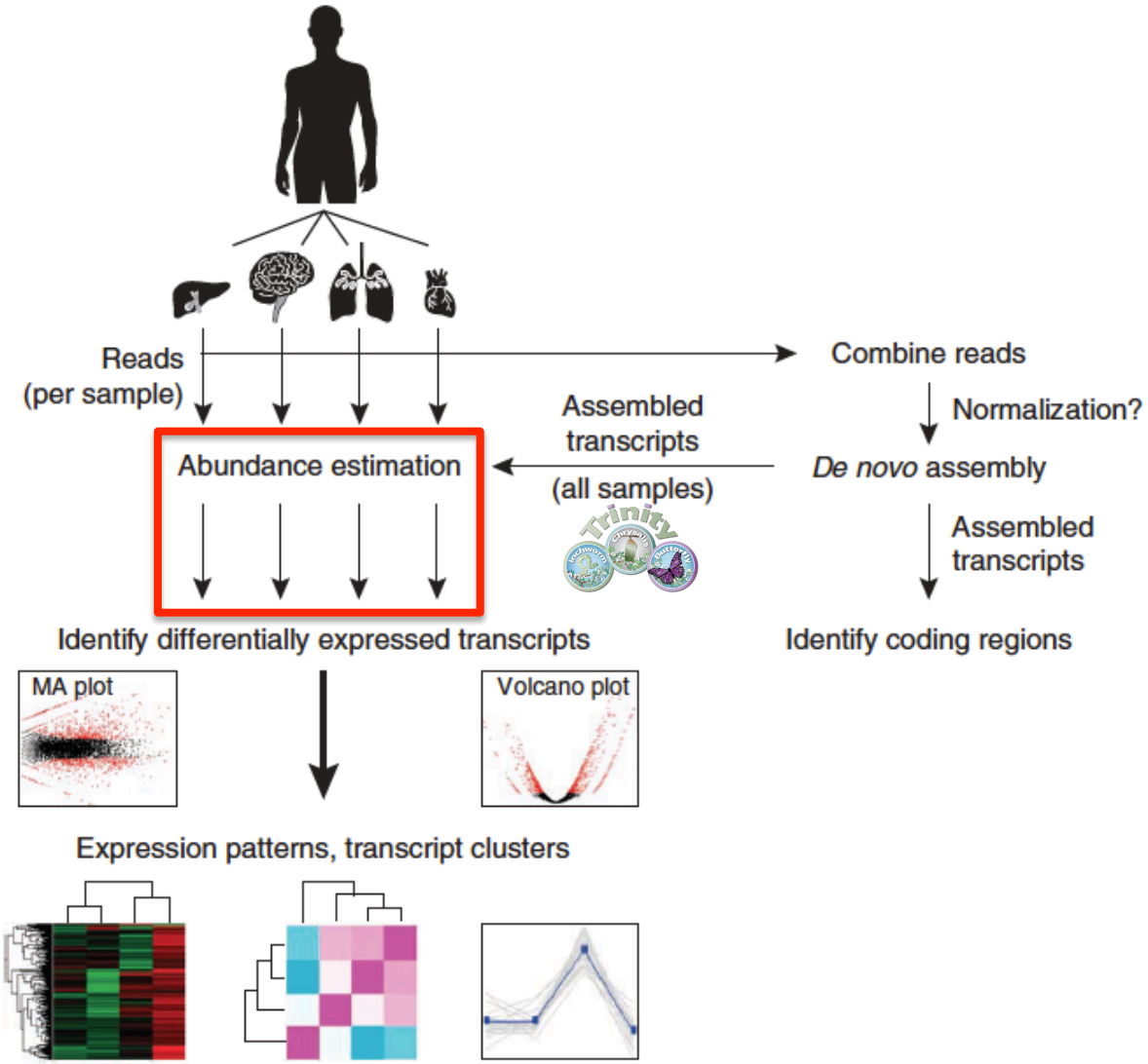
Abundance Estimation
(Aka. Computing Expression Values)

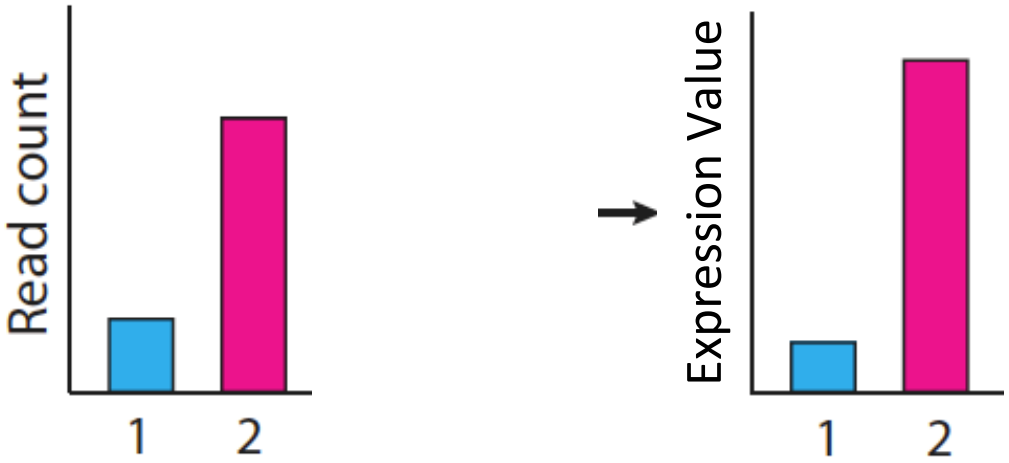
Abundance Estimation

(Aka. Computing Expression Values)



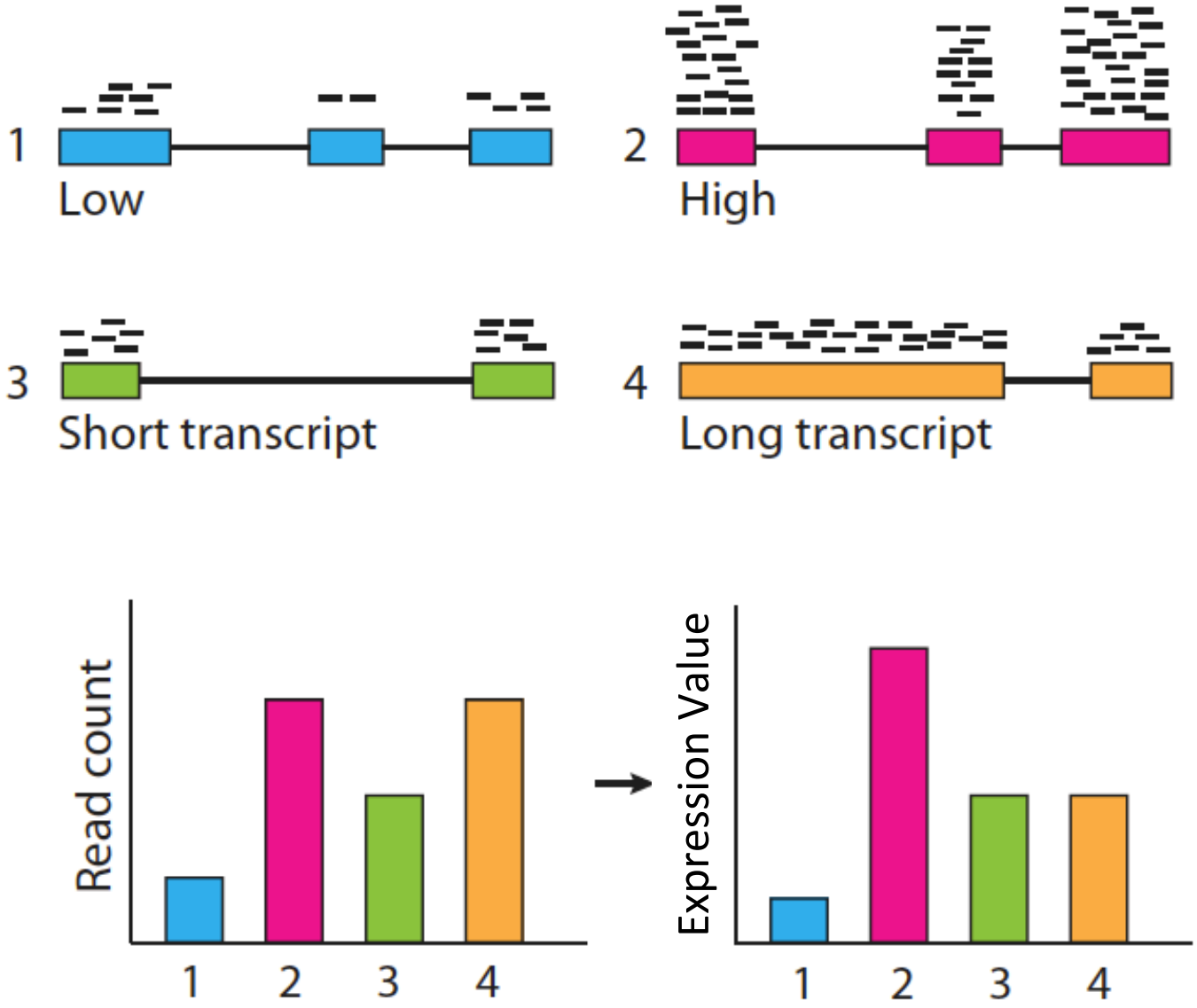
Bioconductor,
& Trinity

Calculating expression of genes and transcripts



Slide courtesy of Cole Trapnell

Calculating expression of genes and transcripts



Slide courtesy of Cole Trapnell

Normalized Expression Values

- Transcript-mapped read counts are normalized for both length of the transcript and total depth of sequencing.

- Reported as: Number of RNA-Seq **F**ragments
  Per **K**ilobase of transcript
  per total **M**illion fragments mapped

FPKM

RPKM (reads per kb per M) used with Single-end RNA-Seq reads
FPKM used with Paired-end RNA-Seq reads.

Transcripts per Million (TPM)

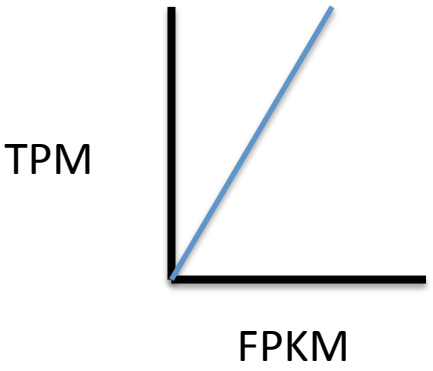
$$TPM_i = \frac{FPKM_i}{\sum_j FPKM} * 1e6$$

Preferred metric for measuring expression

- Better reflects transcript concentration in the sample.
- Nicely sums to 1 million

Linear relationship between TPM and FPKM values.

Both are valid metrics, but best to be consistent.



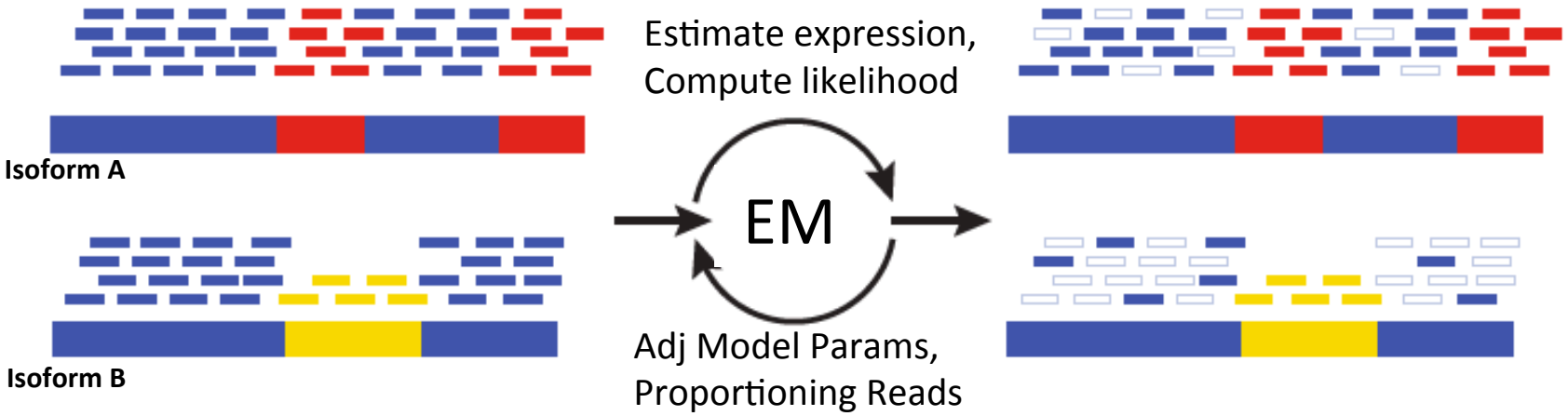
Multiply-mapped Reads Confound Abundance Estimation

Isoform A

Isoform B

Blue = multiply-mapped reads
Red, Yellow = uniquely-mapped reads

Multiply-mapped Reads Confound Abundance Estimation



Blue = multiply-mapped reads
Red, Yellow = uniquely-mapped reads

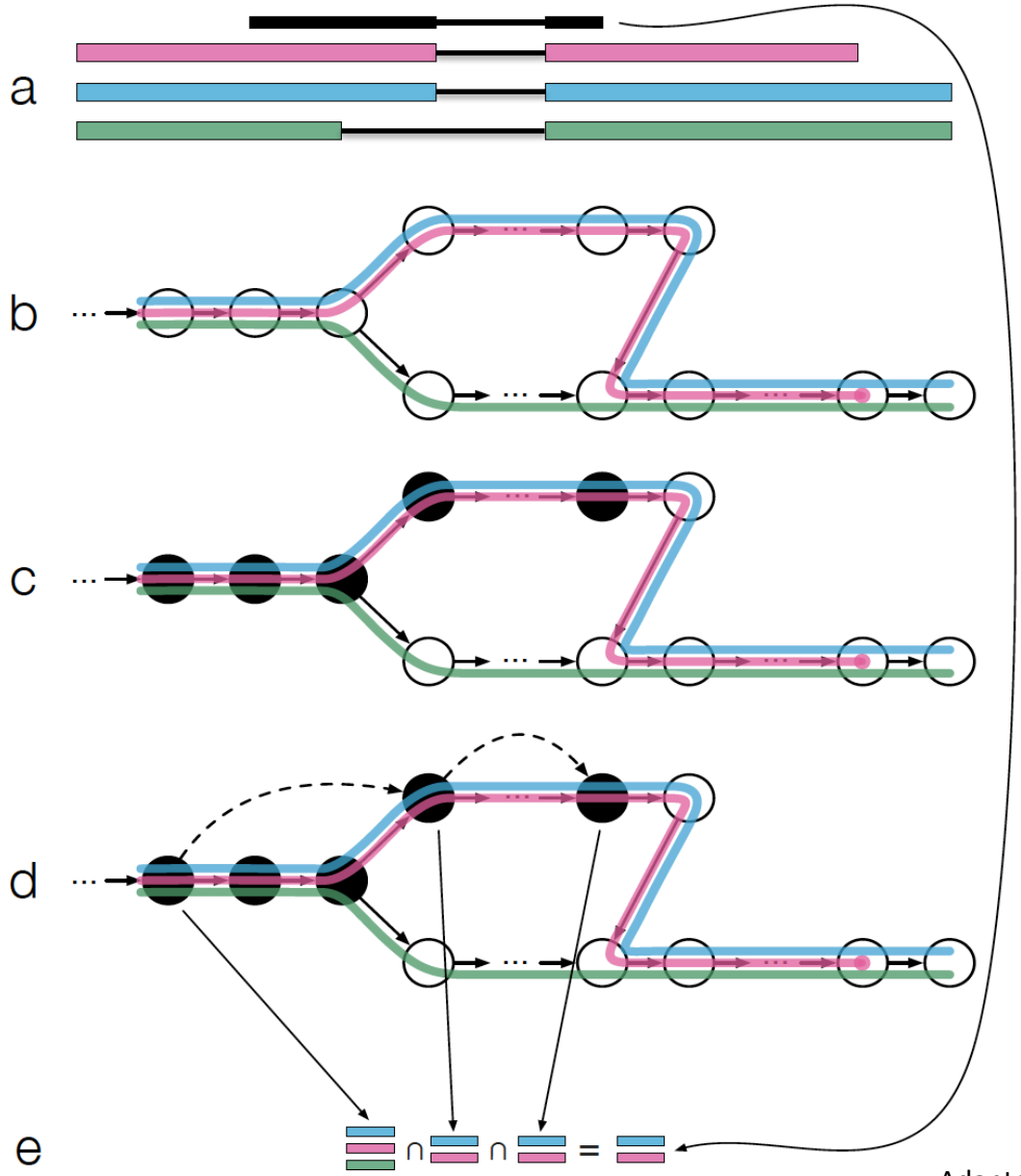
Use Expectation Maximization (EM) to find the most likely assignment of reads to transcripts.

Performed by:

- Cufflinks, String Tie (Tuxedo)
- RSEM, eXpress (genome-free)
- Kallisto, Salmon (alignment-free)

Fast Abundance Estimation Using Pseudo-alignments and Equivalence Classes

(Kallisto software, Bray et al., NBT 2016)

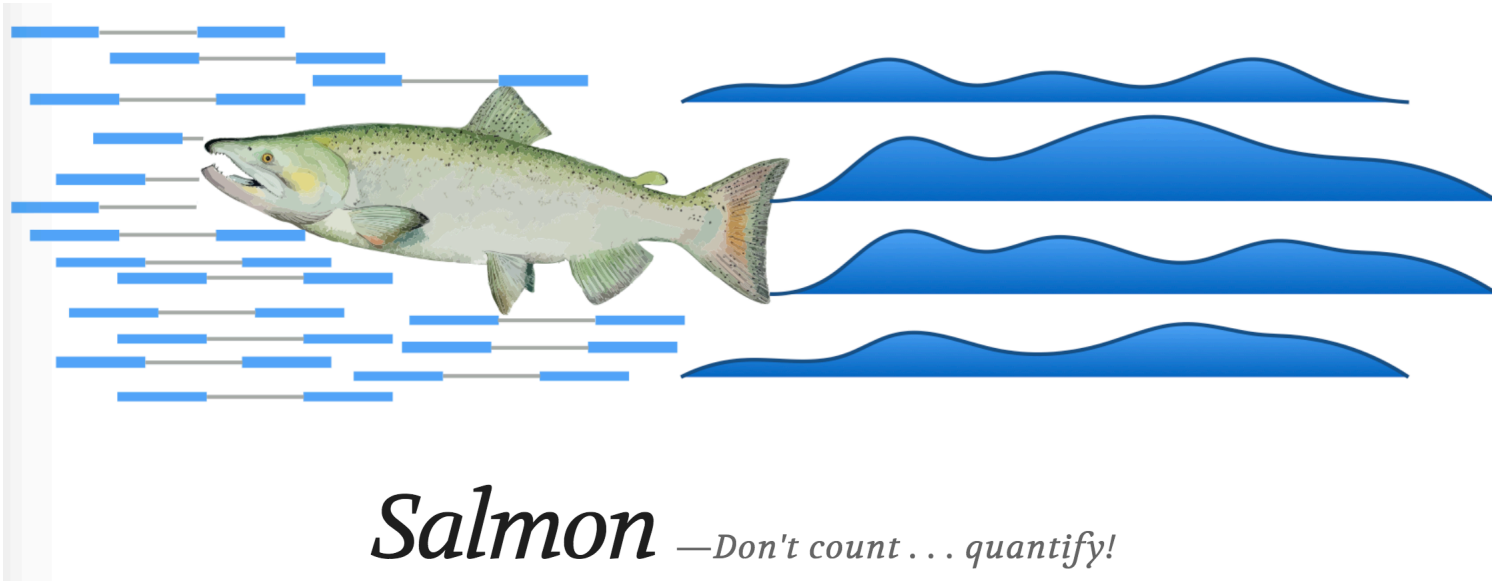


Adapted from Fig 1 from Bray et al.

Expression Quantification Results

(ex. from Kallisto)

target_id	length	eff_length	est_counts	tpm
TRINITY_DN10_c0_g1_i1	334	100.489	13	4186.62
TRINITY_DN11_c0_g1_i1	319	87.9968	0	0
TRINITY_DN12_c0_g1_i1	244	38.2208	2	1693.43
TRINITY_DN17_c0_g1_i1	229	30.2382	5	5351.21
TRINITY_DN18_c0_g1_i1	633	384.493	19	1599.2
TRINITY_DN18_c1_g1_i1	289	65.795	1	491.864
TRINITY_DN19_c0_g1_i1	283	61.0618	10	5299.91



Salmon —*Don't count . . . quantify!*

Uses a suffix array instead of the de Bruijn graph

nature methods

Altmetric: 210 Citations: 42 More detail »

Brief Communication

Salmon provides fast and bias-aware quantification of transcript expression

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

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

Received: 29 August 2016
Accepted: 22 January 2017
Published online: 06 March 2017

https://combine-lab.github.io/salmon/