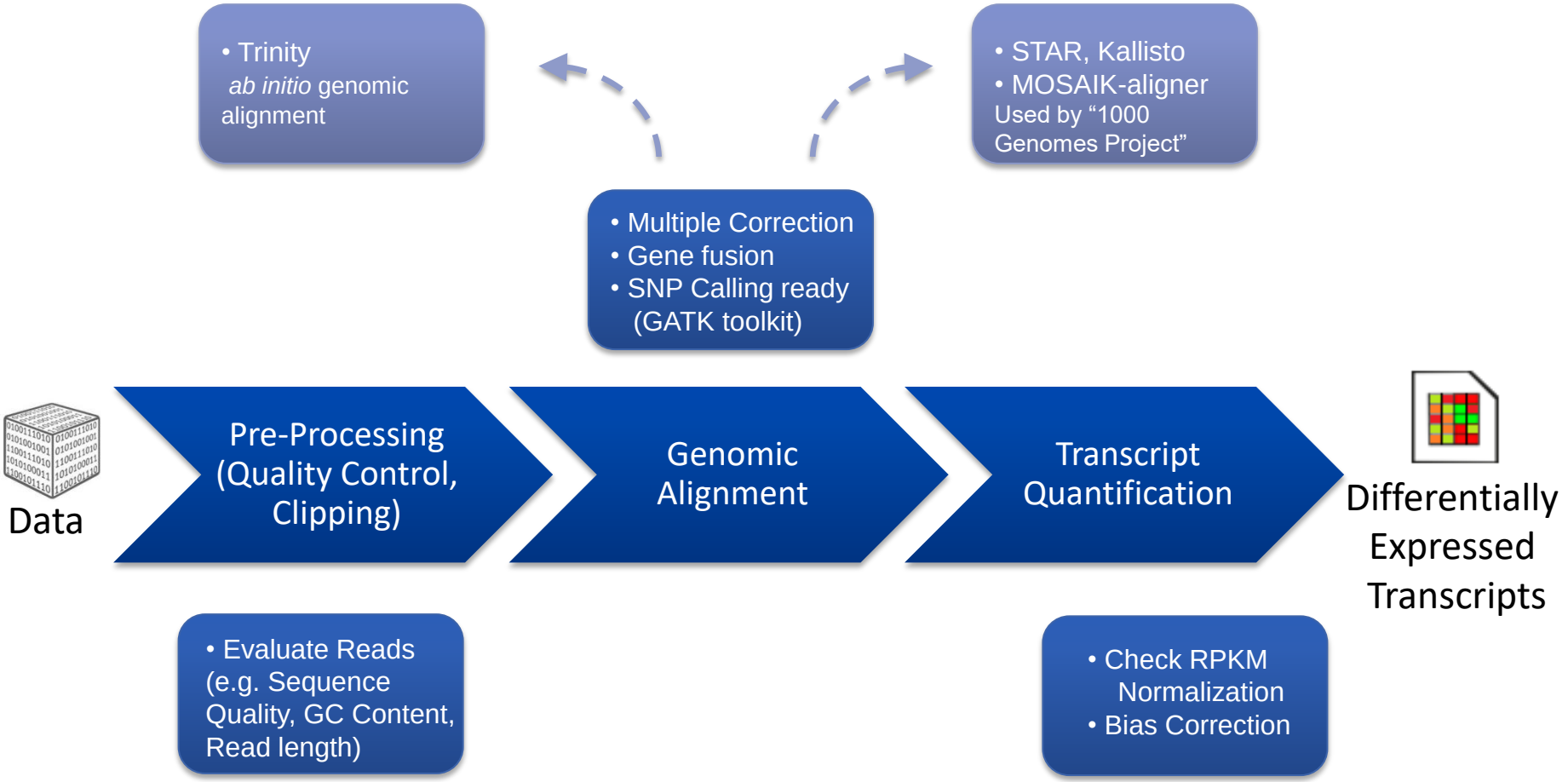
Application and evaluation of different read mapping approaches for genomic alignment

Markus Wolfien and Andrea Bagnacani

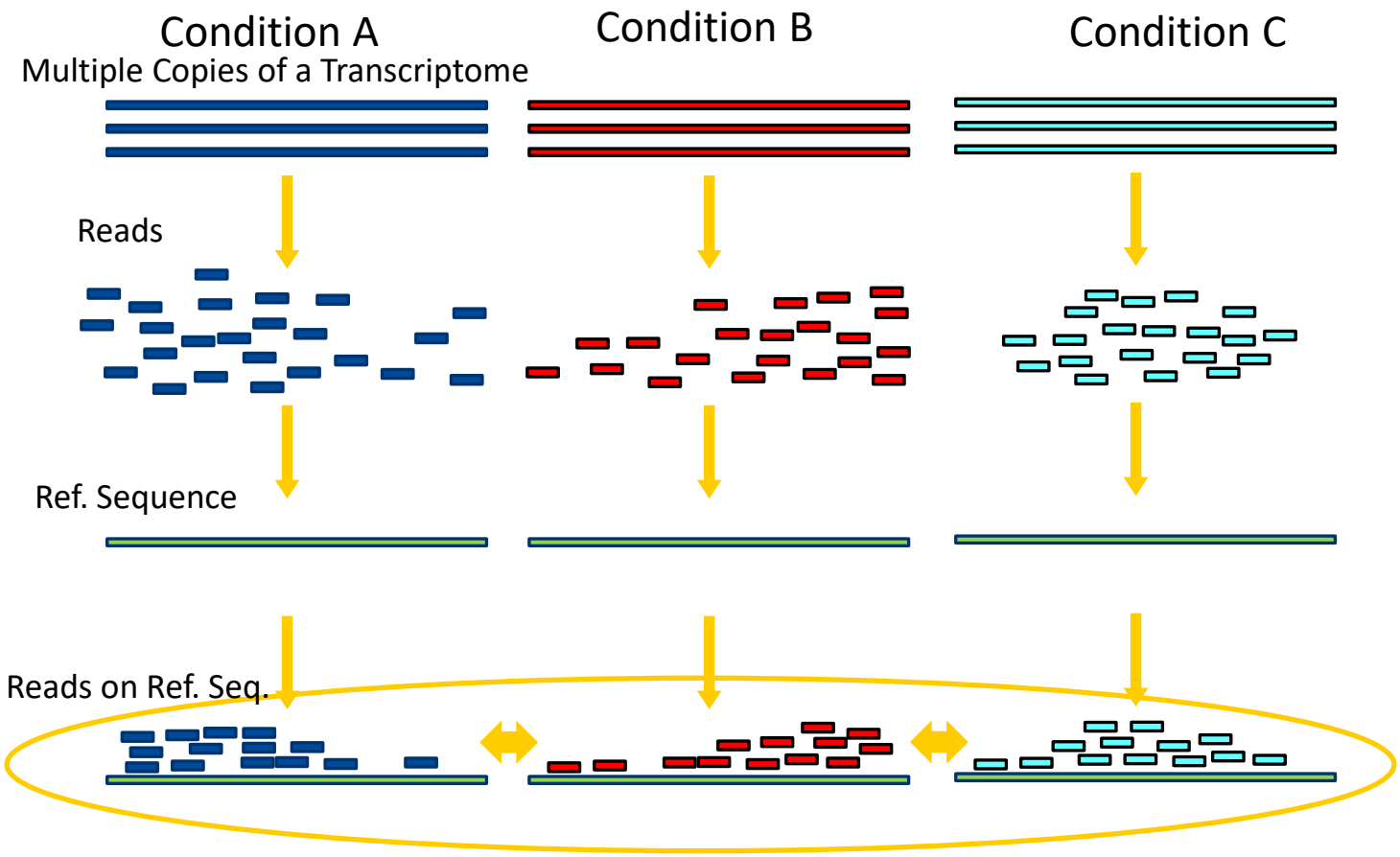
de.NBI Training – 28th June 2018 Jena

www.sbi.uni-rostock.de

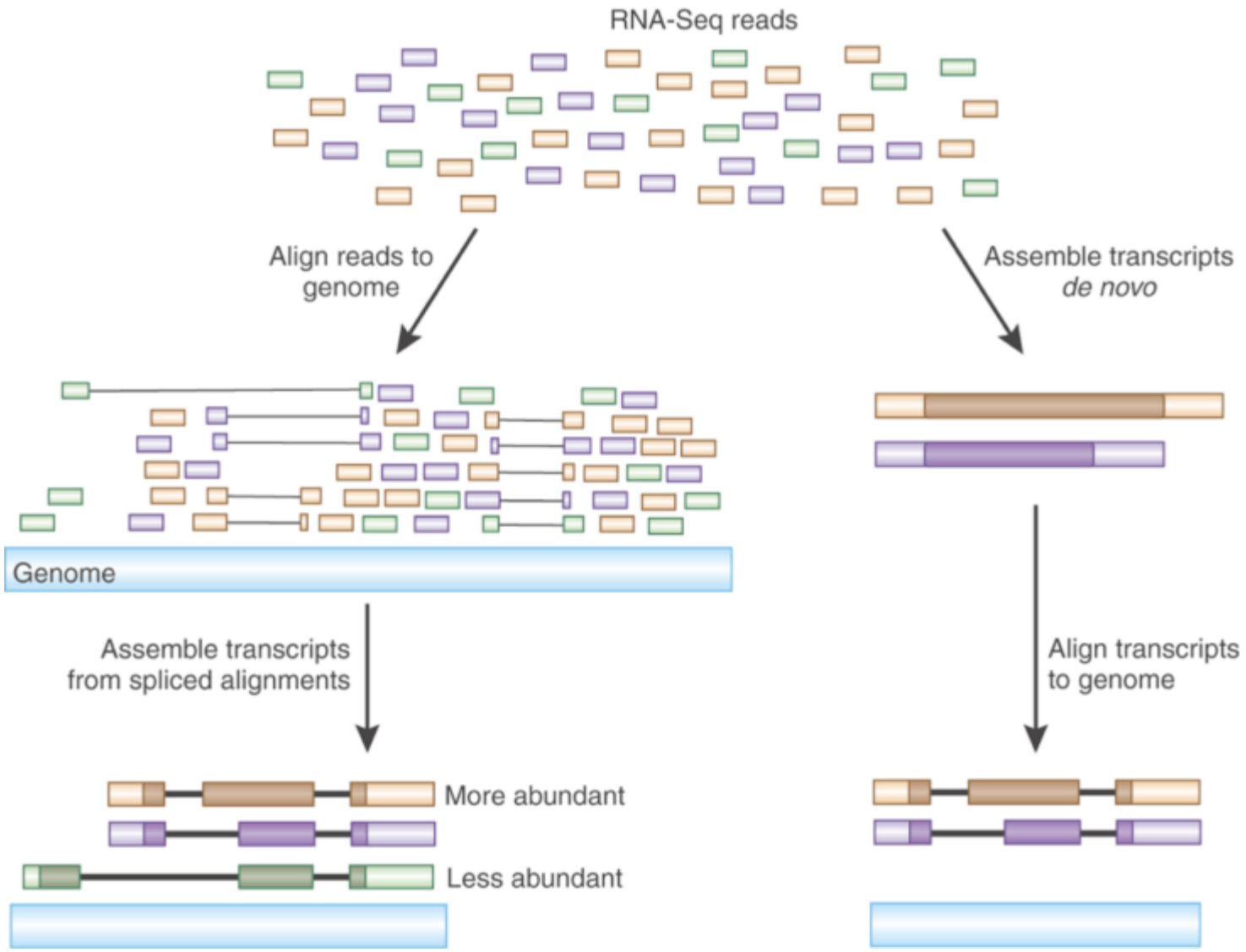
One possible workflow for data processing



One possible workflow for data processing

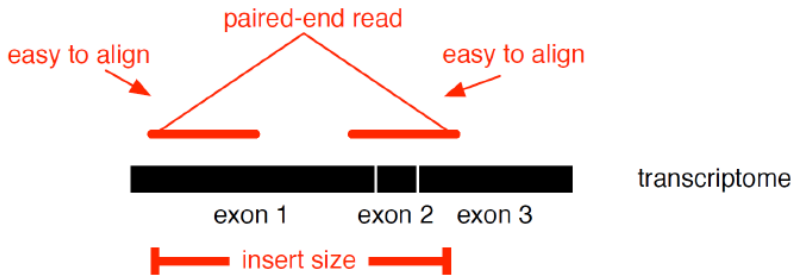


What is mapping?



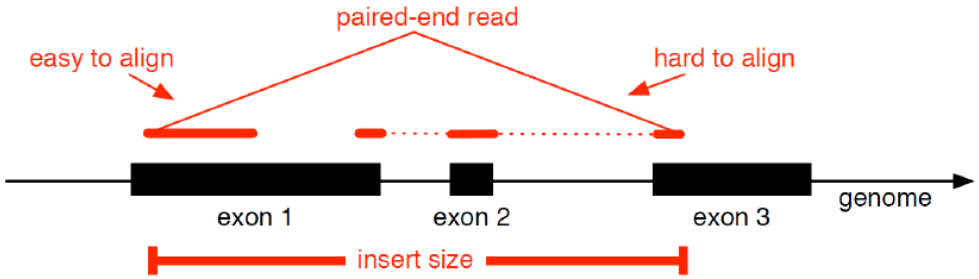
1. Genomic alignment – Splice aware

Transcriptome alignment



- reliable gene models required
- no detection of novel genes

Genome alignment (splice-aware read alignment)

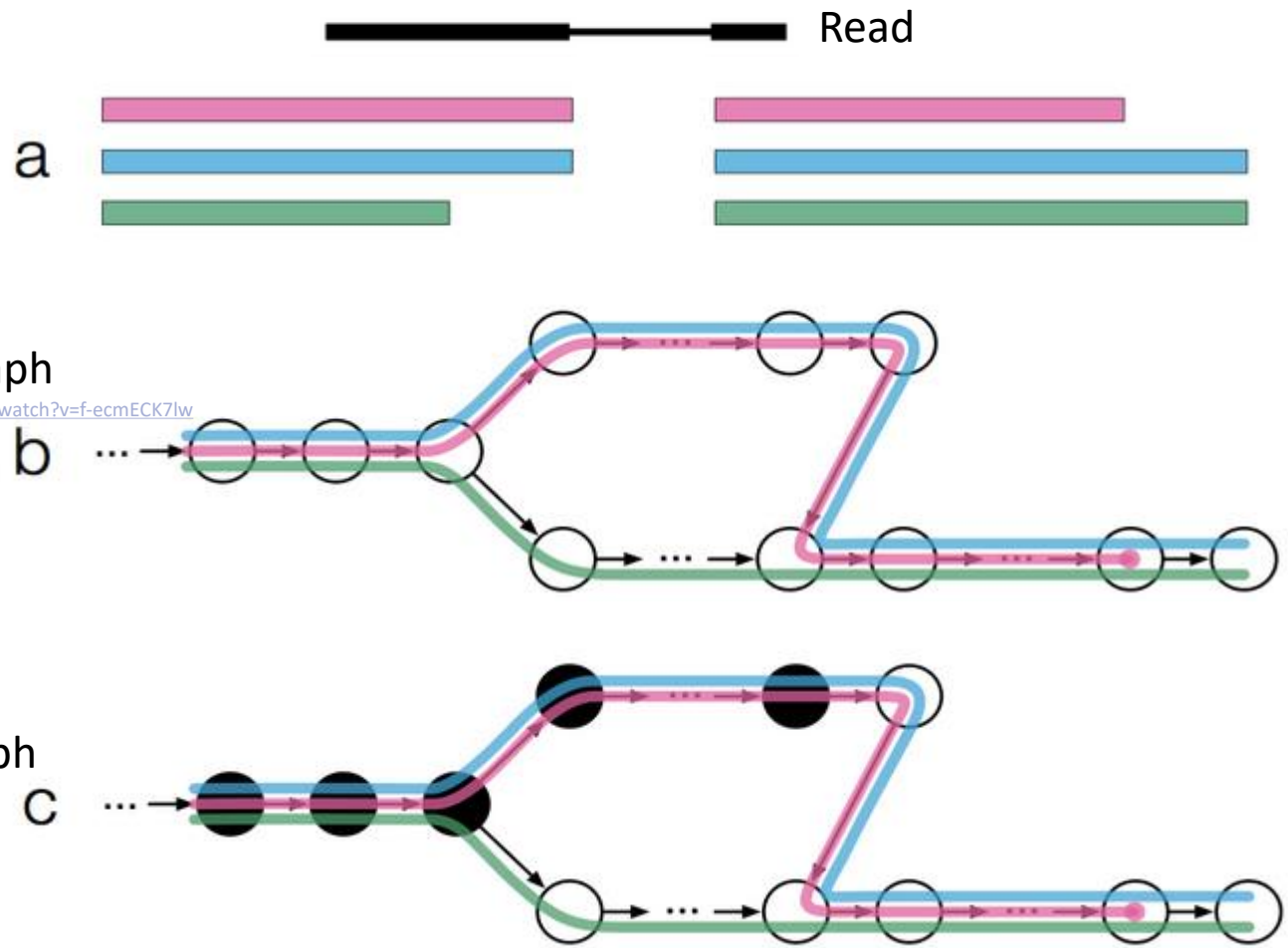


+ detection of novel genes and isoforms

Turro, EMBO, 2012

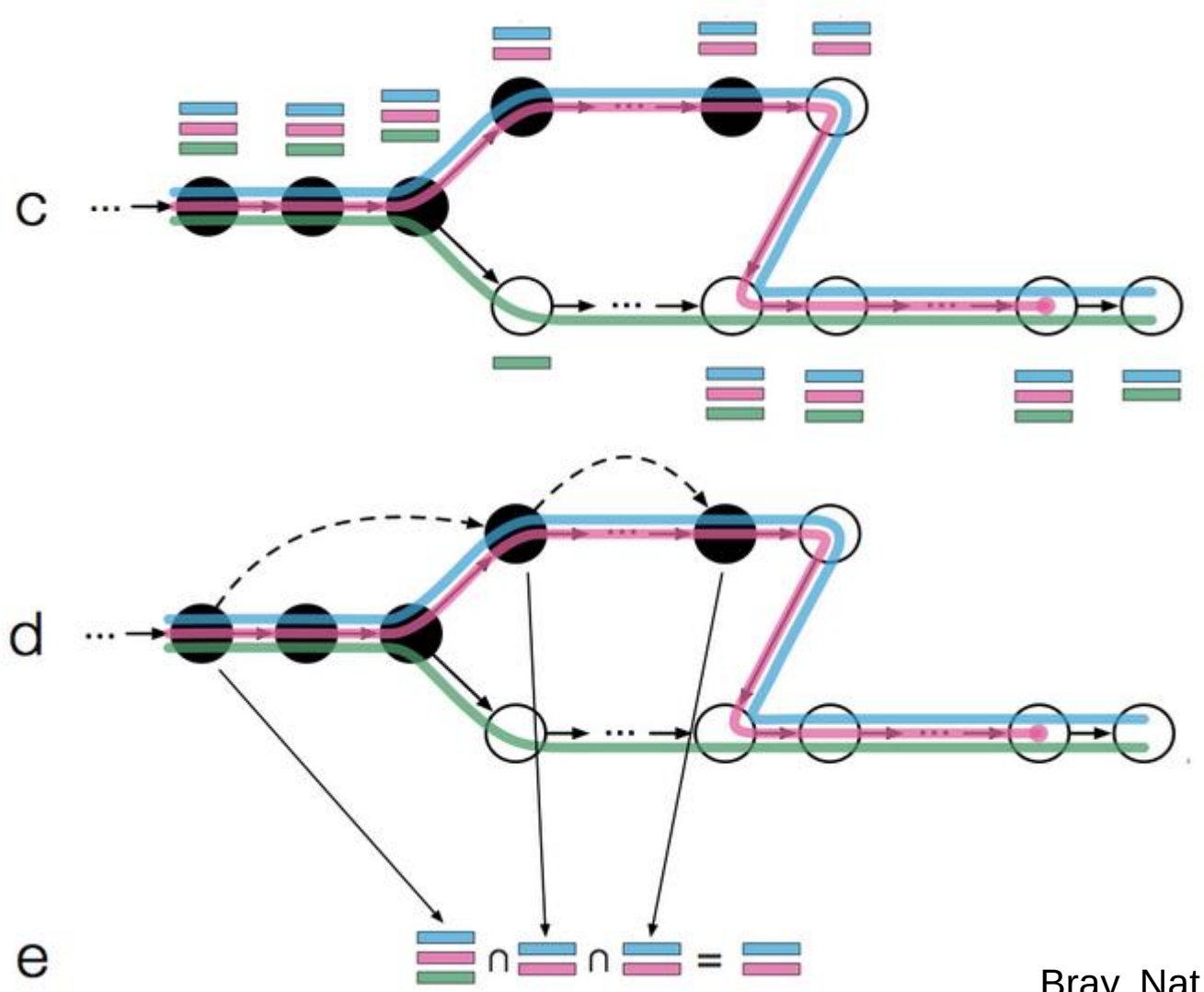
For clinical usage combination of different algorithms possible

2. Genomic alignment - pseudoalignment



Bray, Nat Biotech, 2016

2. Genomic alignment - pseudoalignment



Bray, Nat Biotech, 2016

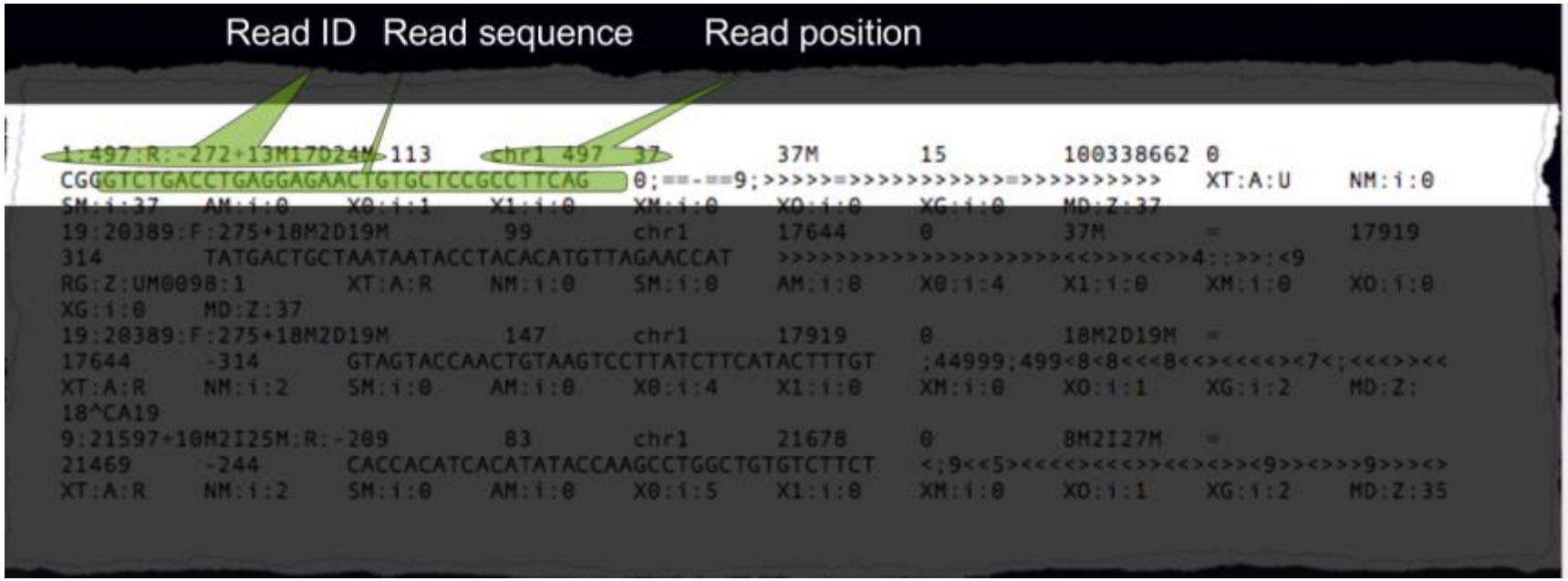
Transcript quantification

- Read counts
  - Count the reads per feature
    - relatively easy: count the number of reads per gene, exon, …
  - How to handle multi-mapping reads (i.e. reads with multiple alignments)?

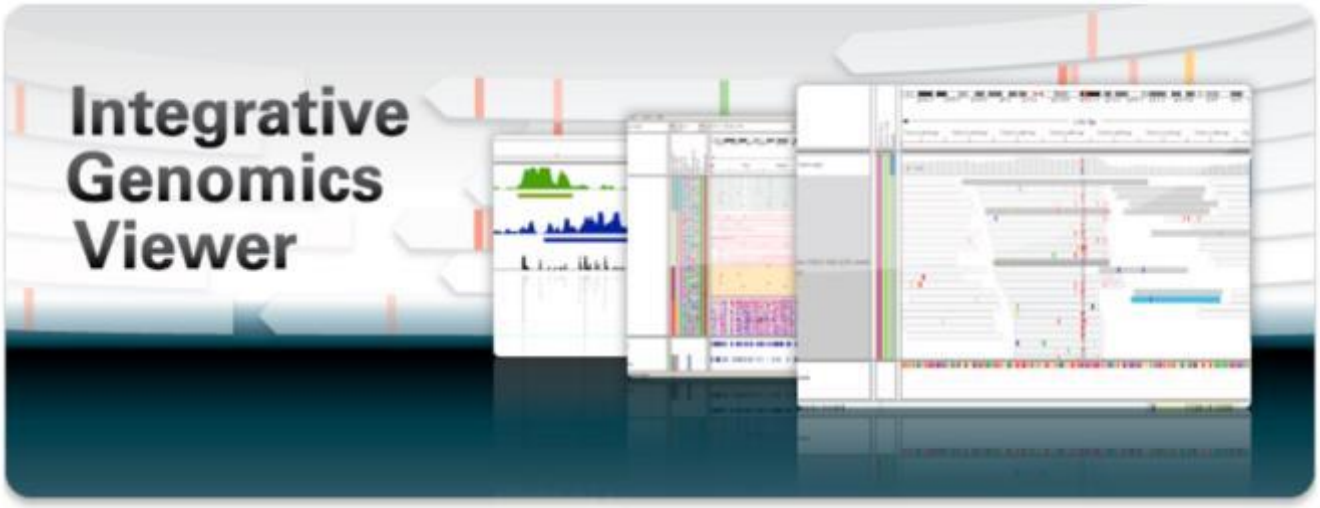
- Normalization - aims to make expression levels comparable across:
  - Features (genes, isoforms, ...)
  - RNA libraries (samples)

- Normalization methods:
  - TPM, RPKM / FPKM (Cufflinks /Cuffdiff) (Mortazavi, Nat Meth, 2008)
  - TMM (edgeR) (Robinson & Oshlack, Genome Biol, 2010)
  - DESeq2 (DESeq2) (Love et al., Genome Biol, 2014)

SAM/BAM file format



Visualization of .bam files



http://software.broadinstitute.org/software/igv/

The James Hutton Institute

Tablet

https://ics.hutton.ac.uk/tablet/

Hands on part 3 10:15 – 11:15

“Application of different read mapping approaches for genomic alignment“

Material: http://galaxyproject.github.io/training-material/topics/sequence-analysis/