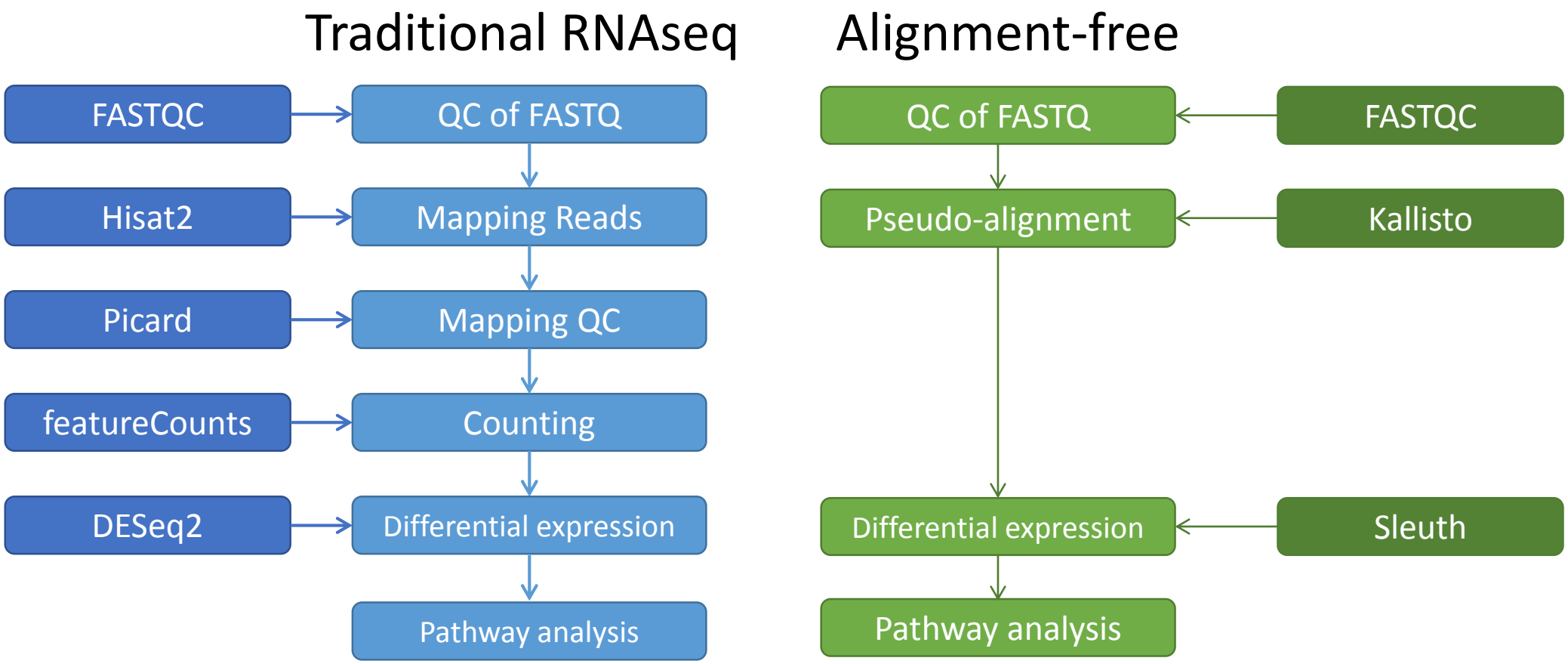
Alignment-free RNAseq

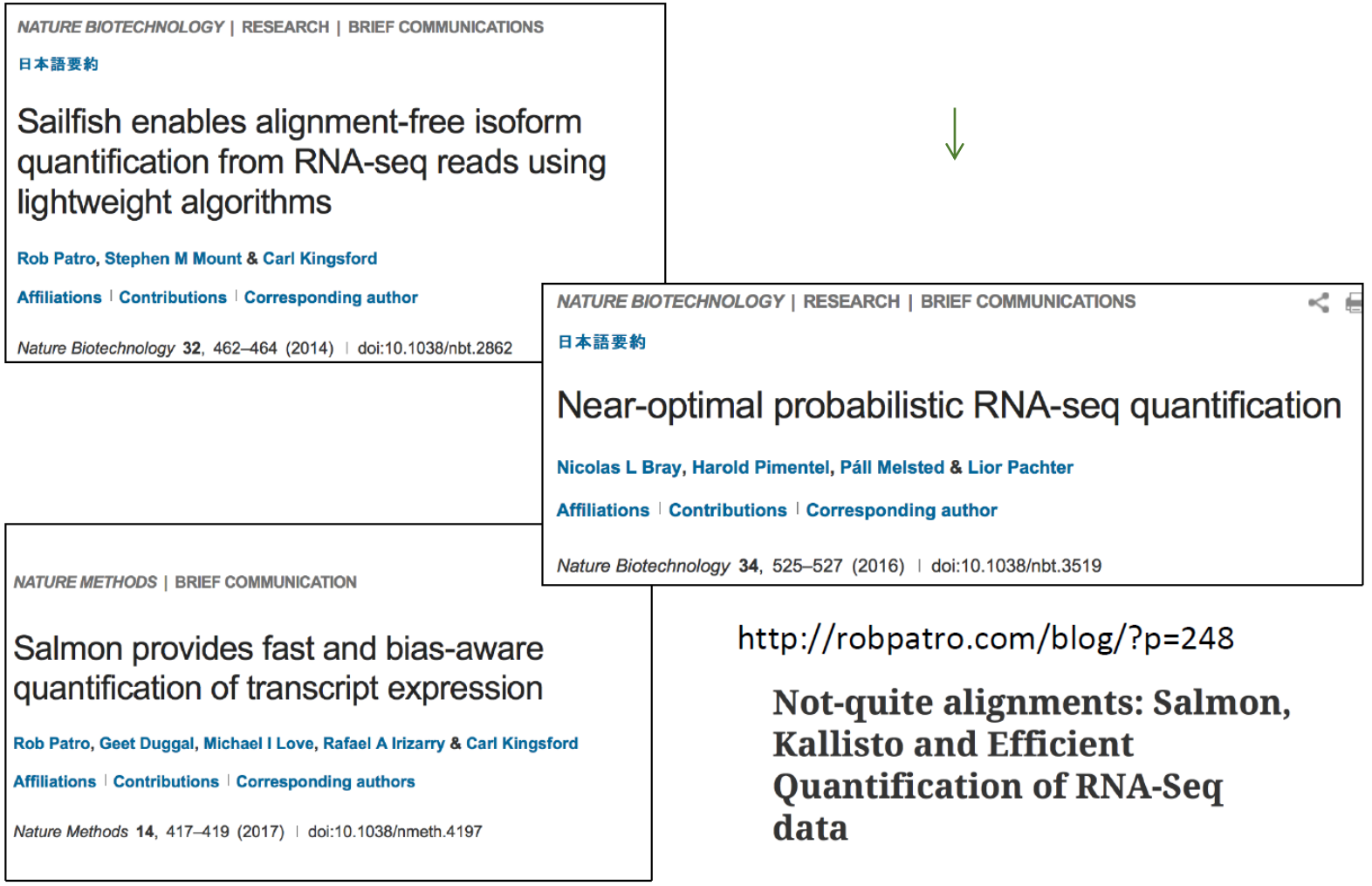
Oxford Biomedical Data Science Training Programme


Traditional RNAseq
FASTQC
QC of FASTQ
Hisat2
Mapping Reads
Picard
Mapping QC
featureCounts
Counting
DESeq2
Differential expression
Pathway analysis
Alignment-free
QC of FASTQ
FASTQC
Pseudo-alignment
Kallisto
Differential expression
Sleuth
Pathway analysis

Alignment-free Quantification

- Also known as Pseudoalignment
- Work directly from Fastq
- Specify **which** transcript the read is from **NOT** where in sequence it is!
- No base level information
- Transcript level quantification
- Output is a quantification table for each sample -> quant.sf
  - Rows = transcripts
  - Quantification Values = Transcripts Per Million
- Very fast – can run on laptop
- Better accuracy then traditional method
- Can only quantify known transcripts
- Does not produce a BAM file – no visualisation

Alignment-free Quantification Tools



NATURE BIOTECHNOLOGY | RESEARCH | BRIEF COMMUNICATIONS

日本語要約

Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms

Rob Patro, Stephen M Mount & Carl Kingsford

Affiliations | Contributions | Corresponding author

Nature Biotechnology 32, 462–464 (2014) | doi:10.1038/nbt.2862

NATURE BIOTECHNOLOGY | RESEARCH | BRIEF COMMUNICATIONS

日本語要約

Near-optimal probabilistic RNA-seq quantification

Nicolas L Bray, Harold Pimentel, Páll Melsted & Lior Pachter

Affiliations | Contributions | Corresponding author

Nature Biotechnology 34, 525–527 (2016) | doi:10.1038/nbt.3519

NATURE METHODS | BRIEF COMMUNICATION

Salmon provides fast and bias-aware quantification of transcript expression

Rob Patro, Geet Duggal, Michael I Love, Rafael A Irizarry & Carl Kingsford

Affiliations | Contributions | Corresponding authors

Nature Methods 14, 417–419 (2017) | doi:10.1038/nmeth.4197

http://robpatro.com/blog/?p=248

Not-quite alignments: Salmon, Kallisto and Efficient Quantification of RNA-Seq data

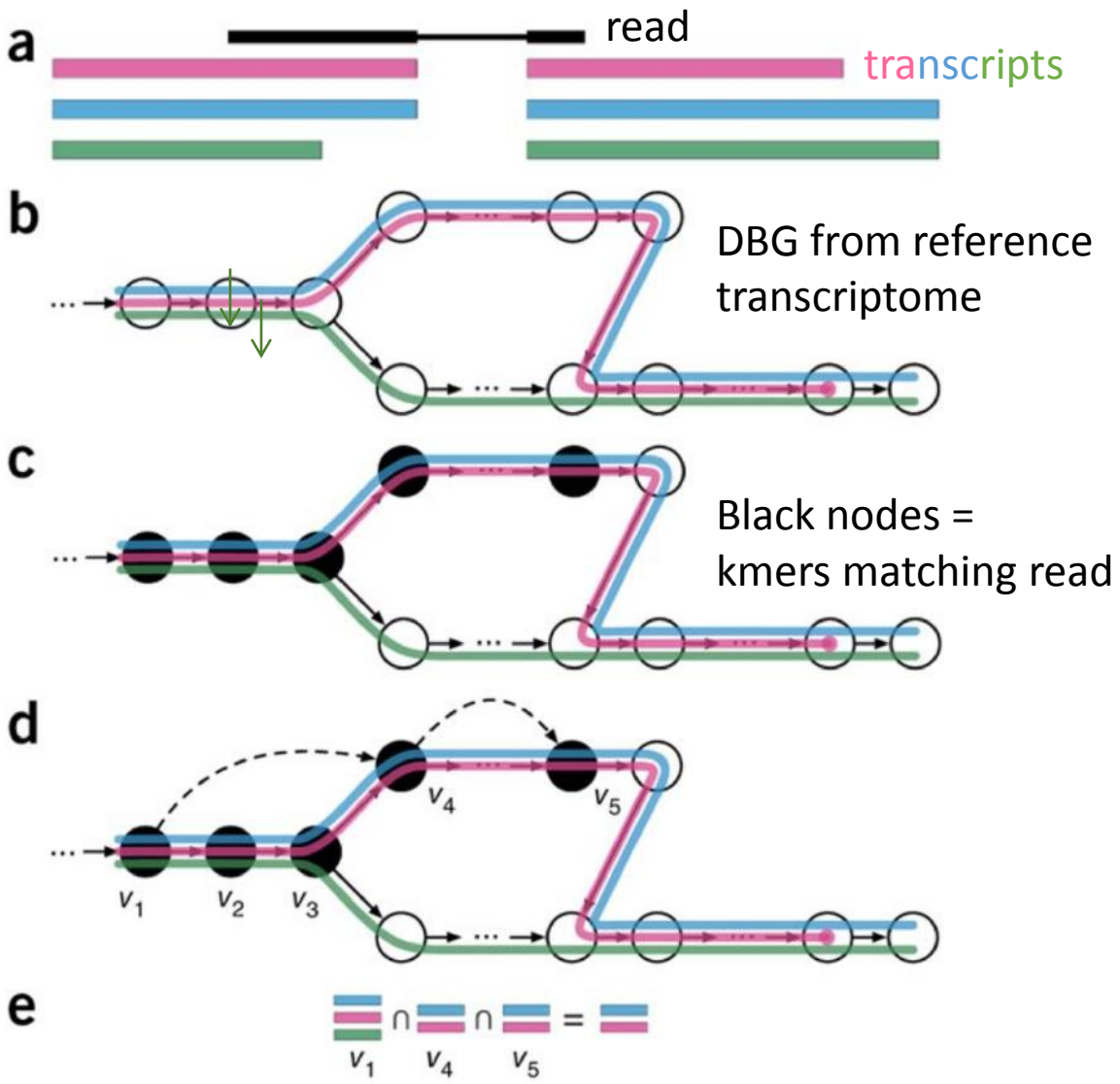
For each read find compatible transcripts

- Build a coloured De-Bruijn graph from all transcripts
  - Each circle/node is a Kmer
  - Each kmer is associated with set of transcripts
    - ‘kmer compatibility class’

- Decompose each read into kmers
  - Compare with graph

- Take the intersection of compatibility classes the read matches
  - V1 = match 3 transcripts
  - V4 = match 2 transcripts
  - Intersection = 2 transcripts

- Fast because it doesn’t not look up every kmer
  - skips to the node in the next equivalence class



Bootstrapping

1) Sample reads randomly with replacement
2) Estimate transcript abundances
3) Repeat 100 times

Output:

- Get a confidence interval for abundance of each transcript

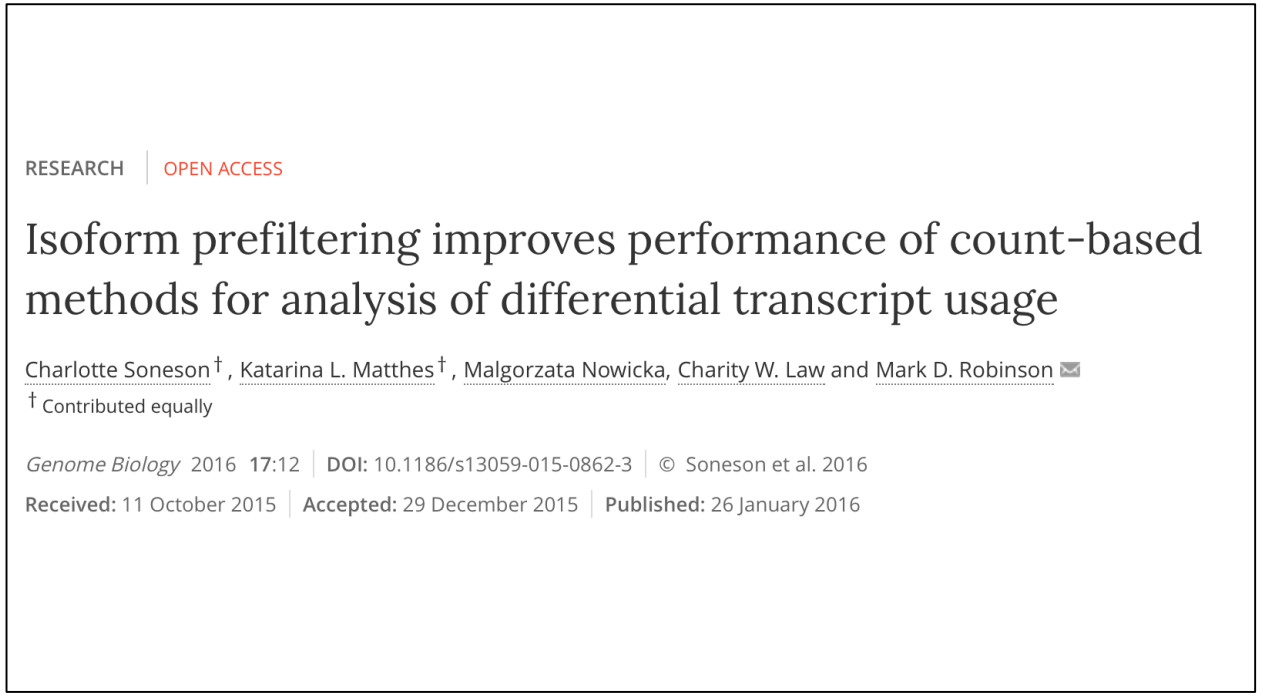
Purpose:

- Identify transcripts with good quality quantification estimates

ONLY POSSIBLE BECAUSE IT'S SO FAST

Gene Annotations

- The quality of the annotations you provide will effect the quality of the output
- What geneset are you using?
  - Ensembl
    - Manually curated
    - Wide range of species
  - Refseq
    - Better for small RNAs
    - Longer first exons



RESEARCH | OPEN ACCESS

Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage

Charlotte Soneson†, Katarina L. Matthes†, Malgorzata Nowicka, Charity W. Law and Mark D. Robinson

† Contributed equally

Genome Biology 2016 **17**:12 | **DOI:** 10.1186/s13059-015-0862-3 | © Soneson et al. 2016

Received: 11 October 2015 | **Accepted:** 29 December 2015 | **Published:** 26 January 2016

Differential Expression Analysis

- Output = TPM counts table
  - need to convert to use with tools requiring raw counts (DESeq2/edgeR)
  - R txiimport package
- Sleuth is a differential expression program built to interface directly with pseudoalignments
- Transcript level differential expression is hard
  - recommend quantifying at gene level

Keeping up to date

- Lior Pactor - https://liorpachter.wordpress.com
- Rob Patro - http://robpatro.com/blog/
- RNASeq blog - http://www.rna-seqblog.com/

- Other people to look out for/keep an eye on:
  - Mike Love, Charlotte Soneson, Wolfgang Huber, Simon Anders, Jeff Leek

- For a book that explains the RNA-Seq workflow and considerations from library prep up to differential expression very well see:

  RNA-seq Data Analysis: A Practical Approach

  Eija Korpelainen, Jarno Tuimala, Panu Somervuo, Mikael Huss, Garry Wong