rQuant: Quantitative Detection of Alternative Transcripts with RNA-Seq Data

Short-SIG, ISMB/ECCB 2009

Regina Bohnert, Jonas Behr, and Gunnar Rätsch



Friedrich Miescher Laboratory
of the Max Planck Society

Tübingen, Germany

June 28, 2009

RNA Sequencing

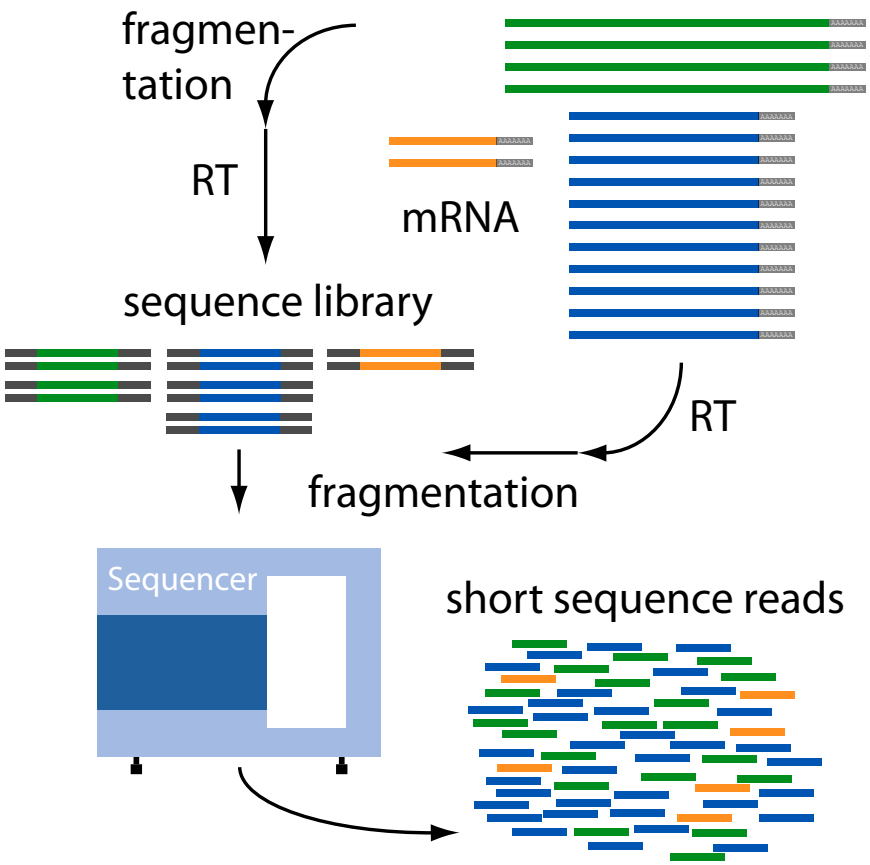


RNA-Seq allows ...

- High-troughput transcriptome measurements
- Qualitative studies
  - Altered expression across tissues, conditions etc.
- Quantitative studies at high resolution
  - Simple approach: naive counting

But: RNA-Seq data has many biases

RNA Sequencing and Biases



Biases due to ...

- cDNA library construction
- Sequencing
- Read mapping

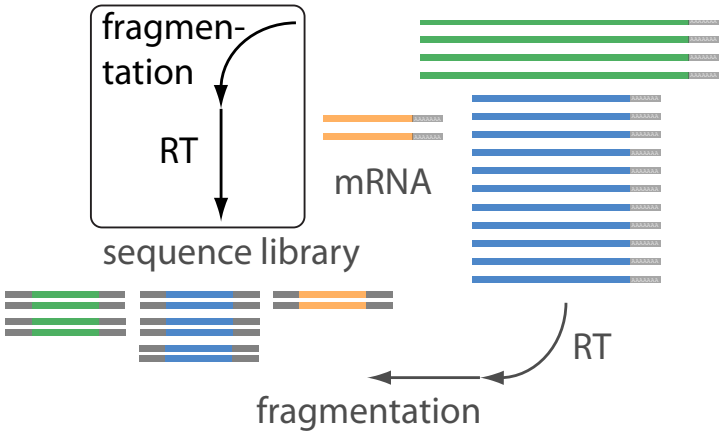


$\Rightarrow$ QPALMA (De Bona et al., 2008)

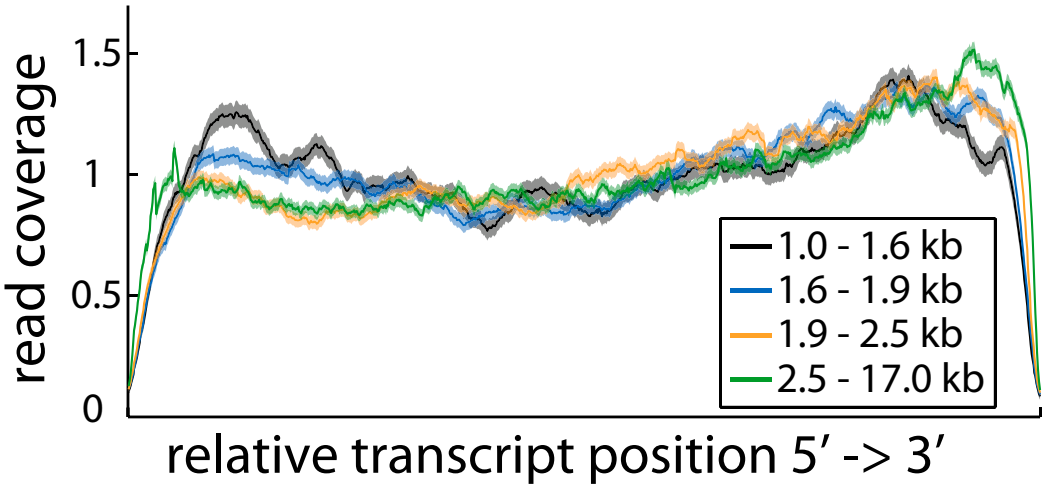
(`www.fml.mpg.de/raetsch/projects/qpalma`)

Priming and Fragmentation Biases

Profile: normalised positional read coverage along the transcript



Transcript profiles for different lengths

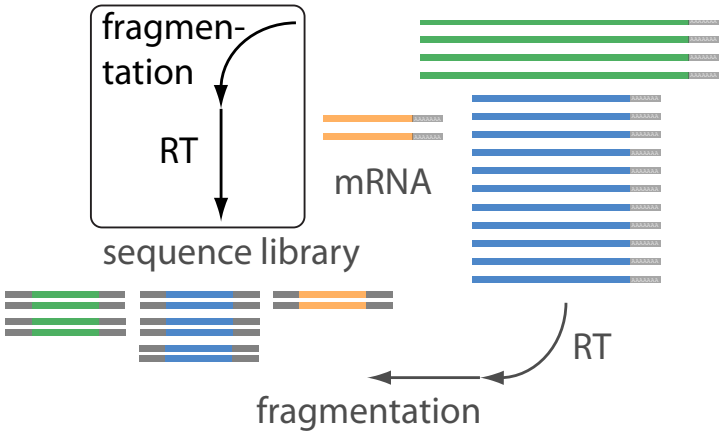


RNA-Seq data (*A. thaliana*)

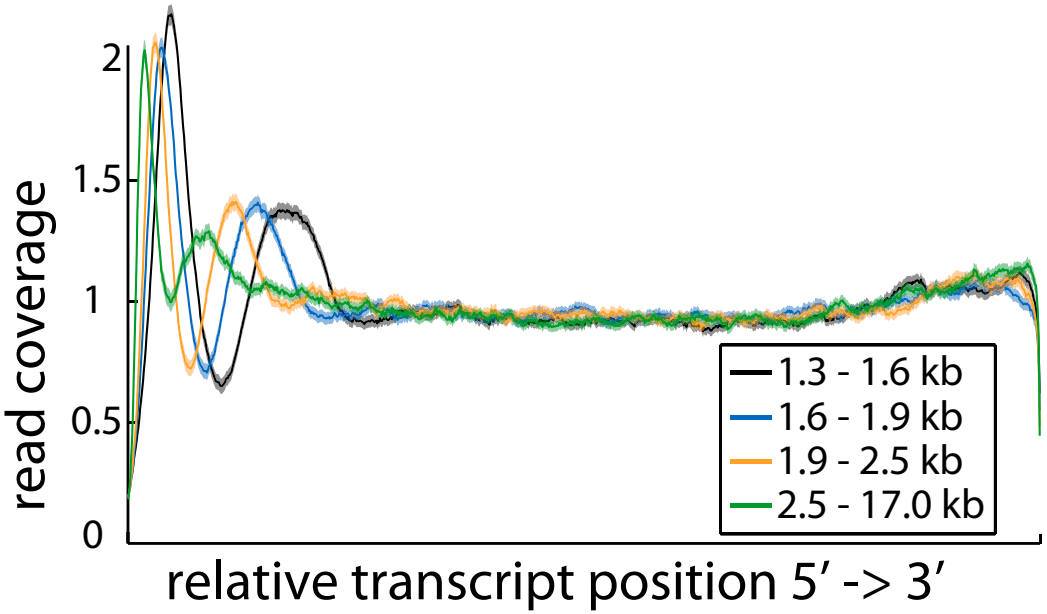
- Chemical RNA fragmentation
- Random priming

Priming and Fragmentation Biases

Profile: normalised positional read coverage along the transcript



Transcript profiles for different lengths

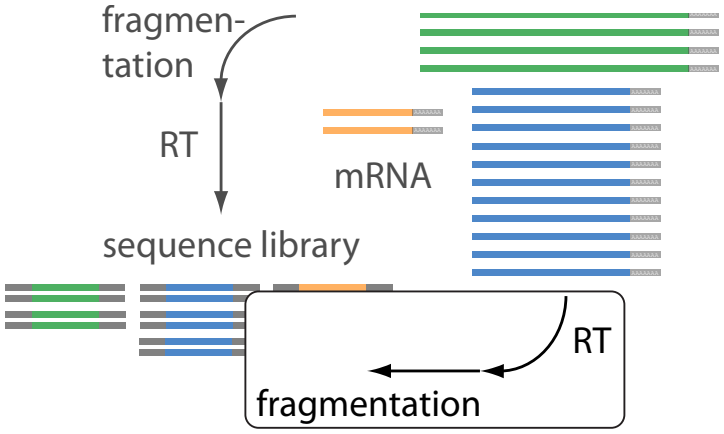


Data generated for *A. thaliana* with Flux Simulator v1.0 (Sammeth, 2009b)

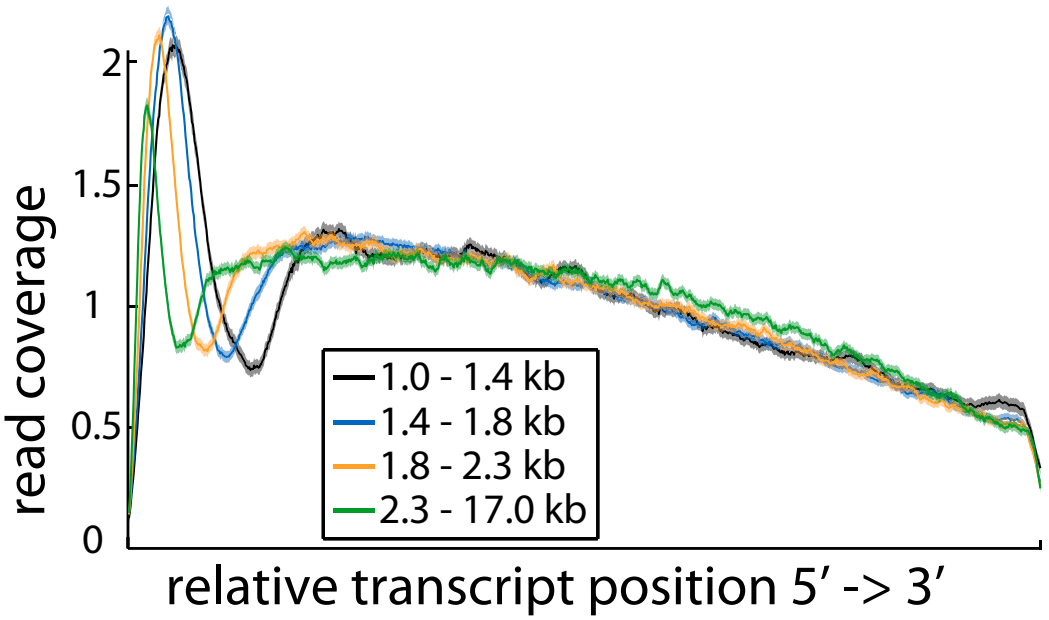
- Physical RNA fragmentation
- Random priming

Priming and Fragmentation Biases

Profile: normalised positional read coverage along the transcript



Transcript profiles for different lengths

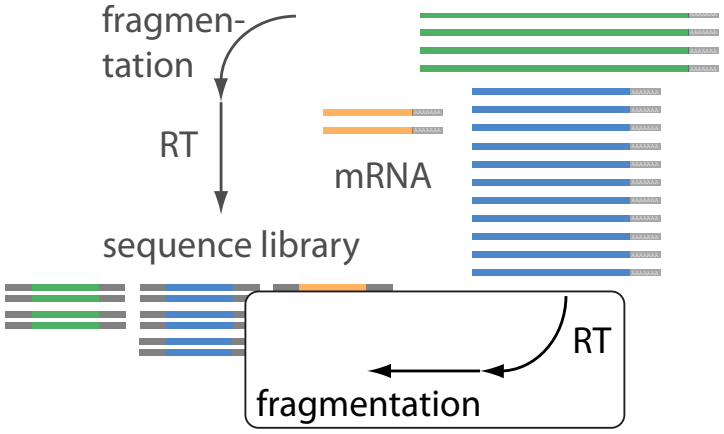


Data generated for *A. thaliana* with Flux Simulator v1.0 (Sammeth, 2009b)

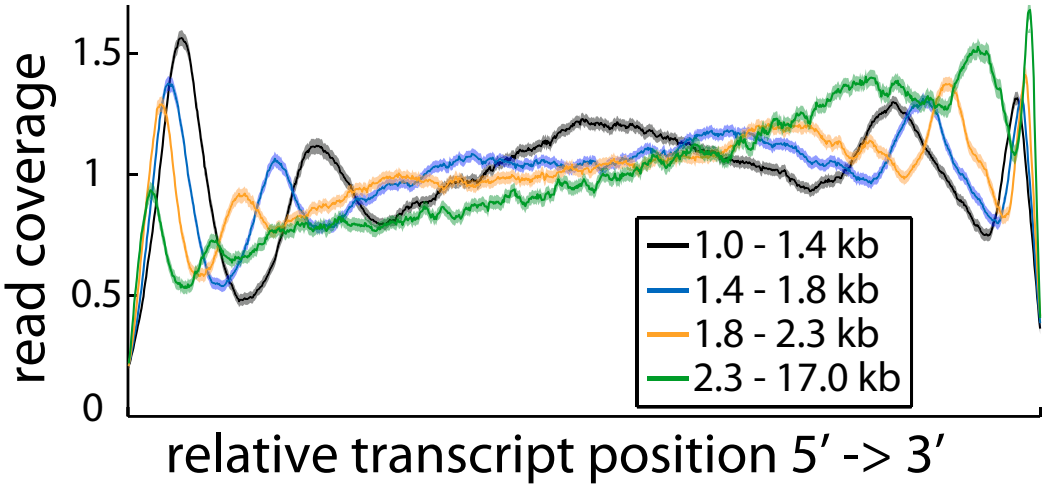
- Random priming
- Chemical cDNA fragmentation

Priming and Fragmentation Biases

Profile: normalised positional read coverage along the transcript



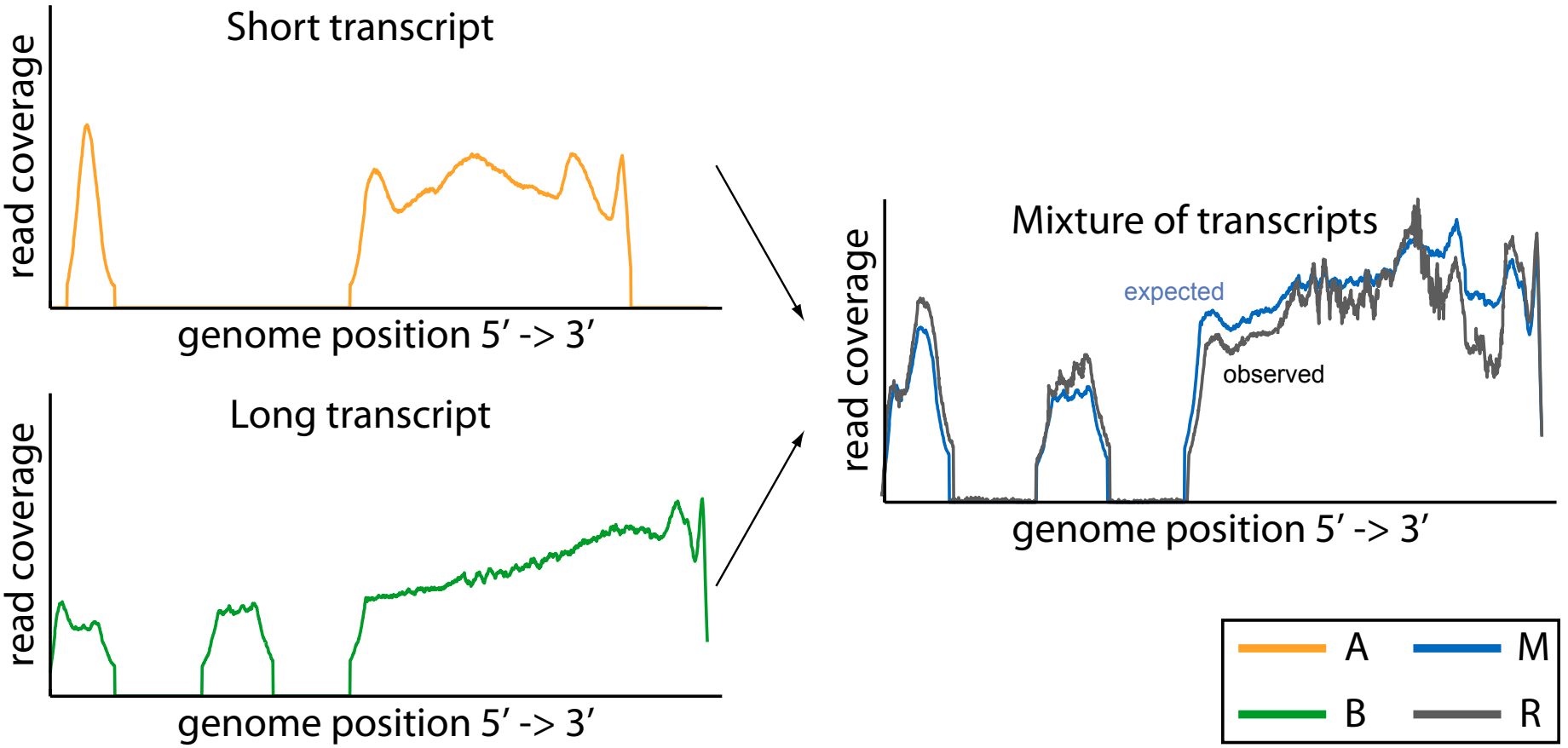
Transcript profiles for different lengths



Data generated for *A. thaliana* with Flux Simulator v1.0 (Sammeth, 2009b)

- Oligo-dT priming
- Physical cDNA fragmentation

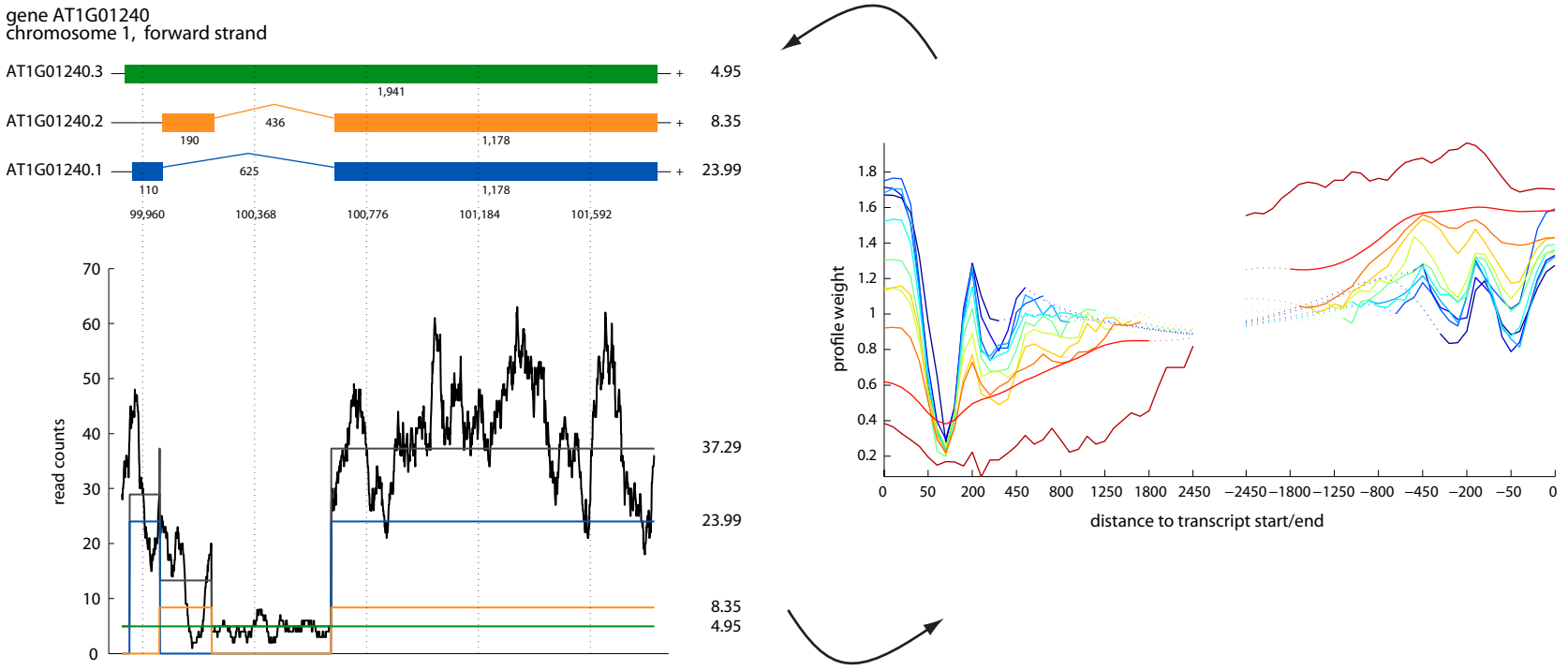
rQuant – Basic Idea



$$M_i = w_A A_i + w_B B_i \quad \Rightarrow \quad \min_{w_A, w_B} \sum_i \ell(M_i, R_i)$$

rQuant – Iterative Algorithm

1. Optimise transcript weights: $\min_{\mathbf{w}} \sum_i \ell\left(\sum_t w^{(t)} p_i^{(t)}, R_i\right)$
2. Optimise profile weights: $\min_{\mathbf{p}} \sum_i \ell\left(\sum_t w^{(t)} p_i^{(t)}, R_i\right)$
3. Repeat 1. and 2. until convergence.



Evaluation I

Our method rQuant: Position-wise, with profiles

compared to

- Position-wise, without profiles
- Segment-wise, without profiles (e.g. Jiang and Wong (2009))
- Segment-wise, with profiles (e.g. Flux Capacitor by Sammeth (2009a))

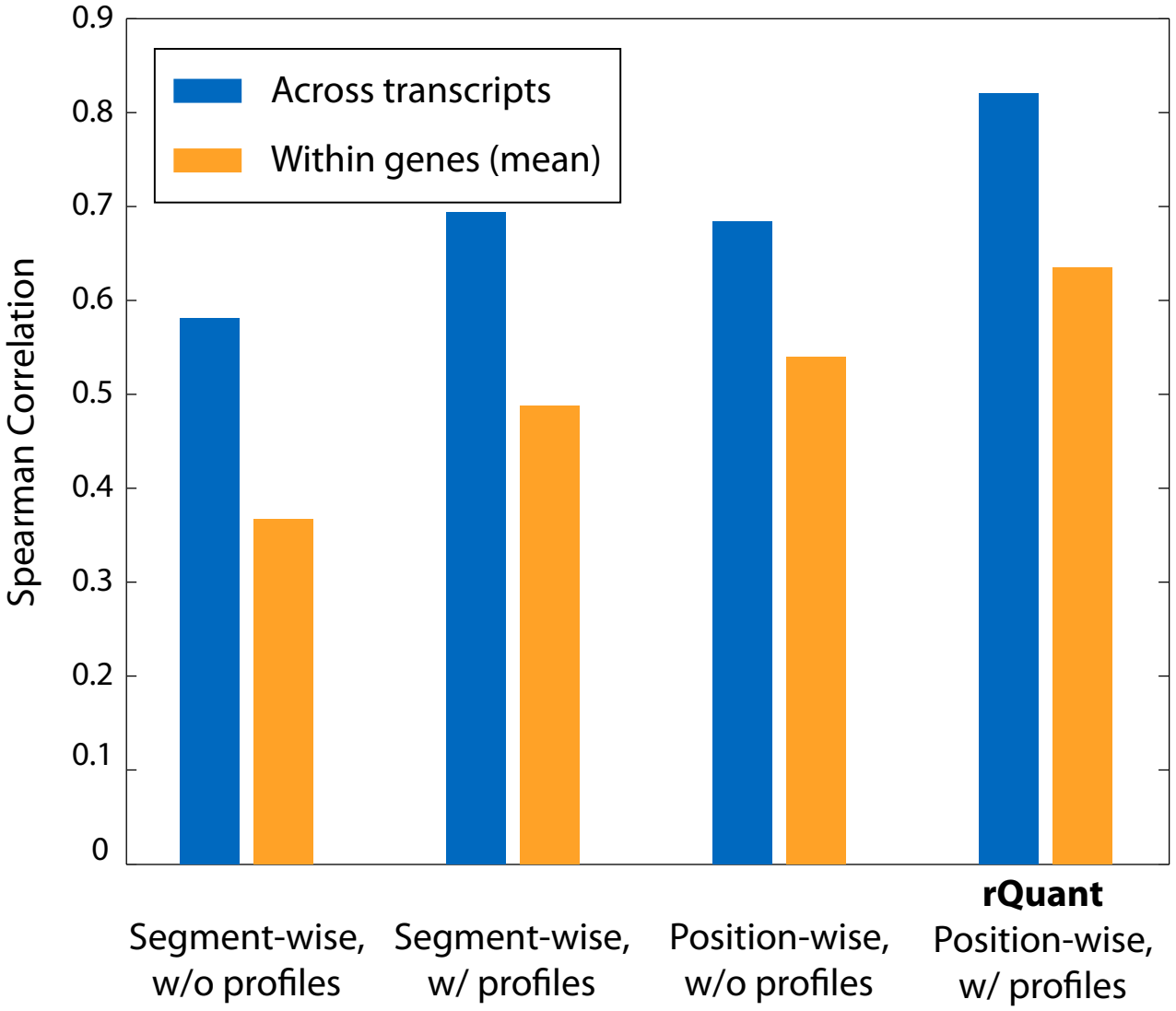
Estimate transcript abundances

- Using simulated data for *A. thaliana*
- Subset of alternatively spliced genes

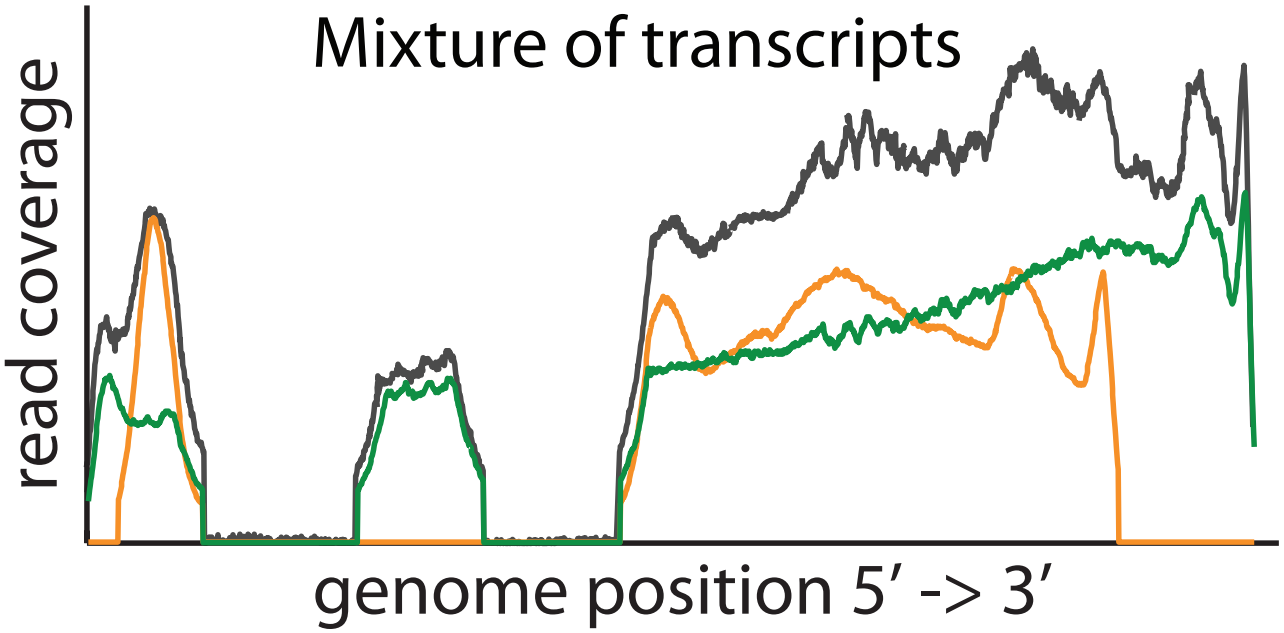
Evaluation: Spearman correlation between

- Simulated RNA expression level and
- Predicted transcript weights

Evaluation II



De Novo Transcript Quantification

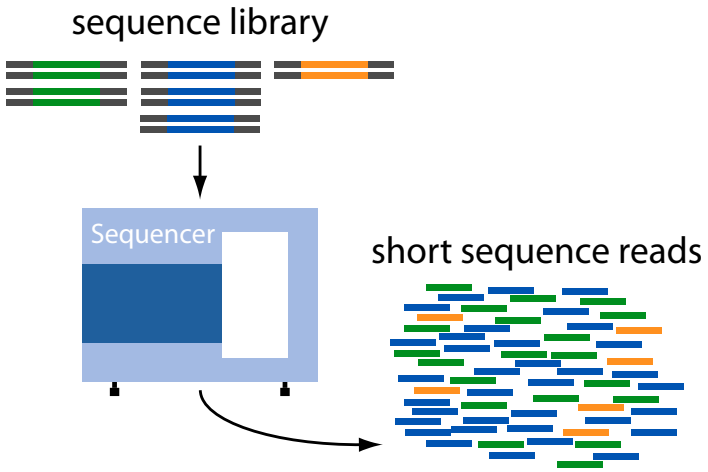


rQuant *de novo*: Simultaneously ...

1. Identify transcripts explaining the observed read coverage best.
2. Determine transcript weights.

Other Biases

PCR Bias



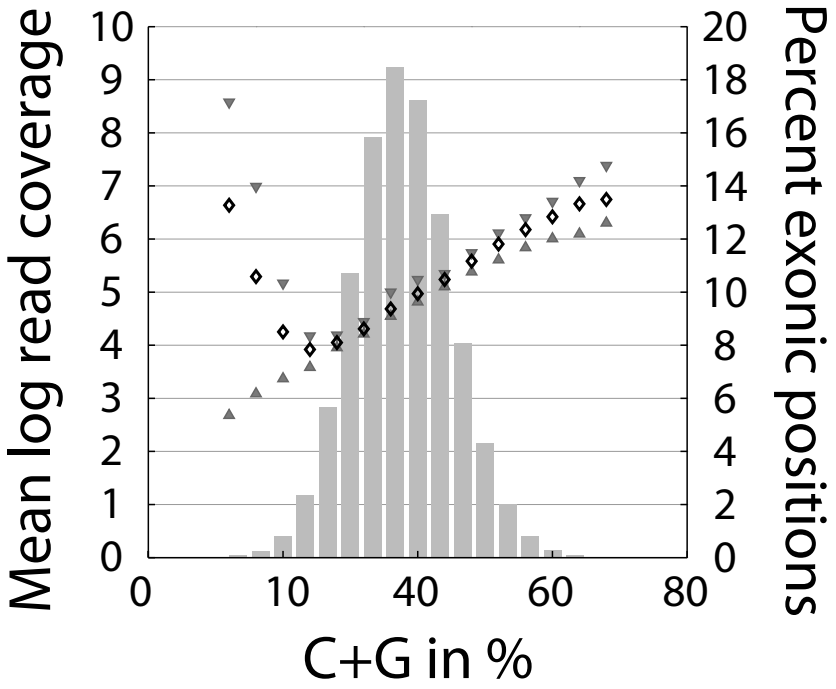


Figure provided by Georg Zeller

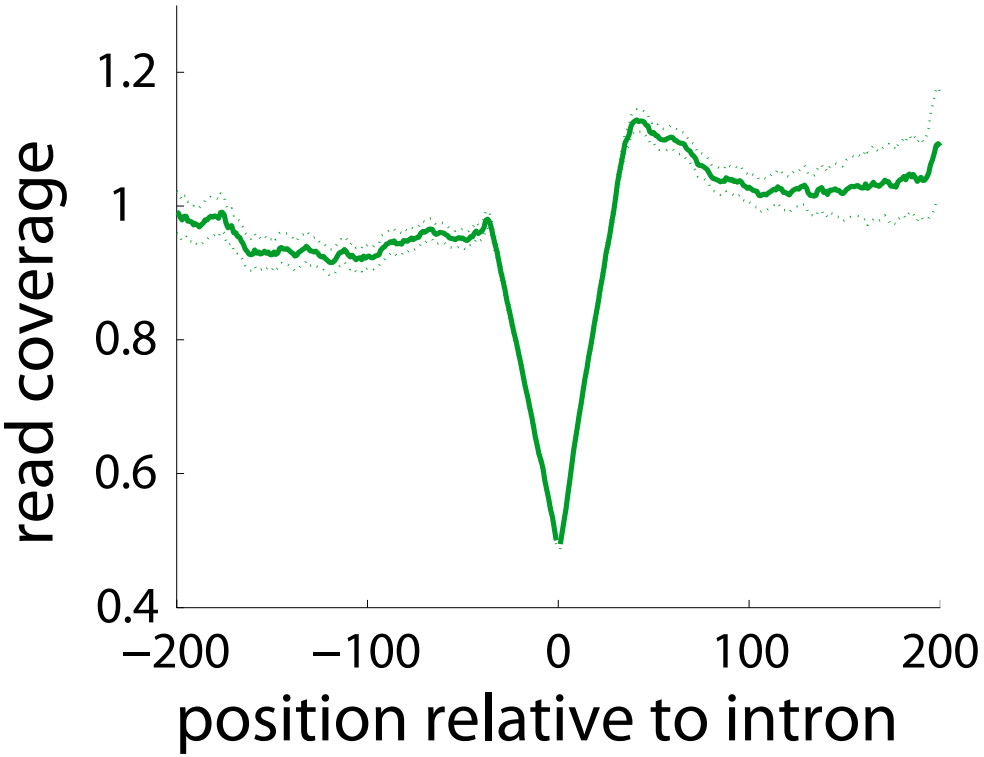
RNA-Seq data (*A. thaliana*)

- Exonic GC content

Other Biases

Read Mapping Bias





RNA-Seq data (*A. thaliana*)

- Exon boundaries

Conclusions



- Biases in RNA-Seq data need to be addressed
- rQuant quantifies transcripts with learned profiles
- Applicable to *de novo* quantitative transcript identification
- rQuant can be extended:
  - Take advantage of confirmed introns
  - Consider ambiguously mapped reads
  - Exploit paired-end data
- More information on
  `http://www.fml.mpg.de/raetsch/suppl/rquant`
- The slides can be found on
  `http://www.fml.mpg.de/raetsch/lectures`
- Poster on Monday (J17)

Thank you for your attention.

References

Fabio De Bona, Stephan Ossowski, Korbinian Schneeberger, and Gunnar Rätsch. Optimal spliced alignments of short sequence reads. *Bioinformatics (Oxford, England)*, 24(16):i174–180, August 2008.

Hui Jiang and Wing Hung Wong. Statistical Inferences for Isoform Expression in RNA-Seq. *Bioinformatics*, February 2009. doi: 10.1093/bioinformatics/btp113.

M. Sammeth. The Flux Capacitor. *Website*, 2009a. http://flux.sammeth.net/capacitor.html.

M. Sammeth. The Flux Simulator. *Website*, 2009b. http://flux.sammeth.net/simulator.html.

G. Schweikert, A. Zien, G. Zeller, J. Behr, C. Dieterich, C.S. Ong, P. Philips, F. De Bona, L. Hartmann, A. Bohlen, N. Krüger, S. Sonnenburg, and G. Rätsch. mGene: Accurate Computational Gene Finding with Application to Nematode Genomes. *Genome Research*, 2009. Under review.