

Serial Expression Analysis: a web tool for the analysis of serial gene expression data

Maria José Nueda¹, José Carbonell², Ignacio Medina^{2,3}, Joaquín Dopazo^{2,4} and Ana Conesa^{2,*}

¹Statistics and Operation Research Department, University of Alicante, ²Bioinformatics and Genomics Department, Centro de Investigaciones Príncipe Felipe (CIPF), ³CIBER de Enfermedades Raras (CIBERER) and ⁴Functional Genomics Node, INB, CIPF, Valencia, Spain

Received February 21, 2010; Revised May 5, 2010; Accepted May 14, 2010

ABSTRACT

Serial transcriptomics experiments investigate the dynamics of gene expression changes associated with a quantitative variable such as time or dosage. The statistical analysis of these data implies the study of global and gene-specific expression trends, the identification of significant serial changes, the comparison of expression profiles and the assessment of transcriptional changes in terms of cellular processes. We have created the SEA (Serial Expression Analysis) suite to provide a complete web-based resource for the analysis of serial transcriptomics data. SEA offers five different algorithms based on univariate, multivariate and functional profiling strategies framed within a user-friendly interface and a project-oriented architecture to facilitate the analysis of serial gene expression data sets from different perspectives. SEA is available at sea.bioinfo.cipf.es.

INTRODUCTION

Serial transcriptomics experiments study the dynamics of gene expression associated with a parameter that defines the course of the series. The most common type of serial experiments are time-course studies, which investigate gene expression changes during development, in response to a given stimulus or due to their cyclic pattern. Series can be short (three—five time points) or long (six or more time points) and involve longitudinal data or independent observations. Although time-course studies are the most usual serial experiment, any transcriptomic data set whose samples can be sequentially ordered within a quantitative factor, such as dosage or OD growth measures, can be defined as serial data.

The statistical assessment of (multi)serial gene expression experiments poses a number of challenges. On one hand, analysis methods should capture the dynamics of the data and propose models for describing the serial behaviour. These models should identify significant variations along the series and evaluate gene expression differences between them when additional factors are considered. Moreover, the analysis of serial transcriptomic data also implies solutions for the interpretation of changes in terms of cellular functions; for example, through the application of functional enrichment strategies to the dynamics of gene expression. Finally, for their effective translation to transcriptomics research, methodologies need to incorporate user-friendly bioinformatics solutions that make algorithms easily accessible to experimentalists.

In recent years, a number of algorithms have been developed for the analysis of time-course microarray data. Many have been implemented as R packages and are available at the Bioconductor Project (www.bioconductor.org). Statistical strategies include linear models (1), empirical Bayes (2), fuzzy algorithms (3) and Bayesian approaches (4).

Other tools are offered as desktop Java applications. The Short Time-series Expression Miner (STEM; 5) assigns gene profiles to predefined clusters and evaluates the significance of the clusters, which can then be functionally interrogated by GO enrichment analysis. BATS (Bayesian Analysis of Time Series), also developed for one series data, applies a Bayesian approach to identify differentially expressed genes and to estimate their profiles (6). GenT χ Warper (<http://www.psb.ugent.be/cbd/papers/gentxwarper/index.htm>) is a tool for the alignment, analysis and mining of gene expression time series based on dynamic time warping techniques. The EDGE is a R-package accompanied by a graphical user interface that implements a splines-based approach to analyse

*To whom correspondence should be addressed. Tel: +34 96 328 96 80; Fax: +34 96 328 97 01; Email: aconesa@cipf.es

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

time expression changes within one or between several conditions (7).

Our group has published different statistical approaches for the analysis of short time-course transcriptomics experiments. These include regression-based (maSigPro; 8) and multivariate approaches (ASCA-genes; 9), as well as specific methodologies for functional and gene-set enrichment analysis (maSigFun, PCA-maSigFun and ASCA-functional; 10). Diverse strategies were developed to address different aspects of the analysis of serial data and were implemented in the R statistical language. The maSigPro method was included in 2007 in the microarray data analysis suite GEPAS (11). In the SEA suite, we have combined and extended these approaches within a novel webserver framework to provide the scientific community with a complete web resource for the analysis of short series transcriptomics. The SEA site offers five different algorithms and a user-friendly environment to evaluate gene expression data for gene-wise and functional-based transcriptomic changes that are general or gene-specific, in single or multiple series. To our knowledge, SEA is the first integrated web-based suite for the analysis of serial gene expression data that combines multivariate, univariate and functional profiling strategies.

THE SEA WEBSITE

The SEA website is based on Web 2.0 technology features such as AJAX and high-level JavaScript libraries (jQuery). The server has been developed in Java language, which calls R scripts for the different statistical algorithms. The server architecture has been designed to support a multi-user environment. Web client interface is served by a Tomcat Java Server front-end that is replicated in two physical machines to guarantee a permanent service availability. Moreover, each front-end responds to every user action (run a job, check results, check data list, etc.) submitted by the client web browser. When a new job is launched, the corresponding data form is processed by one of the server front-ends, a new job instance is created in the user database, and the data are prepared for tool execution. Finally, the job is run in a local cluster controlled by a batch-queuing system (Sun Grid Engine, SGE), thereby ensuring that every job is initiated when all its required computing resources are available. This architecture implies that all tools offer a Linux command line interface that allows an easy execution of jobs in a batch-based high-performance computational environment. The SEA web not only allows tools to be executed by a guest user (no log-in is required), but also offers user session persistence by means of user/password registration. In this case, users can upload and save data in his/her private area and organize jobs in different projects and revisit them at any time. In all cases, data privacy is guaranteed and no exchange between registered and anonymous accounts is possible. The SEA suite includes five different methods for univariate, multivariate and functional analysis of serial gene expression, each of which is available under a different tab, plus two additional tabs for data upload and graphical output.

Three pieces of data are required by the SEA algorithms: expression values, experimental design (covariates) and functional annotation (in the case of running the functional analysis options). These three elements must be provided as tab-delimited files and uploaded into the application before any analysis is started. The expression data file must contain genes in rows and arrays in columns. Array names and gene names should be located in the first row and column, respectively. The covariates file includes experimental design information, presented in a table format with as many columns as arrays and as many rows as experimental factors. Each cell contains the value of the experimental factor in the sample of the array. Finally, the functional annotation is provided in a simple two column—gene[tab]annotation—format, with multiple annotations of the same gene arranged in different rows. Alternatively, users can use annotation data available at the SEA database. In this case, no annotation file upload is necessary.

Once the data is uploaded, these are stored at the user account and can be used in any of the tools contained in the site. Users can visualize uploaded data at the 'Data list' panel on the right side of the user interface. Each analysis method has a separate input-form to select data and indicate specific additional parameters. Optionally, a name and description can be given for the analysis job which is launched by a click on the *Run* button. Data and parameters are then sent to the server, checked for consistency and the job name is returned to the 'Job list' panel where the progress of the computations can be followed. Once completed, the user can access or visit the results page by clicking on the job name. Results are provided as downloadable text (generally lists of significant features) and png files (informative graphs) which are also displayed at the HTML user-interface. In the following, we comment briefly the procedure of each method to clarify the meaning of the different input parameters and the interpretations of the corresponding results.

maSigPro

maSigPro (MicroArray Significant Profiles; 8) applies linear regression to model gene expression in time-course microarray data and selects differentially expressed genes through a two-step algorithm. First, responsive genes are identified by fitting a generic regression model with time as quantitative variable and series as dummy variables, and genes are selected at a given FDR. Second, step-wise regression is applied to selected genes to adjust the model and identify gene-specific variation patterns. At this stage, genes can also be passed through a selection filter based on the R^2 -value of the second regression model. This value measures the goodness of fit, and therefore allows for the selection of genes with consistent expression trends. maSigPro returns lists of genes with statistically significant changes over time and across different series. maSigPro was originally conceived for the analysis of multiple short time-course transcriptomic data, but can be applied in a straight forward manner to the analysis of single series data (i.e. where only one condition is

evaluated) and for other, non-timed, serial experiments based on any quantitative factor.

The SEA implementation of maSigPro first requires the selection of expression data and covariate files. Once a covariate file is selected, it is parsed to identify experimental factors. The user is then prompted to select the quantitative factor (such as time or dosage) and, if more than one factor is present, the qualitative factor (for example, strain or treatment). Note that the qualitative factor can be also numeric, but this will be treated as a categorical rather than quantitative variable. Currently, maSigPro and derived algorithms can only deal with a maximum of two factors at a time. Other parameters to choose are the 'polynomial degree' of the linear regression model, the significance value 'alpha' and the ' R^2 cut-off' value for gene selection.

maSigPro result is basically a set of lists of differentially expressed genes, one for each contrasting series. Each list contains genes with significantly different profiles between a given series and the control series. For example, if the experiment compares the differences in transcriptional responses to a treatment between a mutant strain (S) and the wild type (W) over time, the list named *SvsW* will contain all the genes that respond differently in mutant and wild type. Next to the lists of selected genes, an additional R object is created that can be re-directed to the *maSigVisualization* module for graphical display of the significant profiles (explained subsequently).

ASCA-genes

ASCA-genes (9) is an adaptation of the ASCA method [analysis of variance (ANOVA) simultaneous component analysis] developed by Smilde and co-workers (12) for the analysis of multifactorial experiments in transcriptomics. Basically, ASCA uses ANOVA to decompose data variation associated to experimental factors and principal component analysis (PCA) to discover principal patterns of variation within these factors. ASCA-genes combines this multivariate descriptive analysis of time course expression data with a gene selection procedure. This gene selection is based on two statistics: leverage, which is a measure of the importance of a gene's contribution to the multivariate fitted ASCA-model, and the squared prediction error (SPE), which is an evaluation of the goodness of fit of the model to a particular gene. Threshold values for these parameters can be derived by resampling methods or using the mixture approaches minAS and gamma (Tarazona, S. *et al.*, manuscript in preparation). ASCA-genes have been implemented in experiments comprising one, two or three experimental factors, one of which is typically, but not necessarily, a quantitative factor.

The ASCA-genes input form at SEA requires the selection of expression data and covariate files. No indication of quantitative or qualitative factors is required, as all factors are treated as categorical in the ANOVA decomposition step of the ASCA methodology. Additional parameters permit the interaction between factors to be considered and the second factor to be joined to this interaction if wished. The latter option is recommended when a

study aims to investigate all effects of a given treatment, both at the start value of the experiment and along the series. Additionally, a criterion for selecting significant genes should also be indicated.

The program returns a summary of the distribution of data variability between experimental factors and the percentage of this variability, which is explained by the ASCA model. It should be noted that, due to the ANOVA treatment embedded in ASCA, a separate PCA submodel is generated for each experimental factor or combination of factors. If we take the example of the previous section once again and use ASCA-gene default parameters, two submodels will be returned: time (capturing major time-dependent gene expression changes) and Strain+Time×Strain (capturing the effect of the mutant strain and its interaction with Time). Consequently, further output files are provided for each submodel. These are gene leverage/SPE scatter plots where the importance/error of each gene in the submodel is represented, trajectory plots to indicate major expression trends associated to the submodel, a list of significant genes (those with individual profiles similar to the submodel major trends), and a list of outlier genes (those exhibiting odd but significant changes).

maSigFun

maSigFun (10) derives from the above mentioned maSigPro methodology, and represents an immediate extension of this method to gene sets. In maSigFun, the regression model is not computed gene-wise, as in maSigPro, but rather to the data matrix composed by the expression values of all genes belonging to a gene set or functional class. In this way, one regression model is fitted for each functional category. In this approach, individual genes are considered as different observations of the expression profile of the class, and significant features are typically functional categories with few and highly correlated genes.

SEA input form for maSigFun is similar to that of maSigPro, with the additional feature that annotation data are also required. These can be provided by the user in an annotation file or taken from the site database. Currently, SEA offers access to the most common functional annotation schemas (Gene Ontology, KEGG, InterPro, etc.) for major model species. The output of maSigFun is likewise similar to that of maSigPro, but allows lists of significant functional groups, instead of genes, to be obtained. As in maSigPro, the trajectory charts for significant functional categories can be plotted through the maSigVisualization module.

ASCA-functional

ASCA-functional (10) is a GSA strategy for multifactorial gene expression experiments based on the ASCA modelling described previously. In ASCA-functional, genes are ordered according to their PCA loadings at the selected components of the ASCA submodels, and this ranking is used for gene set enrichment analysis with the partitioning GSA method FatiScan (13). The PCA loading of a gene in the principal component is a

measure of the contribution of that gene to that component, i.e. how well the gene follows the pattern represented by the component. Genes with high positive loadings behave like the component, genes with high negative loadings follow the opposite trajectory and genes with a loading close to zero are not correlated with the component's profile. Therefore, ordering genes by their loadings value is a way of ranking according to a given expression pattern, namely that represented by the component. The program returns a GSA result for each of the loading lists tested. Using again the mutant strain example, let us imagine that the ASCA analysis identified three significant patterns: one global expression trend associated to Time (i.e. general induction of gene expression upon treatment) and two patterns associated to Strain+TimevsStrain (early differences between genotypes and late differences between genotypes). In this case, ASCA-functional would return three lists of significant functional categories: one with functions that are generally regulated in time, one for the differential early response between strains, and one with the functions that are differentially regulated at late time points.

PCA-maSigFun

PCA-maSigFun (10) is a combination of PCA and maSigPro. The method identifies the major gene expression changes within each functional class and evaluates whether these changes are significantly associated to the serial parameters. First, PCA is applied to the gene-expression submatrix associated to the genes belonging to each functional category or gene set. The scores of the relevant principal component(s) of these PCAs are considered as joint expression profile(s) for the gene set, and can thus be regarded as meta-genes of the functional category. maSigPro is then applied to these meta-genes to assess whether the functional patterns are significant changing profiles.

Input and output forms of PCA-maSigFun are equivalent to those of maSigFun, and similarly, results can be redirected to the visualization module for graphical printing.

maSigVisualization

While ASCA-genes and ASCA-functional modules contain their own graphical output, results of the univariate methods are visualized through the maSigVisualization tab. Basically, the visualization module takes a list of significant features, applies a clustering algorithm to divide these into groups of similar elements, and creates trajectory plots for the resulting clusters. In these plots, expression values are averaged by experimental conditions and presented according to the (multi)serial nature of the data. An additional text output file indicates the cluster assignment of each gene or functional class. maSigVisualization accepts and recognizes all the maSig output objects and adjusts the input form to query for required visualization parameters. In all cases, a 'Series to see' is requested, i.e. the list of significant features to be displayed. In the case of maSigPro and maSigFun results, the number of clusters and clustering

algorithms must also be indicated. For PCA-maSigFun, the visualization module returns the profile of each selected category, together with box-plots indicating the importance of each gene in the meta-gene. A gene is considered 'important' if its corresponding loading is greater than a given threshold, which can be defined by the user or computed by re-sampling, minAS and gamma procedures, as in ASCA-genes.

ANALYSIS OF SERIAL DATA WITH SEA

To illustrate the use of SEA and highlight characteristic aspects of the different algorithms, we have applied all five tools to an example of a transcriptomic data set of a plant abiotic stress study. These example data and others are available at the SEA web page. The study investigates the transcriptional response to three different abiotic stressors (salt, cold and heat) in potato seedlings using the NSF 10k potato array (14). A common reference design is used. The data set contains four series (one control and three types of stress: heat, salt and cold), three time points, three replicates per experimental condition and 9124 genes. Gene Ontology functional annotation of this dataset was generated using the Blast2GO software (15).

We begin the analysis by exploring data from a global perspective using the ASCA-genes module. After loading the example data, we select default option values. By selecting 'Include interaction between factors', we allow the model to analyse Time×Treatment interactions and by choosing 'Join interaction with second factor', we study the Treatment and the Time×Treatment effects together. Once the job is completed we can visit the results page by selecting this option on the Job panel. The variability analysis of this data set reveals that a great deal (54.56%) of the variation in the data set is associated to the Time×Treatment factor, while a lower percentage (9.24%) is due solely to the Time factor. This means that changes in gene expression are mostly associated to the time-dependent effects of the different treatments. Furthermore, these variations can be effectively summarized by PCA submodels with one and two components, respectively. Looking at the leverage/SPE plots for both submodels we can observe that most genes have values close to 0 at both parameters and only some present high values and are in the selection areas (Figure 1a). Genes in the red region are well-modeled, as they have high leverages, while genes in the blue area are odd- or bad-modeled, as they present high SPE values. Trajectory plots show the expression pattern of these components. For example, the trajectory plot for the first component of the Treatment+Time×Treatment submodel (Figure 1b) reveals that a large number of genes respond differently in salt/cold and control/heat treatments, with differences starting to emerge at time point 3, peaking at 9 h and being maintained until the end of the experiment. Analysis of the trajectory plots for the remaining two ASCA-genes components indicates that the greatest transcriptional change is at time-point 9, and that early differences between salt and cold treatments can also be found for some genes (Supplementary Figure 1).

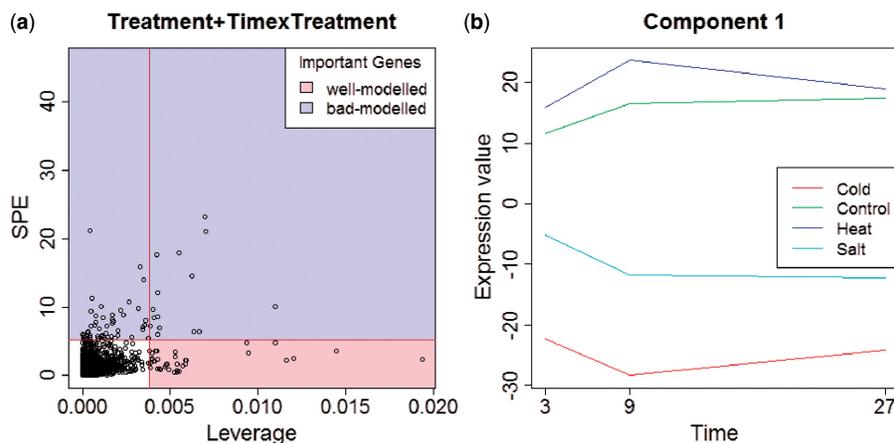


Figure 1. ASCA-genes graphical output for potato stress data. (a) leverage/SPE scatter plot of the ASCA-genes submodel Treatment+Time×Treatment. Vertical and horizontal lines indicate cut-off leverage and SPE values, respectively, for gene selection. Genes in red area have high loadings and follow the expression patterns of the submodel. Genes in blue area have expression changes different from the major patterns. (b) Trajectory plot of the first component of submodel Treatment+Time×Treatment. This component reveals the major transcriptional changes associated with the time-dependent effects of the different stress treatments.

ASCA-genes results suggest that the major transcriptional pattern, depicted in Figure 1b, would be interesting for analysis by GSA. Therefore, we can run ASCA-functional on this data set and visualize the FatiScan results obtained for the genes ranked by the loadings of the first component of the Treatment+Time×Treatment submodel. When this is done, a significant number of enriched functional categories are found. Processes associated to upper ranking positions (up-regulated in cold/salt stresses) relate to ‘protein synthesis and degradation’, ‘lignin biosynthesis’, diverse hormone signaling pathways and response to several stimuli. ‘Photosynthesis’, ‘microtubule-based movement’, ‘RNA binding’ and ‘lypoxigenase activity’ were functions associated to bottom rank genes, i.e. they were down-regulated in these stress conditions. A full list of significant features obtained by ASCA-functional and other SEA modules can be found at the SEA site.

Having obtained a global impression of the transcriptional effects of the three abiotic stressors on potato seedlings, we can now carry out a gene-wise analysis of serial changes using the maSigPro approach. After loading data into the maSigPro module, we choose Time as a quantitative factor, Treatment as a qualitative factor and control as the reference series. Default parameters indicate that second-degree regression model will be applied and genes will be selected at a significance level of 0.05 and a R^2 -value above 0.7. A total of 1315 genes are found to be statistically significant. SEA provides this result by classifying significant genes by series according to the profile contrast considered. In this selection we have 417 genes with changes for the control series, 1300 with differences between cold and control, 766 genes with different profiles in heat and control and 1167 significant genes for the salt versus control comparison. These lists are provided as text files ready for download. We can now red-direct the maSigPro R object to the maSigVisualization module to create plots of the significant gene profiles. Selecting the maSigVisualization tab, the maSigPro result is parsed and

we are asked to select the list of significant genes (*‘Series to see’*) we wish to visualize. For example, we can choose the ColdvsControl series and cluster significant genes in nine groups by hierarchical clustering (*‘hclust’*). Supplementary Figure 2 shows the typical output of maSigPro visualization. For each cluster, a trajectory plot is generated that summarizes gene expression over time and at each treatment for the genes belonging to that group. It should be noted that, although the selected gene list contains only those genes with significant profile differences in the cold vs. control comparison, on the plots all series are included. We can observe in this figure that most populated clusters indicate differences between cold/salt and heat/control treatments, which is the main transcriptional behavior already detected by ASCA-genes. In addition, other types of genes are revealed by this method. For example, cluster 7 includes genes with an exclusive early up-regulation upon cold treatment.

The remaining two modules address further questions concerning functional aspects of the data. By running maSigFun (degree 2, $R^2 = 0.4$, $\alpha = 0.05$) we identified 10 functional categories with consistent expression patterns and significant differences between treatments (data at the SEA site). Finally, we run PCA-maSigFun in order to go a step further in the functional analysis. This tool indicates that a total of 37 functional groups have subsets of genes with highly correlated significant profile changes. Plotting this result using the maSigVisualization module helps to understand the expression profile of the class and the participation of the class members in this profile. Figure 2 shows the PCA-maSigFun graphical output for ‘glutamate metabolic process’. On the left panel, we can observe the trajectory plot for the class, which again reveals the significant treatment differences discussed previously. On the right panel, the gene loadings bar plot indicates that five of the eight members of this class follow the expression pattern displayed on the left in a significant manner, four of them with a positive correlation and one with a negative correlation.

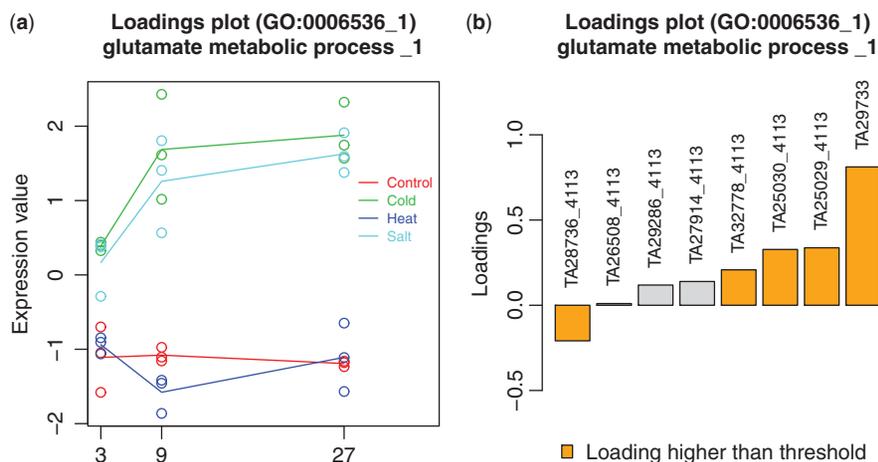


Figure 2. Example of PCA-maSigFun result. The figure shows the expression profile associated with the major transcriptional changes of the genes belonging to the functional category glutamate metabolic process. **(a)** Trajectory plot of the first meta-gene of this functional class reveals gene expression differences between cold/salt and heat/control treatments. **(b)** Bar plot of gene loadings. Orange bars indicate genes with significant contributions to the profile of the first meta-gene, either by positive correlation (loading >0) or negative correlation (loading <0).

CONCLUSIONS

The analysis of multifactorial and serial transcriptomic data sets is a complex task that requires the use of specialized bioinformatics resources that are capable of evaluating data from different perspectives, namely the dynamics of the transcriptional changes and the interaction with experimental factors, the identification and understanding of both global processes and specific features, and the adoption of gene-wise versus gene-set strategies. The SEA website is a compilation of tools with which to investigate all these aspects and constitutes the first dedicated web-based suite for the analysis of serial gene expression data that is especially designed for short series. Although the site relies on elaborated algorithms, we have made a great effort to simplify input and output forms while maintaining the data-mining capabilities of the methodologies. Results are provided as lists of significant features and plots which allow the dynamics of the transcriptional changes and the relationships between trends, genes and functional classes to be explored.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The CIBER Enfermedades Raras and the INB are initiatives of the ISCIII.

FUNDING

Spanish Ministry of Science and Innovation (MICINN, grants BIO BIO2008-04212, BIO2008-04638-E, CEN-2008-1002); Red Temática de Investigación Cooperativa en Cáncer (RTICC, RD06/0020/1019); Instituto de Salud

Carlos III (MICINN). Funding for open access charge: Spanish National Bioinformatics Institute.

Conflict of interest statement. None declared.

REFERENCES

- Smyth,G.K. (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.
- Tai,Y.C. and Speed,T.P. (2006) A multivariate empirical Bayes statistic for replicated microarray time course data. *Ann. Stat.*, **34**, 2387–2412.
- Kumar,J. and Futschik,M.E. (2007) Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics*, **2**, 5–7.
- Aryee,M.J., Gutierrez-Pabello,J.A., Kravnik,I., Maiti,T. and Quackenbush,J. (2009) An improved empirical Bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC Bioinformatics*, **10**, 409.
- Ernst,J. and Bar-Joseph,Z. (2006) STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, **7**, 191.
- Angelini,C., Cuttillo,L., De Canditiis,D., Mutarelli,M. and Pensky,M. (2008) BATS: a Bayesian user-friendly software for analyzing time series microarray experiments. *BMC Bioinformatics*, **9**, 415.
- Leek,J.T., Monsen,E., Dabney,A.R. and Storey,J.D. (2006) EDGE: extraction and analysis of differential gene expression. *Bioinformatics*, **22**, 507–508.
- Conesa,A., Nueda,M.J., Ferrer,A. and Talon,M. (2006) maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics*, **22**, 1096–1102.
- Nueda,M.J., Conesa,A., Westerhuis,J.A., Hoefsloot,H.C., Smilde,A.K., Talon,M. and Ferrer,A. (2007) Discovering gene expression patterns in time course microarray experiments by ANOVA-SCA. *Bioinformatics*, **23**, 1792–1800.
- Nueda,M.J., Sebastian,P., Tarazona,S., Garcia-Garcia,F., Dopazo,J., Ferrer,A. and Conesa,A. (2009) Functional assessment of time course microarray data. *BMC Bioinformatics*, **10**(Suppl. 6), S9.
- Tarraga,J., Medina,I., Carbonell,J., Huerta-Cepas,J., Minguez,P., Alloza,E., Al-Shahrour,F., Vegas-Azcarate,S., Goetz,S.,

- Escobar,P. *et al.* (2008) GEPAS, a web-based tool for microarray data analysis and interpretation. *Nucleic Acids Res.*, **36**, W308–W314.
12. Smilde,A.K., Jansen,J.J., Hoefsloot,H.C., Lamers,R.J., van der Greef,J. and Timmerman,M.E. (2005) ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*, **21**, 3043–3048.
13. Al-Shahrour,F., Arbiza,L., Dopazo,H., Huerta-Cepas,J., Minguez,P., Montaner,D. and Dopazo,J. (2007) From genes to functional classes in the study of biological systems. *BMC Bioinformatics*, **8**, 114.
14. Rensink,W.A., Iobst,S., Hart,A., Stegalkina,S., Liu,J. and Buell,C.R. (2005) Gene expression profiling of potato responses to cold, heat, and salt stress. *Funct. Integr. Genomics*, **5**, 201–207.
15. Conesa,A., Gotz,S., Garcia-Gomez,J.M., Terol,J., Talon,M. and Robles,M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.