality is below the specified level

Practical