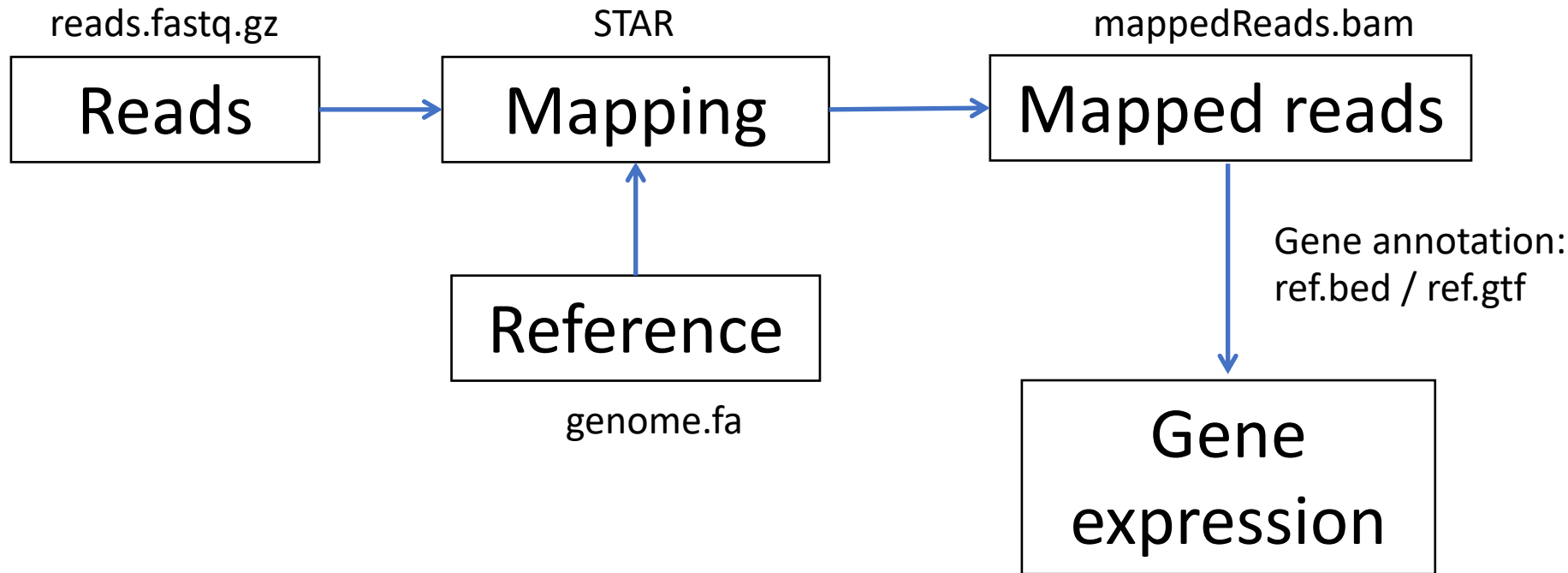


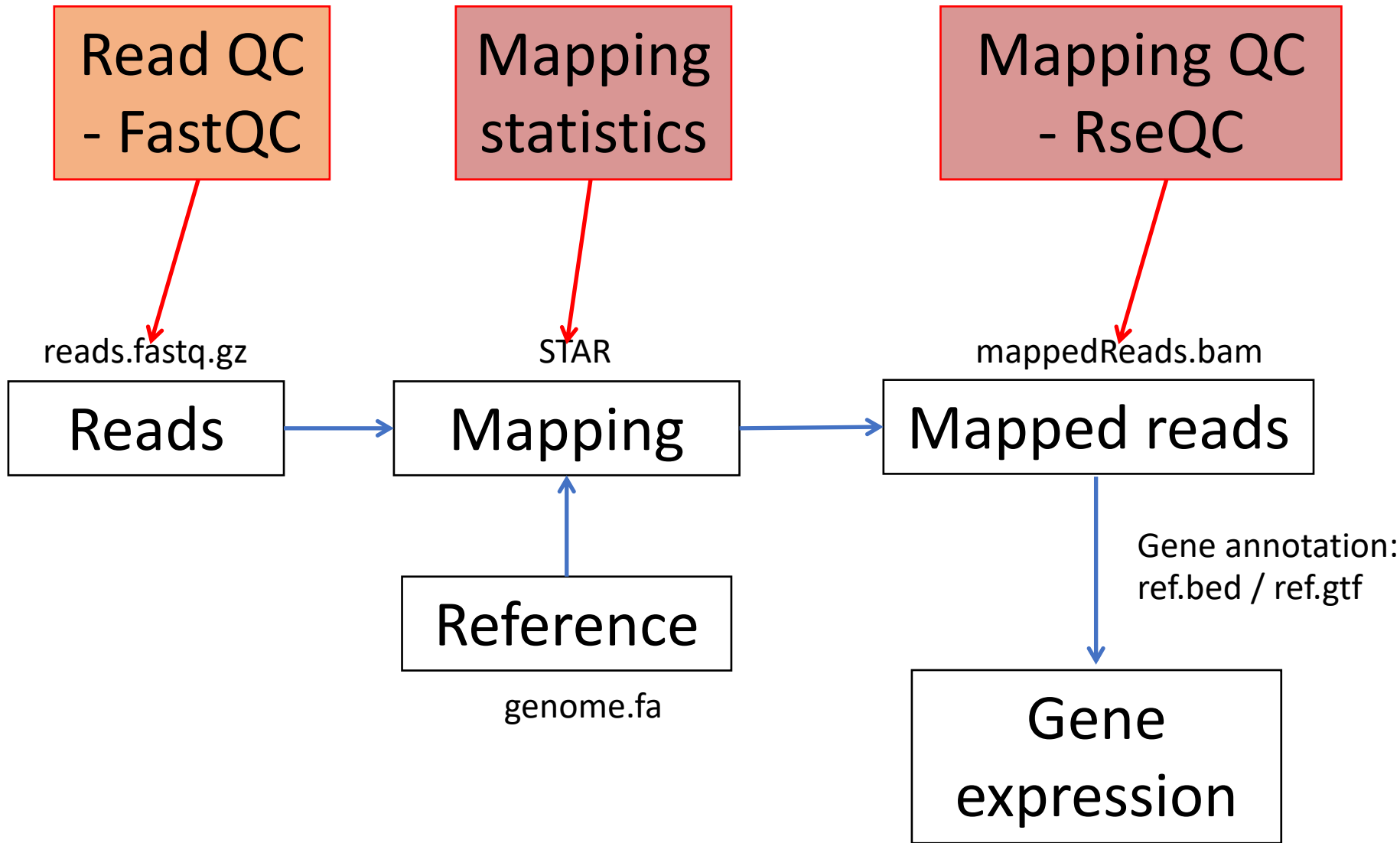
RNA-seq QC analysis

RNA-seq data analysis

Johan Reimegård | 30-November-2020

RNA-seq analysis workflow





What could go wrong?

- RNA quality:
 - Degradation
 - Contaminations (pathogens or other sources)
 - GC-bias
 - Nuclear vs organelle reads
- Library prep:
 - Failed reactions
 - RNA / Adapter ratios – primer dimers
 - Clonal duplicates
 - Chimeric reads
 - Contaminations
- Sequencing:
 - Base calling errors
 - Uncalled bases
 - Low quality bases (3' end)
 - Contaminations
 - Sequence complexity

From samples to reads

- may not be what you think they are

- Mixing samples
 - 30 samples with 5 steps from samples to reads has 24 300 000 potential mix ups of samples
 - Error rate 1/ 100 with 5 steps suggest that one of every 20 sample is mislabeled
- Experiments go wrong
 - 30 samples with 5 steps from samples to reads has 150 potential steps for errors
 - Error rate 1/100 with 5 steps suggest that one of every 20 samples the reads does not represent the sample
- Combine the two error sources and approximately one in every 10 samples is wrong

From samples to reads

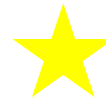
- may not be what you think they are

- Mixing samples
- Experiments go wrong
- How do we understand what went wrong?

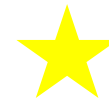
From samples to reads

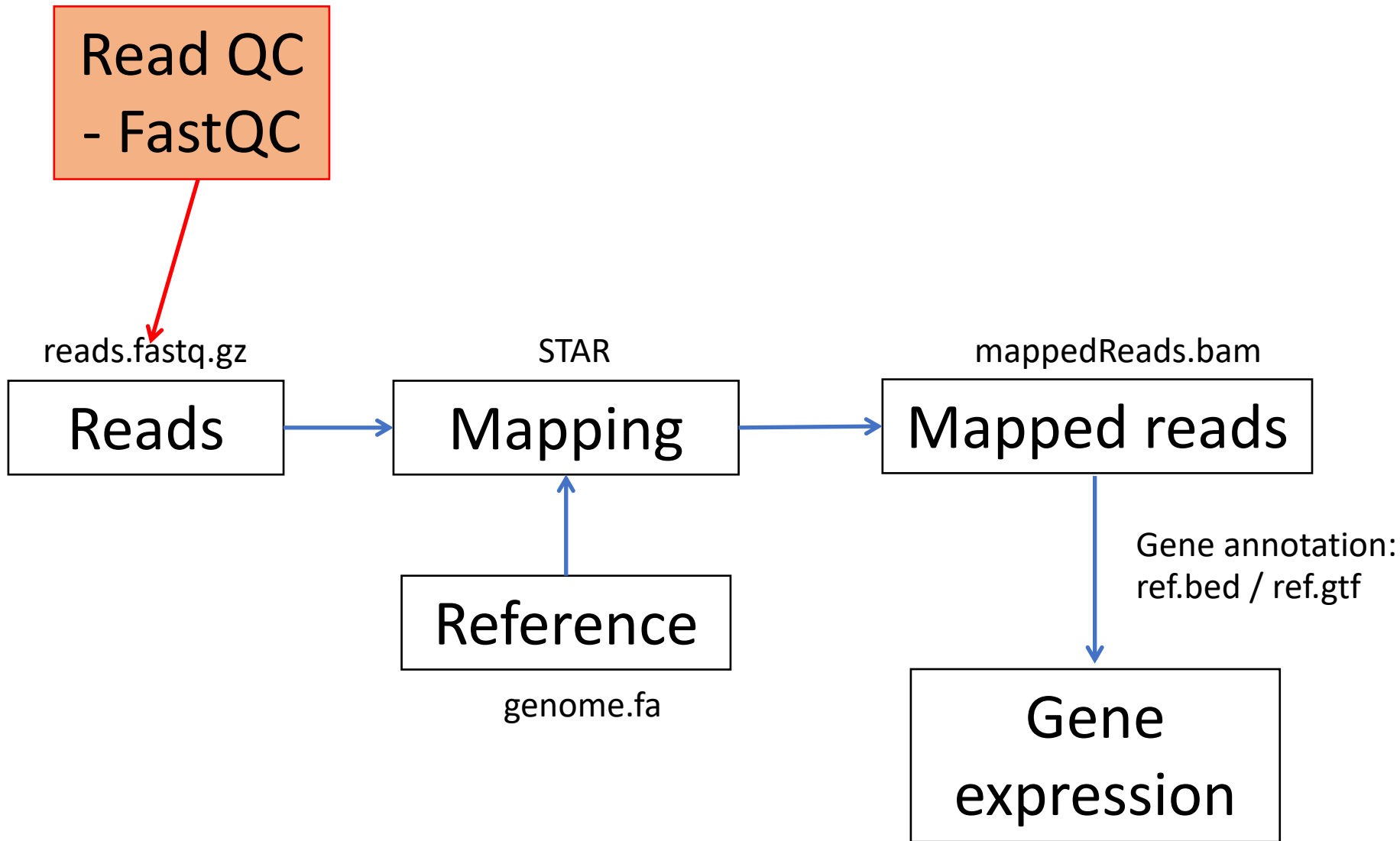
- may not be what you think they are

- Mixing samples
- Experiments go wrong
- How do we understand what went wrong?



BIOINFORMATICS!





Fastq – read file format

Unique identifier

Sequence

```
@SEQ_ID  
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT  
+  
!''*(((((***+))%%%++) (%%%) .1***-+*'')) **55CCF>>>>>CCCCCCC65
```

Sequence quality

Paired end data usually in format sampleX_1.fastq and sampleX_2.fastq with same SEQ_ID for both mate pairs, followed by /1 and /2 (or _f and _r)

Basic read metrics with FastQC

A program that analyses some of the basic metrics on fastq raw read files.

- Quality
- Length
- Sequence bias
- GC content
- Repeated sequences
- Adapter contamination

```
$ fastqc -o outdir seqfile.fastq
# multiple files:
$ fastqc -o outdir seqfile_*.fastq
```

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>



Thank you. Questions?

Johan Reimegård