Quality Checks

QC metrics

Various metrics can give important information about the quality of the library:

-- Total % of reads aligning? % of uniquely mapping reads? % of properly paired PE reads?

-- Genomic origin of reads (exonic, intronic, intergenic)

-- Quantity of rRNA

-- Transcript coverage and 5'-3' bias

How do we compute QC metrics?

- Tools like RNA-SeQC and Qualimap
    - *Input:* alignment file formats (i.e. SAM/BAM)
    - *Output:* summary of the different metrics in an HTML report format

Where do we get this SAM/BAM file from?

- Need to ***align reads to the genome***

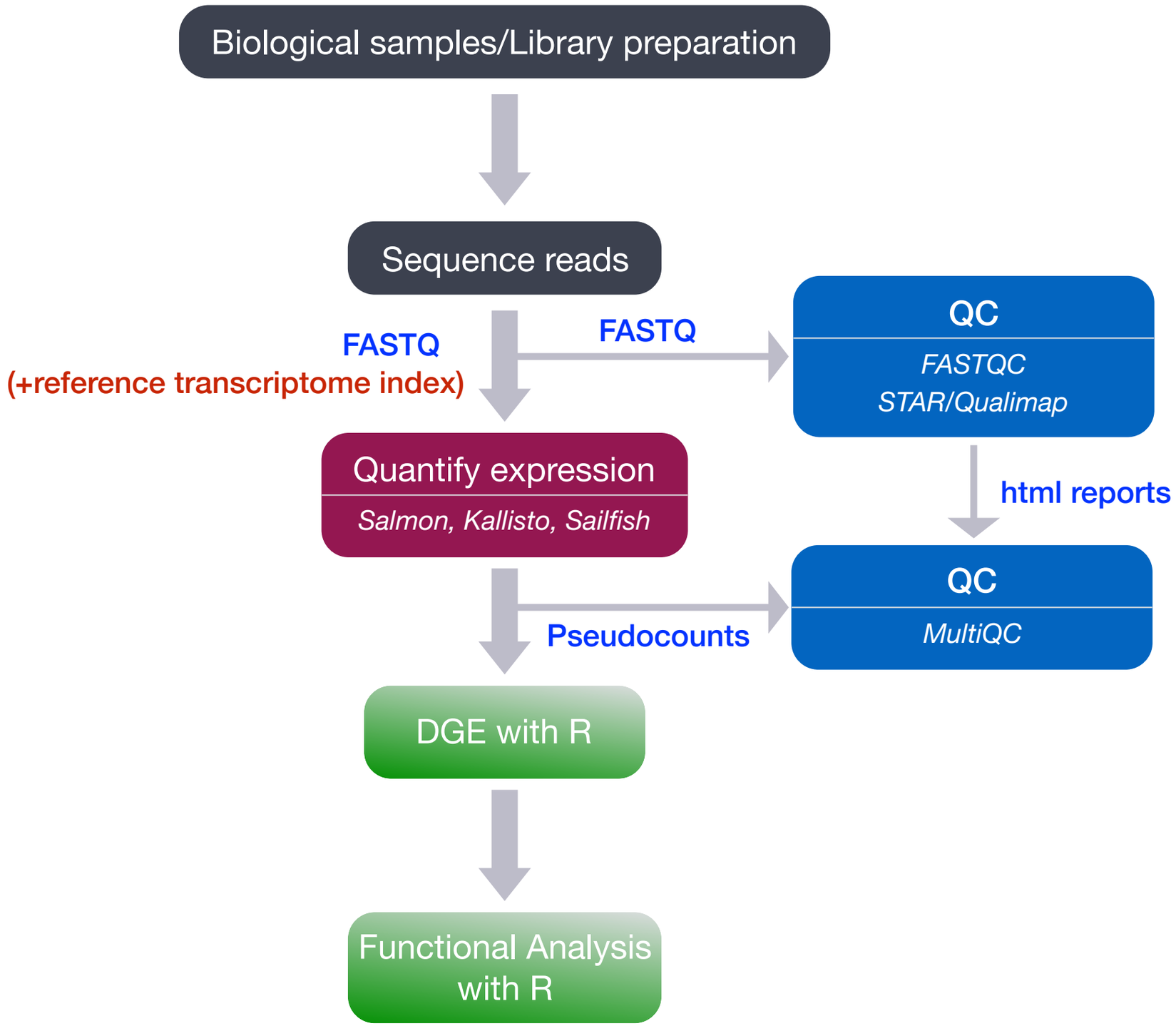
  Splice-aware alignment to the genome

- Genome alignment outputs a SAM/BAM file

SAM/BAM file format

- Sequence Alignment Map (SAM) format contains information on a per-read basis:
  - -- Coordinates of alignment, including strand
  - -- Mismatches
  - -- Mapping information (unique?, properly paired?, etc.)
  - -- Quality of mapping (tool-specific scoring systems)
- BAM: Binary version of SAM alignment format files

More information about SAM/BAM


Biological samples/Library preparation
Sequence reads
FASTQ
(+reference transcriptome index)
FASTQ
QC
FASTQC
STAR/Qualimap
html reports
Quantify expression
Salmon, Kallisto, Sailfish
QC
MultiQC
Pseudocounts
DGE with R
Functional Analysis with R