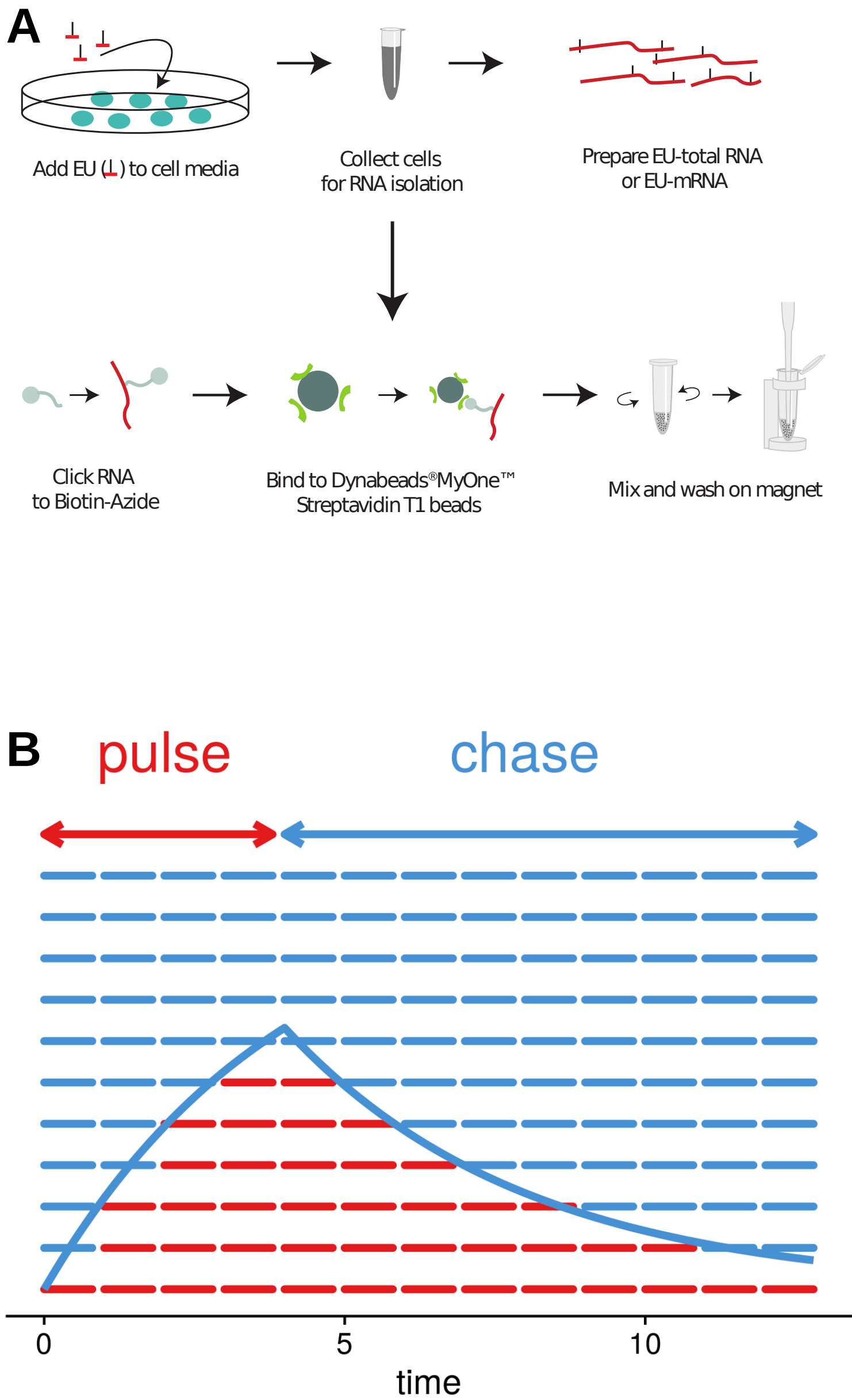


Summary

Gene expression level is defined by rates of RNA synthesis and degradation. Understanding how certain gene levels are regulated by these processes across different experimental conditions helps to gain deeper insights into RNA control mechanisms. Metabolic RNA labelling experiments facilitate to separate newly synthesized (i.e. labelled) from pre-existing RNA. Using this approach, pulse-chase time course experiments can be set up to estimate kinetic rates of RNA turnover from next-generation sequencing data. However, there is no publicly available software for estimating kinetic parameters to date, which is specifically designed to handle fragment count data from RNA sequencing experiments.

Measurement of RNA turnover



RNA labelling: There are two processes involved in RNA metabolism: RNA synthesis and degradation. In order to be able to measure the reaction rates, one needs to distinguish the newly synthesized RNA from the older one. Different approaches were developed, which are based on addition of modified uridine molecules, namely, 4-thiouridine (4sU), 5-ethyniluridine (EU) and 5'-bromo-uridine (BrU). The figure A is adapted from the Click-iT® Kit Thermofisher manual.

References:
Tani H, Akimitsu N. Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling. RNA Biology. 2012;9(10):1233-1238. doi:10.4161/rna.22036.

Pulse-chase experiments allow to measure kinetic rates for RNA synthesis and degradation. After the pulse period, when cells are cultured in the presence of the label, the chase period follows with no labelling (fig. B). Hence, the labelled RNA is being only degraded during the chase time.

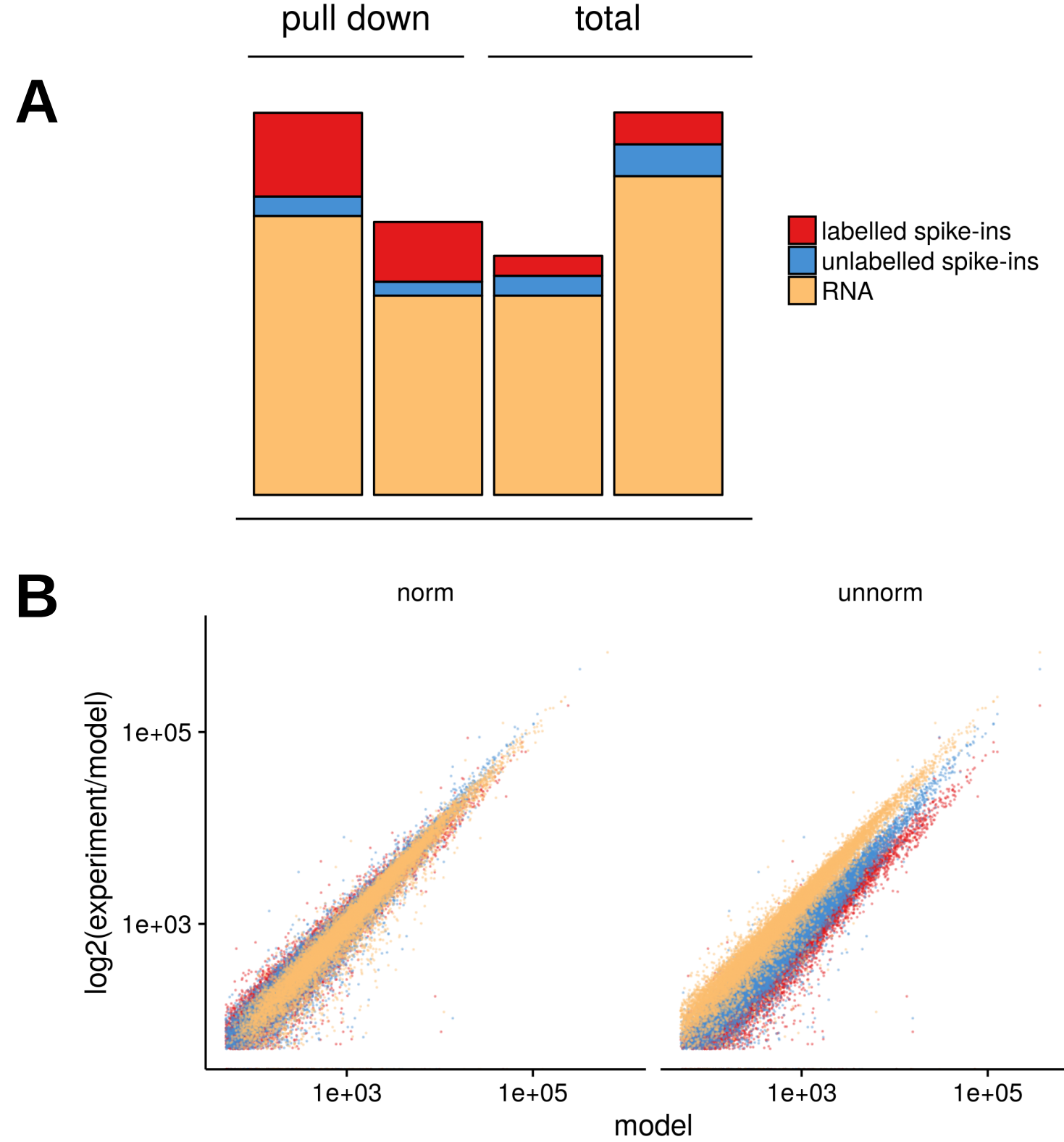
The set-up can be varied in terms of time ranges and types of collected fractions (e.g. all 3 fractions or only total and labelled ones)

Methods

Kinetic model: RNA dynamics can be described by ordinary differential equations, which have a simple analytic solution if the degradation and synthesis rates are assumed to be constant.

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

Stat model: We specifically consider overdispersion in RNA fragment counts by using the negative binomial distribution to model observations from RNA-seq data. The model is fitted using maximum likelihood estimation (MLE) approach.

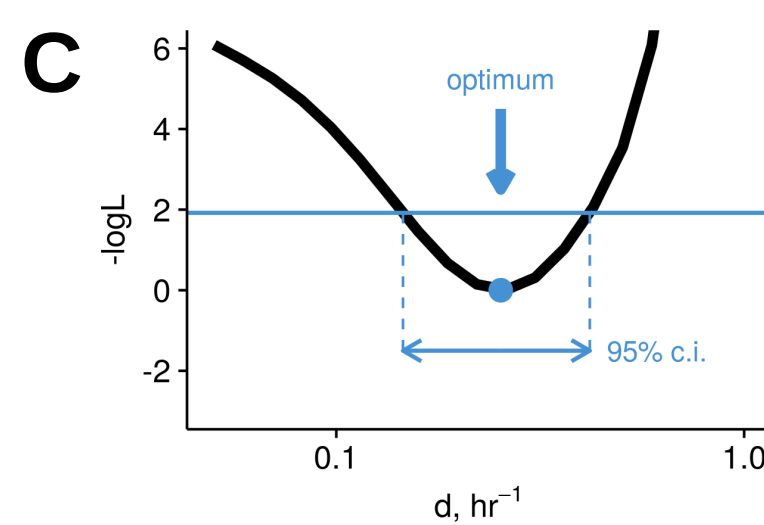


Confidence intervals (CI): The profile likelihood is used to estimate the CI for the model parameters.

In this approach, the parameter p under investigation is fixed to a certain value, while other parameters are left free. The likelihood function $L(p)$ is then a function of one argument and the 95% CI is a set of p such that

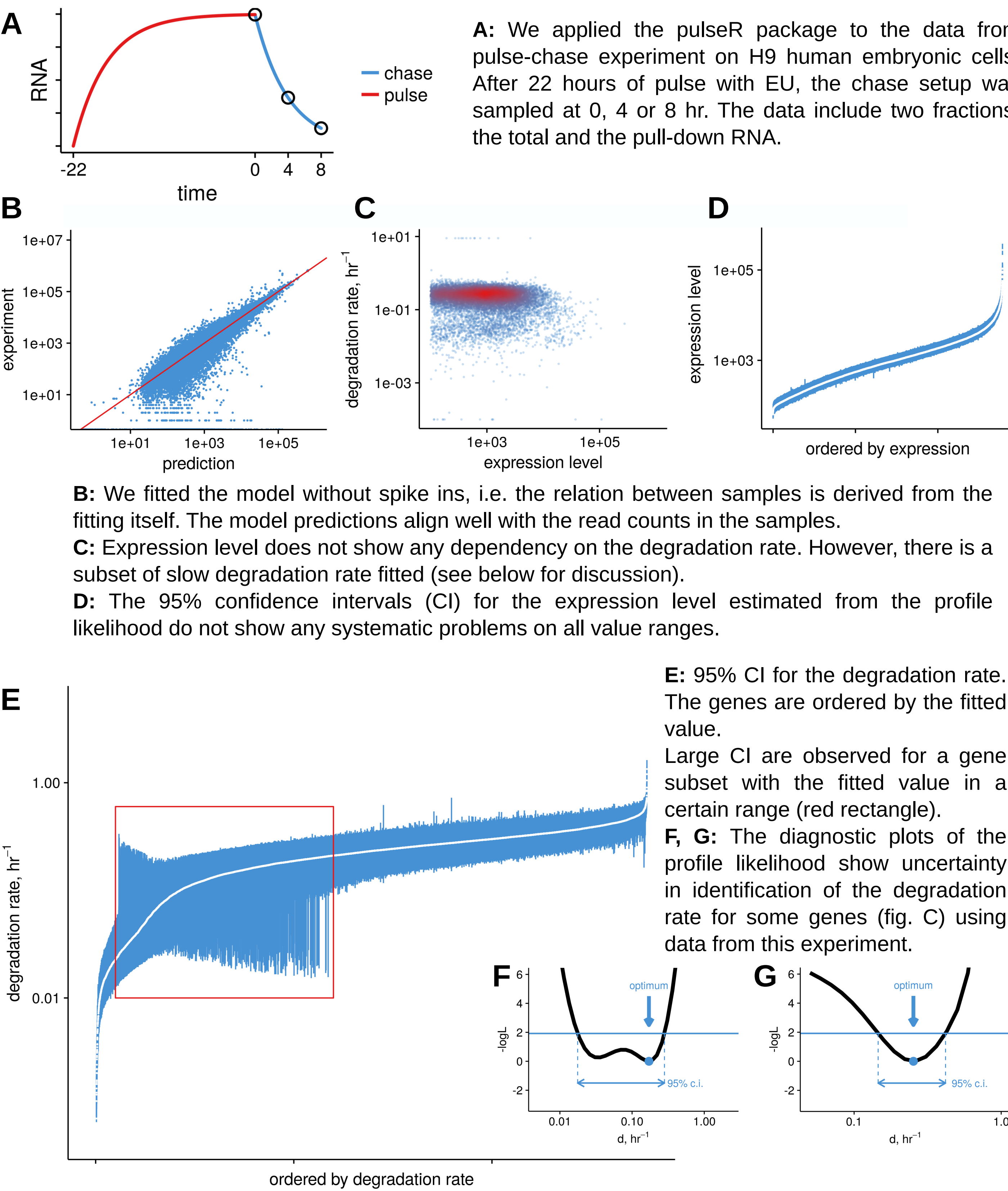
$$2 \frac{\log L(p)}{\log L(p^*)} < \chi_{1,0.95}^2$$

where p^* is the optimal parameter value and the r.h.s is the 95% quantile for the chi-squared distribution with 1 degree of freedom (approx. 3.84), fig. C.



Normalisation: due to difference in sequencing depth and possible cross-contamination (fig. A), it is important to make adjustments in estimations. We use DESeq normalisation if spike ins are provided. Alternatively, we estimate sample relations by MLE. This helps to eliminate sample difference coming from the preparation (fig. B).

Application to H9 cells



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')
)
lbNormFactors <- list(
  total_fraction = .1,
  pull_down = c(.1, .00001)
)
ubNormFactors <- list(
  total_fraction = 10,
  pull_down = c(10, .3)
)
opts <- setBoundaries(list(
  mu = c(.1, 1e8),
  d = c(1e-4, 9),
  size = c(1, 1e6)),
  normFactors = list(lbNormFactors, ubNormFactors)
)
# let conditions be in a data.frame (sample, fraction, time)
pd <- PulseData(counts, conditions, formulas, formulaIndexes,
  groups = ~ fraction + time)
result <- fitModel(pd, initParValues, opts)
```

The typical analysis workflow includes:

definition of the kinetic equations

definition of the fraction content, e.g. the pull-down can be contaminated with the unlabelled fraction

setting lower and upper boundaries for the parameters, as well as fitting options

creation of the PulseData object and, finally, fitting the model

| | pulseR | DRiLL | INSPEcT | DTA | HALO |
|------------------------------|--------|--------|---------|-----|------|
| statistical model | NB | N, BIN | N | — | — |
| spike-ins | + | — | — | — | — |
| several time points | + | + | + | — | — |
| variable design ¹ | + | — | — | — | — |
| non-constant rates | — | + | + | — | — |
| uridine bias | * | — | — | + | + |
| RNA processing | * | + | + | — | — |
| confidence intervals | + | — | — | — | — |
| language | R | MATLAB | R | R | Java |

Table 1. Feature comparison of software packages for parameter estimation in pulse-chase experiments. +: available, -: not implemented, N: normal, NB: negative binomial, BIN: binomial.

*: must be defined by user. ¹: pulse, chase or combination there of experiments.

For references and further information see:

- Alexey Uvarovskii, Christoph Dieterich: pulseR: Versatile computational analysis of RNA turnover from metabolic labeling experiments. Bioinformatics 2017 btx368. doi: 10.1093/bioinformatics/btx368
- Wachutka, Leonhard, and Julien Gagneur. "Measures of RNA metabolism rates: Toward a definition at the level of single bonds." Transcription 8.2 (2017): 75-80.

Acknowledgement

We would like to thank Tobias Jakobi (Dieterich Lab) for the computing support, David Vilchez and Seda Koyuncu (CECAD Cologne), Janine Altmüller and Marek Frantiza (CCG Cologne) for providing the experimental data.