BMC Bioinformatics

Software

Open Access

# The Annotation, Mapping, Expression and Network (AMEN) suite of tools for molecular systems biology

Frédéric Chalmel and Michael Primig*

Address: Institut National de la Santé et de la Recherche Médicale (Inserm) Unité 625, Groupe d'Etude de la Reproduction chez l'Homme et les Mammifères, Institut Fédératif de Recherche 140; Université de Rennes 1, Campus de Beaulieu, F-35042 Rennes, France

Email: Frédéric Chalmel - frederic.chalmel@rennes.inserm.fr; Michael Primig* - michael.primig@rennes.inserm.fr

* Corresponding author

Published: 6 February 2008

*BMC Bioinformatics* 2008, **9**:86 doi:10.1186/1471-2105-9-86

Received: 12 October 2007
Accepted: 6 February 2008

This article is available from: http://www.biomedcentral.com/1471-2105/9/86

© 2008 Chalmel and Primig; licensee BioMed Central Ltd.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

**Background:** High-throughput genome biological experiments yield large and multifaceted datasets that require flexible and user-friendly analysis tools to facilitate their interpretation by life scientists. Many solutions currently exist, but they are often limited to specific steps in the complex process of data management and analysis and some require extensive informatics skills to be installed and run efficiently.

**Results:** We developed the Annotation, Mapping, Expression and Network (AMEN) software as a stand-alone, unified suite of tools that enables biological and medical researchers with basic bioinformatics training to manage and explore genome annotation, chromosomal mapping, protein-protein interaction, expression profiling and proteomics data. The current version provides modules for (i) uploading and pre-processing data from microarray expression profiling experiments, (ii) detecting groups of significantly co-expressed genes, and (iii) searching for enrichment of functional annotations within those groups. Moreover, the user interface is designed to simultaneously visualize several types of data such as protein-protein interaction networks in conjunction with expression profiles and cellular co-localization patterns. We have successfully applied the program to interpret expression profiling data from budding yeast, rodents and human.

**Conclusion:** AMEN is an innovative solution for molecular systems biological data analysis freely available under the GNU license. The program is available via a website at the Sourceforge portal which includes a user guide with concrete examples, links to external databases and helpful comments to implement additional functionalities. We emphasize that AMEN will continue to be developed and maintained by our laboratory because it has proven to be extremely useful for our genome biological research program.

## Background

High-throughput DNA sequencing, microarray-based mRNA expression profiling, proteomics experiments and protein-protein interaction assays have been yielding large and complex datasets that need to be integrated with functional information at the gene- or genome level. Large scale expression profiling using microarrays is among the most popular experimental approaches in genome biology and therefore optimized methods are available for all key analytical steps. They include raw data pre-processing, quality control and normalization [1,2], identification of differentially expressed genes during static or time-course

conditions [3-5], gene clustering [6-8] and searching for significant over- or under-representation of functional annotation in expression clusters [9-11]. The *Bioconductor* project provides numerous software packages developed in *R* that are devoted to high-throughput analysis tasks [12-14]. However, installing and running them requires extensive programming skills that are not yet commonplace among life scientists. To alleviate this problem, programs with a convenient Graphical User Interface (GUI) have been developed that facilitate functional analyses in most cases limited to annotation by the Gene Ontology (GO) consortium [10] or restricted to a set of genomes [15,16]. Other tools correlate expression with chromosomal localization [17-21], protein-protein interaction [22] or pathway data [23].

In an attempt to combine analysis steps many web-based applications have been developed [24-30]. They are free and do not require maintenance work. However, their accessibility and speed depend upon web-traffic, server availability and the specifications of the analyses procedure. Moreover, web-based systems usually provide preconfigured and inflexible approaches to data analysis and often do not include advanced options to combine different types of high-throughput data. In order to address these issues and to allow for integrated exploration of different types of data we have developed the Annotation, Mapping, Expression, and Network (AMEN) program that enables users to explore and analyse multifaceted high-throughput biological data. It includes a suite of tools and algorithms for which parameters can be fine-tuned and analysis steps ordered as required. AMEN covers array data management, analysis and interpretation in a manner similar to EXPANDER 2.0 [29]. However, our software includes more options to combine different types of data and enables users to import, not only genome annotation and transcriptome-, but also proteome and interactome data without species restrictions.

## Implementation

The AMEN software architecture consists of four layers implemented in Tcl/Tk (Figure 1) [31]. The first layer provides modules for uploading, formatting and pre-processing expression, annotation, chromosomal mapping and protein-protein interaction data. The second layer is the user-friendly GUI of the main application window that employs popup menus for *Project*, *Upload data*, *Tools*, *Views* and *Options* functionalities (Figure 2). Six panels provide access to lists of items such as probe IDs, genes, proteins (Group), and data from RNA/protein profiling experiments (Expression), genome annotation (Annotation), protein-protein and protein-DNA interaction (Interaction), chromosomal localization (Mapping) as well as the output of the statistical module (Statistic). Function buttons below each panel enable users to scroll item lists (Up/Down), mark items (Select all/Deselect), change the content of a panel (Add/Remove), change item names (Modify) or change the file content (Edit). Selected items in each panel highlighted in yellow are combined into different workflows by the user. For example, selecting mouse genes (Annotation panel) showing differential expression in testis (DET) and peak signals in somatic Sertoli cells (SO) compared to mitotic (MI), meiotic (ME) and post-meiotic (PM) germ cells (Group panel) and Spermatogenesis data (Expression panel) (see Figure 2) yields a graphical display of RNA profiling signals generated by the module controlled via the *Views* > *Expression data* > *Profiles* menu (Figure 3). Selecting protein network data (Interaction panel) enables users to display interaction patterns (see Figure 4 in reference [32]). Selecting chromosomal localization (Mapping panel) and statistical items enables users to correlate expression and mapping information or to reveal a link between transcriptional patterns and roles in biological processes (see reference [33] for more details). In the background, the third layer automatically creates and runs scripts for the statistical computing environment R which execute statistical calculations and clustering methods implemented in Bioconductor packages. The fourth layer displays the output based on Tk scripts and the graph rendering software GraphViz.

Our program requires Tcl/Tk, R and GraphViz programs available for all frequently used operating systems such as MS Windows, UNIX, Linux and Mac OS X to be preinstalled. Detailed downloading and installation instructions are accessible via the Sourceforge website. We also provide a Windows version of the software that preinstalls Tcl/Tk, R and GraphViz embedded into it. Note that AMEN supersedes goCluster, a much simpler tool previously developed in our laboratory [34]. We have decided to discontinue goCluster because it lacked key analysis features and the cost for further development and maintenance outweighed the benefit for our lab and the community. We emphasize that AMEN is frequently updated because this software is a key tool for our ongoing biomedical studies. Its structure greatly facilitates the implementation of new modules. Indeed, a single Tcl code line is sufficient to include additional functionalities into the main GUI.

## Results and Discussion

### *Data uploading*

The typical workflow involves five types of modules available in the current release: data uploading and preprocessing, statistical filtration, clustering, functional mining, and visualization. Data are imported and combined within an analysis project using the main application window (Figure 2). It includes six panels corresponding to different input data: items (such as

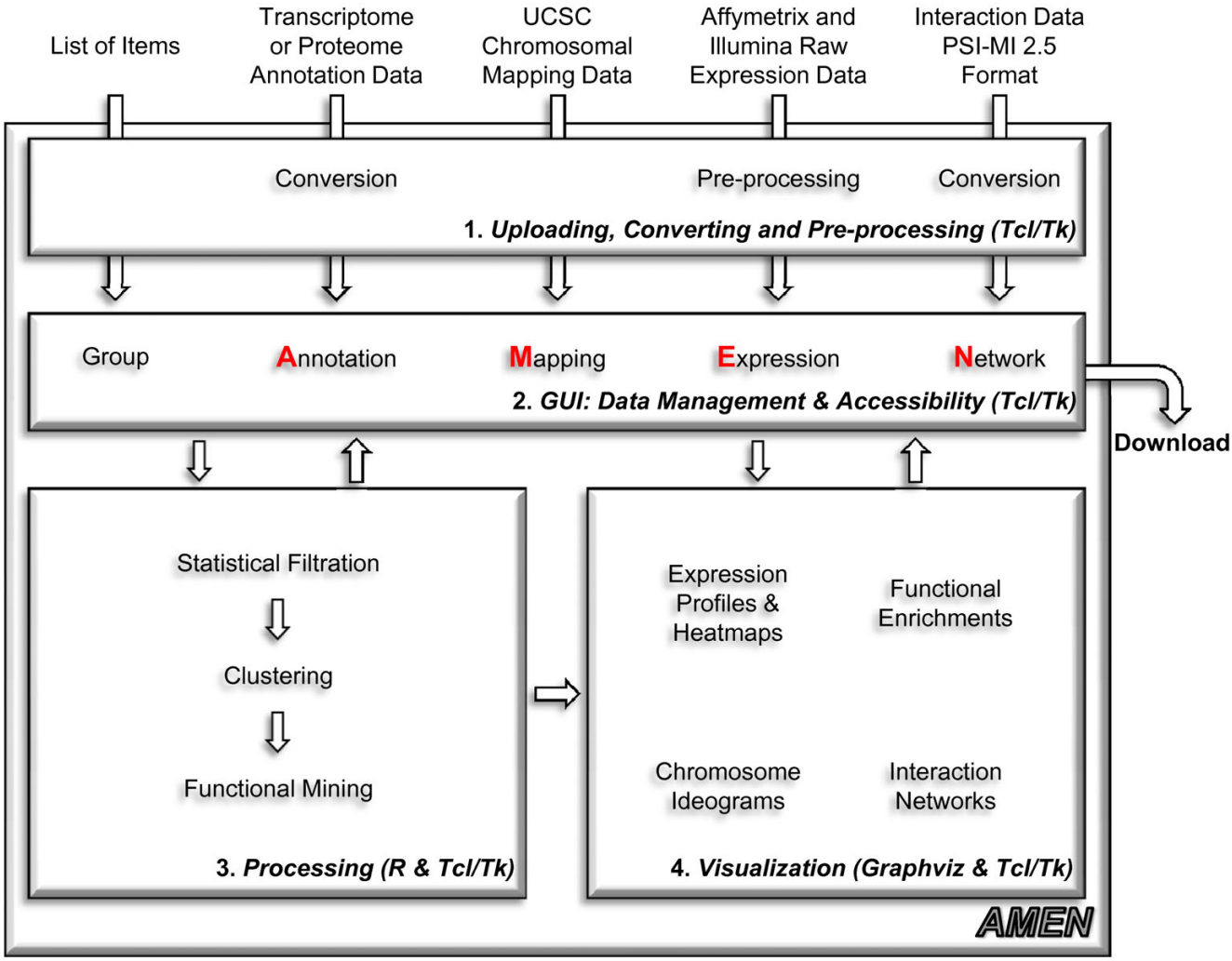



**Figure 1**
**The AMEN architecture**. A flow-chart diagram of the software and work-flow is shown.

genes, transcripts, proteins or probe identifiers), expression signal, functional annotation, (protein-protein) interaction, chromosomal location and statistical data. During this process items (such as probe set IDs) are automatically associated with other data types (including gene symbols, chromosomal position and GO terms). This interface makes it easy to access the data and to design an optimal analysis procedure. Data are input as text files in tab-delimited compatible format to ensure compatibility with all operating systems.

*Group data*

Users can upload pre-selected lists of items (called "main entries", such as probe, transcript, gene, or protein identifiers) or they can obtain such lists via statistical filtration, clustering, and/or visualization modules (see Figure 3 and 4).

*Expression data*

Currently, expression data quality control and normalization modules are implemented for commercial Affymetrix high density oligonucleotide microarrays (GeneChips) and Illumina Gene Expression BeadArrays. Methods for background correction, variance stabilization and normalization methods are MAS5.0, RMA, GCRMA and RSN [35]. It is also possible to upload pre-normalized expression datasets as long as they are represented as tab-delimited matrices whose rows and columns contain main items (usually probe identifiers) and experimental condition names, respectively.

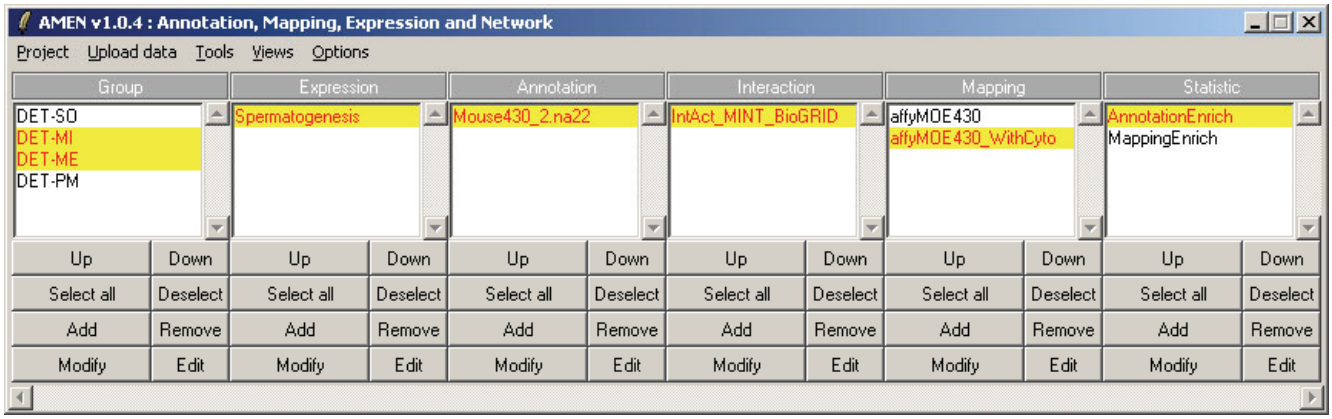



**Figure 2**

**The Main Application Window**. A screen shot of the main application window is given. A possible analysis strategy for mammalian testicular expression data is shown in the six data type panels as indicated. Four groups (clusters) of genes are defined as differentially expressed in testis and somatic (DET-SO), mitotic (DET-MI), meiotic (DET-MI) and post-meiotic (DET-PM) depending on peak expression in Sertoli cells, spermatogonia, spermatocytes and spermatids, respectively. The expression dataset (Spermatogenesis) was obtained with a GeneChip covering approximately 25000 protein-coding mouse genes for which an appropriate annotation file is selected (Mouse430_2.na22). To visualize the interaction network of proteins falling into two selected clusters (DET-MI and ME) information from three sources is combined (IntAct_MINT_BioGRID). To display the chromosomal localization of selected genes falling into given expression clusters files with gene coordinates are available with and without cytological bands (affyMOE430, affyMOE430_WithCyto). Users can choose from statistical analysis of GO term enrichment in clusters (AnnotationEnrich) or gene enrichment on chromosomes (MappingEnrich).

*Annotation, Interaction and Mapping data*

Functional information for transcriptome (Affymetrix and Illumina CSV annotation files), proteome (International Protein Index, EBI and NCBI whole-proteome data files), interactome (Proteomics Standards Initiative-Molecular Interactions 2.5 files used by IntAct, MINT, and BioGRID) and chromosomal mapping analysis (PSL chromosomal location files from the UCSC web site) is imported and converted into the appropriate file format using a straightforward procedure [36-43].

### *Statistical filtration*

These modules output lists of significantly differentially expressed items (represented as transcripts or probe identifiers) identified within a given set of samples. Users can select transcripts showing strong variations across experimental conditions via threshold parameters including expression level cut-off, standard deviation or fold-change. Furthermore, a similarity search module helps retrieve groups of co-expressed transcripts using a specific user-defined pattern. Once a set of target transcripts is identified, the permutation (randomization), moderated t-test (empirical Bayes approach) and non-parametric rank-based statistic methods are employed to determine if changes in signal intensity are reproducible and significant. These methods are implemented in *multtest*, *samr*, *limma* and *RankProducts R* packages, respectively [44-46]. False positives are taken into account by adjusting p-values according to the Hommel (control of the family wise error rate) or Benjamini/HochBerg [determination of the False Discovery Rate (FDR)] multiple testing correction methods [47,48].

### *Clustering*

Clustering methods are used to classify items based on their overall degree of similarity across the experimental conditions. These algorithms are notably critical for the identification of genes that are co-expressed (showing similar patterns of transcription), co-regulated (sharing common promoter elements) or that play roles in a particular biological process. Users can choose between three hierarchical clustering modules: HCLUST (hierarchical), AGNES (AGglomerative NESting) and DIANA (DIvisive ANAlysis) [8,49]. Four supervised partitioning methods include k-means [50], PAM (Partitioning Around Medoids), FANNY (Fuzzy Analysis Clustering) and CLARA (Clustering LARge Applications) [49]. We also included two unsupervised clustering modules called MCLUST (Model-based CLUstering) and HOPACH (Hierarchical Ordered Partitioning and Collapsing Hybrid) that automatically determine the number of clusters in a given dataset [51,52]. Finally, to estimate the quality of the classification or to help identify the optimal number of clusters that yields the best separation of different expression patterns the silhouette plot method is available [53].

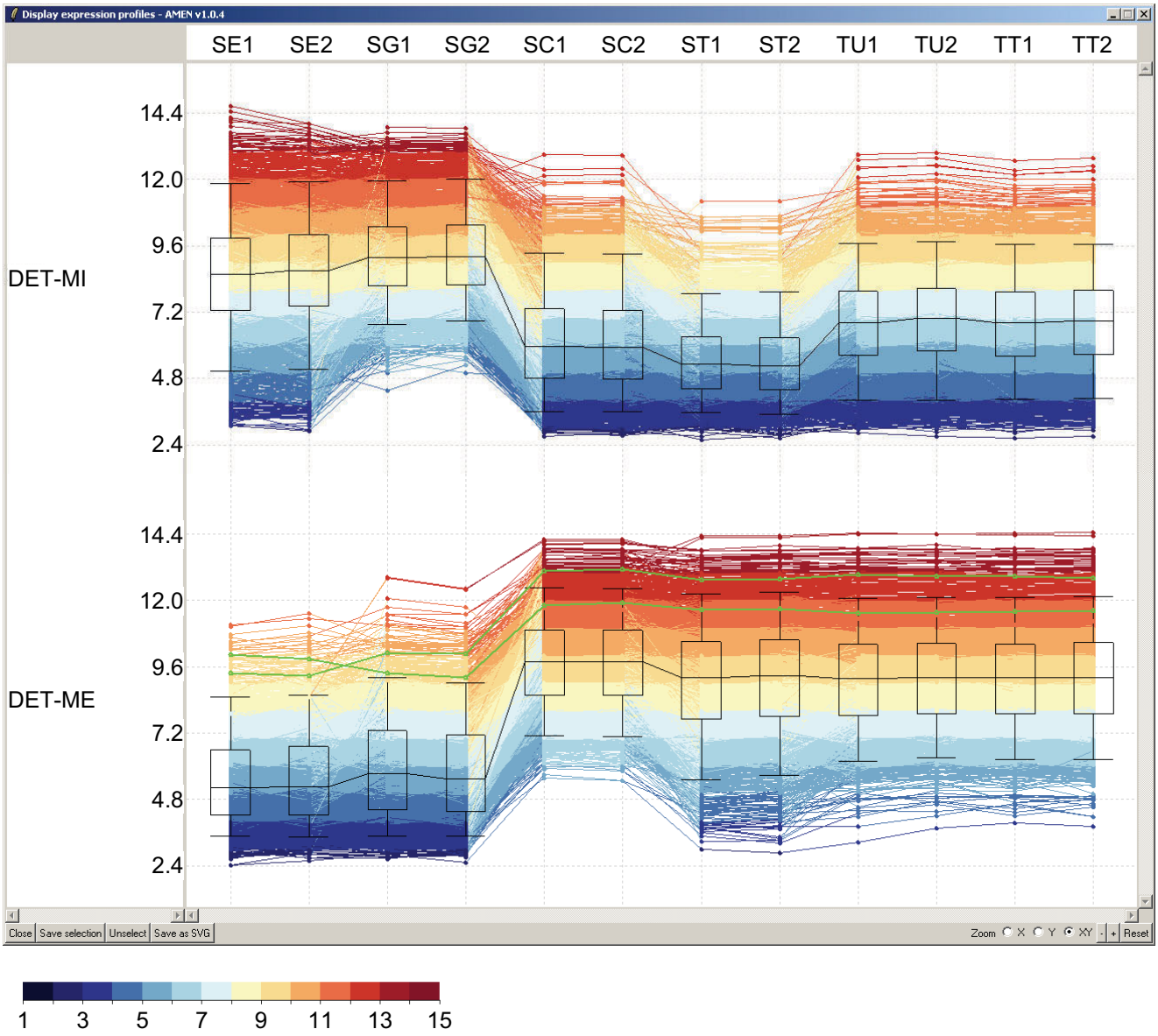



**Figure 3**

**Graphical display of expression profiling data**. Log2-transformed expression signal intensities are plotted against sample names on the Y- and X-axis, respectively. The signal distribution and the median are shown for each sample by a box plot. Data obtained for genes classified as differentially expressed in testis (DET) and showing peak transcription in mitotic (MI) or meiotic (ME) germ cells are displayed (see [33]). Sample names given in duplicate are Sertoli cells (SE), spermatogonia (SG), spermatocytes (SC), spermatids (ST), seminiferous tubules (TU), and total testis (TT). Lines and columns correspond to probe set ids and samples. Expression signals are shown in red (high) or blue (low) as indicated in the scale bar. Green lines represent expression profiles selected by the user.

### *Functional mining*

Expression clusters are validated and further analysed by searching for over- or under-represented functional annotation terms associated with the items (genes) in these clusters using hypergeometric or binomial statistical tests. *p*-values are adjusted using multiple testing correction methods as described above. Functional information is most often provided by the GO consortium [9] but, in principle, AMEN can process data from any source of information present in the uploaded annotation files such as InterPro protein domains, biochemical pathways, chromosomal mapping data or other information provided by the user.

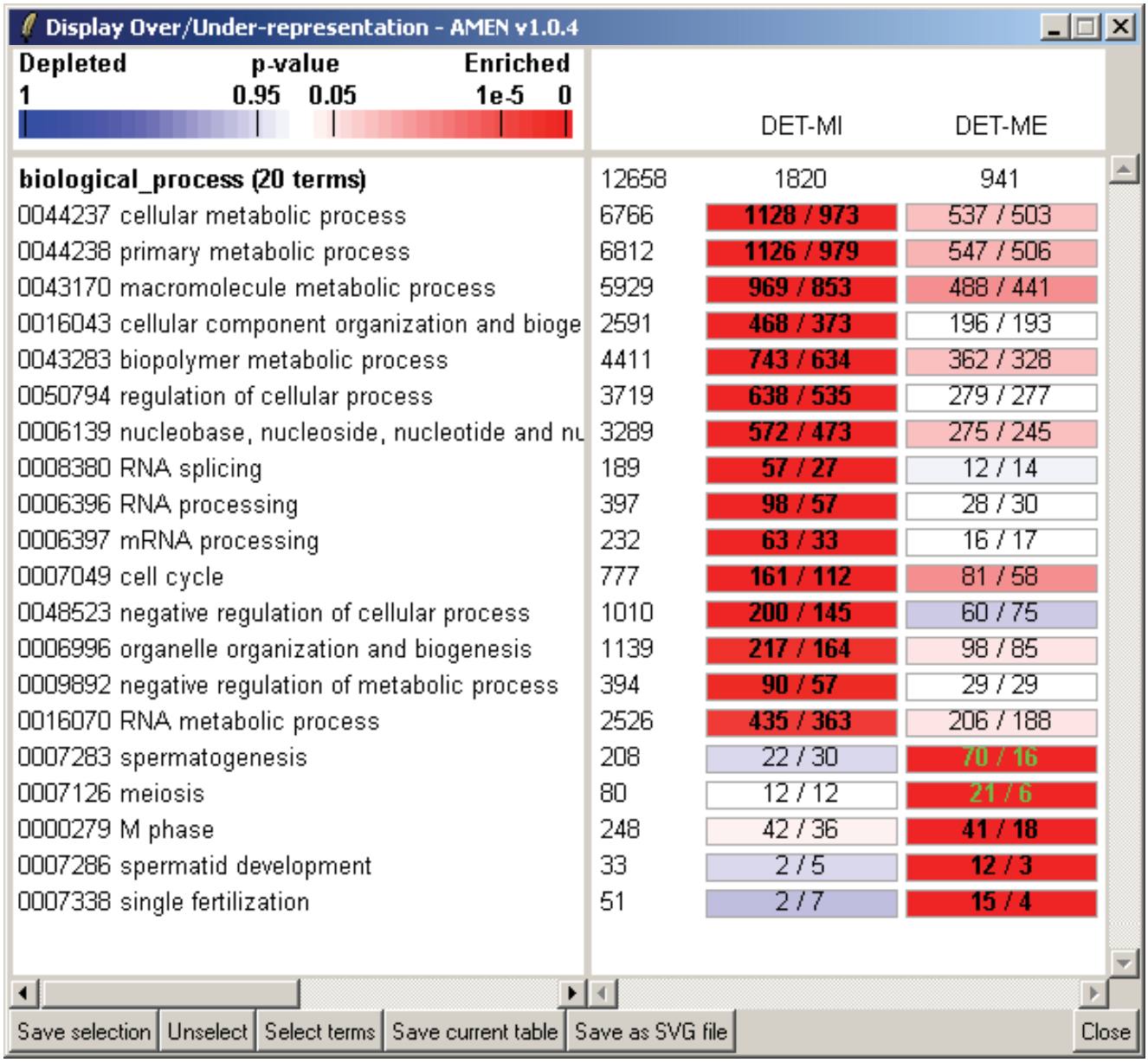



**Figure 4**

**Graphical output of GO term analysis**. An example of over-represented GO terms form the biological process category associated with genes from the DET-MI and -ME expression clusters is shown. The names of expression cluster and the numbers of genes are indicated on top of each column. The number of loci associated with a given GO term is shown to the left of the columns. Numbers of loci as observed and expected are given within color-coded rectangles with red and blue indicating over- and under-representation, respectively, according to the scale bar on top of the GO terms. Numbers in bold or green indicate significantly over-represented terms or genes selected by the user. To obtain the output shown we used an FDR-adjusted *p*-value of < 0.001, an OSIR > 0.1, and the minimum number of genes associated with one term was set to be > 10.

Note that ontology vocabularies have a hierarchical structure such that an item (e.g. gene) is associated with multiple redundant terms. To reduce the annotation term output we employ the Ontology Specific Information Rate (OSIR): $OSIR = (n\text{-}m)/n$, where $n$ and $m$ are the numbers of items associated with a given over-represented term (parent node) and associated to its subordinate over-represented terms (child nodes) respectively. The minimal

*OSIR* threshold value is typically set between 0.05 and 0.20 which means that if less than 5% or 20% of the genes associated with a given parent node are not related with its child nodes, the parent node is eliminated.

### *Data visualization*

Four types of visualization modules are currently implemented. First, users can display expression data as false color-coded heat maps or as graphs (Figure 3). Second, a color-coded graphical module to display significantly over- or under-represented functional annotation terms among clusters (Figure 4). It is possible for this output to contain data from multiple experiments and simultaneously display as distinct columns. Alternatively, it is possible to display over-represented *GO* terms and their related parent nodes as directed acyclical graphs. Third, a module is included to create chromosome ideograms according to the International Standard on Cytogenetic Nomenclature (ISCN) (Figure 5). This functionality helps reveal correlations between expression patterns and chro-




**Figure 5**

**Chromosomal ideogram representation**. An ISCN ideogram of the mouse X chromosome is shown (column 1). The chromosomal localization of genes in the DET-MI expression cluster is marked by red (plus or top DNA strand) and blue (minus or bottom DNA strand) lines (column 2). A color coded heat map (see scale bar in Figure 3) shows expression signals for each sample (column 3). The numbers of mapped genes within consecutive regions of 10 Mbp are plotted on the X- and Y-axes, respectively (column 4). Color coded bars show the numbers of observed loci with red and blue indicating over- or under-representation. Grey bars represent the number of loci falling into a given region by chance. Red arrows mark regions that are enriched in loci (FDR-adjusted *p*-value < 0.001). The remaining columns 5–7 show that the X-chromosome is devoid of meiotic genes falling into the DET-ME cluster.

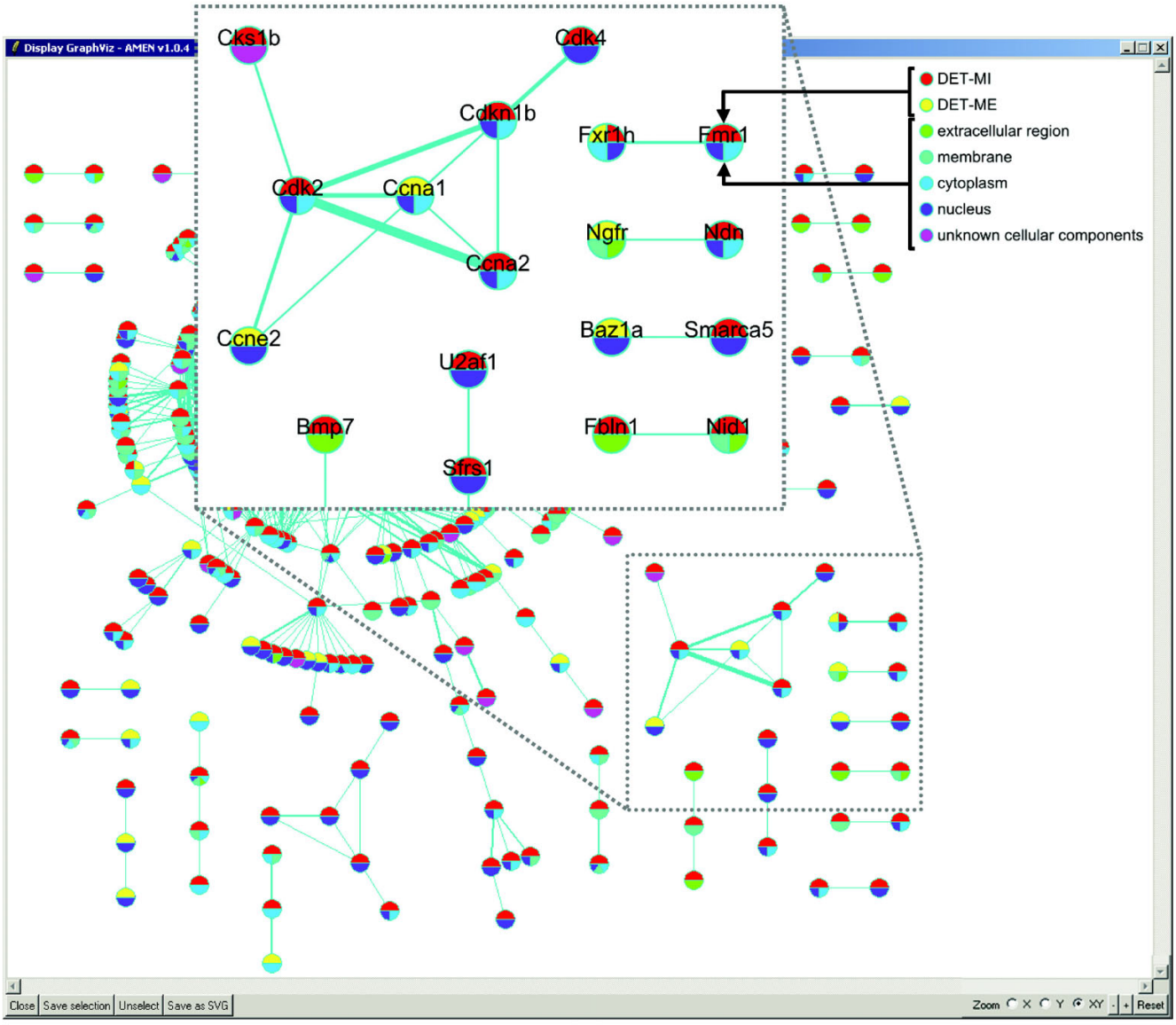



**Figure 6**

**Display of Protein-protein interaction networks**. A global view of protein-protein interactions based on combined mouse, rat and human data retrieved from IntAct, MINT and BioGRID databases is shown. Blue lines connecting nodes (proteins) represent direct physical interactions. Line thickness increases with the number of published observations supporting the predicted interaction. Nodes are color coded to indicate the expression cluster the protein belongs to (top half) and the sub-cellular component to which it localizes (bottom half) as shown.

mosomal mapping of selected target items. It includes ISCN ideograms together with heat maps of expression data and histograms showing observed and expected numbers of genes in a given region. Finally, the complete set of *GraphViz* tools to draw network graphs of protein-protein or other types of interactions is available (Figure 6). Nodes representing biological items (proteins, genes) are color-coded to facilitate the interpretation of relationships between expression clusters, sub-cellular location and interaction data.

The visualization modules output interactive and clickable images providing detailed information for each cluster and gene that can be manually selected for further analysis. They also provide manual zooming (in the X, Y or both directions) and panning features enabling users to focus on specific results of interest. Finally, users can employ the Scalable Vector Graphics format (SVG) for viewing with all web browsers and further processing with SVG editors such as Inkscape or Adobe Illustrator [54,55].

### *Data export*

AMEN provides a module to export selected lists of items as an HTML table file which can be opened and further processed in spreadsheet applications such as MS Excel. Users can select different types of annotation, mapping and expression data and determine their order within the file to be exported.

### *Application of AMEN*

The program was critical for our study of the testicular expression program in human and rodents where a clear correlation between germline expression and reproductive function was established [33]. Note that this study included work on the negative correlation between meiotic gene expression and X-chromosome localization (see Additional file 1). Furthermore, we have used AMEN to compare testicular transcriptome and proteome data and to explore the protein-protein interaction network of gene products differentially expressed between testicular somatic and germ cells [32].

Ongoing work includes the expression signature of high infertility risk associated with undescended testes (Hadziselimovic et al., in revision), high-throughput analysis of mRNAs and proteins present in residual bodies (Rolland et al. and Brun et al., unpublished) as well as enrichment of functional annotation among the target genes of Abf1, an essential budding yeast DNA binding transcription factor (U. Schlecht and M. Primig, in press), and Ume6, a regulator involved in mitotic repression of meiotic genes in *S. cerevisiae* (T. Walther and M. Primig, in preparation). Our software is thus suitable for molecular systems biological data analysis combining data on DNA, mRNA and proteins across different species.

### *Comparison to other solutions*

Since AMEN is a freely available standalone molecular systems biology analysis tool we have compared it to typical examples of such software and not to web server-based applications that, in our opinion, are often less flexible and less complete than locally installed programs. Most available local solutions are R packages such as affylmGUI and illuminaGUI which provide quality control (QC), pre-processing, statistical tests and clustering applicable to Affymetrix GeneChip and Illumina BeadArray data, respectively [56,57]. BRB-ArrayTools is an MS Excel plug-in providing advanced statistical tests for the identification of differentially expressed genes, GO term enrichment and the option to expand functionalities using external R-scripts [58]. AMDA provides various QC, normalization, statistical and clustering functionalities and also includes a GUI, as well as GO term and KEGG enrichment [59]. As compared to these solutions our software has useful additional features such as an elaborate main application window facilitating work-flow management, sophisticated graphical output of (for example) GO term enrichment, cross-microarray platform compatibility, proteomics data import functionality, protein-protein and protein-DNA network data processing, and chromosomal localization and enrichment (Additional file 1). Finally, the graphical output of AMEN is interactive and enables users to sub-select and save lists of items.

### *Future development*

AMEN is regularly updated with new functionalities and modules. We intend to include, in the near future, data management for molecular pathway databases such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG) in order to display metabolic pathways combined with protein-protein interaction and expression data [60]. We also plan to implement additional data pre-processing and normalization algorithms for the most recent generations of all-exon [61] and tiling microarrays [62,63] as well as a novel Principle Component Analysis (PCA) statistical module. Finally, we will integrate AMEN with MIMAS [64], our own solution for array data management and annotation to provide our laboratory and the community with a complete package for storing, describing, analysing and interpreting high-throughput data.

## Conclusion

AMEN facilitates the design and execution of optimized procedures for processing, analysis and interpretation of multifaceted high-throughput data. Key advantages include: an intuitive GUI, flexible design of transcriptome and proteome analyses strategies; and convenient interactive graphical output of results on expression signals, chromosomal mapping, functional annotation and network interactions. The modular structure allows for easy extension and customization. We will continue development and support of AMEN as an integral part of our long-term biomedical research program. The source-code is freely available for members of the bioinformatics community who wish to add their own functionalities.

## Availability and requirements

- **Project name:** AMEN
- **Project home page:** http://sourceforge.net/projects/amen
- **Operating system(s):** Platform independent
- **Programming language:** Tcl/Tk, R, GraphViz
- **Other requirements:** ActiveTcl version 8.4.16.0, R version 2.6.0, GraphViz version 2.14.1 or higher
- **License:** GNU GPL

## Authors' contributions

FC initiated, developed the software and drafted the manuscript. MP contributed to the concept and wrote the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional file 1

*Comparison of AMEN and other solutions. Comparison of features implemented in AMEN and other standalone solutions for high-throughput data analysis and interpretation. Corresponding references are given in the main text. An asterisk indicates that the program includes a given feature while a minus is put when the functionality is lacking.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2105-9-86-S1.doc]

## Acknowledgements

We thank A. Lardenois, J. Moore and A. Gattiker for stimulating discussions, O. Collin for beta testing on Mac OS X and R. Houlgatte for critical reading of the manuscript. This work was supported by the Institut National de la Santé et de la Recherche Médicale (Inserm), the Swiss Institute of Bioinformatics (SIB) and Région Bretagne grant No R07077NN. Funding to pay the Open Access publication charges for this article was provided by Inserm.

## References

1. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19(2):**185-193.
2. Carvalho B, Bengtsson H, Speed TP, Irizarry RA: **Exploration, Normalization, and Genotype Calls of High Density Oligonucleotide SNP Array Data.** *Biostatistics* 2006.
3. Park T, Yi SG, Lee S, Lee SY, Yoo DH, Ahn JI, Lee YS: **Statistical tests for identifying differentially expressed genes in time-course microarray experiments.** *Bioinformatics* 2003, **19(6):**694-703.
4. Pan W: **A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments.** *Bioinformatics* 2002, **18(4):**546-554.
5. Bar-Joseph Z: **Analyzing time series gene expression data.** *Bioinformatics* 2004, **20(16):**2493-2503.
6. Wicker N, Dembele D, Raffelsberger W, Poch O: **Density of points clustering, application to transcriptomic data analysis.** *Nucleic Acids Res* 2002, **30(18):**3992-4000.
7. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR: **Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation.** *Proc Natl Acad Sci U S A* 1999, **96(6):**2907-2912.
8. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci U S A* 1998, **95(25):**14863-14868.
9. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25(1):**25-29.
10. Khatri P, Draghici S: **Ontological analysis of gene expression data: current tools, limitations, and open problems.** *Bioinformatics* 2005, **21(18):**3587-3595.
11. Rivals I, Personnaz L, Taing L, Potier MC: **Enrichment or depletion of a GO category within a class of genes: which test?** *Bioinformatics* 2007, **23(4):**401-407.
12. **The R Project for Statistical Computing** [http://www.r-project.org]
13. **The Bioconductor project** [http://www.bioconductor.org]
14. Reimers M, Carey VJ: **Bioconductor: an open source framework for bioinformatics and computational biology.** *Methods Enzymol* 2006, **411:**119-134.
15. Blom EJ, Bosman DW, van Hijum SA, Breitling R, Tijsma L, Silvis R, Roerdink JB, Kuipers OP: **FIVA: Functional Information Viewer and Analyzer extracting biological knowledge from transcriptome data of prokaryotes.** *Bioinformatics* 2007, **23(9):**1161-1163.
16. Scheer M, Klawonn F, Munch R, Grote A, Hiller K, Choi C, Koch I, Schobert M, Hartig E, Klages U, Jahn D: **JProGO: a novel tool for the functional interpretation of prokaryotic microarray data using Gene Ontology information.** *Nucleic Acids Res* 2006, **34(Web Server issue):**W510-5.
17. Awad IA, Rees CA, Hernandez-Boussard T, Ball CA, Sherlock G: **Caryoscope: an Open Source Java application for viewing microarray data in a genomic context.** *BMC Bioinformatics* 2004, **5:**151.
18. Menten B, Pattyn F, De Preter K, Robbrecht P, Michels E, Buysse K, Mortier G, De Paepe A, van Vooren S, Vermeesch J, Moreau Y, De Moor B, Vermeulen S, Speleman F, Vandesompele J: **arrayCGHbase: an analysis platform for comparative genomic hybridization microarrays.** *BMC Bioinformatics* 2005, **6:**124.
19. Stanley SM, Bailey TL, Mattick JS: **GONOME: measuring correlations between GO terms and genomic positions.** *BMC Bioinformatics* 2006, **7:**94.
20. Toedling J, Schmeier S, Heinig M, Georgi B, Roepcke S: **MACAT--microarray chromosome analysis tool.** *Bioinformatics* 2005, **21(9):**2112-2113.
21. Turkheimer FE, Roncaroli F, Hennuy B, Herens C, Nguyen M, Martin D, Evrard A, Bours V, Boniver J, Deprez M: **Chromosomal patterns of gene expression from microarray data: methodology, validation and clinical relevance in gliomas.** *BMC Bioinformatics* 2006, **7:**526.
22. Vlasblom J, Wu S, Pu S, Superina M, Liu G, Orsi C, Wodak SJ: **GenePro: a Cytoscape plug-in for advanced visualization and analysis of interaction networks.** *Bioinformatics* 2006, **22(17):**2178-2179.
23. Cerami EG, Bader GD, Gross BE, Sander C: **cPath: open source software for collecting, storing, and querying biological pathways.** *BMC Bioinformatics* 2006, **7:**497.
24. Hokamp K, Roche FM, Acab M, Rousseau ME, Kuo B, Goode D, Aeschliman D, Bryan J, Babiuk LA, Hancock RE, Brinkman FS: **ArrayPipe: a flexible processing pipeline for microarray data.** *Nucleic Acids Res* 2004, **32(Web Server issue):**W457-9.
25. Kapushesky M, Kemmeren P, Culhane AC, Durinck S, Ihmels J, Korner C, Kull M, Torrente A, Sarkans U, Vilo J, Brazma A: **Expression Profiler: next generation--an online platform for analysis of microarray data.** *Nucleic Acids Res* 2004, **32(Web Server issue):**W465-70.
26. Psarros M, Heber S, Sick M, Thoppae G, Harshman K, Sick B: **RACE: Remote Analysis Computation for gene Expression data.** *Nucleic Acids Res* 2005, **33(Web Server issue):**W638-43.
27. Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z: **CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis.** *Nucleic Acids Res* 2006, **34(Web Server issue):**W498-503.
28. Romualdi C, Vitulo N, Del Favero M, Lanfranchi G: **MIDAW: a web tool for statistical analysis of microarray data.** *Nucleic Acids Res* 2005, **33(Web Server issue):**W644-9.
29. Shamir R, Maron-Katz A, Tanay A, Linhart C, Steinfeld I, Sharan R, Shiloh Y, Elkon R: **EXPANDER--an integrative program suite for microarray data analysis.** *BMC Bioinformatics* 2005, **6:**232.
30. Vaquerizas JM, Conde L, Yankilevich P, Cabezon A, Minguez P, Diaz-Uriarte R, Al-Shahrour F, Herrero J, Dopazo J: **GEPAS, an experiment-oriented pipeline for the analysis of microarray gene expression data.** *Nucleic Acids Res* 2005, **33(Web Server issue):**W616-20.
31. **Tcl Developper Site** [http://www.tcl.tk/]

32. Chalmel F, Lardenois A, Primig M: **Toward understanding the core meiotic transcriptome in mammals and its implications for somatic cancer.** *Ann N Y Acad Sci* 2007, **1120:**1-15.
33. Chalmel F, Rolland AD, Niederhauser-Wiederkehr C, Chung SS, Demougin P, Gattiker A, Moore J, Patard JJ, Wolgemuth DJ, Jegou B, Primig M: **The conserved transcriptome in human and rodent male gametogenesis.** *Proc Natl Acad Sci U S A* 2007, **104(20):**8346-8351.
34. Wrobel G, Chalmel F, Primig M: **goCluster integrates statistical analysis and functional interpretation of microarray expression data.** *Bioinformatics* 2005, **21(17):**3575-3577.
35. Zakharkin SO, Kim K, Mehta T, Chen L, Barnes S, Scheirer KE, Parrish RS, Allison DB, Page GP: **Sources of variation in Affymetrix microarray experiments.** *BMC Bioinformatics* 2005, **6:**214.
36. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G: **MINT: the Molecular INTeraction database.** *Nucleic Acids Res* 2007, **35(Database issue):**D572-4.
37. Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B, Thorneycroft D, Zhang Y, Apweiler R, Hermjakob H: **IntAct--open source resource for molecular interaction data.** *Nucleic Acids Res* 2007, **35(Database issue):**D561-5.
38. Kersey PJ, Duarte J, Williams A, Karavidopoulou Y, Birney E, Apweiler R: **The International Protein Index: an integrated database for proteomics experiments.** *Proteomics* 2004, **4(7):**1985-1988.
39. Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, Rosenbloom KR, Rhead B, Raney BJ, Pohl A, Pedersen JS, Hsu F, Hinrichs AS, Harte RA, Diekhans M, Clawson H, Bejerano G, Barber GP, Baertsch R, Haussler D, Kent WJ: **The UCSC genome browser database: update 2007.** *Nucleic Acids Res* 2007, **35(Database issue):**D668-73.
40. Labarga A, Valentin F, Anderson M, Lopez R: **Web services at the European bioinformatics institute.** *Nucleic Acids Res* 2007, **35(Web Server issue):**W6-11.
41. **The UCSC Genome Browser Site** [http://genome.ucsc.edu]
42. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets.** *Nucleic Acids Res* 2006, **34(Database issue):**D535-9.
43. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2007, **35(Database issue):**D5-12.
44. Breitling R, Herzyk P: **Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data.** *J Bioinform Comput Biol* 2005, **3(5):**1171-1189.
45. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98(9):**5116-5121.
46. Wettenhall JM, Smyth GK: **limmaGUI: a graphical user interface for linear modeling of microarray data.** *Bioinformatics* 2004, **20(18):**3705-3706.
47. Hommel G: **A stagewise rejective multiple test procedure based on a modified Bonferroni test.** *Biometrika* 1988, **75(1):**383-386.
48. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *JR Stat Soc Ser* 1995, **57(1):**289-300.
49. Kaufman L, Rousseeuw PJ: **Finding Groups in Data: An Introduction to Cluster Analysis.** New York: Wiley; 1990.
50. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22(3):**281-285.
51. van der Laan MJ, Pollard KS: **A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap.** *Journal of Statistical Planning and Inference* 2003, **117:**275-303.
52. Yeung KY, Fraley C, Murua A, Raftery AE, Ruzzo WL: **Model-based clustering and data transformations for gene expression data.** *Bioinformatics* 2001, **17(10):**977-987.
53. Schlecht U, Demougin P, Koch R, Hermida L, Wiederkehr C, Descombes P, Pineau C, Jegou B, Primig M: **Expression profiling of mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility.** *Mol Biol Cell* 2004, **15(3):**1031-1043.
54. **Inkscape: Open Source Scalable Vector Graphics Editor** [http://www.inkscape.org]
55. **SVG.org** [http://www.svg.org]
56. Schultze JL, Eggle D: **IlluminaGUI: graphical user interface for analyzing gene expression data generated on the Illumina platform.** *Bioinformatics* 2007, **23(11):**1431-1433.
57. Wettenhall JM, Simpson KM, Satterley K, Smyth GK: **affylmGUI: a graphical user interface for linear modeling of single channel microarray data.** *Bioinformatics* 2006, **22(7):**897-899.
58. Xu X, Zhao Y, Simon R: **Gene Set Expression Comparison kit for BRB-ArrayTools.** *Bioinformatics* 2007, **24(1):**137-139.
59. Pelizzola M, Pavelka N, Foti M, Ricciardi-Castagnoli P: **AMDA: an R package for the automated microarray data analysis.** *BMC Bioinformatics* 2006, **7:**335.
60. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res* 2006, **34(Database issue):**D354-7.
61. Elvidge G: **Microarray expression technology: from start to finish.** *Pharmacogenomics* 2006, **7(1):**123-134.
62. Bertone P, Gerstein M, Snyder M: **Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery.** *Chromosome Res* 2005, **13(3):**259-274.
63. Mockler TC, Chan S, Sundaresan A, Chen H, Jacobsen SE, Ecker JR: **Applications of DNA tiling arrays for whole-genome analysis.** *Genomics* 2005, **85(1):**1-15.
64. Hermida L, Schaad O, Demougin P, Descombes P, Primig M: **MIMAS: an innovative tool for network-based high density oligonucleotide microarray data management and annotation.** *BMC Bioinformatics* 2006, **7:**190.



Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral