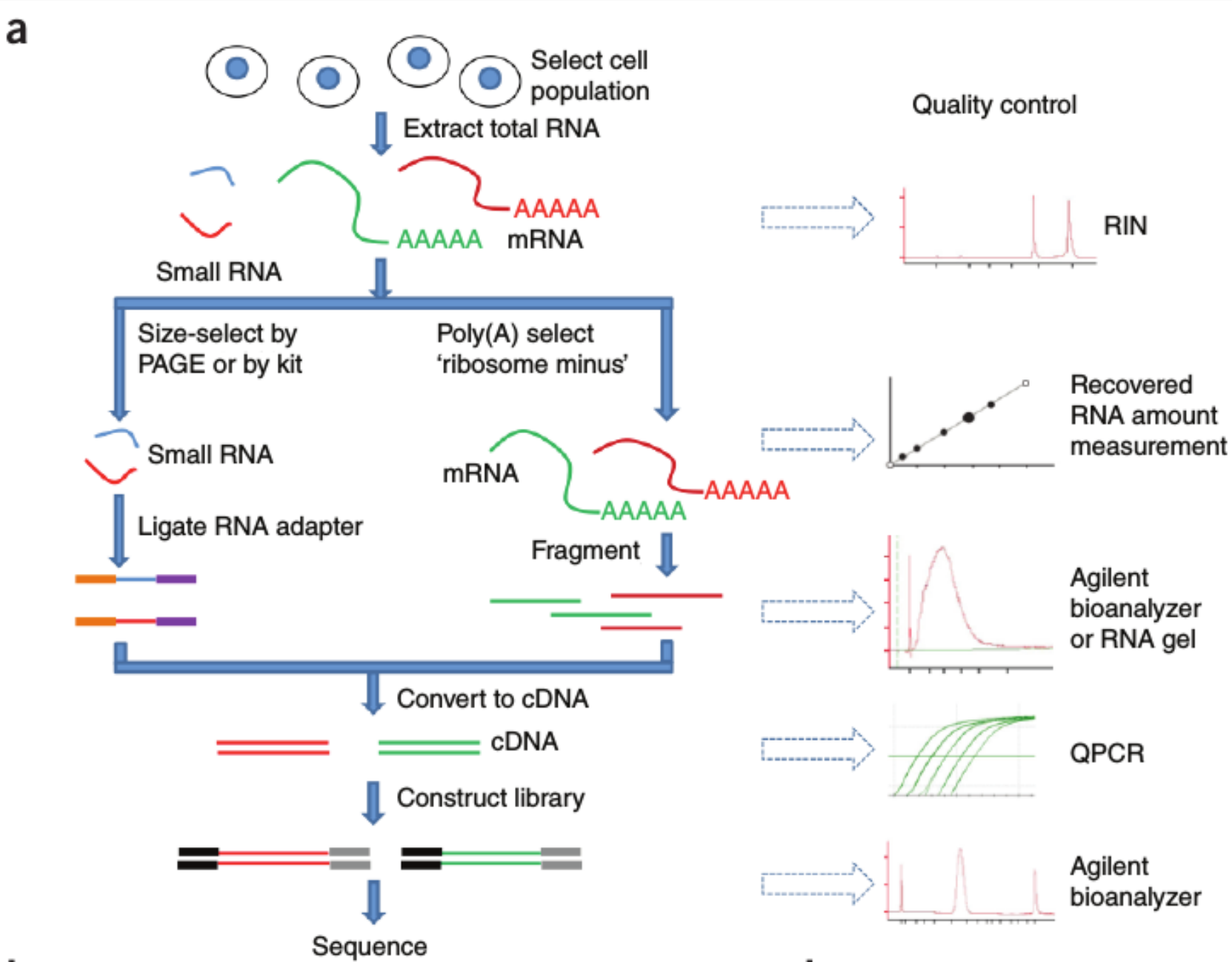
Near-optimal probabilistic RNA-seq quantification

Bray and Pachter *et al.* Nature biotechnology(2016) doi:10.1038/nbt.3519

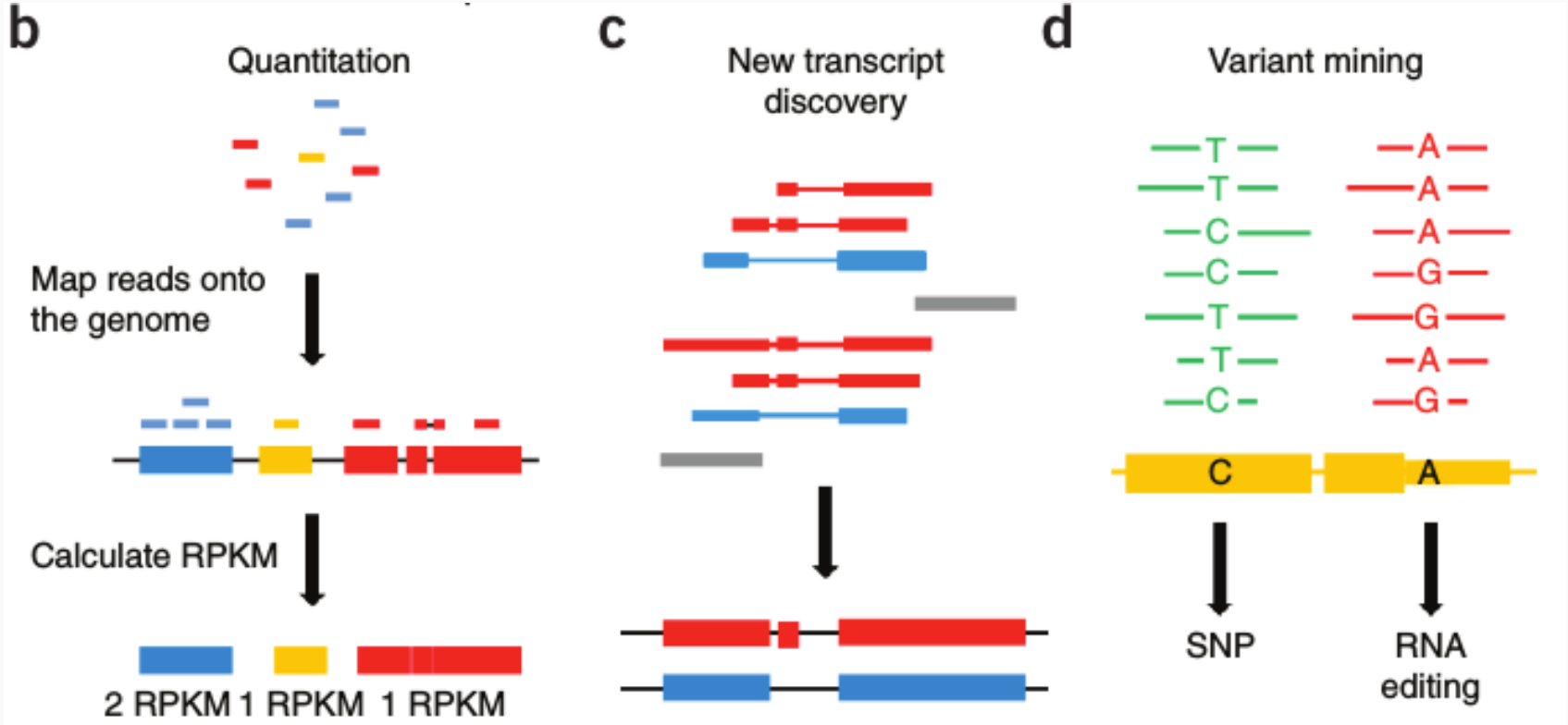
Saket Choudhary

September 25, 2016

RNA-Seq Workflow



RNA-Seq Workflow



Zheng and Mortazavi(2012)

Motivation

- First two steps in typical RNA-Seq processing pipeline:

Motivation

- First two steps in typical RNA-Seq processing pipeline:
  - Alignment

Motivation

- First two steps in typical RNA-Seq processing pipeline:
  - Alignment
  - Quantification

Motivation

- First two steps in typical RNA-Seq processing pipeline:
  - Alignment
  - Quantification
- Alignments are slow and probably not so important

It’s all about compatible transcripts

- Circumvent alignment step – Use information from $k-mers$
- Pseudoalignment: Find *compatible* transcripts for a read, without pinpointing where exactly it aligns

Method I

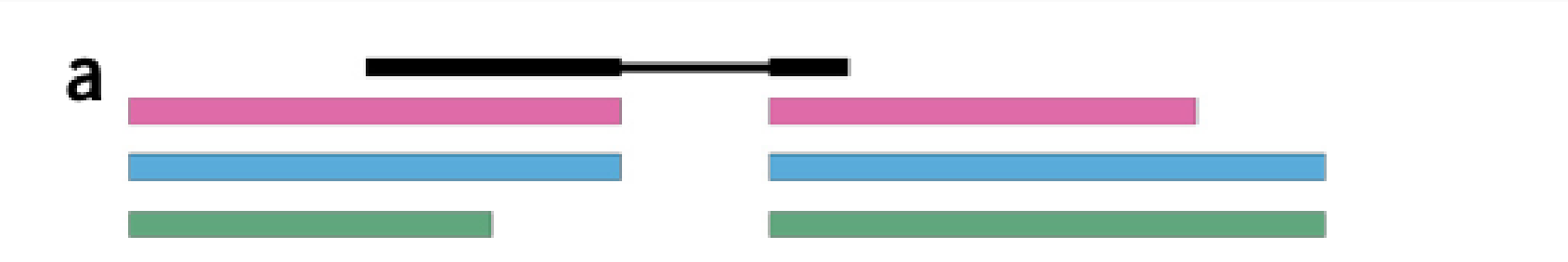


Figure 1: Reads and overlapping transcripts

Method II

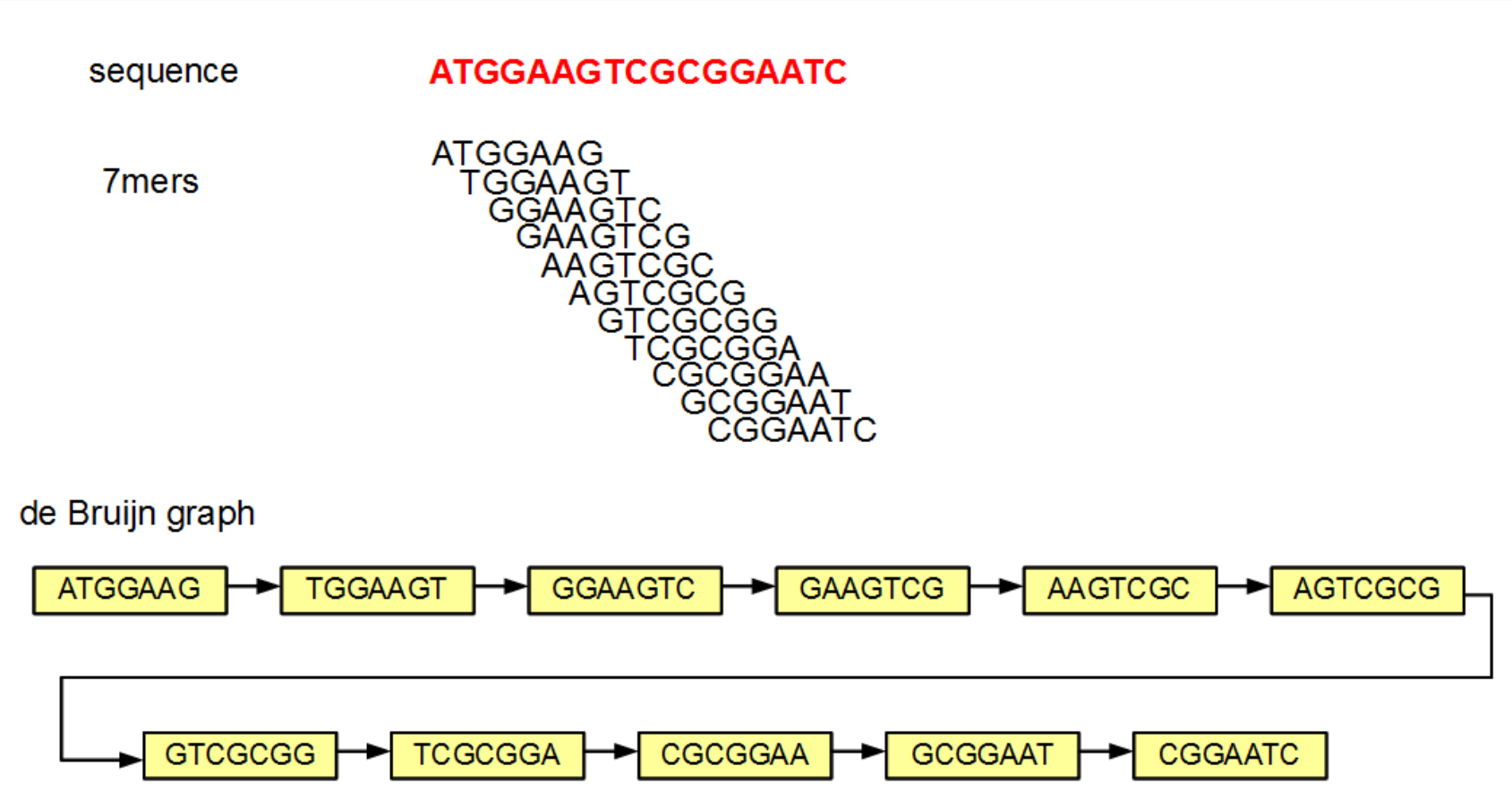


Figure 2: de Bruijn Graph

Method III

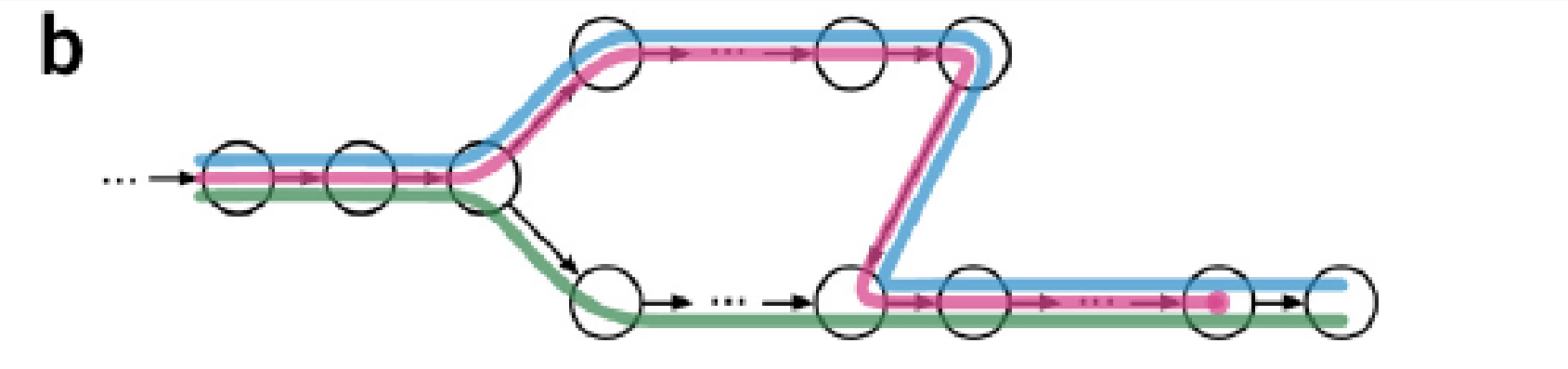


Figure 3: Transcriptome - de Bruijn Graph. Node = k — *mers*, Path = Transcript

Method IV

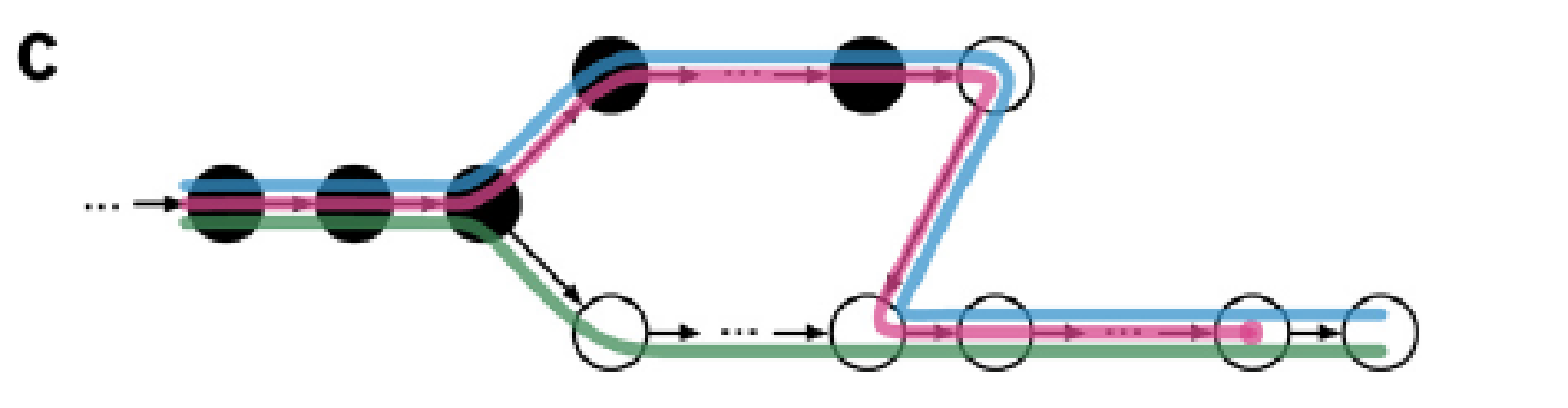


Figure 4: *k — mers* in read = black nodes

Method V

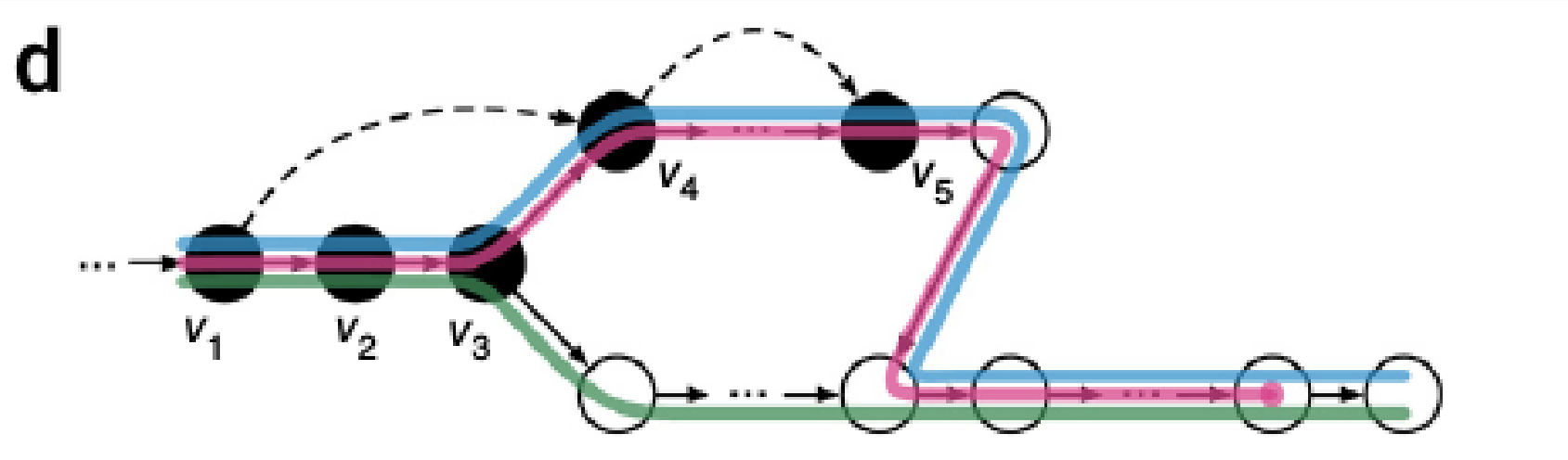


Figure 5: Nodes can be skipped if $k - mers$ did arise from blue transcript

Method VI

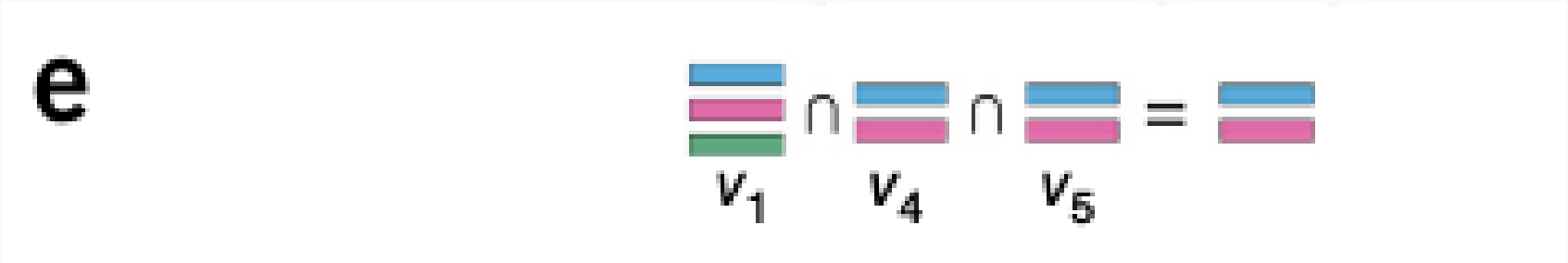


Figure 6: Intersection of k-compatibility class

Method VII

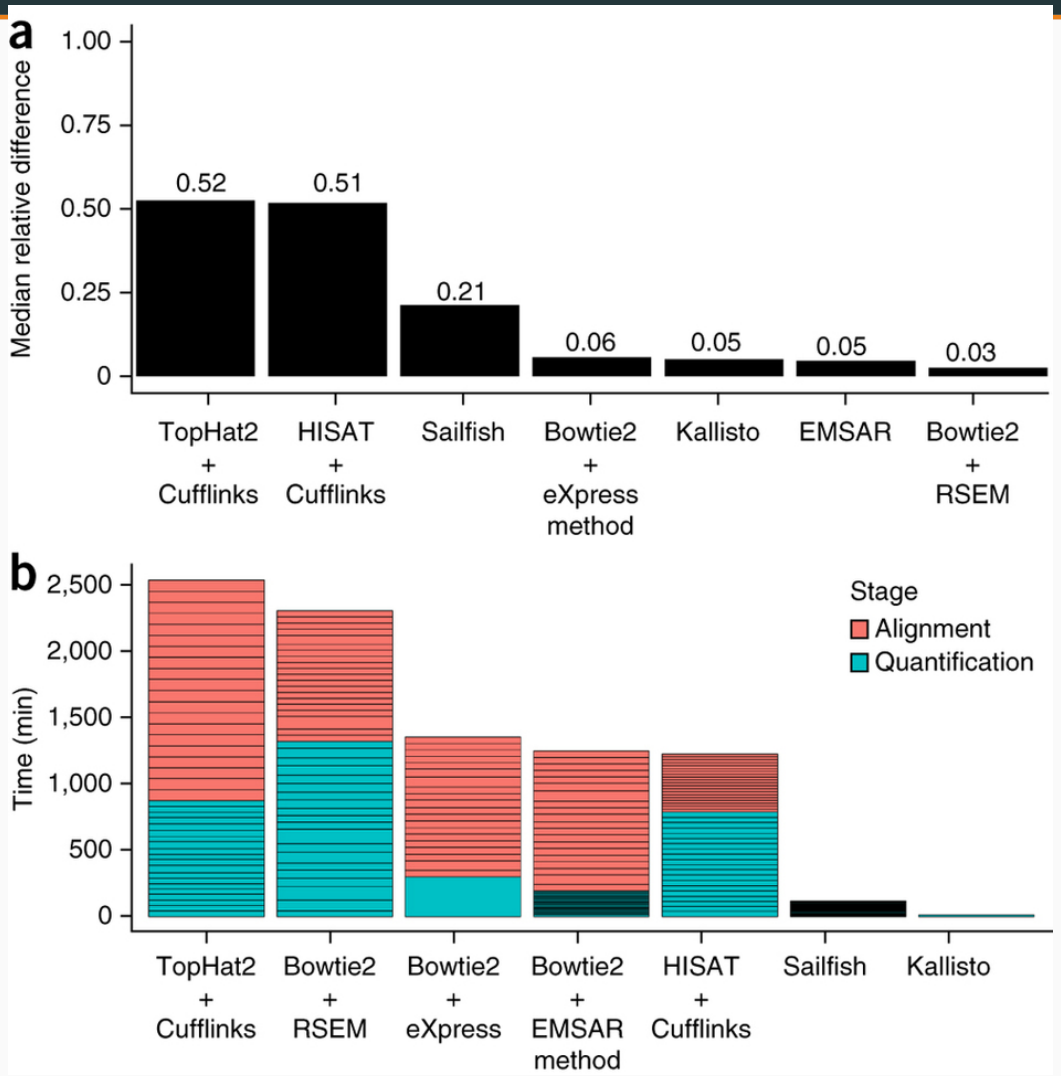


Figure 7: Quantification over a cup of coffee

Questions?

notes

- Better than *Sailfish* that looks up $k - mers$ in reads into $k - mers$ of transriptome

notes

- Better than *Sailfish* that looks up $k-mers$ in reads into $k-mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns

notes

- Better than *Sailfish* that looks up $k-mers$ in reads into $k-mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns
- Key Idea: Find comtabile transcript for a read, not where it exactly aligns

notes

- Better than *Sailfish* that looks up $k-mers$ in reads into $k-mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns
- Key Idea: Find comtabile transcript for a read, not where it exactly aligns
- Psuedoalignment: Subset $S \subset T$ such that read r is compatible.

notes

- Better than *Sailfish* that looks up $k - mers$ in reads into $k - mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns
- Key Idea: Find comtabile transcript for a read, not where it exactly aligns
- Psuedoalignment: Subset $S \subset T$ such that read r is compatible.
- Hash k-mers of reads and have a de-bruijn graph of transcriptome assembly handy

notes

- Better than *Sailfish* that looks up $k-mers$ in reads into $k-mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns
- Key Idea: Find comtabile transcript for a read, not where it exactly aligns
- Psuedoalignment: Subset $S \subset T$ such that read r is compatible.
- Hash k-mers of reads and have a de-bruijn graph of transcriptome assembly handy
- T-DBG: nodes are k-mers , each transcript corresponds to a path and path cover induces a k-comptability class for each k-mer

notes

- Better than *Sailfish* that looks up $k-mers$ in reads into $k-mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns
- Key Idea: Find comtabile transcript for a read, not where it exactly aligns
- Psuedoalignment: Subset $S \subset T$ such that read r is compatible.
- Hash k-mers of reads and have a de-bruijn graph of transcriptome assembly handy
- T-DBG: nodes are k-mers , each transcript corresponds to a path and path cover induces a k-comptability class for each k-mer
- T-DBG: Colors correspond to transcripts, node corresponds to k-mers, every k-mer receives a color for each transcript it occurs in

notes

- Better than *Sailfish* that looks up $k - mers$ in reads into $k - mers$ of transriptome
- Pseudoalignment: Find compatible transcript for a read, not where it exactly aligns
- Key Idea: Find comtabile transcript for a read, not where it exactly aligns
- Psuedoalignment: Subset $S \subset T$ such that read r is compatible.
- Hash k-mers of reads and have a de-bruijn graph of transcriptome assembly handy
- T-DBG: nodes are k-mers , each transcript corresponds to a path and path cover induces a k-comptability class for each k-mer
- T-DBG: Colors correspond to transcripts, node corresponds to k-mers, every k-mer receives a color for each transcript it occurs in
- Hash table stores mapping of each k-mer to the contig it is contained in