Published online 31 May 2008 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* ***W341–W346***
doi:10.1093/nar/gkn318

Babelomics: advanced functional profiling of transcriptomics, proteomics and genomics experiments

Fátima Al-Shahrour[1], José Carbonell[1], Pablo Minguez[1,2], Stefan Goetz[1,2], Ana Conesa[1], Joaquín Tárraga[1,3], Ignacio Medina[1,2], Eva Alloza[1], David Montaner[1,3] and Joaquín Dopazo[1,2,3,*]

[1]Department of Bioinformatics, Centro de Investigación Príncipe Felipe (CIPF), Autopista del Saler 16, E46013 Valencia, [2]CIBER de Enfermedades Raras (CIBERER), Valencia, E46013 and [3]Functional Genomics Node, INB, CIPF, Valencia, E46013, Spain

Received January 31, 2008; Revised April 19, 2008; Accepted May 7, 2008

ABSTRACT

We present a new version of Babelomics, a complete suite of web tools for the functional profiling of genome scale experiments, with new and improved methods as well as more types of functional definitions. Babelomics includes different flavours of conventional functional enrichment methods as well as more advanced gene set analysis methods that makes it a unique tool among the similar resources available. In addition to the well-known functional definitions (GO, KEGG), Babelomics includes new ones such as Biocarta pathways or text mining-derived functional terms. Regulatory modules implemented include transcriptional control (Transfac, CisRed) and other levels of regulation such as miRNA-mediated interference. Moreover, Babelomics allows for sub-selection of terms in order to test more focused hypothesis. Also gene annotation correspondence tables can be imported, which allows testing with user-defined functional modules. Finally, a tool for the '*de novo*' functional annotation of sequences has been included in the system. This allows using yet unannotated organisms in the program. Babelomics has been extensively re-engineered and now it includes the use of web services and Web 2.0 technology features, a new user interface with persistent sessions and a new extended database of gene identifiers. Babelomics is available at http://www.babelomics.org

INTRODUCTION

Over the last few years, due to the popularization of high-throughput methodologies such as transcriptomics (microarrays), proteomics, large-scale genotyping, ultra high-throughput sequencing, etc. (1), the possibility of obtaining experimental data has increased drastically. There has been a parallel demand in methods for the interpretation of the results, which involves translating these data into useful biological knowledge. The methods and strategies used for this interpretation are in continuous evolution and new proposals are constantly arising (2).

Initially, methods for testing the enrichment in functional annotations in a set of pre-selected genes with respect to a reference set have been developed by different groups in the last years. Most recently, a new family of methods pioneered by the GSEA (3) that aim to detect the behaviour of modules or sets of genes related by any biological property of interest (function, regulation, etc.) have been proposed and successfully applied to different biological systems (2,4).

Although some of the components of Babelomics have been working since 2003 [e.g. the FatiGO algorithm was first published in 2004 (5)] the idea of assembling different methods (functional enrichment and gene set enrichment methods) with a large number of functional module definitions crystallized as the Babelomics project, which was first published in 2005 (6,7). This new version of Babelomics includes new methods and new module definitions of different nature (functional, regulatory, phenotypic, etc.). Also a new tool for the '*de novo*' functional annotation of sequences, the Blast2GO (8), has been included in the system. This allows analysing organisms yet unannotated within the program. In terms of technology, Babelomics has been re-engineered and transformed to web services technology and Web 2.0 features have been included in its design. Babelomics has a new interface that allows the definition of persistent sessions and asynchronous use (a program can be left running and users can come back later to see the results).

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

Table 1. Functional profiling data analysis webtools

Tool	URL	Analysis type[a]	References	Citations[b]
GSEA	http:/www.broud.mit.edu/gsea/	GSA	(3,32)	1013
DAVID	http://www.DAVID.niaid.nih.gov	FE	(33)	504
GOMiner	http://discover.nci.nih.gov/gominer/	FE	(34,35)	408
Babelomics	http://www.babelomics.org	FE, GSA	(5–7,27)	402
MAPPFinder	http://www.GenMAPP.org	FE	(36)	379
GOStats	http://gostat.wehi.edu.au/	FE	(25)	249
Ontotools	http://vortex.cs.wayne.edu/ontoexpress/	FE	(37,39–42)	223
GOTM	http://genereg.ornl.gov/gotm/	FE	(43)	164
FunSpec	http://funspec.med.utoronto.ca webcite	FE	(44)	100
GeneMerge	http://www.oeb.harvard.edu/hartl/lab/publications/GeneMerge.html	FE	(45)	96
FuncAssociate	http://llama.med.harvard.edu/Software.html	FE, GSA	(38)	91
GOToolBox	http://gin.univ-mrs.fr/GOToolBox	FE	(26)	74
GFINDer	http://www.medinfopoli.polimi.it/GFINDer/	FE	(46,47)	49
WebGestalt	http://bioinfo.vanderbilt.edu/webgestalt/	FE	(48)	46
GOAL	http://microarrays.unife.it	GSA	(49)	25
Pathway Explorer	https://pathwayexplorer.genome.tugraz.at/	FE	(50)	25
PLAGE	http://dulci.biostat.duke.edu/pathways/	GSA	(51)	18
t-profiler	http://www.t-profiler.org/	GSA	(52)	12
WebBayGO	http://blasto.iq.usp.br/~tkoide/BayGO/	FE	(53)	10
JProGO	http://www.jprogo.de/	GSA	(54)	7
ADGO	http://array.kobic.re.kr/ADGO	GSA	(55)	3
GeneTrail	http://genetrail.bioinf.uni-sb.de/	GSA	(56)	3
GAZER	http://integromics.kobic.re.kr/GAzer/index.faces	GSA	(57)	–
PathExpress	http://bioinfoserver.rsbs.anu.edu.au/utils/PathExpress	FE	(58)	–

[a]Type of analysis: FE, functional enrichment, GSA, gene set analysis.
[b]Citations are taken from Scholar Google as of 24 January 2008.

The availability of different flavours of functional enrichment and gene set enrichment methods that use gene modules of different nature makes of Babelomics a unique tool among other available resources of similar characteristics. Babelomics is among the most widely used web tools for functional profiling (Table 1). During the last year Babelomics has registered an average of 200 experiments analysed per day. A map of daily usage can be found here: http://bioinfo.cipf.es/access_map/map.html. With the novelties included in terms of both technology and methods, we can anticipate an even higher rate of use in the new release of Babelomics.

SOURCES OF BIOLOGICAL INFORMATION FOR DEFINING GENE MODULES

Functional profiling methods depend upon the definition of gene modules based on biological properties of interest, whose differential behaviour is analysed. Different types of gene modules can be used depending on the type of biological information used. Babelomics uses the definition of functional modules biological terms extracted from the GO (9), KEGG (10), BioCarta (http://www.biocarta.com/) and Interpro (11) repositories.

Text mining technologies offer the possibility of defining new types of functional modules. Typically, the relationships between different biomedical entities and genes are estimated on the basis of their co-occurrences in sentences (12). Babelomics includes two interesting functional aspects: disease-related and chemical terms. Gene modules associated with diseases or disease symptoms as well as gene modules associated to different chemical entities (drugs, toxics, etc.) can be defined in this way. In opposition to the conventional gene modules, the modules defined by co-citation scores are not discrete, but continuous entities [see (13) for details].

Information of regulatory nature can be found in different repositories such as Transfac (14) or CisRed (15). This information can be used to define regulatory modules. Details on the way the modules are defined can be found in (6). Also, levels of regulation other than purely transcriptional can be considered by including information of miRNAs, whose role in the negative regulation of the expression of their target genes has recently been demonstrated (16). miRNA target genes can be considered a special type of regulatory modules. Information on miRNA target genes can be extracted from miRBase (17).

Gene expression data already available in databases can be employed to define tissue- or phenotype-specific gene expression profiles that can be used to check the similarity to the experimental observation to the profile of healthy or diseased tissues. SAGE Tag libraries from the Cancer Genome Anatomy Project (comprising a total of 279 human libraries of 29 tissues and 190 mouse libraries from 26 tissues, available at http://cgap.nci.nih.gov/SAGE) and gene expression data from the Genomics Institute of the Novartis Foundation (comprising a total of 79 human tissues and 61 mouse tissues with normal histology available at http://wombat.gnf.org/index.html) were used here.

Although some module definitions are restricted to humans, the most common ones (GO, Interpro and Transfac) are available for several model organisms (*Anopheles gambiae*, *Arabidopsis thaliana*, *Bos taurus*, *Caenorhabditis elegans*, *Danio rerio*, *Drosophila*

melanogaster, *Gallus gallus*, *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *Saccharomyces cerevisiae*), and have been cross-referenced using Ensembl (18) identifiers.

In addition to the modules available, Babelomics allow users to define their own gene modules. In a straightforward manner, a file containing the correspondences between genes and annotations (user-defined module labels) can be provided to the program. This allows using customized versions of conventional annotations or even defining completely new annotations so as Babelomics can build up gene modules based on such annotations.

TESTING STRATEGIES IN FUNCTIONAL PROFILING

Functional enrichment

Functional enrichment methods aim to test the enrichment in any given biological property within a group of genes previously selected in a high-throughput experiment. Typical pre-selection processes applied in this first step are differential gene expression, gene co-expression (clustering) or predictive signatures (19). The biological properties of interest are represented by gene modules as defined above. Enrichment of the selected genes in a given module is tested in a second, independent step, by means of Fisher's exact test for 2×2 contingency tables (20,21). Other similar tests, such as the hypergeometric, χ^2 and binomial, can also be applied and are considered to give similar results (20). When the entity tested for enrichment is not a discrete class but a continuous variable, such as in the case of text mining-based bioentities, the test used is Kolmogorow-Smirnov. Since many tests are conducted in order to check all the gene modules, adjustment for multiple testing, such as FDR (22) of others, is used here. See reviews in (2).

Gene set analysis (GSA)

There are different methods (collectively known as GSA methods), which provide a more sensitive approach to functional profiling. GSEA (3) that use a non-parametrical version of a Kolmogorov-Smirnov test, is the most popular among them. There are a considerable number of GSA methods (2,4), many of them available as tools (Table 1). Babelomics implements a second generation version of GSA methods, the segmentation test Fatiscan (23), which has the advantage of being independent from both the type of experiment that generated the data and from the experimental design. This major advantage allows applying it to different microarray-based experimental designs (case control, multiclass, survival, etc.) or even to other type of data (e.g. large-scale genotyping, evolutionary analysis, etc.) (24).

Gene modules defined using text mining-derived functional annotations related to medical terms and chemical compounds can also be used under a GSA perspective. Babelomics includes such a tool (13), being this new addition a novelty unique to the this package.

Testing through the GO hierarchy

Another important aspect particular to gene modules defined using GO annotations is the way in which the structure of the ontology is taken into account. Many programs test each GO module independently, which do not respect dependencies between the GO terms. This constitutes a major drawback given that the true path rule (each term in GO shares all the annotations of all of its descendants) is ignored in this case. Other programs partly circumvent the problem by selecting a particular level of the GO hierarchy and analyse the gene annotations at this level (5,25), use a 'slim' ontology which is a reduced set of terms with more informative content (26) or even try to find the optimal and more informative level for each case (27). Here the enrichment analysis is carried out at different levels and finally, the deepest (the more detailed definition) level at which significance is found is reported. This strategy increases the power of the enrichment tests (2,28).

TECHNICAL IMPROVEMENTS

Internal re-engineering and the new session interface

Babelomics has been completely re-engineered and now it is based on SOAP web services and on new Web 2.0 technology features such as AJAX. This has facilitated the design of a new interface that allows asynchronous use, as well as projects, jobs and user management. Thus, the users can choose between the traditional anonymous sessions without login in (as in previous versions) or to log into the new environment with username and password. This new environment offers persistent sessions in which data keep stored as well as different facilities for tracking of the operations performed. Both options are free.

Also the code for the functional enrichment and the GSA modules has been improved with an evident increase in their speeds.

File and data formats and new data conversion facilities

Babelomics can be used directly from the GEPAS (29,30) to produce the functional annotation of microarray experiments. In this case, GEPAS sends the data in the required format without the necessity of user's intervention. Babelomics can be used alone. In this case there are two data formats for the two main types of analysis. For functional enrichment two lists of gene or protein identifiers are required. These are text files with a gene or protein identifier per line. In the case of GSA the input is a text file with two columns (separated by tabulators). The first column contains gene or protein identifiers and the second one contains the value that represents the hypothesis to be tested (e.g. differential expression according to a *t*-test, survival according to a Cox regression, association to a disease according to an association test for SNPs, etc.).

Babelomics accepts all the standard gene and protein identifiers, which makes it suitable for the analysis of proteomics experiments too. Actually, an improved tool for protein and gene ID conversion including a large

number of species and databases has been implemented. More species and gene references have been added and now the converter tool supports >10 species and >40 ID references for human (including SNP and orthologous information). Besides the web interface a public web service API is provided, allowing anyone to access the data from their code.

Also gene annotation correspondence tables can be imported, which allows the use of user-defined functional modules. Again the file format is very simple. It is a text file with two columns separated by tabulators. The first column contains gene identifiers and the second one contains the corresponding annotations, from which gene modules will be built up by Babelomics. Such annotations are arbitrary names and define categorical classes.

FUNCTIONAL ANNOTATION OF UNKNOWN SEQUENCES

A major drawback of many functional analysis tools is that they can only be applied to organisms with functional information available in public databases often leaving aside non-model organisms. In order to overcome this limitation we included a new module within Babelomics which allows the functional annotation of (novel) sequence data in an automatic and high-throughput manner. This module integrates the Blast2GO annotation method (8), which provides homology-based transfer of Gene Ontology terms to uncharacterized sequences. Blast2GO uses BLAST (31) on FASTA-formatted DNA or protein sequences to find homologues. Multiple hits per query sequence are then mapped to their existing GO annotations. An annotation rule further selects the functional terms appropriate for the query sequence from the pool of candidate annotations. The rule takes into account the degree of homology, the length and hit coverage percentage of the matching region, the annotation evidence of existing functional information and the hierarchical structure of the gene ontology. All these parameters can be adjusted by the user permitting customized annotation configurations.

Once the sequences have been annotated they are available in the user session to be used in any of the Babelomics options.

GENE MODULE-RELATED TOOLS

Filters

Babelomics allows for sub-selection of gene annotations, in which gene modules are based, in order to test hypothesis in a more focused and sensitive manner. Removing from the analysis modules whose testing is unnecessary and superfluous increases the power of the tests in the multiple-testing adjustment step. An interactive component allows selecting subsets of annotations in any of the repositories used based on keywords and on the size of the gene module defined by them. In the particular case of GO, there is the possibility of using the level of the DAG and the evidence code as filtering criteria.

GO-Graphviewer

Another interesting feature is the graphical viewer of the GO hierarchy. This tool generates joined gene ontology graphs to create overviews of the functional context of groups of sequences. Interactive graph visualization allows the navigation of large and unwieldy graphs often generated when trying to biologically explore large sets of sequence annotations. Zoom and graph navigation is provided through the tool.

Graph colouring and highlighted information content are provided through a colour scale proportional to annotation weight. A term annotation weight can be computed as the number of genes annotated to that GO term or as an annotation confluence score. This confluence score (Node-Score) takes into account the number of genes converging at one GO term and penalizes by the distance to the term where each sequence actually was annotated. Assigned sequences and Node-Scores can be also displayed at the terms level.

COMPARISON TO OTHER AVAILABLE TOOLS

There are a large number of web-based tools for functional profiling, which demonstrates the importance of this issue and the demand of tools to address it. We have used Scholar Google citations as a measure of the impact of each tool in scientific community and of its dissemination. This index constitutes an indirect estimation of the citation in papers. Table 1 shows (for weekly updated table see http://bioinfo.cipf.es/docus/tools-citations/functional_profiling) that the most popular tool by far is the GSEA (3,32), designed for GSA. Only four tools surpass the threshold of 400 citations (GSEA (3,32), DAVID (33), GOminer (34,35) and Babelomics (6,7)), and only three more have received over 200 citations [MAPPfinder (36), GOStat (25) and Ontotools (37)]. The first four web tools monopolize the 60% of the total number of citations, proportion that rises up to 80% if the next three are considered. It is worth noting that only two tools offer the possibility of performing both, functional enrichment and GSA, types of analysis: Babelomics (6,7) and FuncAssociate (38). Obviously, any citation index is affected by the date in which the paper was published. Consequently, GSA methods, which are newer, are affected by this fact.

CONCLUSIONS

Babelomics is a long-term project that started in 2004 with the publication of the FatiGO (5), which is now a constituent part of the package. Later, different methods (functional enrichment and gene set enrichment methods) along with a number of functional module definitions were assembled as the prototype (6,7) of today's Babelomics project. This project aims to provide the scientific community with an advanced set of tools for the functional profiling of high-throughput transcriptomic, genomic and proteomic data, without renouncing to a user-friendly and intuitive use. As the Functional Genomics node of the Spanish Institute of Bioinformatics

(INB, http://www.inab.org) and being part of the Spanish Network of Cancer (RTICC, http://www.rticcc.org) and the Network of Centres for Research in Rare Diseases (CIBERER, http://www.ciberer.es) we have a direct contact with researchers who provided us much of the feedback necessary to make Babelomics a useful tool.

This new version of Babelomics includes new methods and new module definitions of different nature (functional, regulatory, phenotypic, etc.). Also a new tool for the '*de novo*' functional annotation of sequences, the Blast2GO (8), has been included in the system. Innovative visualization methods and new interfaces have been implemented in order to improve the presentation of the results, an important aspect in the analysis of genome scale experiments.

Babelomics is running in a high-end cluster with 10 dedicated Intel XEON Quad-Core CPUs at 2.0 GHz (summing up a total of 40 cores) with a large amount of RAM (total 60 GB).

Although there are many alternatives for the functional profiling of high-throughput experiments (Table 1), Babelomics is a widely used tool which offers a combination of features that makes it unique among other similar resources available.

ACKNOWLEDGEMENTS

This work was supported by grants from the Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER) ISCIII, and projects BIO2005-01078 from the Spanish Ministry of Education and Science and the National Institute of Bioinformatics (www.inab.org), a platform of Genoma España. Funding to pay the Open Access publication charges for this article was provided by grant BIO2005-01078 from the Spanish Ministry of Education and Science.

Conflict of interest statement. None declared.

REFERENCES

1. Hoheisel,J.D. (2006) Microarray technology: beyond transcript profiling and genotype analysis. *Nat. Rev. Genet.*, **7**, 200–210.
2. Dopazo,J. (2006) Functional interpretation of microarray experiments. *OMICS*, **10**, 398–410.
3. Mootha,V.K., Lindgren,C.M., Eriksson,K.F., Subramanian,A., Sihag,S., Lehar,J., Puigserver,P., Carlsson,E., Ridderstrale,M., Laurila,E. *et al.* (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.*, **34**, 267–273.
4. Goeman,J.J. and Buhlmann,P. (2007) Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, **23**, 980–987.
5. Al-Shahrour,F., Diaz-Uriarte,R. and Dopazo,J. (2004) FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, **20**, 578–580.
6. Al-Shahrour,F., Minguez,P., Tarraga,J., Montaner,D., Alloza,E., Vaquerizas,J.M., Conde,L., Blaschke,C., Vera,J. and Dopazo,J. (2006) BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments. *Nucleic Acids Res.*, **34**, W472–W476.
7. Al-Shahrour,F., Minguez,P., Vaquerizas,J.M., Conde,L. and Dopazo,J. (2005) BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res.*, **33**, W460–W464.
8. 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.
9. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
10. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
11. Mulder,N.J., Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Binns,D., Bork,P., Buillard,V., Cerutti,L., Copley,R. *et al.* (2007) New developments in the InterPro database. *Nucleic Acids Res.*, **35**, D224–D228.
12. Krallinger,M. and Valencia,A. (2005) Text-mining and information-retrieval services for molecular biology. *Genome Biol.*, **6**, 224.
13. Minguez,P., Al-Shahrour,F., Montaner,D. and Dopazo,J. (2007) Functional profiling of microarray experiments using text-mining derived bioentities. *Bioinformatics*, **23**, 3098–3099.
14. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
15. Robertson,G., Bilenky,M., Lin,K., He,A., Yuen,W., Dagpinar,M., Varhol,R., Teague,K., Griffith,O.L., Zhang,X. *et al.* (2006) cisRED: a database system for genome-scale computational discovery of regulatory elements. *Nucleic Acids Res.*, **34**, D68–D73.
16. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
17. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
18. Flicek,P., Aken,B.L., Beal,K., Ballester,B., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cunningham,F., Cutts,T. *et al.* (2008) Ensembl 2008. *Nucleic Acids Res.*, **36**, D707–D714.
19. Allison,D.B., Cui,X., Page,G.P. and Sabripour,M. (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.*, **7**, 55–65.
20. Rivals,I., Personnaz,L., Taing,L. and Potier,M.C. (2007) Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, **23**, 401–407.
21. Fisher,L. and van Belle,G. (1993) *Biostatistics: A Methodology for the Health Sciences*. Wiley, New York.
22. Benjamini,Y. and Yekutieli,D. (2001) The control of false discovery rate in multiple testing under dependency. *Ann. Stat.*, **29**, 1165–1188.
23. Al-Shahrour,F., Diaz-Uriarte,R. and Dopazo,J. (2005) Discovering molecular functions significantly related to phenotypes by combining gene expression data and biological information. *Bioinformatics*, **21**, 2988–2993.
24. 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.
25. Beissbarth,T. and Speed,T.P. (2004) GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, **20**, 1464–1465.
26. Martin,D., Brun,C., Remy,E., Mouren,P., Thieffry,D. and Jacq,B. (2004) GOToolBox: functional analysis of gene datasets based on Gene Ontology. *Genome Biol.*, **5**, R101.
27. Al-Shahrour,F., Minguez,P., Tarraga,J., Medina,I., Alloza,E., Montaner,D. and Dopazo,J. (2007) FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res.*, **35**, W91–W96.
28. Khatri,P. and Draghici,S. (2005) Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, **21**, 3587–3595.

29. Herrero,J., Al-Shahrour,F., Diaz-Uriarte,R., Mateos,A., Vaquerizas,J.M., Santoyo,J. and Dopazo,J. (2003) GEPAS: a web-based resource for microarray gene expression data analysis. *Nucleic Acids Res.*, **31**, 3461–3467.
30. Montaner,D., Tarraga,J., Huerta-Cepas,J., Burguet,J., Vaquerizas,J.M., Conde,L., Minguez,P., Vera,J., Mukherjee,S., Valls,J. *et al.* (2006) Next station in microarray data analysis: GEPAS. *Nucleic Acids Res.*, **34**, W486–W491.
31. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
32. Subramanian,A., Tamayo,P., Mootha,V.K., Mukherjee,S., Ebert,B.L., Gillette,M.A., Paulovich,A., Pomeroy,S.L., Golub,T.R., Lander,E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
33. Dennis,G. Jr., Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.*, **4**, P3.
34. Zeeberg,B.R., Feng,W., Wang,G., Wang,M.D., Fojo,A.T., Sunshine,M., Narasimhan,S., Kane,D.W., Reinhold,W.C., Lababidi,S. *et al.* (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.*, **4**, R28.
35. Zeeberg,B.R., Qin,H., Narasimhan,S., Sunshine,M., Cao,H., Kane,D.W., Reimers,M., Stephens,R.M., Bryant,D., Burt,S.K. *et al.* (2005) High-Throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of Common Variable Immune Deficiency (CVID). *BMC Bioinformatics*, **6**, 168.
36. Doniger,S.W., Salomonis,N., Dahlquist,K.D., Vranizan,K., Lawlor,S.C. and Conklin,B.R. (2003) MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.*, **4**, R7.
37. Draghici,S., Khatri,P., Bhavsar,P., Shah,A., Krawetz,S.A. and Tainsky,M.A. (2003) Onto-Tools, the toolkit of the modern biologist: Onto-Express, Onto-Compare, Onto-Design and Onto-Translate. *Nucleic Acids Res.*, **31**, 3775–3781.
38. Berriz,G.F., King,O.D., Bryant,B., Sander,C. and Roth,F.P. (2003) Characterizing gene sets with FuncAssociate. *Bioinformatics*, **19**, 2502–2504.
39. Khatri,P., Bhavsar,P., Bawa,G. and Draghici,S. (2004) Onto-Tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments. *Nucleic Acids Res.*, **32**, W449–W456.
40. Khatri,P., Desai,V., Tarca,A.L., Sellamuthu,S., Wildman,D.E., Romero,R. and Draghici,S. (2006) New Onto-Tools: Promoter-Express, nsSNPCounter and Onto-Translate. *Nucleic Acids Res.*, **34**, W626–W631.
41. Khatri,P., Sellamuthu,S., Malhotra,P., Amin,K., Done,A. and Draghici,S. (2005) Recent additions and improvements to the Onto-Tools. *Nucleic Acids Res.*, **33**, W762–W765.
42. Khatri,P., Voichita,C., Kattan,K., Ansari,N., Khatri,A., Georgescu,C., Tarca,A.L. and Draghici,S. (2007) Onto-Tools: new additions and improvements in 2006. *Nucleic Acids Res.*, **35**, W206–W211.
43. Zhang,B., Schmoyer,D., Kirov,S. and Snoddy,J. (2004) GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *BMC Bioinformatics*, **5**, 16.
44. Robinson,M.D., Grigull,J., Mohammad,N. and Hughes,T.R. (2002) FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics*, **3**, 35.
45. Castillo-Davis,C.I. and Hartl,D.L. (2003) GeneMerge—post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics*, **19**, 891–892.
46. Masseroli,M., Galati,O. and Pinciroli,F. (2005) GFINDer: genetic disease and phenotype location statistical analysis and mining of dynamically annotated gene lists. *Nucleic Acids Res.*, **33**, W717–W723.
47. Masseroli,M., Martucci,D. and Pinciroli,F. (2004) GFINDer: Genome Function INtegrated Discoverer through dynamic annotation, statistical analysis, and mining. *Nucleic Acids Res.*, **32**, W293–W300.
48. Zhang,B., Kirov,S. and Snoddy,J. (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.*, **33**, W741–W748.
49. Volinia,S., Evangelisti,R., Francioso,F., Arcelli,D., Carella,M. and Gasparini,P. (2004) GOAL: automated Gene Ontology analysis of expression profiles. *Nucleic Acids Res.*, **32**, W492–W499.
50. Mlecnik,B., Scheideler,M., Hackl,H., Hartler,J., Sanchez-Cabo,F. and Trajanoski,Z. (2005) PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res.*, **33**, W633–W637.
51. Tomfohr,J., Lu,J. and Kepler,T.B. (2005) Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics*, **6**, 225.
52. Boorsma,A., Foat,B.C., Vis,D., Klis,F. and Bussemaker,H.J. (2005) T-profiler: scoring the activity of predefined groups of genes using gene expression data. *Nucleic Acids Res.*, **33**, W592–W595.
53. Vencio,R.Z., Koide,T., Gomes,S.L. and Pereira,C.A. (2006) BayGO: Bayesian analysis of ontology term enrichment in micro-array data. *BMC Bioinformatics*, **7**, 86.
54. Scheer,M., Klawonn,F., Munch,R., Grote,A., Hiller,K., Choi,C., Koch,I., Schobert,M., Hartig,E., Klages,U. *et al.* (2006) JProGO: a novel tool for the functional interpretation of prokaryotic microarray data using Gene Ontology information. *Nucleic Acids Res.*, **34**, W510–W515.
55. Nam,D., Kim,S.B., Kim,S.K., Yang,S., Kim,S.Y. and Chu,I.S. (2006) ADGO: analysis of differentially expressed gene sets using composite GO annotation. *Bioinformatics*, **22**, 2249–2253.
56. Backes,C., Keller,A., Kuentzer,J., Kneissl,B., Comtesse,N., Elnakady,Y.A., Muller,R., Meese,E. and Lenhof,H.P. (2007) GeneTrail—advanced gene set enrichment analysis. *Nucleic Acids Res.*, **35**, W186–W192.
57. Kim,S.B., Yang,S., Kim,S.K., Kim,S.C., Woo,H.G., Volsky,D.J., Kim,S.Y. and Chu,I.S. (2007) GAzer: gene set analyzer. *Bioinformatics*, **23**, 1697–1699.
58. Goffard,N. and Weiller,G. (2007) PathExpress: a web-based tool to identify relevant pathways in gene expression data. *Nucleic Acids Res.*, **35**, W176–W181.