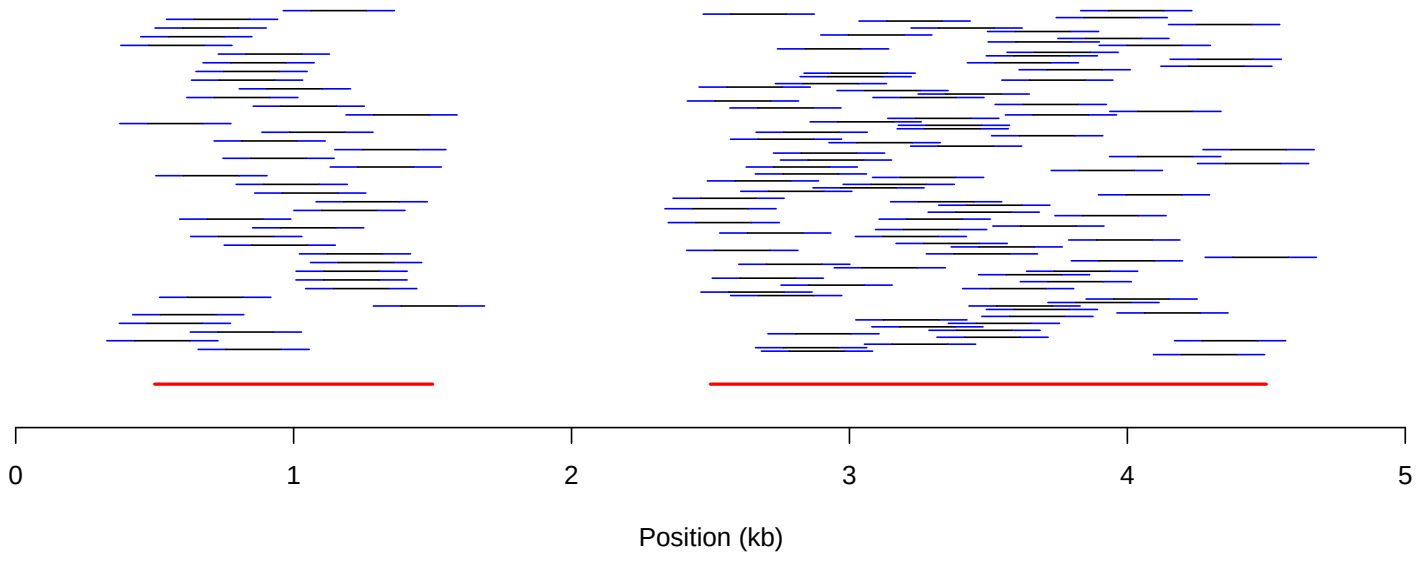
RNA-Seq and differential expression analysis

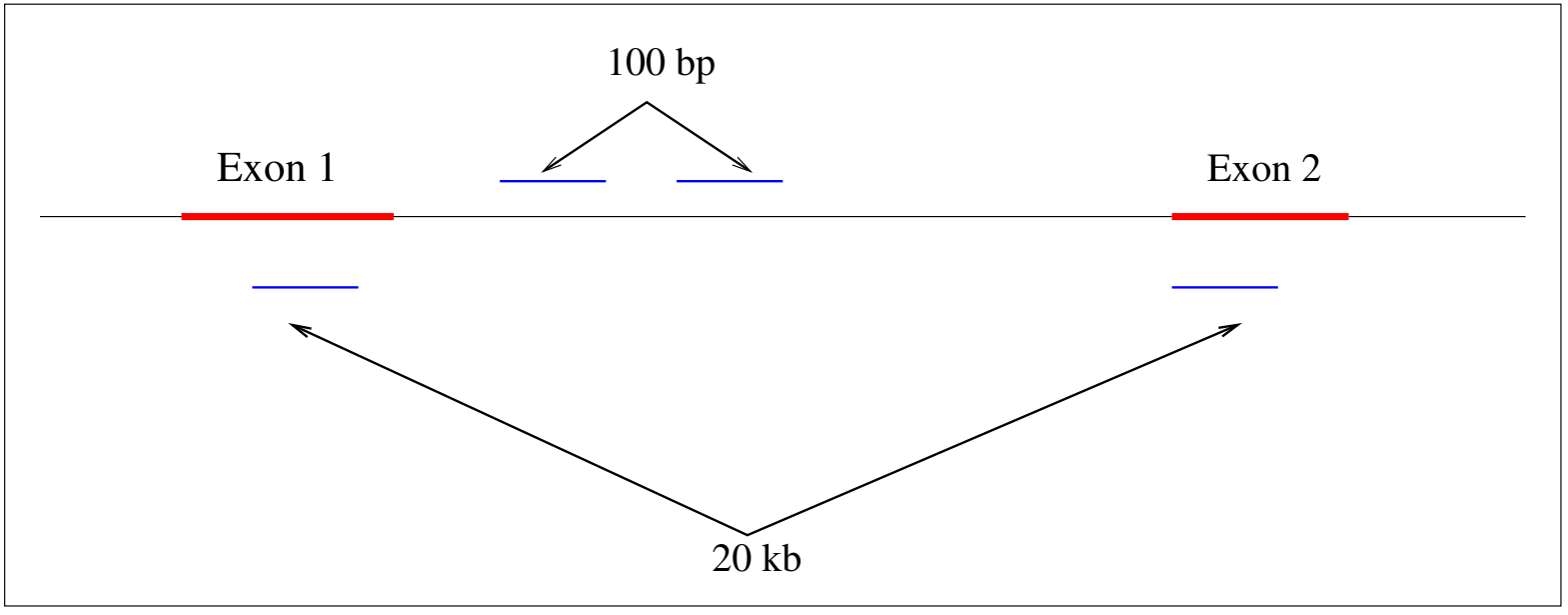
Vincent Plagnol

UCL

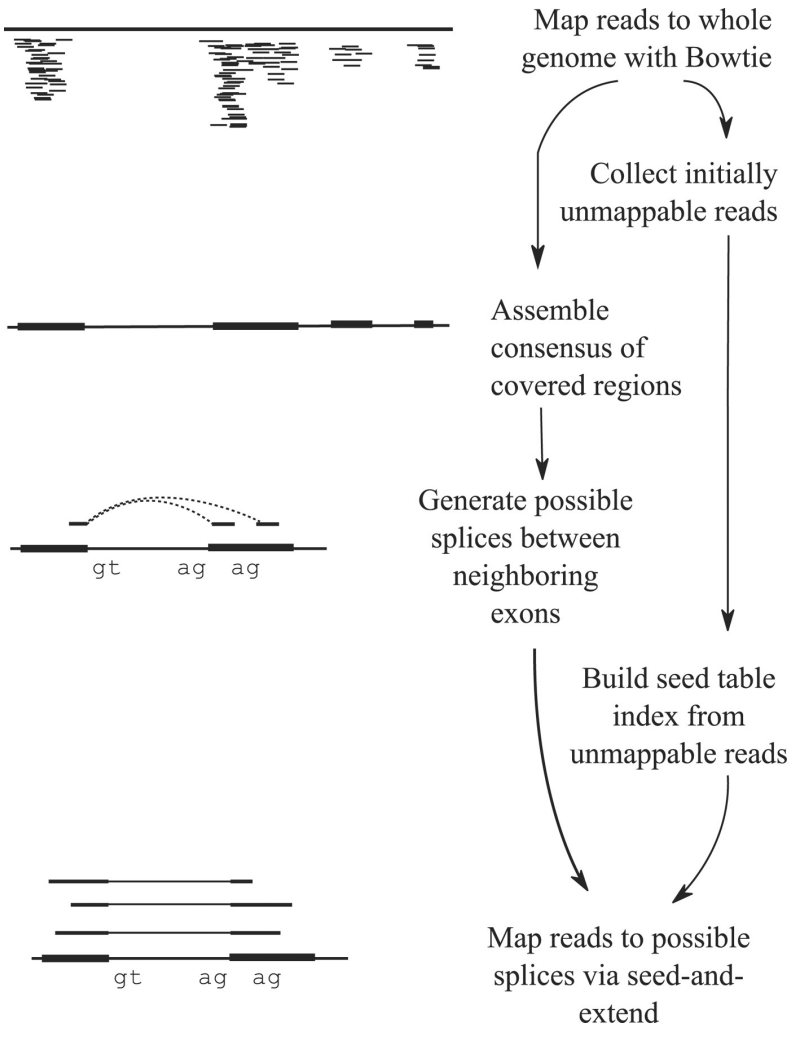
Concept of RPKM



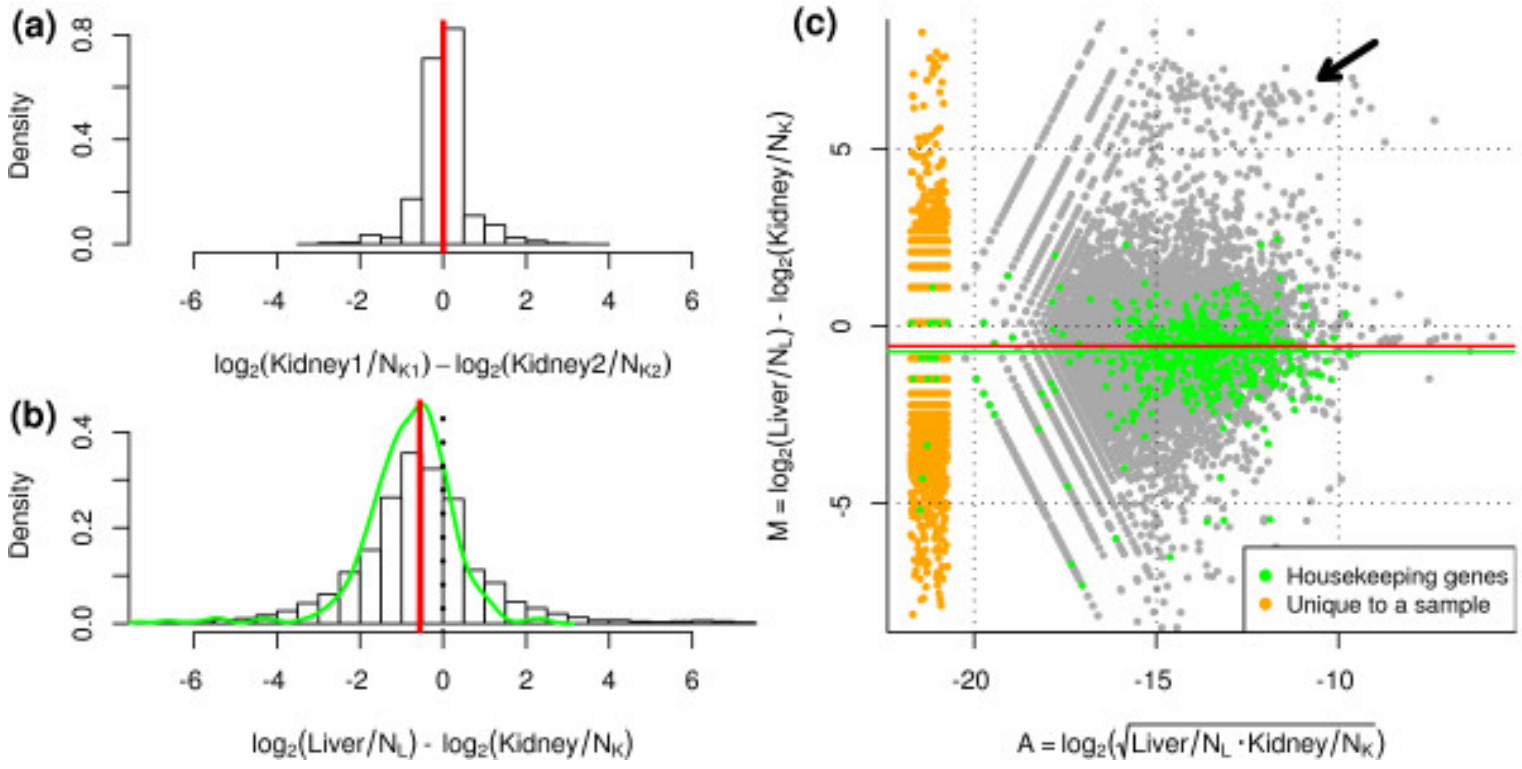
Alignment is a much bigger challenge



Tophat mapping strategy



Normalization is a more difficult problem than it seems



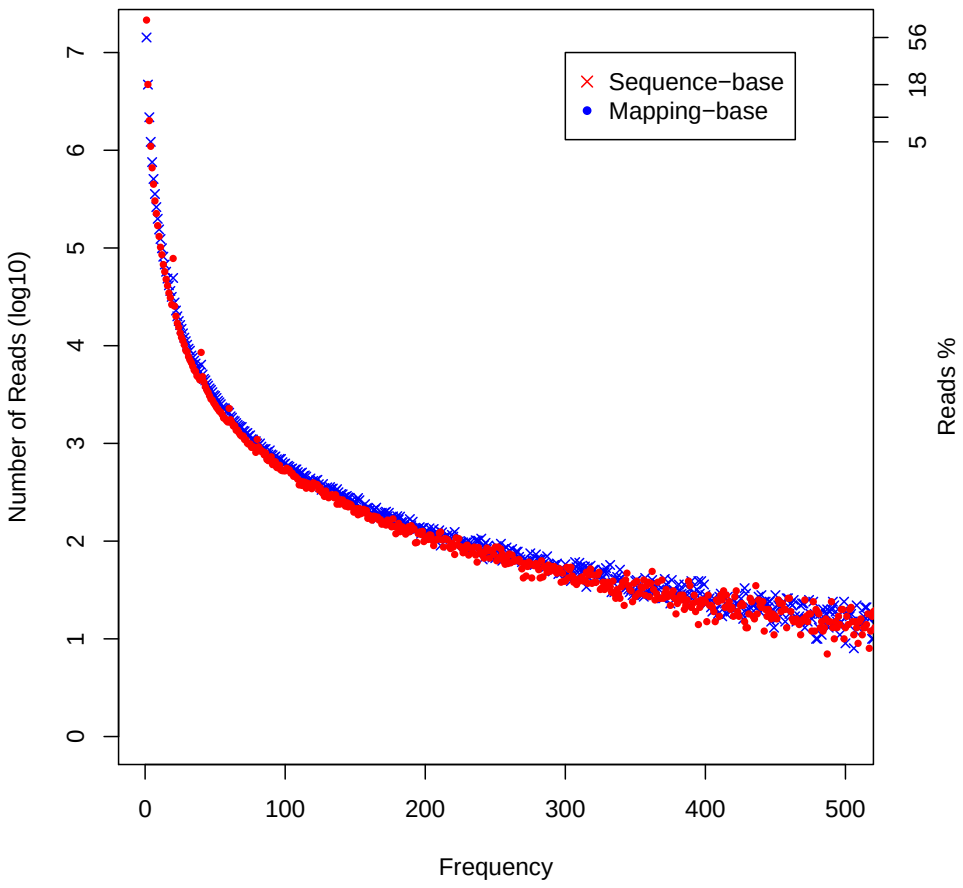
The black arrow highlights the set of prominent genes that are largely attributable for the overall bias in log-fold-changes.

Robinson, Oshlack, Genome Biology 2010

RNA-Seq is fiddly

- Many more things can go wrong.
- And what goes wrong can easily create subtle biases that look like case control differences.
- I very strongly recommend:
  - Careful inspection of PCA plots to identify any technical issue.
  - Using tools like PEER or SVA-Seq that help mitigate these technical issues.

Should I remove duplicates in RNA-Seq data?



RNA-Seq workflow

- What you really need is:
  - A choice of aligner (tophat, STAR...)
  - A choice of differential expression analysis tool (edgeR, DESeq, DESeq2)
- I use tophat and DESeq2 but this is slightly old school.
- STAR, or the new tools like Callisto, are probably better alternatives.
  - They are, however, RAM hungry.

An overview of DEXSeq output