Gene expression

anota: analysis of differential translation in genome-wide studies

Ola Larsson1,∗,† Nahum Sonenberg1 and Robert Nadon2,3,∗

1Department of Biochemistry, McGill University, Montreal, Quebec H3A 1A3, 2Department of Human Genetics, McGill University, Montreal, Quebec H3A 1B1 and 3McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada

Associate Editor: Trey Ideker

ABSTRACT

Summary: Translational control of gene expression has emerged as a major mechanism that regulates many biological processes and shows dysregulation in human diseases including cancer. When studying differential translation, levels of both actively translating mRNAs and total cytosolic mRNAs are obtained where the latter is used to correct for a possible contribution of differential cytosolic mRNA levels to the observed differential levels of actively translated mRNAs. We have recently shown that analysis of partial variance (APV) corrects for cytosolic mRNA levels more effectively than the commonly applied log ratio approach. APV provides a high degree of specificity and sensitivity for detecting biologically meaningful translation changes, especially when combined with a variance shrinkage method for estimating random error. Here we describe the anota (analysis of translational activity) R-package which implements APV, allows scrutiny of associated statistical assumptions and provides biologically motivated filters for analysis of genome wide datasets. Although the package was developed for analysis of differential translation in polysome microarrays or ribosome-profiling datasets, any high-dimensional data that result in paired controls, such as RNP immunoprecipitation-microarray (RIP-CHIP) datasets, can be successfully analyzed with anota.


Received on January 25, 2011; revised on March 2, 2011; accepted on March 9, 2011

1 INTRODUCTION

Translational control of gene expression acts primarily at the initiation step of translation prior to peptide bond formation and controls how many ribosomes are initiated per mRNA (Mathews et al., 2007), thereby enabling instantaneous control of gene expression. Studies in many biological model systems and human diseases have documented that such regulation can be mRNA specific and can have large biological impact (Silvera et al., 2007), thereby enabling instantaneous control of gene expression. Studies in many biological model systems and human diseases have documented that such regulation can be mRNA specific and can have large biological impact (Silvera et al., 2007). Neither of these data types, however, is independent of total cytosolic mRNA levels. In order to study differential translation per se, it is therefore necessary to correct the translationally active mRNA data for confounding total cytosolic mRNA levels. We have recently shown that analysis of partial variance (APV) in combination with the Random Variance Model (RVM) (Wright and Simon, 2003) is an effective method for such analysis that circumvents problems with the commonly applied log ratio approach in combination with t-tests or analysis of variance (ANOVA)s (Larsson et al., 2010). In APV, a hierarchical per gene linear regression is performed on translational activity data, with cytosolic mRNA data and contrast vectors (which define membership in sample classes such as genotype, treatment or disease) as covariates. Class comparison effects are estimated by calculating differences between class intercepts (Larsson et al., 2010).

As with all statistical procedures, it is important to evaluate model assumptions. An interpretive problem arises in high-throughput contexts, however, because of the large number of individual statistical tests. In genome-wide analysis of translational activity, for example, a number of genes are expected to fall outside of the assumption space simply by chance. If the number of assumption violations does not exceed chance levels, the data are behaving as expected of a random variable and analysis can safely proceed. Individual genes with serious violations of assumptions, however, should nonetheless be interpreted with caution. Thus, genome-wide application of APV to identify differential translation requires a set of methods that assess the genome-wide performance of APV, relates it to chance outcomes and provides filtering capabilities based on statistical and biological criteria. Here we provide the anota (analysis of translational activity) statistical package, which is implemented in the R statistical computing language and contains the analytical approaches and diagnostic outputs to guide such analysis.

2 IMPLEMENTATION

Identification of differential translation within anota is a two-step process. Model assumptions are assessed, followed by identification of differential translation.

2.1 Assessment of model assumptions

Model assumptions are examined first to determine if the data are appropriate for anota (Fig. 1A). If there are more violations of
particular class comparisons (Fig. 1B). Both the omnibus and the custom contrasts analyses are performed using APV with or without RVM. We recommend the former when there are few replicates for its superior statistical power (Larsson et al., 2010). When applied, RVM’s assumption that gene variances are random variables from an inverse gamma distribution should be examined in anota for both the omnibus and the contrast tests. The RVM $P$-values for the interactions described above (Point iii) should also follow a uniform distribution under the assumption of a common slope among the classes.

There are two additional biologically motivated slope considerations when using anota. Slopes $>1$ should not occur insomuch as the actively translated mRNAs are a fraction of all mRNAs. Similarly, although slopes $<0$ are biologically plausible, they indicate complex regulatory mechanisms which are usually not the primary focus. To address this, anota tests whether slopes are significantly $>1$ or $<0$ and reports a $P$-value that can be used for filtering. Unrealistic slopes will likely be more common in datasets with few replicates/sample classes and it is therefore especially important to consider the slope when evaluating results from such studies. Filtering of genes based on slope and many other user-defined criteria is done using the anotaPlotSigGenes function which generates both a tabular summary and per mRNA graphical representations of the APV analysis together with key statistics. This output can easily be used to examine the performance of APV for individual genes of interest. The output also includes intercepts for each class which could be used, for example, in clustering of translational activity levels. Nominal and various types of false discovery rate (FDR)-adjusted $P$-values can also be produced and exported for further examination within R or other statistical software.

2.3 Work flow

Translationally active mRNA data and cytosolic mRNA data will typically be analyzed within anota using four main functions: anotaPerformQc, anotaResidOutlierTest, anotaGetSigGenes and anotaPlotSigGenes (Fig. 1A and B).

ACKNOWLEDGEMENTS

We thank Richard Simon and George Wright for permission to incorporate some of their original RVM code into anota.

Funding: O.L. was initially supported by a fellowship from the Knut and Alice Wallenberg foundation. Partial support was provided by Le Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) Grant 119258 (to R.N.).

Conflict of interest: none declared.

REFERENCES


