

# tRanslatome: an R/Bioconductor package to portray translational control

Toma Tebaldi<sup>1,\*</sup>, Erik Dassi<sup>1,†</sup>, Galena Kostoska<sup>1,‡</sup>, Gabriella Viero<sup>2</sup> and Alessandro Quattrone<sup>1</sup>

<sup>1</sup>Laboratory of Translational Genomics - Centre for Integrative Biology, University of Trento, Via delle Regole 101, 38123 Mattarello (TN) and <sup>2</sup>Institute of Biophysics CNR – Via alla Cascata 56/C, 38123 Povo (TN), Italy

Associate Editor: Inanc Birol

## ABSTRACT

**Summary:** High-throughput technologies have led to an explosion of genomic data available for automated analysis. The consequent possibility to simultaneously sample multiple layers of variation along the gene expression flow requires computational methods integrating raw information from different ‘-omics’. It has been recently demonstrated that translational control is a widespread phenomenon, with profound and still underestimated regulation capabilities. Although detecting changes in the levels of total messenger RNAs (mRNAs; the transcriptome), of polysomally loaded mRNAs (the translome) and of proteins (the proteome) is experimentally feasible in a high-throughput way, the integration of these levels is still far from being robustly approached. Here we introduce tRanslatome, a new R/Bioconductor package, which is a complete platform for the simultaneous pairwise analysis of transcriptome, translome and proteome data. The package includes most of the available statistical methods developed for the analysis of high-throughput data, allowing the parallel comparison of differentially expressed genes and the corresponding differentially enriched biological themes. Notably, it also enables the prediction of translational regulatory elements on mRNA sequences. The utility of this tool is demonstrated with two case studies.

**Availability and implementation:** tRanslatome is available in Bioconductor.

**Contact:** t.tebaldi@unitn.it

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on July 2, 2013; revised on October 10, 2013; accepted on October 29, 2013

## 1 INTRODUCTION

High-throughput (‘-omics’) measurements of macromolecule variations in the cell offer the possibility to comprehensively understand how the cellular processes are regulated and to reveal how different layers of control are coordinated in producing a physiologically coherent response. These measurements are also invaluable to understand how the loss of this coordination contributes to disease origin. The establishment of high-

throughput technologies and the consequent explosion of available data allow us to reach a ‘systems’ understanding of the variations in gene expression only when a parallel evolution of algorithms and data mining techniques is achieved. This eventually enables to suggest and prioritize potential mechanistic processes. Nonetheless, the integration of ‘-omics’ data, ranging from epigenetic chromatin remodeling to the dynamics of transcription, translation and protein activities, still requires considerable experimental and computational developments.

In this context, the low correlation observed between messenger RNA (mRNA) and protein levels is an unsolved issue (Vogel and Marcotte, 2012). Recently we showed that the analysis of the translome, an intermediate level between the transcriptome and the proteome formed by mRNAs engaged with polysomes, provides substantial and somewhat surprising new information (Tebaldi *et al.*, 2012). This and other examples (Colman *et al.*, 2013; Schwanhäusser *et al.*, 2011; Vogel *et al.*, 2010) show how the integration of ‘-omics’ data can provide a biologically relevant outcome.

Here we present tRanslatome, a new Bioconductor package for the analysis of differential profiles coming from transcriptome, translome and proteome studies. tRanslatome will help to study mRNA and protein variations in an exhaustive way, providing specific tools for the comparison of polysomal mRNA with total mRNA or protein data.

## 2 DESCRIPTION AND USAGE

tRanslatome is a complete platform for the analysis and pairwise comparison of two ‘-omics’ levels implemented as a Bioconductor package. It is developed to compare translome data with transcriptome and/or proteome data. A general overview of the functions offered by tRanslatome is given in Figure 1A. The package is conceptually organized in three modules, described as follows:

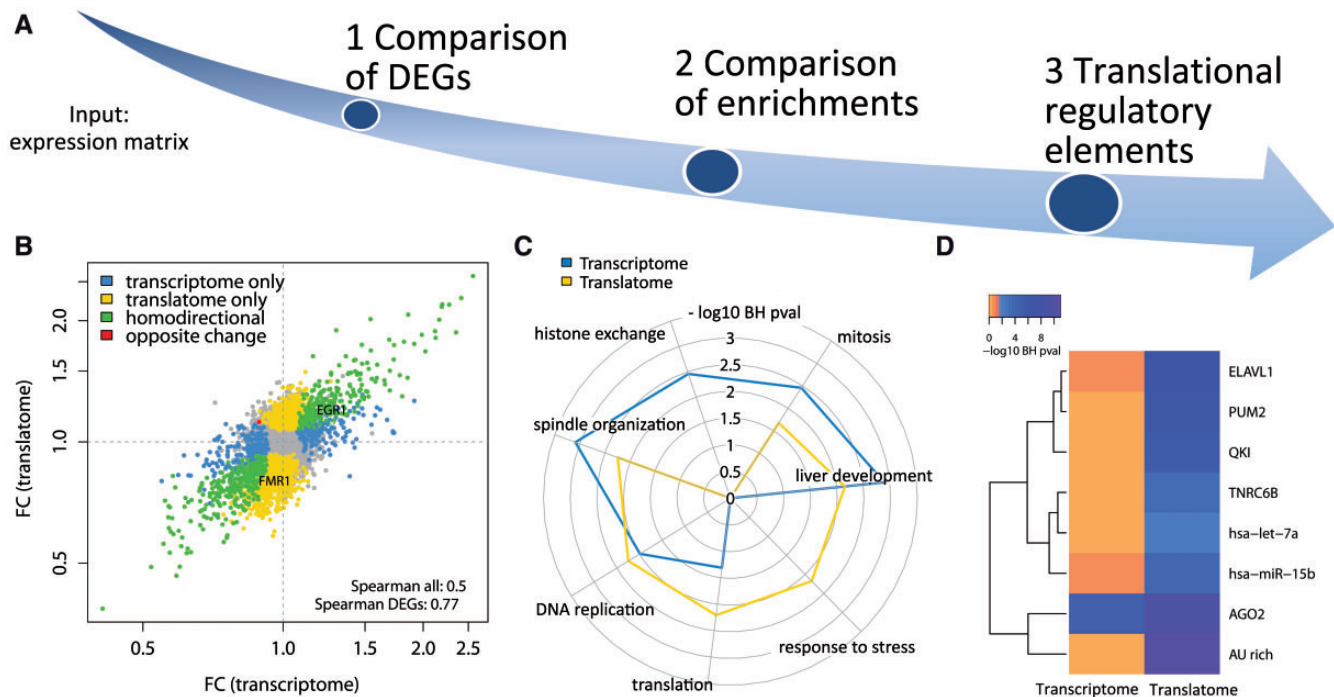
### 2.1 DEGs detection

The only input required by tRanslatome is an expression matrix containing either read counts (from next-generation sequencing data) or normalized signals (from microarray or proteome experiments). To select differentially expressed genes (DEGs), the package offers, in the same computational environment, the integrative analysis of established and emerging statistical methods: (i) DEseq (Anders and Huber, 2010) and edgeR

\*To whom correspondence should be addressed.

†Present address: Department of Information Engineering and Computer Science, University of Trento, via Sommarive 5, 38123 Povo (TN), Italy.

‡The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.



**Fig. 1.** Outline of tRanslatome workflow and graphical outputs. **(A)** General overview of the three modules provided by tRanslatome. **(B)** Scatterplot of fold changes. Each gene is mapped according to the fold change in the transcriptome and the translatoome. Different classes of DEGs are color labeled. The Spearman correlation coefficients are displayed for all genes and for all DEGs. **(C)** Radar plot of the top enriched GO biological process terms for the transcriptome and the translatoome DEGs. **(D)** Heatmap of the top enriched post-transcriptional regulators for the transcriptome and the translatoome DEGs. The color scale is based on the  $-\log_{10}$  of the enrichment  $P$ -value, calculated with a Fisher test

(Robinson *et al.*, 2010), specifically implemented for the analysis of next-generation sequencing data; (ii) significance analysis of microarrays (SAM) (Tusher *et al.*, 2001), developed for the analysis of microarray data; (iii)  $t$ -test, RankProd (Breitling *et al.*, 2004), linear models and moderated  $t$ -test (Smyth, 2004), suitable to deal with general quantitative data; and (iv) methods dealing specifically with the comparison of translatoome and transcriptome data, e.g. ANOTA (Larsson *et al.*, 2011) and translational efficiency, derived from the ratio of polysomal and subpolysomal signals (Powley *et al.*, 2009) or the ratio of ribosome protected fragments and RNA-seq reads (Ingolia *et al.*, 2011). These techniques are described more exhaustively in the documentation and in the Supplementary Material.

To study all the relevant differences arising from the two ‘-omics’ levels, tRanslatome offers a variety of graphical outputs, helping the quality assessment and interpretation of the results. The minus-average plots aid the identification of intensity-dependent patterns, whereas the standard deviation plot can help the selection of the best method for DEGs identification (see Supplementary Figs 4 and 5). The graphics also include scatterplots, displaying changes in the expression of genes in terms of fold changes at both levels (Fig. 1B), and histograms, showing a detailed representation of all the DEGs classes (Supplementary Figs 2 and 3).

## 2.2 Gene Ontology enrichment comparison

One of the most frequent applications of the Gene Ontology (GO) is enrichment analysis, i.e. the identification of significantly

overrepresented GO terms in a given gene set (Ashburner *et al.*, 2000). tRanslatome includes the detection and comparison of GO terms, resuming information about cellular components, molecular functions and biological processes associated to DEGs detected from the two ‘-omics’ levels. Multiple choices are offered for the overrepresentation test, which exploits the GO ‘tree’ structure by means of the Bioconductor package ‘topGO’: these choices can satisfy the need for either more general or more specific biological themes. To simplify the inspection of the results and to effectively represent the differences in the enrichment of ontological terms, the corresponding radar plots and heatmaps can be produced (Fig. 1C and Supplementary Figs 7 and 8). tRanslatome also provides methods for a sensitive comparison of the similarity between enriched GO terms, including the semantic similarity scores (Wang *et al.*, 2007) between terms at each level, and the global similarity score between the two levels.

## 2.3 Enrichment analysis of post-transcriptional regulatory elements

As tRanslatome focuses on the study of global translational controls, enrichment analysis of RNA binding proteins and microRNA binding sites or other RNA regulatory motifs (e.g. AU-rich elements) can be performed on the lists of DEGs. This analysis allows the user to identify possible regulatory factors responsible for the translational regulation of genes in the experiment under consideration (Fig. 1D). The list of genes regulated by each post-transcriptional element is obtained from the

recently established Atlas of UTR Regulatory Activity (AURA) database (Dassi *et al.*, 2012). The method computes a Fisher test *P*-value indicating whether binding sites for each regulator are significantly enriched in the DEGs lists. The annotations from AURA will be updated on every release of tRanslatome. Users can also specify a custom annotation file in place of the one provided by default.

### 3 EXAMPLES

A worked example, derived from data on differentiated versus undifferentiated human hepatocytes (Parent *et al.*, 2007) is used to generate the panels contained in Figure 1. Detailed explanations of this example, along with a second example dealing with the comparison of the proteome and the transcriptome between two human cell lines (Stevens and Brown, 2013), are provided in the Supplementary Material.

### 4 CONCLUSION

tRanslatome allows a user-friendly comparison and integration of data generated from two ‘-omics’ measurements, empowering the discovery of regulatory mechanisms underlying the uncoupling processes among the transcriptome, the translome and the proteome.

*Funding:* This work is supported by the University and Scientific Research Services of the Autonomous Province of Trento.

*Conflicts of interest:* none declared.

### REFERENCES

Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.

Ashburner, M. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Breitling, R. *et al.* (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.*, **573**, 83–92.

Colman, H. *et al.* (2013) Genome-wide analysis of host mRNA translation during hepatitis C virus infection. *J. Virol.*, **87**, 6668–6677.

Dassi, E. *et al.* (2012) AURA: Atlas of UTR Regulatory Activity. *Bioinformatics*, **28**, 142–144.

Ingolia, N.T. *et al.* (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.

Larsson, O. *et al.* (2011) anota: Analysis of differential translation in genome-wide studies. *Bioinformatics*, **27**, 1440–1441.

Parent, R. *et al.* (2007) Mammalian target of rapamycin activation impairs hepatocytic differentiation and targets genes moderating lipid homeostasis and hepatocellular growth. *Cancer Res.*, **67**, 4337–4345.

Powley, I.R. *et al.* (2009) Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. *Genes Dev.*, **23**, 1207–1220.

Robinson, M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Schwanhäusser, B. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.

Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.

Stevens, S.G. and Brown, C.M. (2013) *In silico* estimation of translation efficiency in human cell lines: potential evidence for widespread translational control. *PLoS One*, **8**, e57625.

Tebaldi, T. *et al.* (2012) Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. *BMC Genomics*, **13**, 220.

Tusher, V.G. *et al.* (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA.*, **98**, 5116–5121.

Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.*, **13**, 227–232.

Vogel, C. *et al.* (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.*, **6**, 400.

Wang, J.Z. *et al.* (2007) A new method to measure the semantic similarity of GO terms. *Bioinformatics*, **23**, 1274–1281.