Next generation sequencing

RNA-seq

Stig Uggerhøj Andersen, PhD

Department of Molecular Biology

University of Aarhus

Lecture overview

- Applications and data scale
- Sample preparation
- Strand specificity
- Single-cell sequencing
- Spliced alignments
- Quantification and normalization
- Long read RNA-seq

Applications and data scale

Counting or Profiling

- **10 million total reads** of **50 bp** length from poly-A selected RNA will give performance better than any microarray

Studying Alternative Splicing or quantifying cSNPs for most transcripts

- Deeper profiling of **50 to 100 million reads**, with read lengths of **50 to 100 bps**, from poly-A selected RNA using mRNA-Seq assay

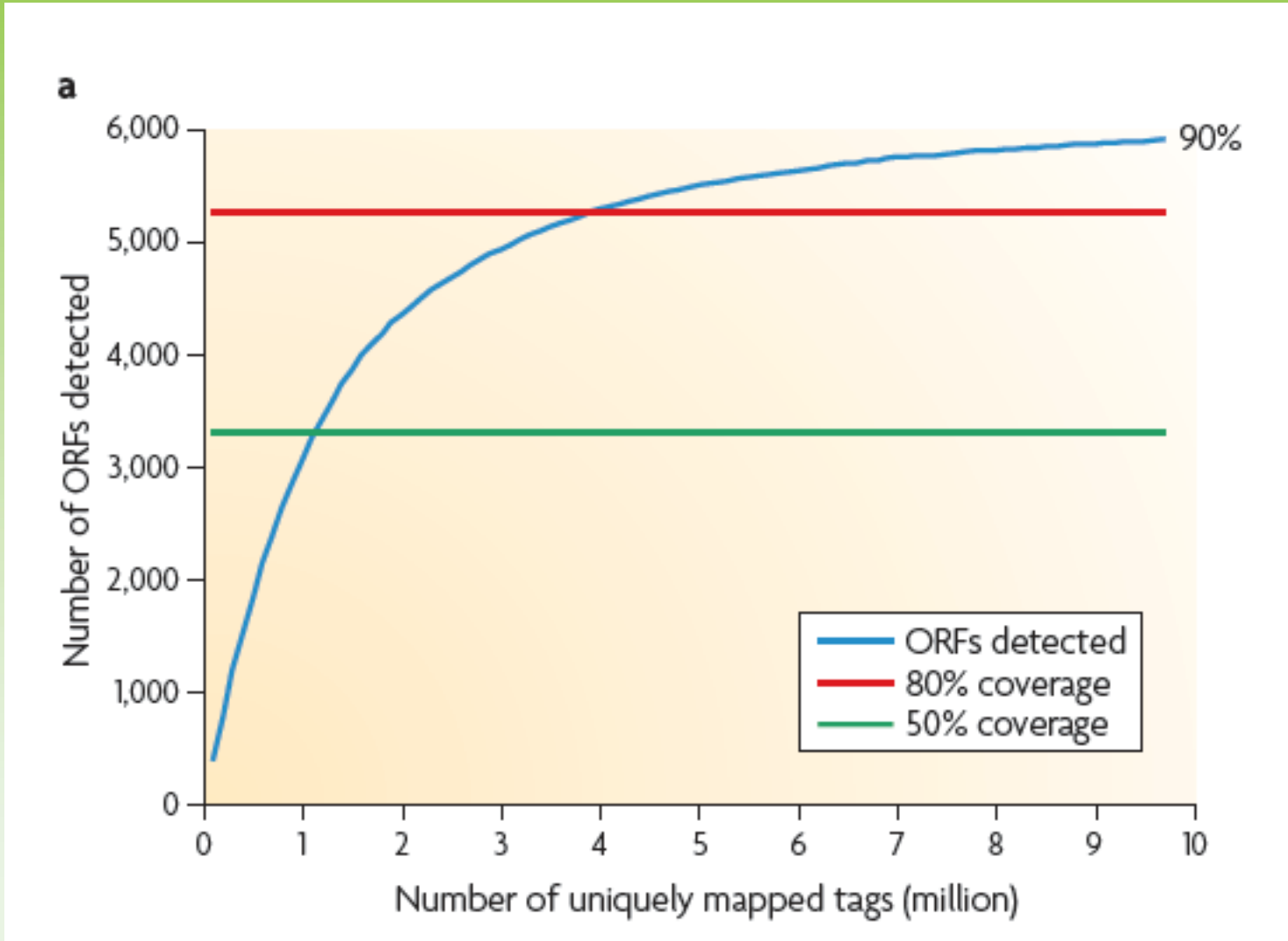
Complete Annotation of an entirely New Transcriptome

- ~500 Million reads of 100 bp read length from multiple tissues
- Normalized stranded mRNA-Seq
- Normalized stranded Total RNA-Seq for looking at ncRNAs
- Small RNA-Seq for microRNAs

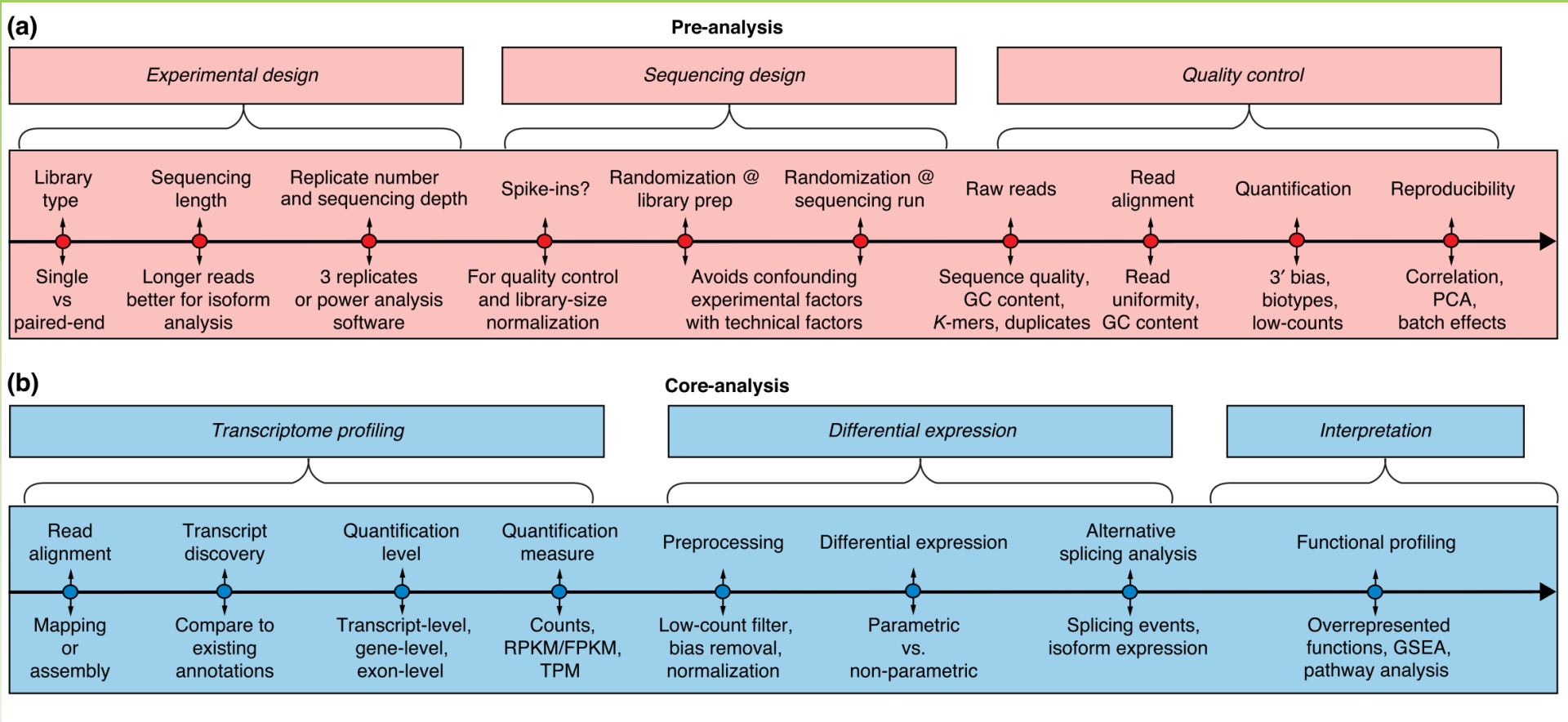
Consider Hifi reads for annotation of a new transcriptome

You should profile ~15 million full length transcripts at average 2kb length

Applications and data scale



RNA-seq overview



PMID: 26813401

Experimental design

- Beware of batch effects
  - circadian rhythm (time of day)
  - sampling method
  - sequencing strategy (adapters, lanes)
- Record all information and check for co-variation
- Three replicates to keep statistical analysis options open

Sample preparation - isolating mRNA

Classes of RNA Molecules in Human Cells

- **Ribosomal RNA - rRNA**
  - 28 S
  - 18 S
  - 5.8 S
- **Non-coding RNA - ncRNA**
  - tRNA
  - snoRNA
  - lincRNA
  - miRNA
  - Many, many others…
- **Mitochondrial RNA - mtRNA**
- **Messenger RNA - mRNA**
  - Highly expressed transcripts (>10,000 copies per cell)
  - Rarely expressed transcripts (~1 copy per cell)

Sample preparation

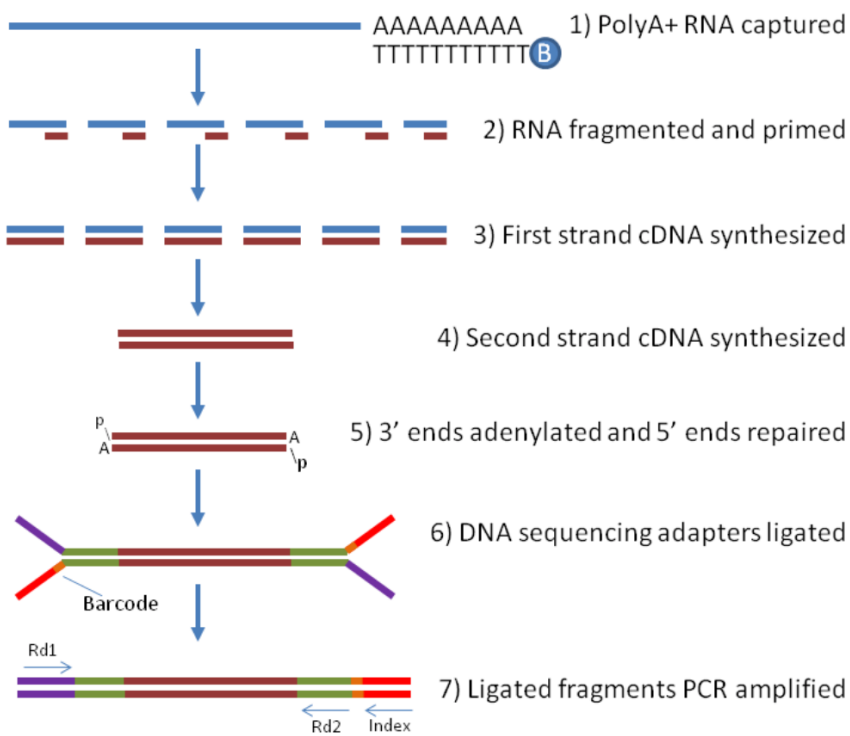
Enrichment of RNA or cDNA samples by selection or counterselection

Hybridization based

- oligo-dT
  - *selects pA+ RNA*
- Ribominus (Plus)
  - *counterselection of ribosomal RNAs (Plus includes mitochondrial rRNAs)*
- Capture arrays/bead capture
  - *for selection of an array of 'hand-picked' targets*
- Custom-made selection or counterselection strategies
  - *e.g. through hybridization to biotinylated RNA or DNA molecules*
- DSN-normalization
  - *elimination of cDNAs arising from highly abundant RNAs in total RNA samples by a differential hybridization approach coupled with duplex-specific nuclease*

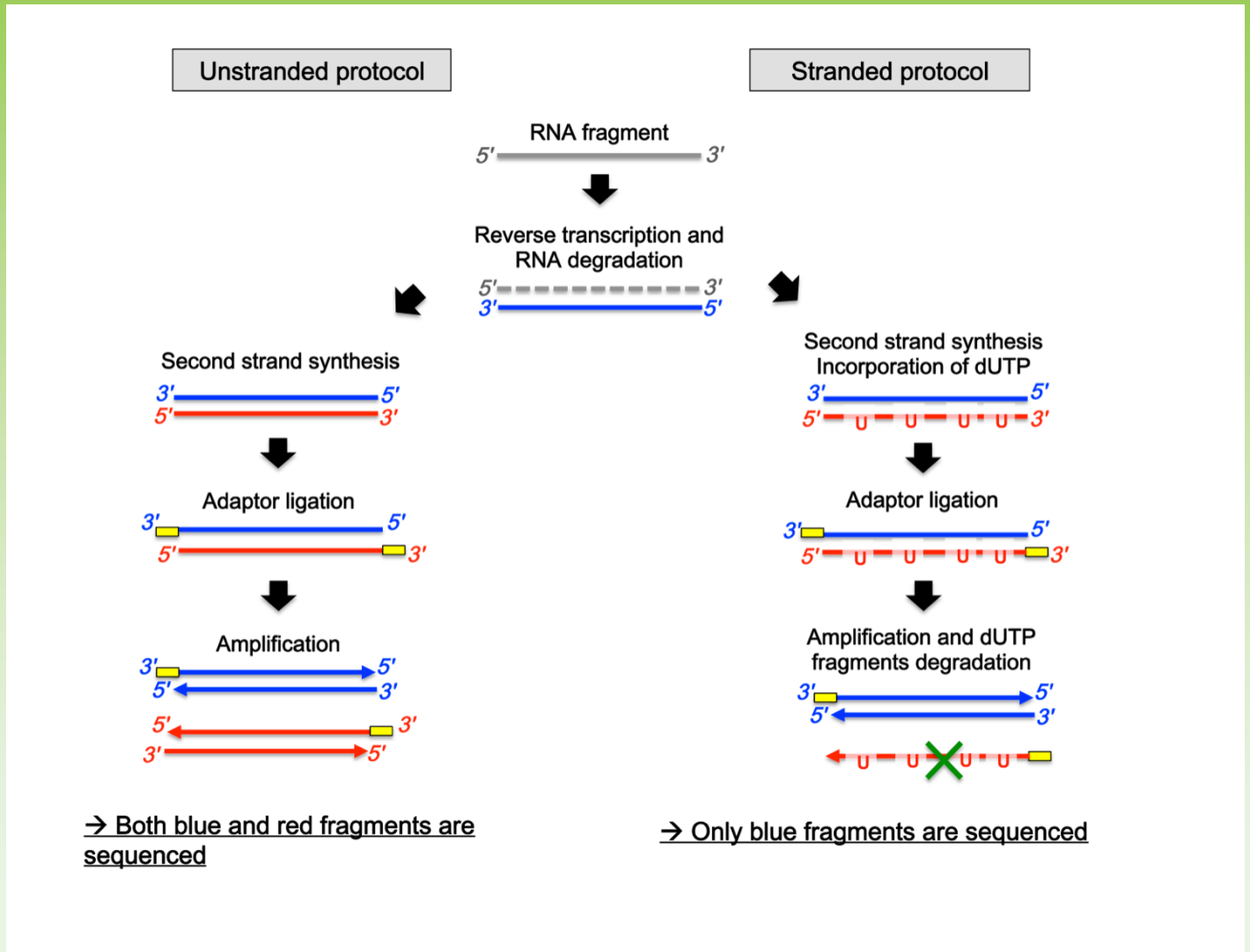
Sample preparation – Illumina sequencing

Illumina TruSeq stranded poly(A) library preparation steps

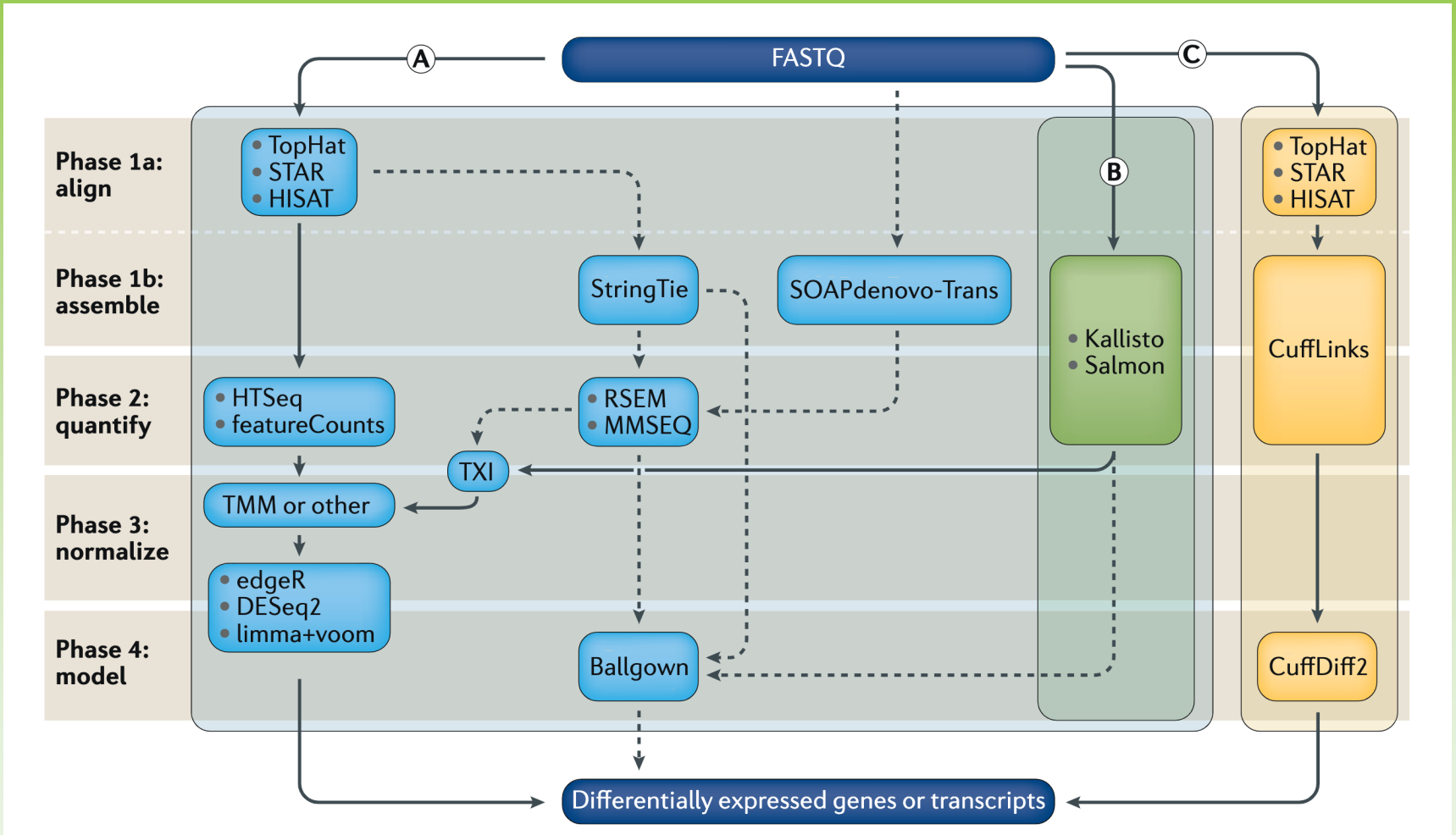


from https://www.labome.com/method/RNA-seq.html(https://www.labome.com/method/RNA-seq.html)

Stranded RNA-seq



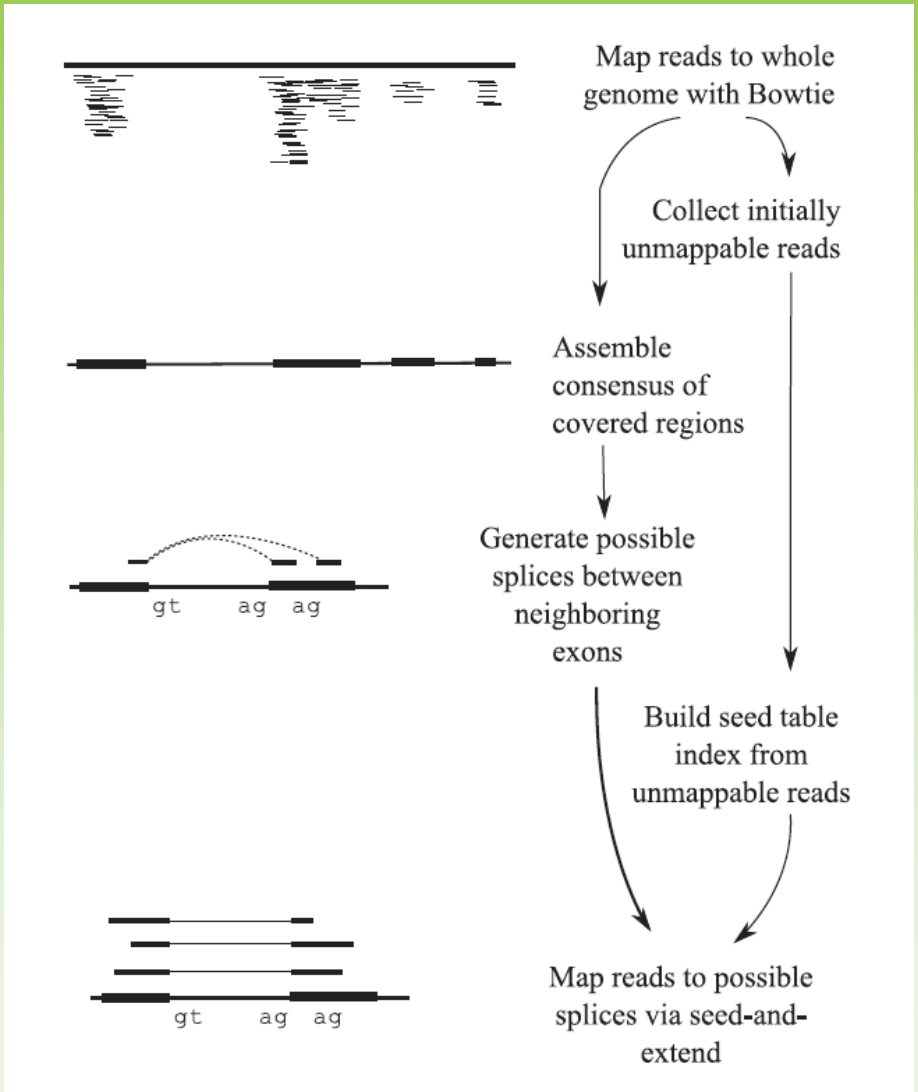
Differential gene expression analysis



10.1038/s41576-019-0150-2

Spliced read alignment problem

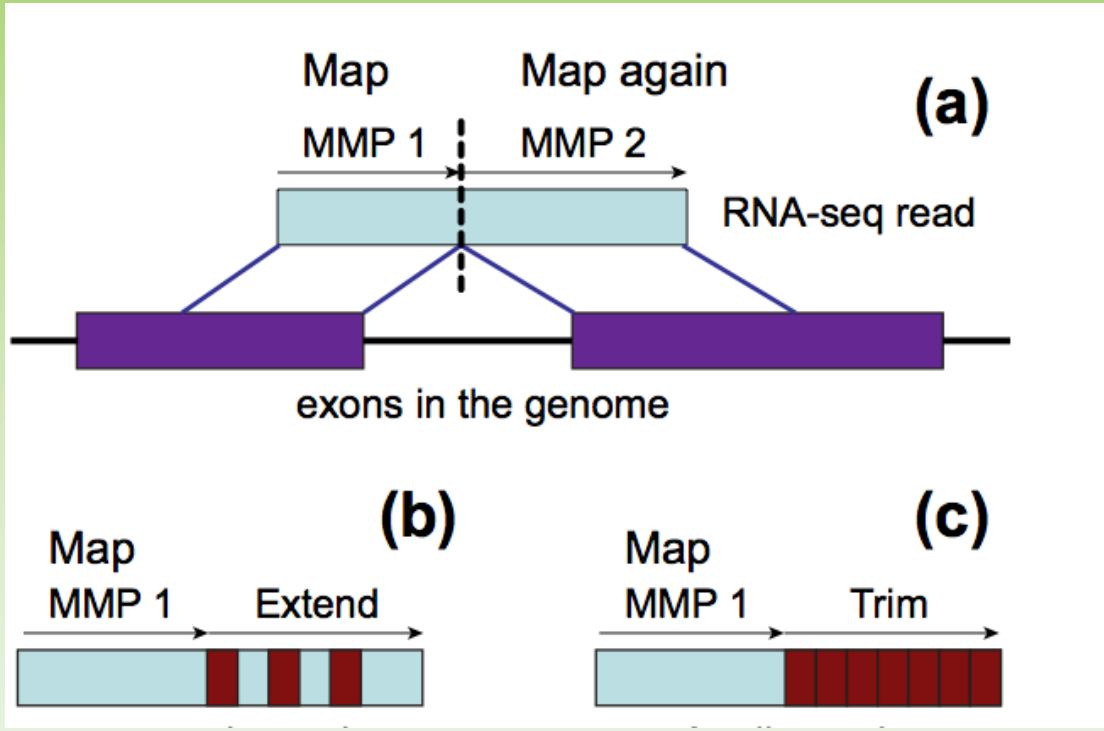
Tophat
> 3000 citations



PMID: 19289445

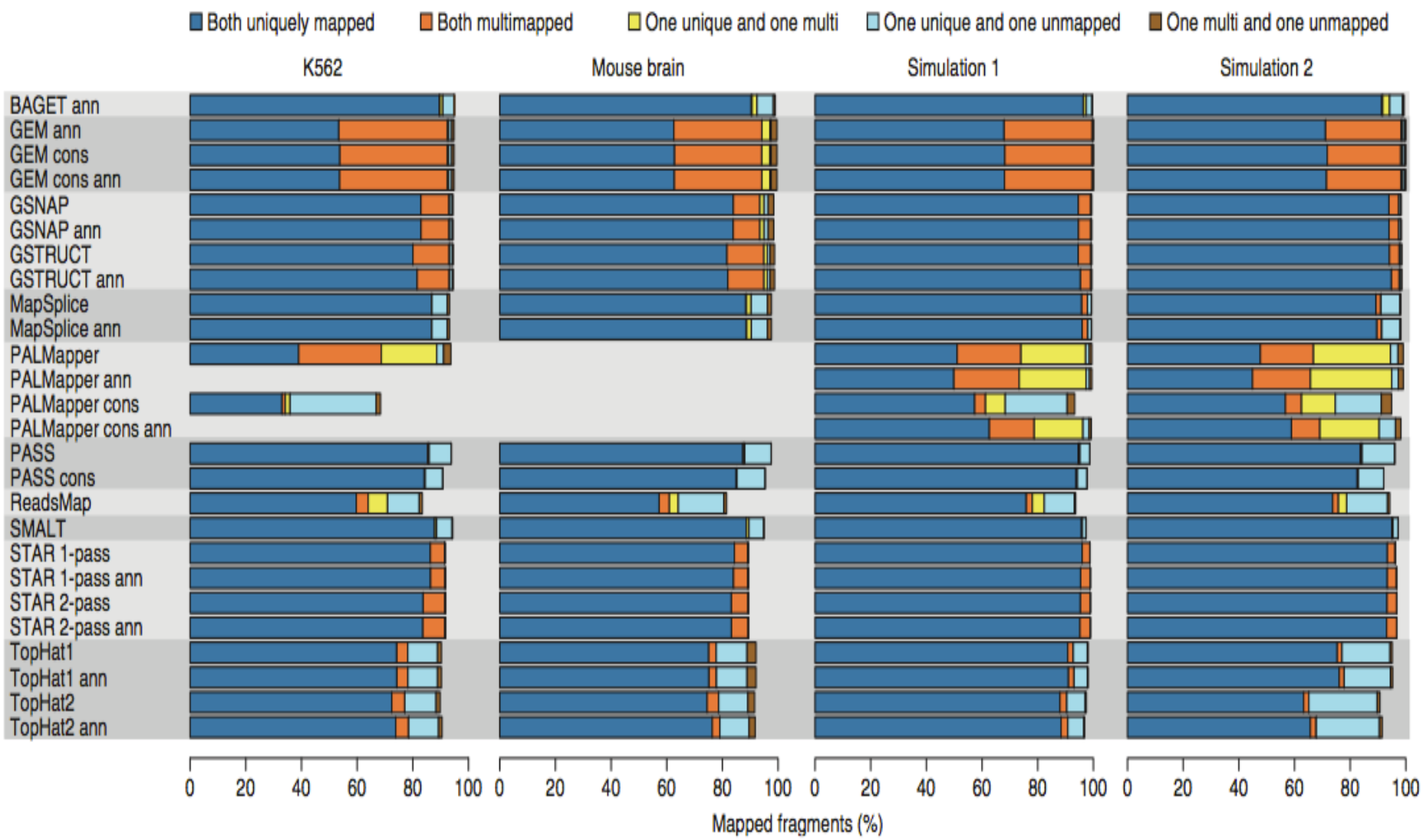
Spliced read alignment problem

STAR >2000 citations



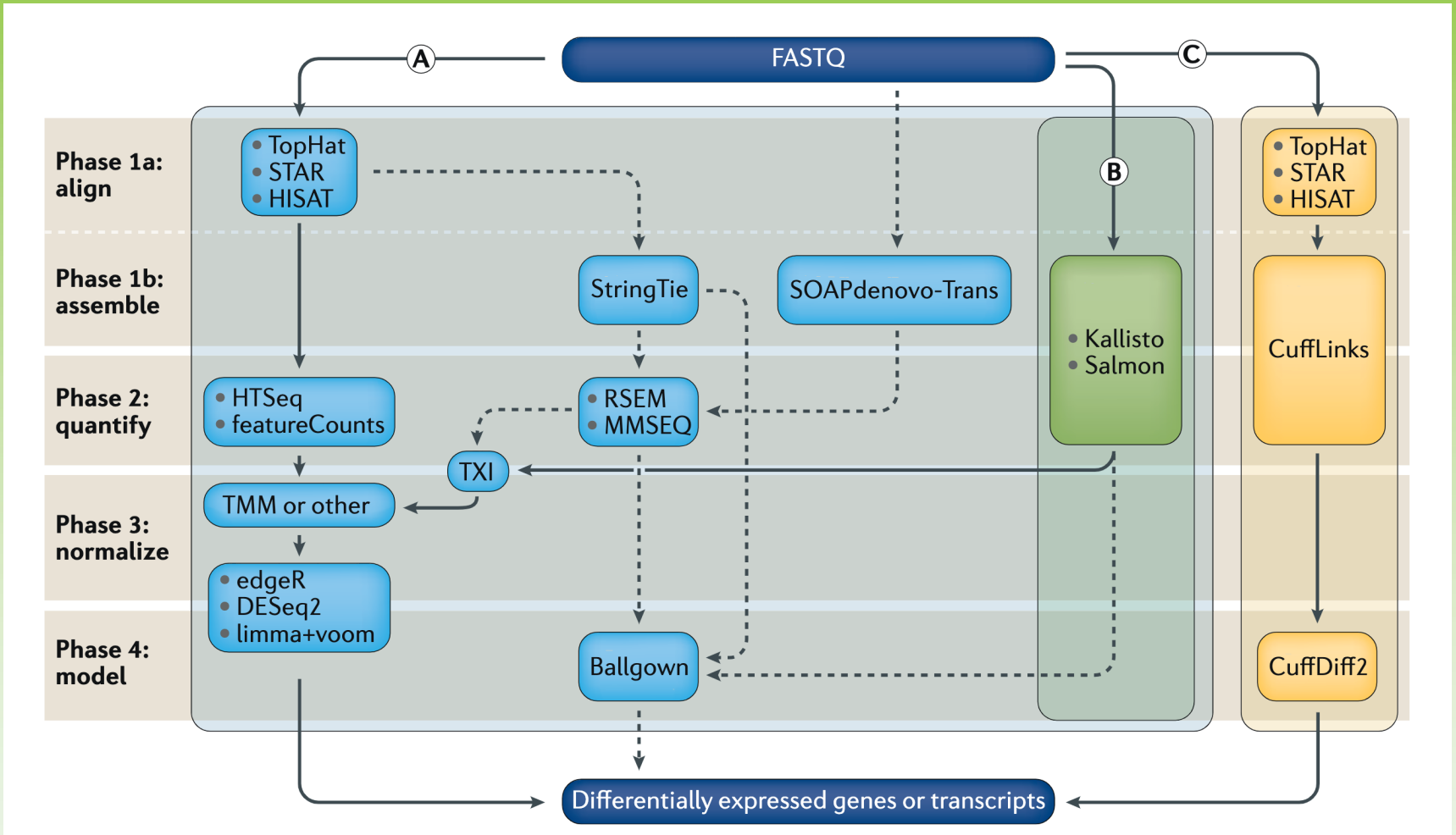
PMID: 23104886

Spliced read aligners



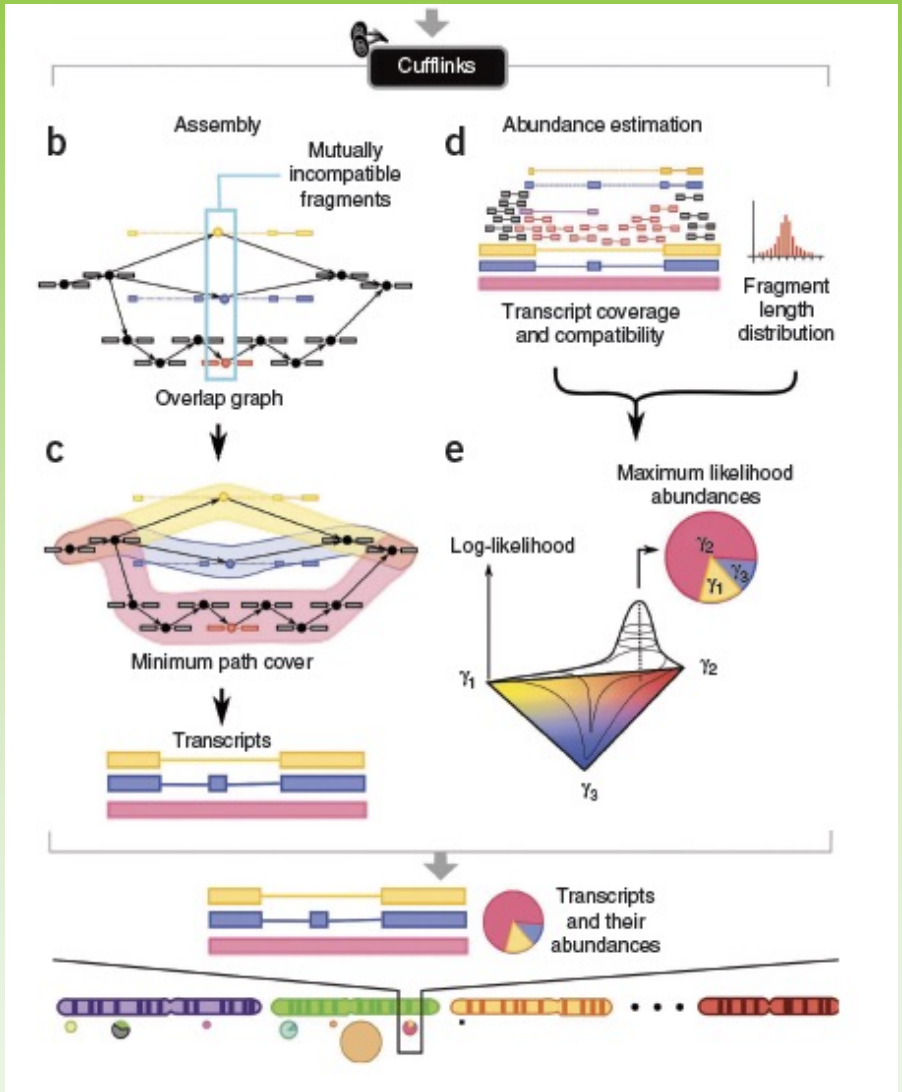
PMID: 24185836

Differential gene expression analysis



10.1038/s41576-019-0150-2

Cufflinks isoform quantification

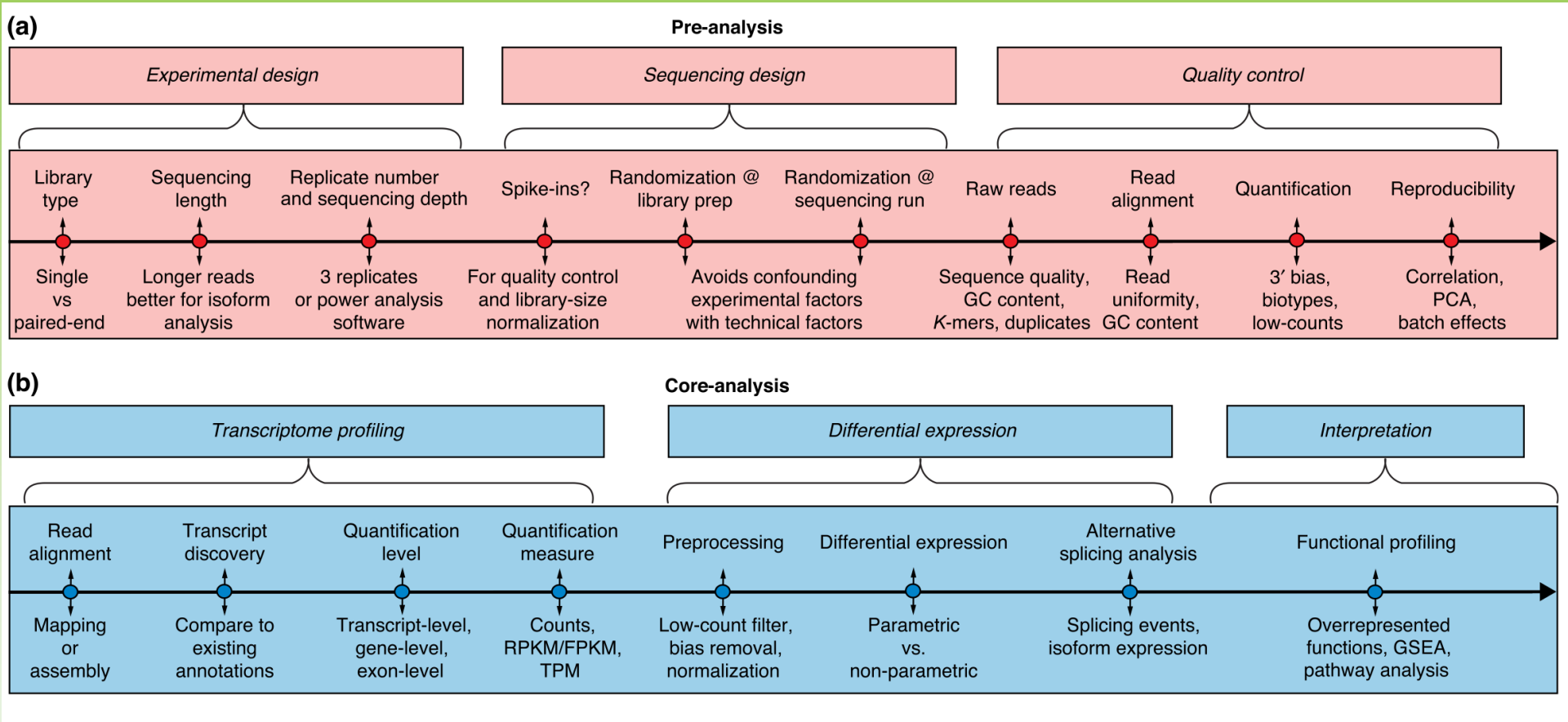


PMID: 20436464

The *de novo* way

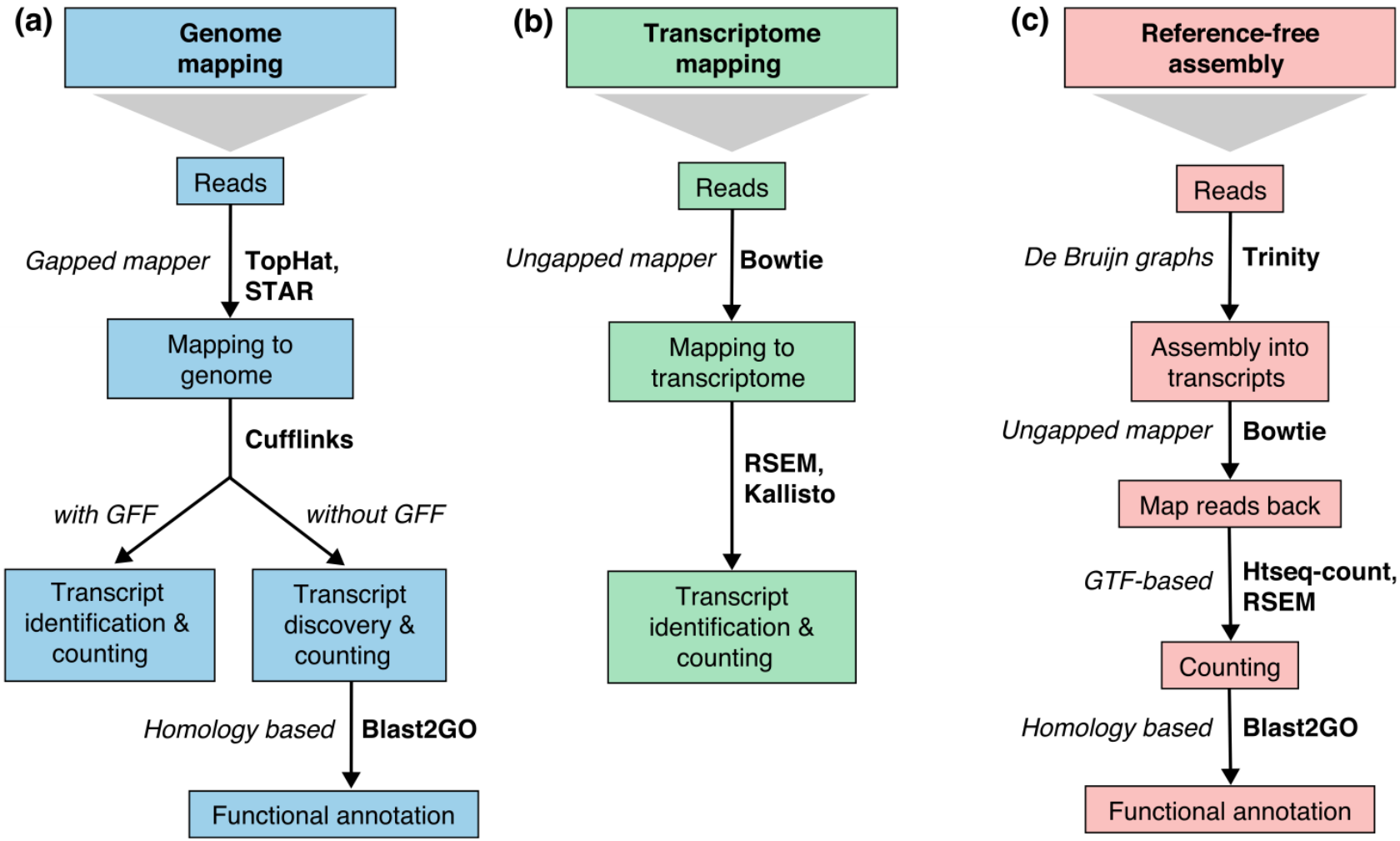
- Trans-ABySS (PMID: 20935650)
- Trinity (PMID: 21572440)
- Mapping reads more sensitive
- *de novo* has no problems with intron/exon structure
- With long reads, de novo assembly of transcripts is becoming obsolete (most transcripts are significantly shorter than HiFi reads).

RNA-seq overview



PMID: 26813401

RNA-seq overview



PMID: 26813401

Nanopore RNA-seq

Which kit do I choose?

Start here

Already prepared cDNA Library?

- NO → Interested in modifications?
- YES → Direct cDNA Sequencing Kit

Interested in modifications?

- YES → Direct RNA Sequencing Kit
- NO → Poly-A enriched?

Poly-A enriched?

- YES → ≥100 ng sample?
- NO → PCR-cDNA Sequencing Kit

≥100 ng sample?

- YES → Direct cDNA Sequencing Kit
- NO → PCR-cDNA Sequencing Kit

Direct RNA Sequencing Kit

Preparation time	115 mins
Input requirement	500 ng RNA (poly-A⁺)
RT required	Optional
PCR required	No
Read length	Equal to RNA length
Typical throughput	● ○ ○
Typical number of reads	1 million
Multiplexing options	In development

Buy now https://store.nanoporetech.com/direct-rna-sequencing-kit.html

PCR-cDNA Sequencing Kit

Preparation time	165 mins
Input requirement	1 ng RNA (poly-A⁺)
RT required	Yes
PCR required	Yes
Read length	Enriched for full-length cDNA
Typical throughput	● ● ●
Typical number of reads	7–12+ million
Multiplexing options	Yes

Buy now https://store.nanoporetech.com/cdna-pcr-sequencing-kit.html

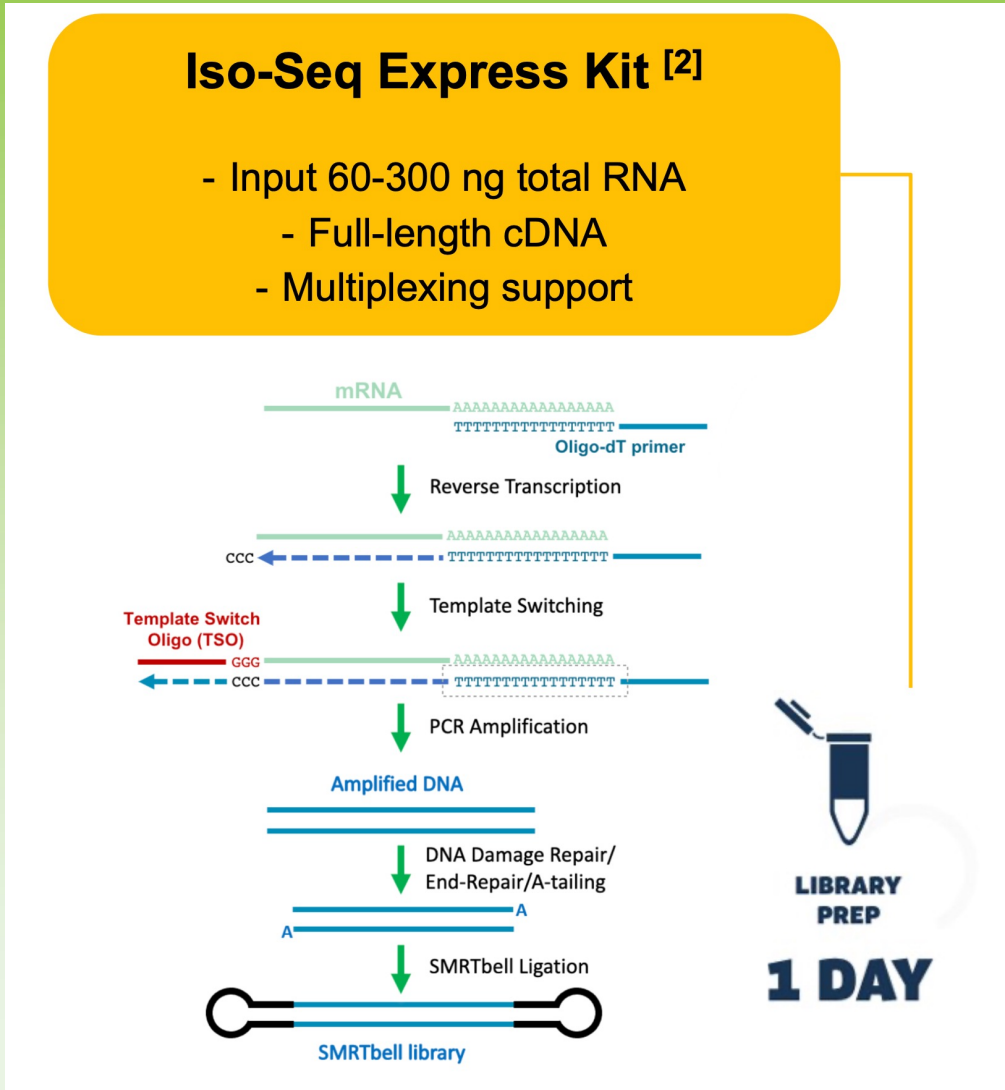
Direct cDNA Sequencing Kit

Preparation time	275 mins
Input requirement	100 ng RNA (poly-A⁺)
RT required	Yes
PCR required	No
Read length	Enriched for full-length cDNA
Typical throughput	● ● ○
Typical number of reads	5–10 million
Multiplexing options	Yes

Buy now https://store.nanoporetech.com/direct-cdna-sequencing-kit.html

https://nanoporetech.com/sites/default/files/s3/literature/Nanopore-cDNA-Guide.pdf

PacBio RNA-seq



https://www.pacb.com/wp-content/uploads/Underwood-AGBT-2020-A-complete-solution-for-full-length-transcript-sequencing-using-the-PacBio-Sequel-II-System.pdf

PacBio RNA-seq

Sequel II System

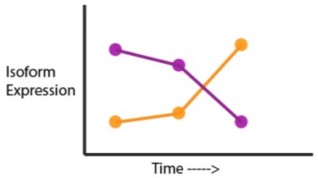
- 1 SMRT Cell 8M for whole transcriptome
- Up to 4 million full-length reads
- Accuracy 99-99.9%

	Unique Genes	Unique Transcripts	Unique ORFs
Single Cell, Human Brain Organoid	14,737	60,815	34,697
Single Cell, Human Cell Line	17,767	237,951	89,399
Bulk, UHRR [3]	16,328	183,689	60,649
Bulk, Alzheimer Brain [4]	17,670	162,290	80,539

https://www.pacb.com/wp-content/uploads/Underwood-AGBT-2020-A-complete-solution-for-full-length-transcript-sequencing-using-the-PacBio-Sequel-II-System.pdf

PacBio RNA-seq

Main Bioinformatics Tools

		Input	Output
Sequencing			subreads.bam
Iso-Seq Analysis		subreads.bam or ccs.bam	Collapsed unique transcripts (GFF, FASTA)
Transcript Classification		Unique transcripts Reference genome Annnotation (GTF) CAGE Peak Junction data...	Transcript classification Junction classification Figures
Functional Annotation		SQANTI output	Annotated GTF
Differential Analysis		Experimental design Annotated GTFs	

https://www.pacb.com/wp-content/uploads/Underwood-AGBT-2020-A-complete-solution-for-full-length-transcript-sequencing-using-the-PacBio-Sequel-II-System.pdf