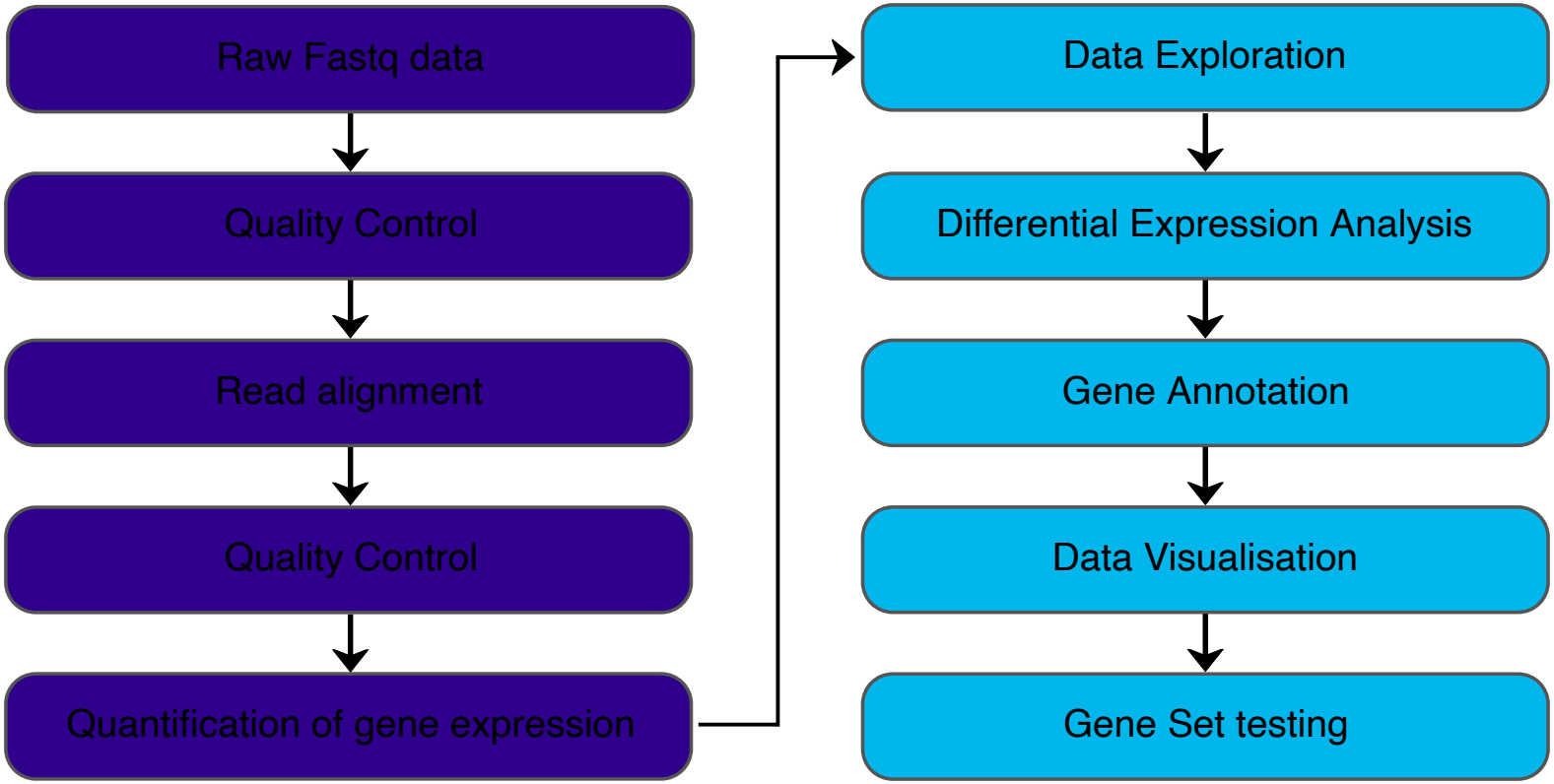
CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

QC of Aligned Reads

March 2021

Differential Gene Expression Analysis Workflow



QC of aligned reads

- Alignment Rate
- Duplication Rate
- Insert Size
- Genomic location of reads
- Transcript coverage

QC of aligned reads - Alignment Rate

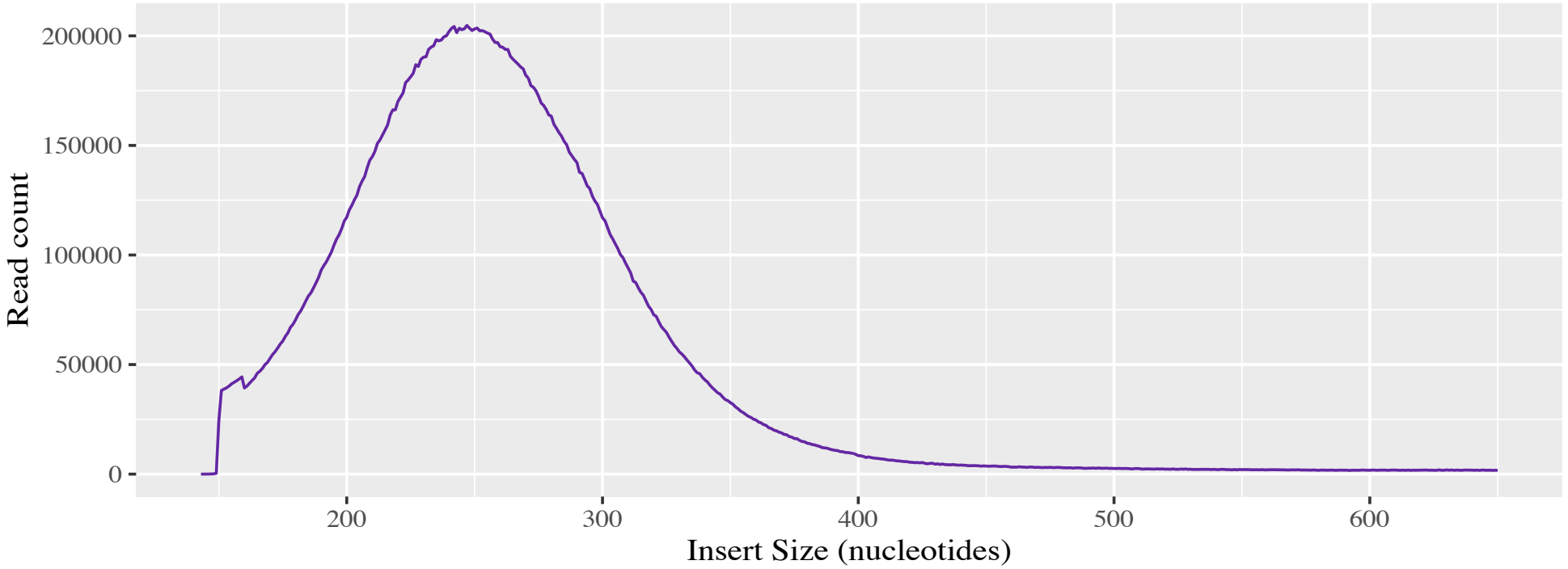
- Depends on:
  - Quality of Reference Genome
  - Quality of library prep and sequencing
  - For human and mouse > 95%

QC of aligned reads - Duplication Rate

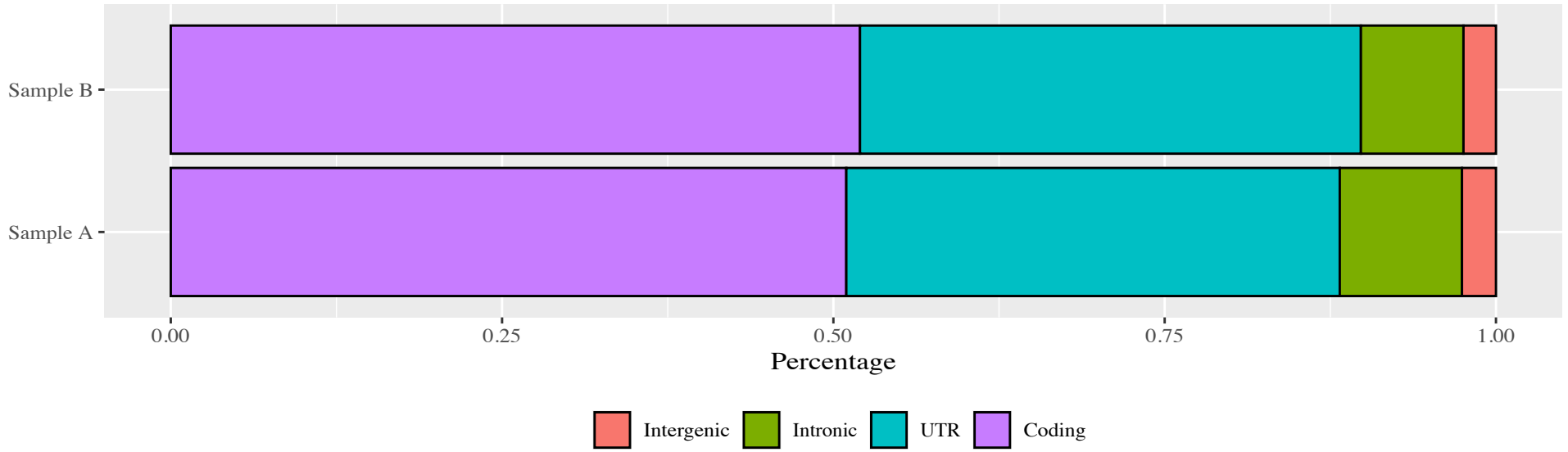
- Human exome is ~30 Mb therefore there are < 30 million possible reads
- Duplication rates in RNAseq can be > 40%

QC of aligned reads - Insert Size

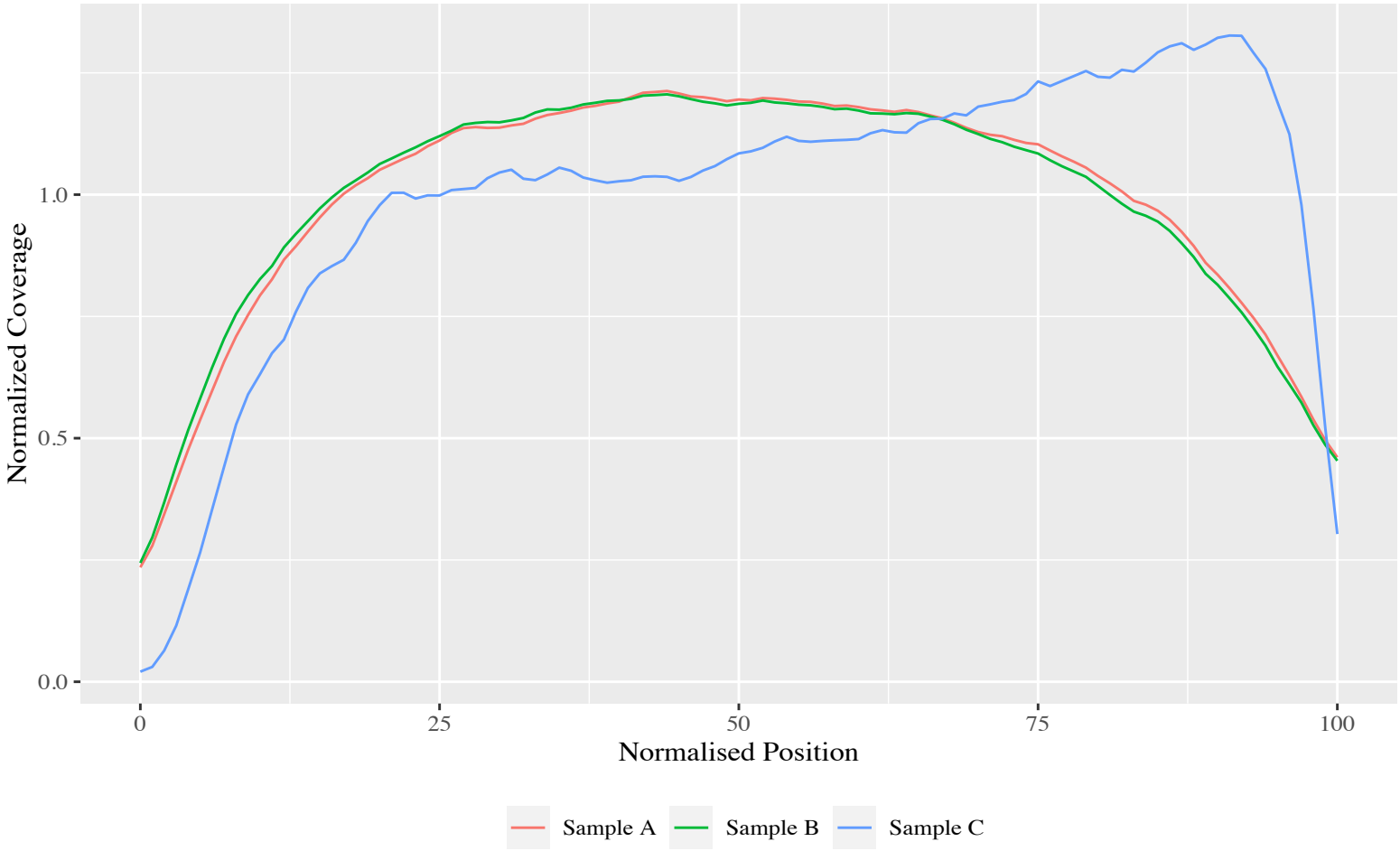
- Insert size is the length of the fragment of mRNA from which the reads are derived



QC of aligned reads - Genomic location of reads



QC of aligned reads - Transcript coverage



QC Goals

- Ensure the experiment generated the expected data
- Check is the sequencing depth and alignment rates are similar across samples
- Identify poor alignment parameters (sample quality, library prep ?)
- Discover contamination from another organism or from DNA
- Identify biases present in the data