Analysis of RNA kinetics

using

Alexey Uvarovskii, Christoph Dieterich

University Hospital Heidelberg

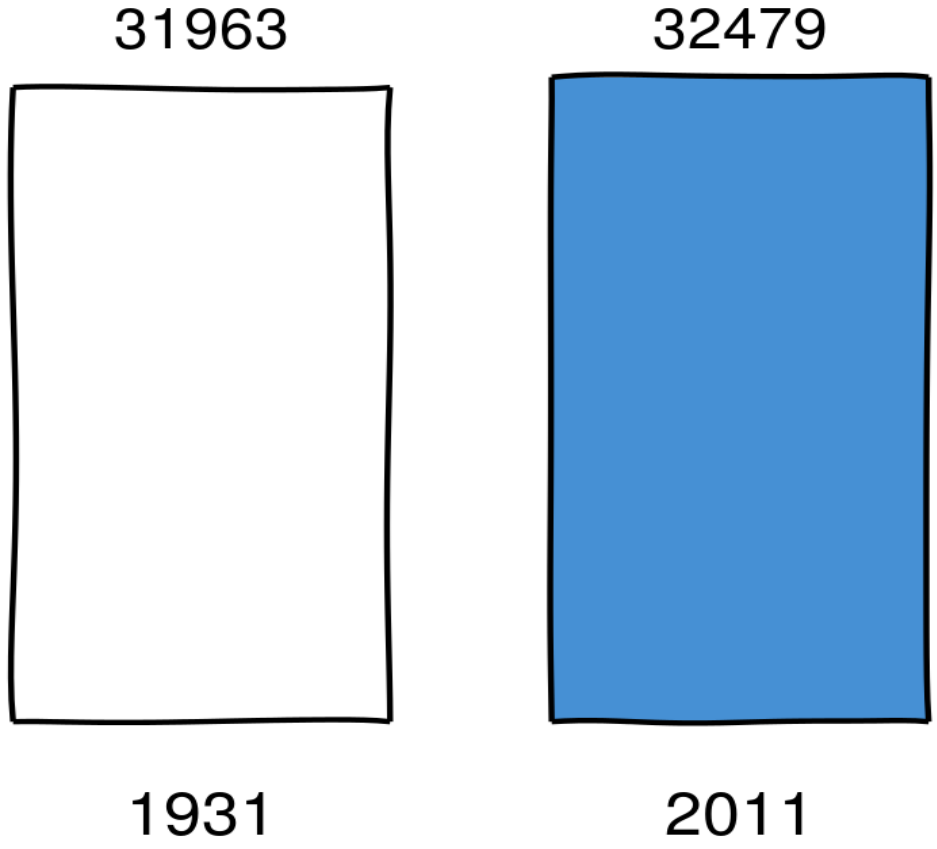
Acknowledgement

Tobias Jakobi (Dieterich Lab) - computing support

David Vilchez, Seda Koyuncu (CECAD Cologne),
Janine Altmüller, Marek Franitza (CCG Cologne) - experimental data

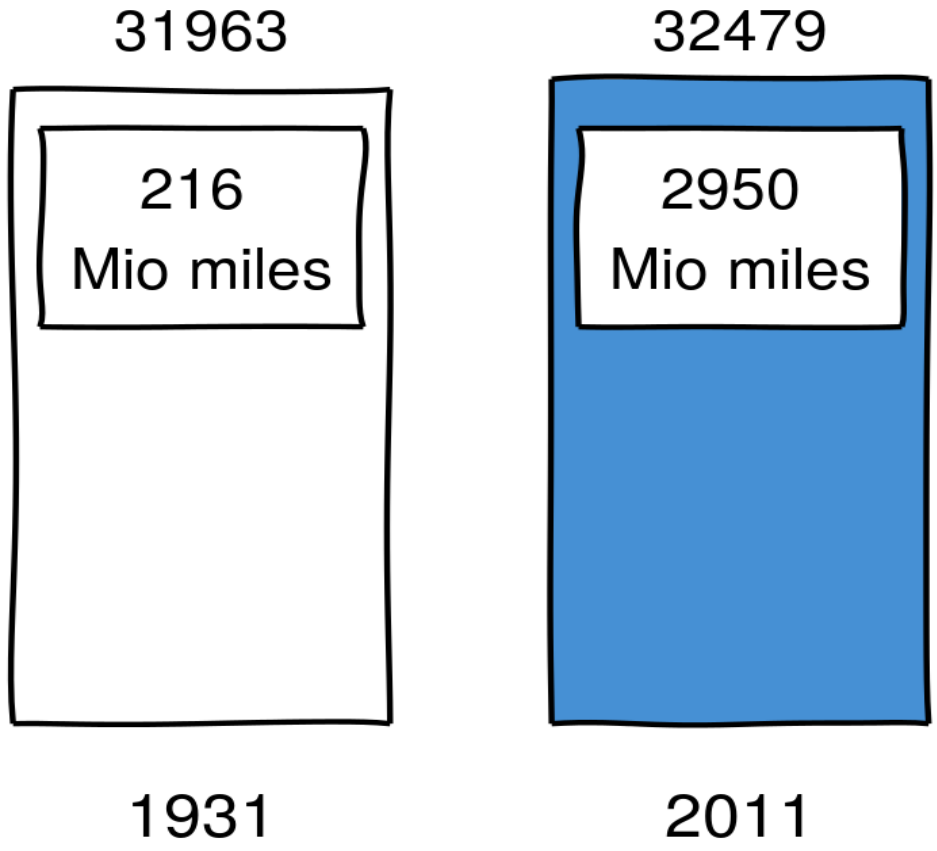
Is there a difference?

Number of motor vehicle deaths in the US (wiki)



Is there a difference?

Number of motor vehicle deaths in the US (wiki)

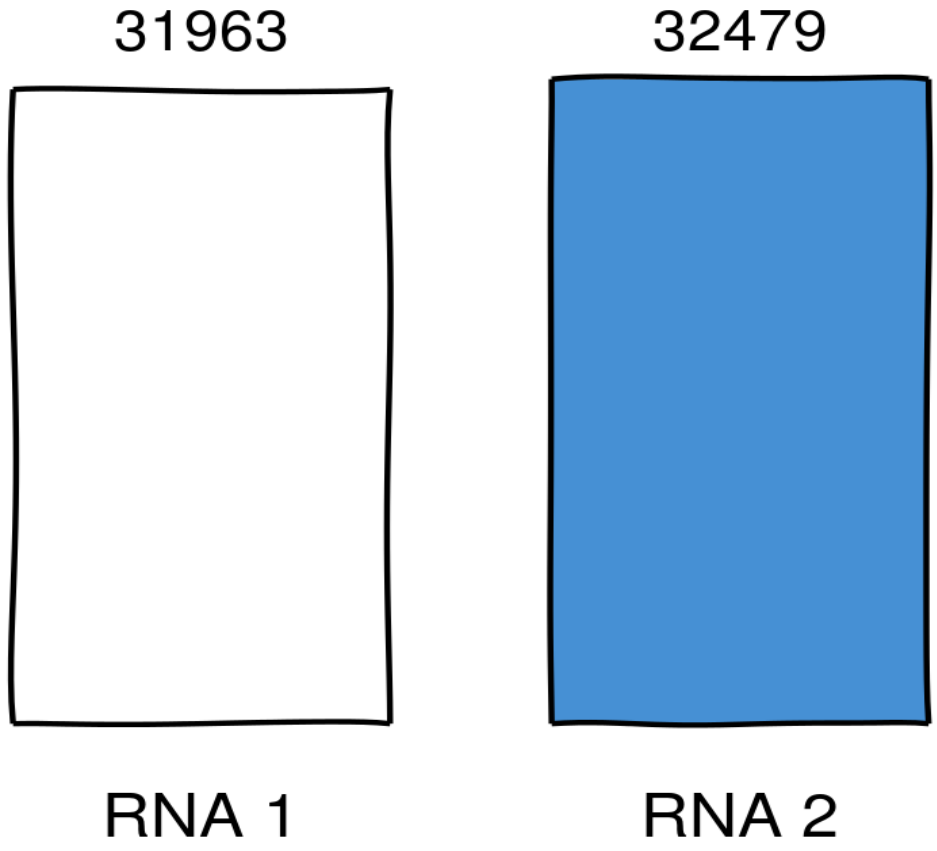


Same number, different context

I bet you think about

Rates

Reads in RNA-seq



RNA level is a balance

[synthesis] [degradation]

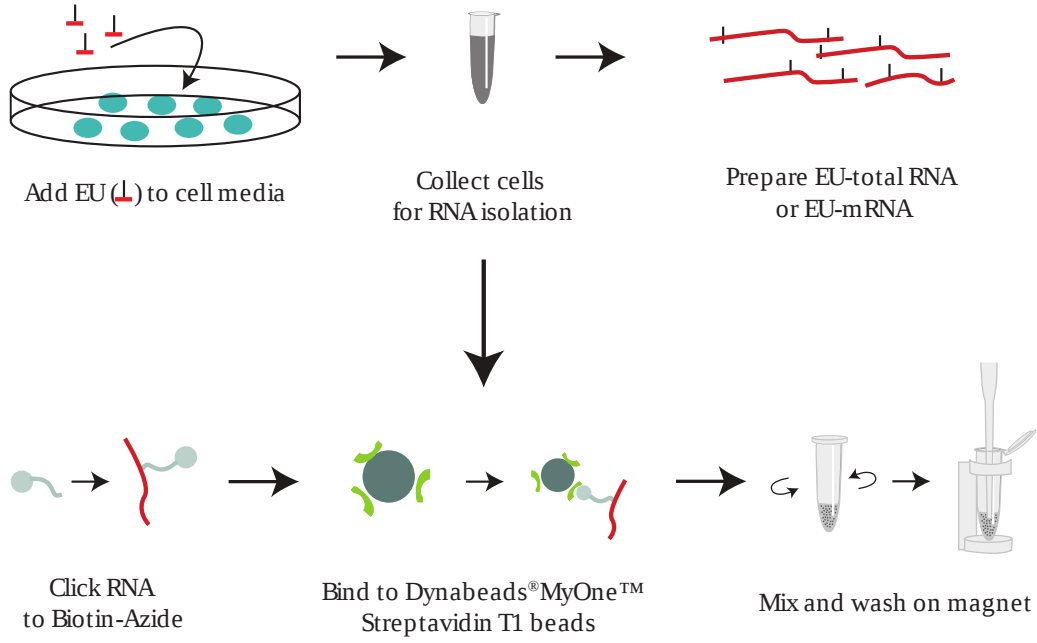
RNA level is a balance

$$\frac{d[RNA]}{dt} = +[synthesis] - [degradation] \cdot [RNA]$$

$$[steady\ state\ RNA] = \frac{[synthesis]}{[degradation\ ]}$$

Nascent RNA can be traced

EU = 5-ethyniluridine

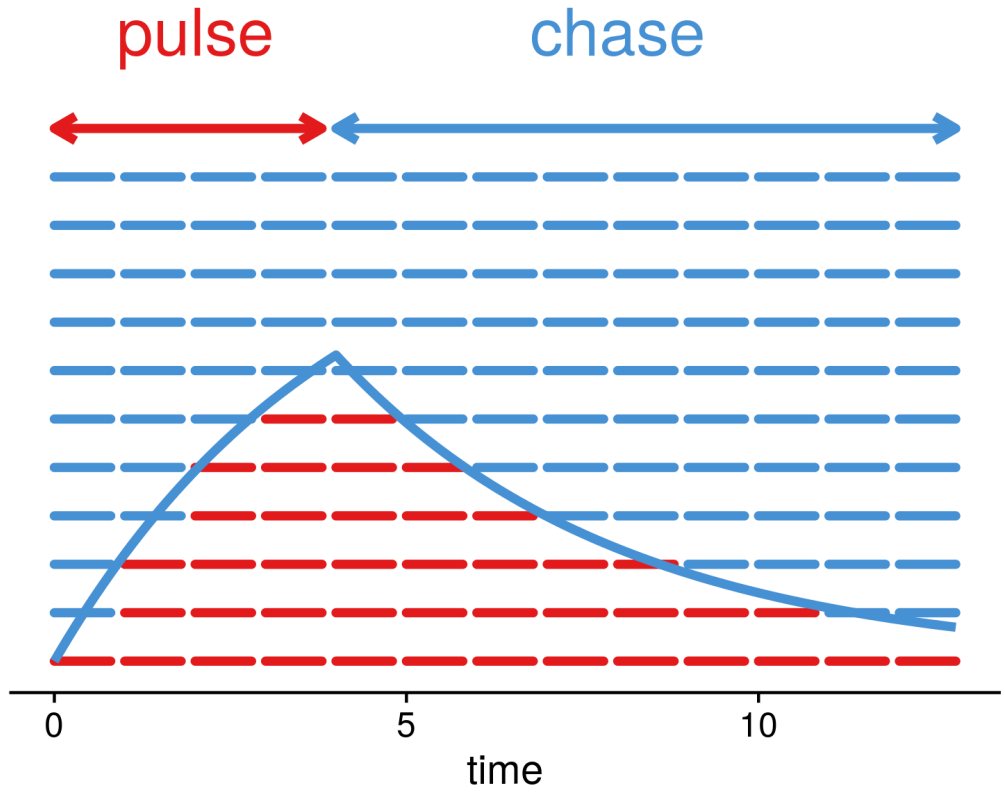


from the Click-iT® Kit Thermofisher manual

usually followed by RNA-seq

Pulse-chase experiment

a way to measure RNA kinetics



Background

- rates are also interesting
- new RNA can be traced
- pulse-chase RNA-seq

Analysis

- kinetic model
- stat model
- normalisation

`pulseR` to help

Alexey Uvarovskii, Christoph Dieterich; pulseR: Versatile computational analysis of RNA turnover from metabolic labeling experiments. Bioinformatics 2017 btx368. doi: 10.1093/bioinformatics/btx368

Kinetic model

defined by the setup, e.g. pulse labelling is

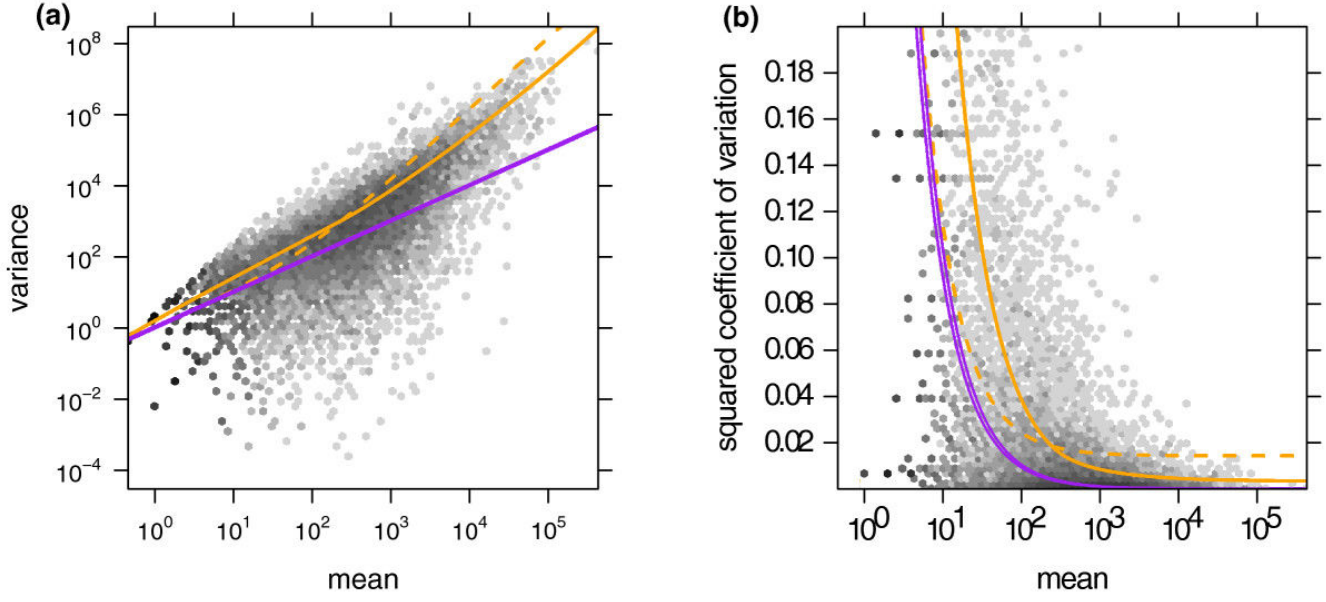
$$[\text{total}] = T \equiv \text{const}$$

$$[\text{pull down}] = T \cdot \left(1 - e^{-dt}\right)$$

$$\text{t} = 1, 2, 4\,\text{hr}$$

Stat model

Negative binomial distribution



purple: no overdispersion, yellow: with overdispersion

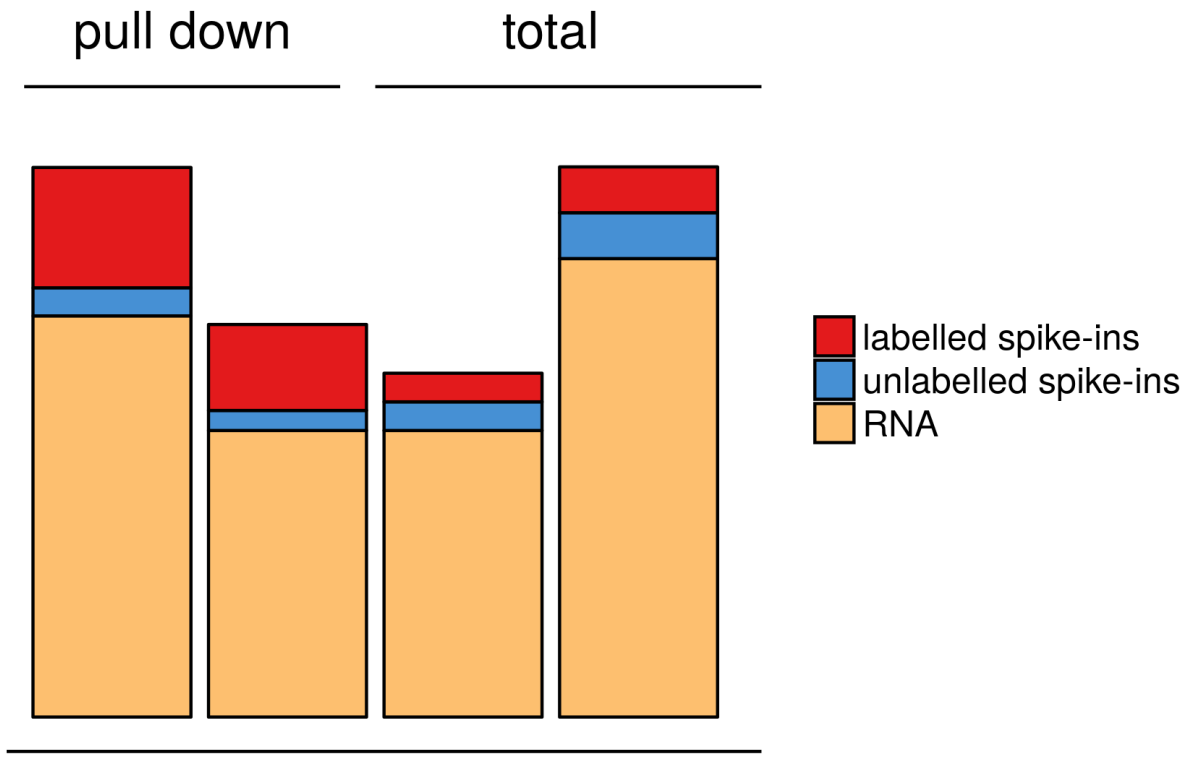
Anders, Simon, and Wolfgang Huber. "Differential expression analysis for sequence count data." Genome biology 11.10 (2010): R106.

Normalisation

$$[\text{total}] = [\text{labelled}] + [\text{unlabelled}]$$

$$[\text{pull down}] = ?[\text{labelled}] + ?[\text{unlabelled}]$$

Normalisation using spike-ins



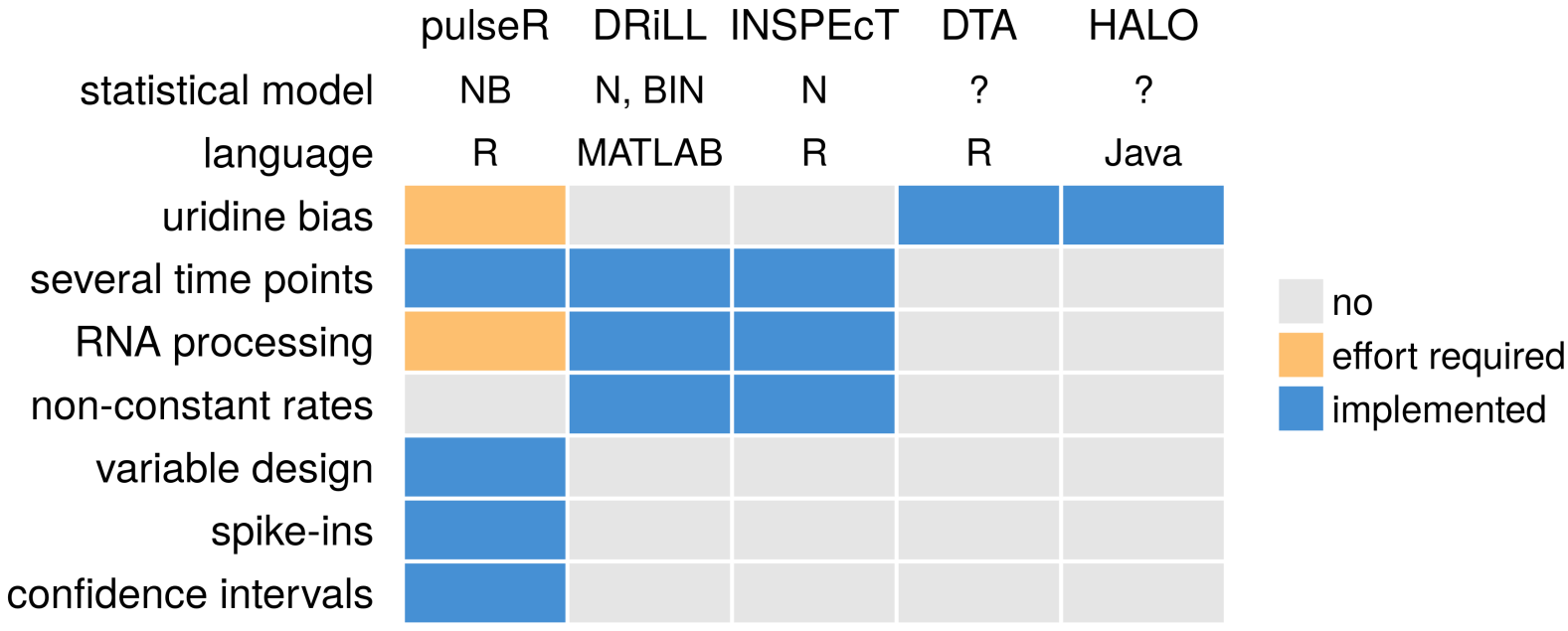
In `pulseR`

there are two options for normalisation:

using spike-ins (DESeq) absolute synthesis rate

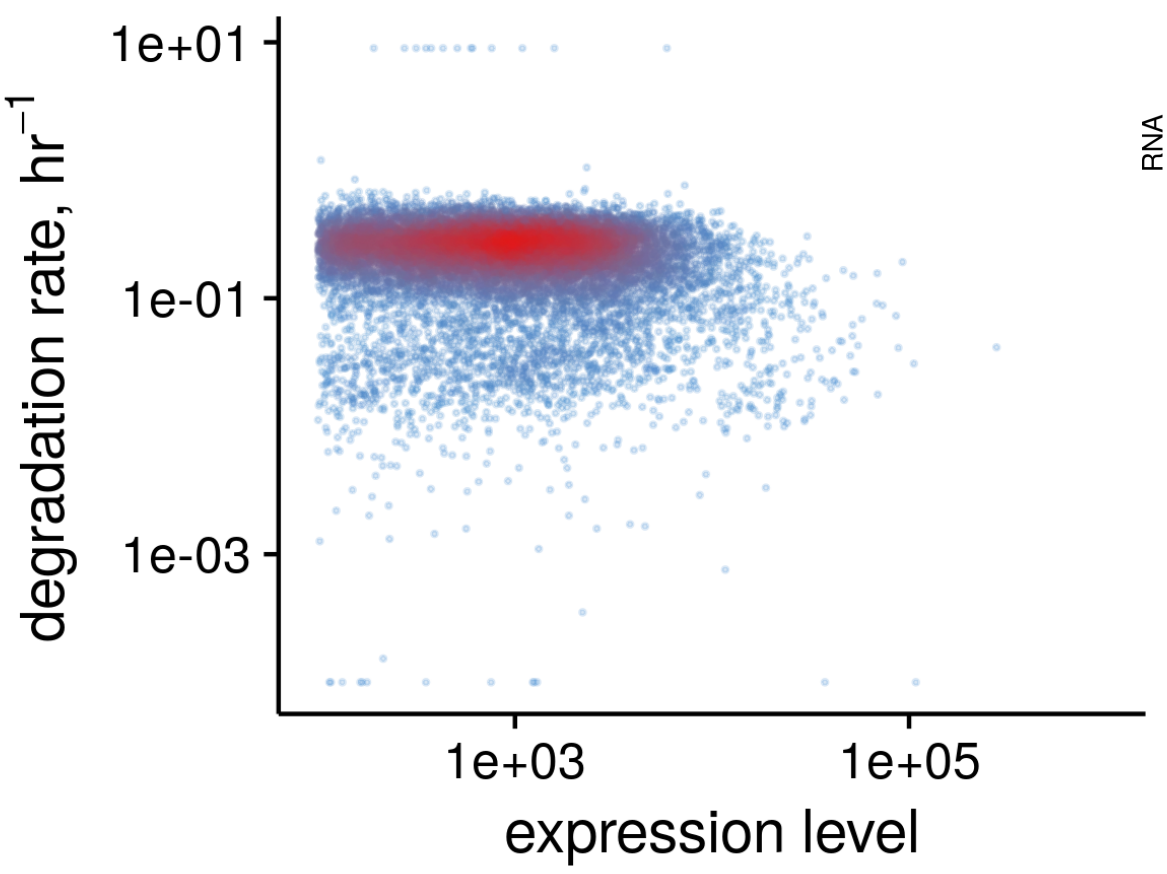
by MLE fitting no spike-ins needed

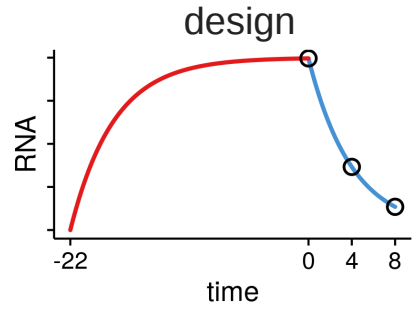
Alternatives



N: normal, NB: negative binomial, BIN: binomial.

EU pulse-chase on H9 cells



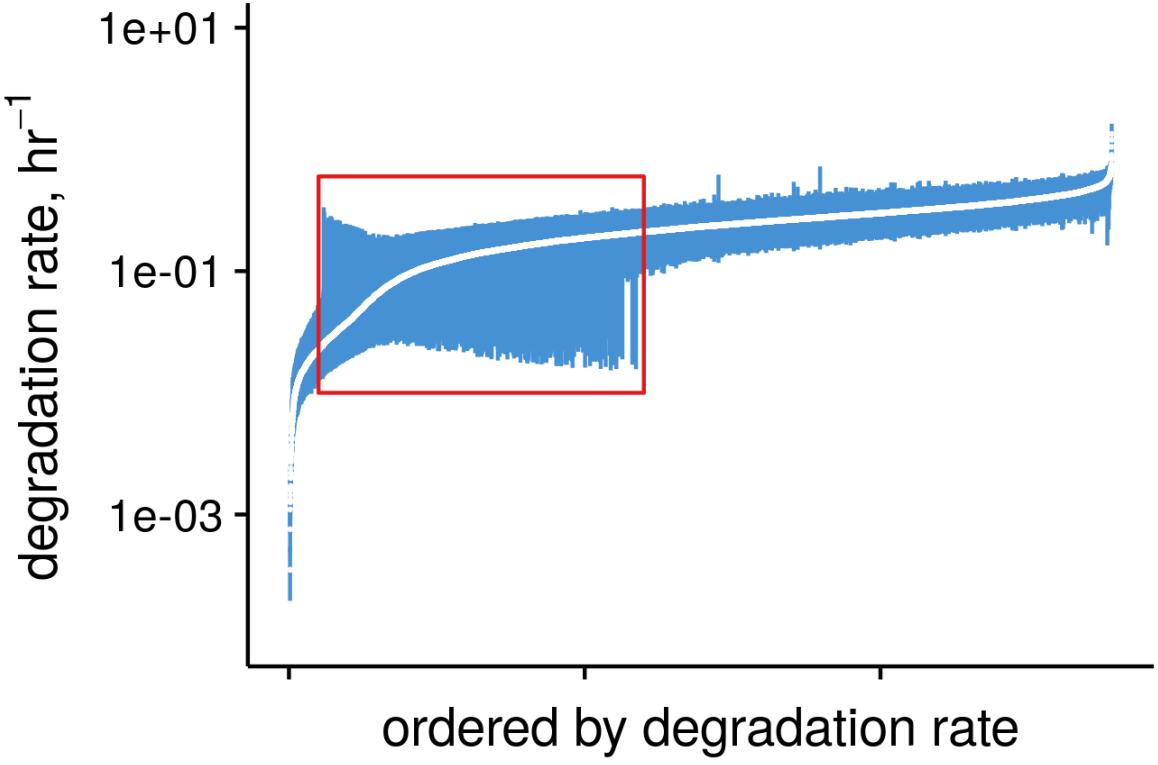


A number is not enough

`pulseR` can estimate confidence intervals for you

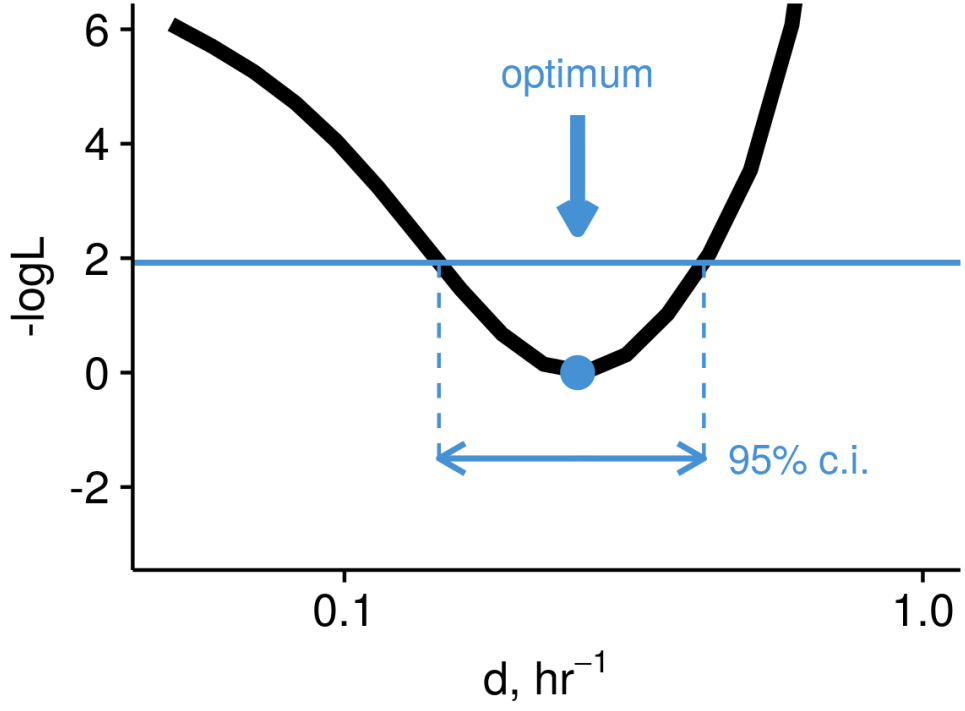
diagnostics and comparisons

Confidence intervals

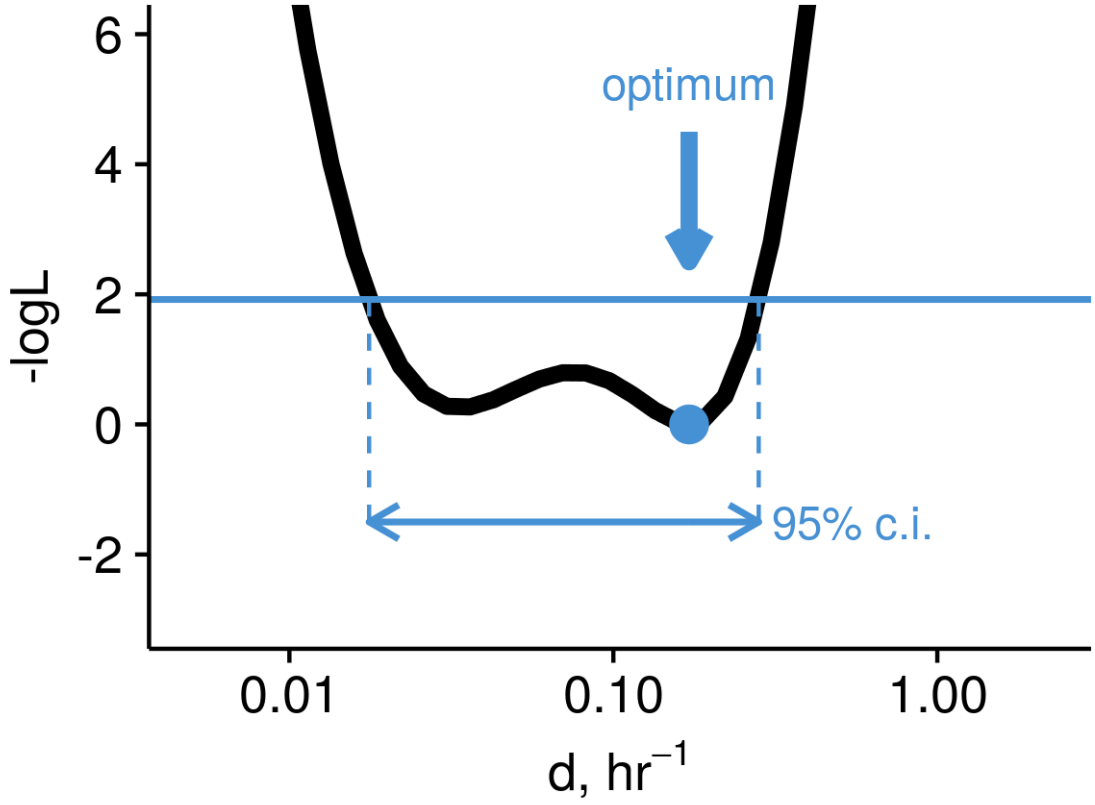


Profile likelihood

$$\frac{\log L(d)}{\log L(d_{optimal})} < \frac{1}{2}\chi^2_{0.95,1\ d.f.} \approx 1.92$$



Uncertainty in fit



The workflow

```
library(pulseR)
# put math here
formulas <- MeanFormulas(
  total = mu,
  labelled = mu * (1 - exp(-d*22)) * exp(-d*time),
  unlabelled = mu * (1 - exp(-d*time) * (1 - exp(-d * 22)))
)
# define the fractions
formulaIndexes <- list(
  total_fraction = 'total',
  pull_down      = c('labelled', 'unlabelled'))
```

```
pd <- PulseData(counts, conditions, formulas, formulaIndexes,
  groups = ~ fraction + time)
result <- fitModel(pd, initValues, opts)
```

`pulseR` allows to

- estimate kinetic rates from RNA-seq
- flexible analysis (spike-ins, cross-contamination, etc.)
- diagnostics with profile likelihood

Poster A-271

An open post-doc position

dieterichlab.org

github.com/dieterich-lab/pulseR

a.uvarovskii@uni-heidelberg.de