Gene Expression

mRNAStab - A web application for mRNA stability analysis

Andrei Alic[1], José E. Pérez-Ortín[2], Joaquín Moreno[2] and Vicente Arnau[1, *]

[1]Department of Informatics and [2]Department of Biochemistry & Molecular Biology, Universitat de València, 46100 Burjassot. Spain.

ABSTRACT

Eukaryotic gene expression is regulated both at the transcription and the mRNA degradation levels. The implementation of functional genomics methods that allow the simultaneous measurement of transcription (TR) and degradation (DR) rates for thousands of mRNAs is a huge improvement in this field. One of the best established methods for mRNA stability determination is Genomic Run-On (GRO). It allows the measurement of DR, TR and mRNA levels during cell dynamic responses. Here we offer a software package that implements two new algorithms to determine mRNA stabiliiesy during dynamic GRO experiments.

Availability: The program mRNAStab is freely accessible at http://mRNAStab.uv.es/.

Implementation: mRNAStab is written in C, PHP and R

Contact: vicente.arnau@uv.es

1 INTRODUCTION

mRNA is the messenger molecule transferring information from genes to proteins. It is by its own nature and function quite unstable (see Pérez-Ortín *et al*., 2007 for discussion). There is a variable turnover that depends on synthesis (transcription rate, TR) conducted by RNA polymerase II and degradation rate (DR) conducted by a complex system of RNases and other proteins or RNAs (Parker, 2012). The importance of TR in determining the levels of mRNAs is widely documented, however, the parallel importance of DR in it is only starting to be recognized (Pérez-Ortín *et al.*, 2007).

There are several methods currently available for DR determination. Some of them are able to measure mRNA stabilities genome-wide (see Pérez-Ortín *et al*, 2011 and Supplementary data for a critical description and comparison). The most popular although prone to mistakes and biases is the transcription shut-off and subsequent determination of mRNA decay: assuming first-order kinetics, the degradation constant (k_d) and the mRNA half-life (HL= ln2/ k_d) can be calculated.

The TR and mRNA concentrations can be determined genome-wide by using a genomic run-on protocol (GRO) which allows to estimate the HLs of mRNAs even under non steady-state conditions (García-Martínez *et al.*, 2004; Pérez-Ortín *et al.*, 2007). This possibility is extremely interesting because it allows to determine TR, mRNA levels and HLs in a single experiment during the fast transcriptional response to an imposed environmental change (Castells-Roca *et al.*, 2011 and references therein). This is the simplest protocol currently available for monitoring mRNA turnover in dynamic situations. However, it involves complex numerical analysis and costly computing. Now we describe here the mRNAStab package which offers two alternative algorithms (programs "StepK" and "RampK") for determining mRNA HLs from data generated by GRO experiments. Additionally, it also includes a program ("Shutoff") to calculate k_d from simple decay experiments (e.g. Wang *et al.*, 2002). All three programs are devised to cope with genome-wide data (TR and/or [mRNA] time sequences for thousands of genes).

2 WHAT THE PROGRAMS DO

Typically, the GRO experiment delivers values of TR (TR_0, TR_1, TR_2…) and mRNA concentration ($[mRNA]_0$, $[mRNA]_1$, $[mRNA]_2$…) simultaneously determined at discrete times (t_0, t_1, t_2. …). By assuming a linear variation of TR, the following relation between consecutive pointwise determinations (e,g., values of TR and [mRNA] at t_0 and t_1) has been demonstrated to hold (Pérez-Ortín *et al.*, 2007; Marín-Navarro *et al.*, 2011; see Suppl. data):

$$p - TR_1 \cdot k_d + [mRNA]_1 \cdot k_d^2 = (p - TR_0 \cdot k_d + [mRNA]_0 \cdot k_d^2) \cdot \exp[- k_d \cdot (t_1 - t_0)] \quad \text{Eq. 1}$$

where p is the slope of the variation of TR:

$$p = (TR_1 - TR_0)/(t_1 - t_0) \quad \text{Eq. 2}$$

Because Eq.1 cannot be algebraically solved for k_d, the StepK program solves it numerically using a bisection algorithm. Hence, the StepK program will deliver a k_d value for each time interval (i.e. a mean value of k_d along this lapse) in between two consecutive samplings in which TR and [mRNA] have been determined.

Values for k_d obtained through the StepK program jump stepwise from one interval to the next. A more realistic change would be to assume that k_d varies also linearly in between sample points as TR does. Under these conditions, the rate of mRNA change would be:

$$d[mRNA]/dt = TR_0 + p \cdot (t - t_0) - [(k_d)_0 + q \cdot (t - t_0)] \cdot [mRNA] \quad \text{Eq. 3}$$

*To whom correspondence should be addressed.

where q is the (unknown) rate of k_d change with time, assumed constant for the time interval under consideration. For the first interval (between t_0 and t_1) the initial value of k_d [i.e., $(k_d)_0$] is calculated supposing an initial steady state (since time 0 is usually taken before the onset of the environmental change triggering the transcriptional response) as TR_0 divided by $[mRNA]_0$. The RampK program integrates numerically Eq. 3 for the considered time interval by Euler's method taking a random initial q value. This value is contrasted by checking the difference between the calculated and experimental value of mRNA concentration at the end of the interval, and iteratively refined through a bisection algorithm. Once the value of q has been determined to a chosen precision, the final value of $(k_d)_1$ (i.e., the value of k_d at t_1) is calculated as:

$$(k_d)_1 = (k_d)_0 + q \cdot (t_1 - t_0) \qquad \text{Eq. 4}$$

and this is used as initial value of k_d for integrating the next interval (between t_1 and t_2).

Therefore, under slightly different assumptions, both programs (StepK and RampK) deliver a time course for k_d from experimental determinations of TR and mRNA concentration sampled at a sequence of time points. StepK estimates a mean value of k_d for each interval while RampK gives the k_d values at the interval extremes assuming a linear time course between them. It might be noted that the starting experimental values must be in compatible units (e.g. [mRNA] in molecules/cell and TR in molecules/minute) to yield HLs in conventional time units (e.g. minutes, in this case).

Shutoff is an additional program included in this website to treat data of mRNA decay after transcriptional arrest (Wang *et al*., 2002). Thus, data will be mRNA concentrations at different time points. Shutoff calculates k_d fitting these data to an exponential decay by minimizing square deviations.

3 STRUCTURE OF THE WEBSITE

We offer a novel implementation of a web site having two new algorithms described earlier. The user has the chance to work with a set of well known tools, all gathered in a friendly web application. The algorithms run on our server. It introduces the notion of *Experiment* which contains one or more *Operations*, including *Input*, *Output* and *Settings* as child notions for the latter.

The input file must have a TR and/or a RA set of columns, each one having the time in the header. If one of the types is missing, neither StepK nor RampK can run. Furthermore, the first column always contains the unique names of the genes analyzed. The application consists of 3 layers: the UI, the algorithms and the storage, built using the YII framework. It can be viewed as 3 tier architecture (see Suppl. Figure 1), employing different languages at each level. The first one uses PHP with JavaScript/HTML/CSS. It controls the algorithms execution written in R (UPGMA, Sota, Kmeans – not implemented by us – and Shutoff), C (StepK and RampK) and PHP (Filtering). The last layer consists of the file system (storing of the inputs/outputs of each algorithm) and the database (store the successfully executed operations for a registered user).

4 USAGE AND APPLICATION

One can create an account or to use the application anonymously. Both cases offer the same features when it comes to capabilities. The difference between them lies in the possibility to save the experiments with their history, inputs and results when an account is used. A user must upload his/her file and then he/she can use the algorithms to process it. The results (visual or text) can be downloaded onto the user's computer. There is also a FAQ section where more details about the algorithms and the workflow can be found. This website was tested under Mozilla Firefox, Google Chrome and Microsoft Internet Explorer (because of their market share).

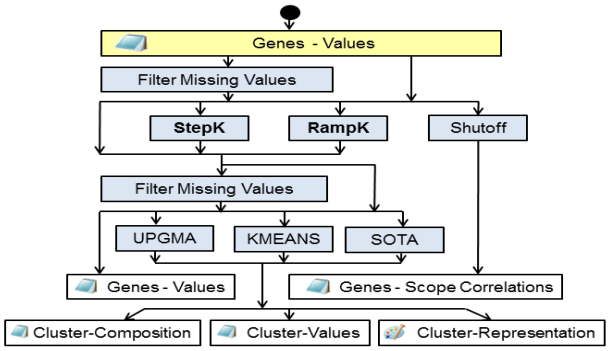

Fig. 1. Application working schema of the mRNAStab site.

5 CONCLUSIONS

We offer a web application to analyze the stability of the genes from microarray data. It contains two new applications capable of estimating the degradation constant (k_d) and, therefore, the mRNA stability for a list of time points of TR and mRNA.

ACKNOWLEDGEMENTS

Funding: J.E.P-O. is supported by grants from the Spanish MCINN (BFU2010-21975-C03-01), and from the Regional Valencian Government (PROMETEO 2011/088). Research by J.M. is supported by grant BFU2009-11965 from MCeI-DGI. Andrei Alic was supported by an Erasmus Grant offered by the Technical University of Cluj-Napoca, Romania.
Conflict of interest: none declared.

REFERENCES

Castells-Roca, J. *et al*., (2011). Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities, *PLoS One* **6**, e17272.

García-Martínez J. *et al*., (2004). Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms, *Mol Cell* **15**, 303-313.

Marín-Navarro, J., *et al*. (2011). Global estimation of mRNA stability in yeast. *Meth. Mol. Biol.* **734**, 3-23.

Parker R. (2012) RNA degradation in *Saccharomyces cerevisiae*. *Genetics*. **191**, 671-702.

Pérez-Ortín J.E., *et al*., (2007) Genomics and gene transcription kinetics in yeast, *Trends Genet* **23,** 250-257.

Pérez-Ortín J.E., *et al*., (2011) A genomic view of mRNA turnover in yeast. *C R Biol*. **334**, 647-654.

Wang Y, *et al*.(2002) Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci. USA*. **99**, 5860-5865.