Measurement of genome-wide RNA synthesis and decay rates with Dynamic Transcriptome Analysis (DTA)

Björn Schwalb, Daniel Schulz, Mai Sun, Benedikt Zacher, Sebastian Dümcke, Dietmar Martin, Patrick Cramer, Achim Tresch

1 Gene Center Munich and Department of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany.

Associate Editor: Prof. Ivo Hofacker

ABSTRACT

Summary: Standard transcriptomics measures total cellular RNA levels. Our understanding of gene regulation would be greatly improved if we could measure RNA synthesis- and decay rates on a genome-wide level. To that end, the Dynamic Transcriptome Analysis (DTA) method has been developed. DTA combines metabolic RNA labeling with standard transcriptomics to measure these rates in a precise and non-perturbing manner. Here we present the open source R/Bioconductor software package DTA. It implements all required bioinformatics steps that allow the accurate absolute quantification and comparison of RNA turnover.

Availability: DTA is part of R/Bioconductor. To download and install DTA refer to http://bioconductor.org/packages/2.10/bioc/html/DTA.html.

Contact: schwalb@lmb.uni-muenchen.de

1 INTRODUCTION

Total RNA levels are a consequence of RNA synthesis and decay. These individual contributions can be monitored by Dynamic Transcriptome Analysis (DTA, Miller et al., 2011, Amorim et al., 2010, Friedel et al., 2010). DTA requires culturing cells in the presence of a labeling substrate (e.g. 4sU or 4tU). During a short, non-perturbing RNA labeling pulse, cells incorporate 4tUTP into newly transcribed RNA instead of uridine. This setup yields three types of RNA fractions: total cellular RNA, newly transcribed labeled RNA and pre-existing unlabeled RNA. The quantification of these fractions on microarrays or by RNA-seq is used to estimate RNA synthesis and decay rates on a genome-wide scale, assuming exponential decay. For each RNA g, the synthesis rate \( \mu_g \) and the decay rate \( \lambda_g \) are estimated from the equations

\[
e^{-\lambda_g t} = \frac{U_g}{T_g} = 1 - \frac{L_g}{T_g}, \quad T_g = \frac{\mu_g}{\lambda_g + \alpha}, \quad \alpha = \frac{\log(2)}{CCL},
\]

where \( t \) is the labeling duration, \( CCL \) the cell cycle length, \( L_g, U_g, T_g \) are the measurements of the labeled, unlabeled, and total RNA fractions, respectively. \( \alpha \) can be set to 0 in the case of primary cells (e.g. macrophages). It has been shown that DTA has higher sensitivity and higher temporal resolution in detecting gene regulatory changes than standard transcriptomics (Miller et al., 2011). One of the technical obstacles of standard transcriptomics is the unknown normalization factor between samples, i.e. wild-type and mutant cells. Variations in RNA extraction efficiencies and amplification steps in the biochemical protocol introduce differences in the global intensity levels. The estimation of the normalization factors limits the precision of DTA. We have extended DTA to comparative DTA (cDTA), to eliminate this obstacle (Sun et al., 2011), cDTA allows for direct comparison of RNA synthesis and decay rates between samples. This is done by mixing an internal normalization standard of S.pombe (Sp) cells to samples of wild-type or perturbed S.cerevisiae (Sc) cells. Here we present the R/Bioconductor package DTA. It implements all required methods to estimate RNA synthesis and decay rates from pre-processed microarray or RNA-Seq measurements of any organisms that are obtained via the DTA or the cDTA protocol.

2 IMPLEMENTATION AND AVAILABILITY

DTA is implemented in R (version ≥ 2.14, Ihaka et al., 1996) and is part of Bioconductor (version ≥ 2.10, Gentleman et al., 2004). To download and install DTA and all its dependencies refer to http://bioconductor.org/packages/2.10/bioc/html/DTA.html.

Other methods have been published by Friedel et al., 2010 and Rabani et al., 2010 to infer RNA decay rates from gene expression data. So far, only Friedel et al. 2010 have provided a Java implementation of their estimation procedure. The DTA package is the first to cover all aspects of the RNA labeling technique: Bias correction, detailed visualisation of quality control aspects and proper handling of the results. It contains simulation methods and functions for error estimation based on replicates and resampling. Further it contains 4 example datasets for Yeast, Mouse and Human (Miller et al., 2011, Sun et al., 2011 and Doelken et al., 2008).

3 USAGE AND APPLICATION

DTA can be used for genome-wide synthesis and decay rate estimation under all kinds of perturbations and for all kinds of organisms. The rate estimation procedure can be easily accessed via the core function DTA.estimate. Its input consists of a matrix containing the normalized measurements of the total, labeled and unlabeled RNA fractions (datamat) and a description of the
consider the parameter essential to check the reliability of the results. RStudio users should of quality control plots, if then the estimation procedure to yield the reported median wild-type mRNA

Note that it is advisable to experimentally determine the median cell cycle length (ccl) of a sample if appropriate as a correction factor. Additionally the ratio of labeled to total mRNA can be passed to the estimation procedure to yield the reported median wild-type mRNA half-lives of (Miller et al., 2011). The same ratio is then used to estimate the mRNA half-lives of the rpb1-N488D strain (Pol). This is reasonable as the data has been preprocessed via added Sp cells (DTA.normalize). To yield Figure 1, just execute the following:

```r
> r = Sc.ensg.reliable
> x = log2(Pol$d$r[r]/Wt$d$r[r])
> y = log2(Pol$s$r[r]/Wt$s$r[r])
> heatscatter(x,y,xlim=c(-4,1),ylim=c(-4,1))
```

**Fig. 1.** Heat scatter plot of the changes in RNA decay rates (log fold, x-axis) vs. the changes in RNA synthesis rates (log fold, y-axis) in the rpb1-N488D strain vs. its isogenic wild-type. The parallel offsets of the black lines to their grey counterparts indicate the median changes of synthesis rates (horizontal), decay rates (vertical), or total RNA levels (diagonal) by a factor of 0.31, 0.25, and 0.77, respectively. The coloring indicates the local point density.

**ACKNOWLEDGEMENTS**

We thank Kerstin C. Maier, Martin Seizl, Laurent Larivi`ere, Nicole Pirkl and Stefanie Etzold (Gene Center Munich) for stimulating discussions and great experimental work.

**Funding:** AT was supported by the LMUexcellent guest professorship “Computational Biochemistry”. PC was supported by the Deutsche Forschungsgemeinschaft, SFB646, TR5, FOR1068, NIM, EMBO, ERC, BIN, a LMUexcellent research professorship “Molecular systems biology of gene regulation”, and the Jung-Stiftung.

**REFERENCES**


The raw data is first normalized, mapped to Yeast ORF IDs and then the estimation procedure DTA.estimate generates a series of quality control plots, if check = TRUE. These are absolutely essential to check the reliability of the results. RStudio users should consider the parameter RStudio.

```r
> cDTA.datamat = DTA.normalize(Raw.datamat, Sp.affy.reliable)
> Sc.datamat = DTA.map.it(cDTA.datamat, map = Sc.affy2ensg)
> Wt = DTA.estimate(Wt.phenomat,Sc.datamat, Sc.tnumber,Sc.ensg.reliable,ccl = 93.5, LtoTratio = 0.05,check = FALSE)
> Pol = DTA.estimate(Pol.phenomat,Sc.datamat, Sc.tnumber,Sc.ensg.reliable,ccl = 149.8, LtoTratio = 0.05,check = FALSE)
```

4 CONCLUSION

The *DTA* package delivers straightforward methods to estimate RNA synthesis and decay rates from pre-processed microarray or RNA-Seq measurements that are obtained via the *DTA/cDTA* protocol. The *DTA* package fulfills the high standard of the *Bioconductor* platform, regarding documentation and usability. It can therefore be easily incorporated in R scripts for pre-processing and further statistical analysis of the results can readily be carried out by other methods within the R/Bioconductor programming environment.