

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

## Finding the needle in the haystack

ArticleInfo		
ArticleID	:	4319
ArticleDOI	:	10.1186/gb-2002-3-2-reports2008
ArticleCitationID	:	reports2008
ArticleSequenceNumber	:	20
ArticleCategory	:	Web report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2001-11-27 Received : 2001-11-27 OnlineDate : 2002-1-21
ArticleCopyright	:	BioMed Central Ltd2002
ArticleGrants	:	

Alessandro Guffanti

## Abstract

The Expression Profiler is an ensemble of web-based computational resources (still under development) for clustering gene-expression data with different algorithms and distance measures, obtaining graphical displays of the results (EPCLUST) and linking the cluster with annotation resources (URLMAP)

## Content

The Expression Profiler is an ensemble of web-based computational resources (still under development) for clustering gene-expression data with different algorithms and distance measures, obtaining graphical displays of the results (EPCLUST) and linking the cluster with annotation resources (URLMAP). The latter is a URL mapper that uses Sequence Accession numbers or other identifiers within clusters to create on-the-fly links to external resources such as KEGG, PFAM, SPEXS (a tool for extracting patterns from selected sequence datasets), a SWISS-PROT browser, a Gene Ontology (GO) browser and MEDLINE. It is also possible to match the patterns against sequences stored on the server and make direct graphical comparison with expression profiles (PATMATCH), and to visualize the information content of the patterns using SEQUENCE LOGO, a beautiful and useful sequence logo generator that starts from aligned or unaligned patterns. Sequence logos are an intuitive graphical way of representing consensus sequences or patterns. Clustering is very fast and the number of available options is noticeably larger than similar PC-based solutions. Another interesting tool is GENOMES, a full genome-sequence and expression-data extractor (limited at present to *Saccharomyces cerevisiae* open reading frames).

## Navigation

Expression Profiler has strictly functional navigation, so a certain sequence of operations must be followed in a given order: load a set of expression values, calculate the distance matrix, perform clustering, link with an external site, and so on. The first phase in the clustering procedure consists of the data upload. The data must be in a standard tabular format, with columns corresponding to different experiments and rows to different genes. Uploaded data can be selected and then stored in a folder on the server for subsequent use or directly carried on to cluster analysis. The clustering procedures can be based on hierarchical (producing a dendrogram with the expression data) or K-means (partitioning the

expression data set into k clusters) clustering methods with a wide choice of distance-measurement methods and parameters. Standard options are well-tuned, however, to ensure a good preliminary calculation. The output of hierarchical clustering is a nice GIF image containing the classic red and green graph, as used by Michael Eisen (Lawrence Orlando Berkeley National Laboratories, Berkeley, USA), side by side with the tree. By clicking on the various branches it is possible to jump to the various sub-trees (clusters), and it is also possible to cut the tree according to a given threshold. When requesting K-means clustering, the output corresponds directly to the clusters.

## Reporter's comments

### Timeliness

The software is in continuous development and the latest available version at the time of writing was 1 December 2001.

### Best feature

EPCLUST (the clustering and analysis part of the Expression Profiler) is very fast, rich in options and produces nice GIF images that can be downloaded and used for presentation. The author gives prompt answers to any question; a mailing list (ep-users) is available for discussions.

### Worst feature

The lack of an extensive, centralized tutorial sometimes makes it hard to follow all the possible paths and to understand all the possibilities of the software. Some options allow one to make a wrong input and hence will not work; this may happen when selecting a subset of genes or linking to PATMATCH from the clustering results. The EP:PPI (Protein-Protein Interaction) feature is not ready yet, and should therefore not be included in the main page.

### Wish list

The site is functional and useful, but the enormous number of options explained in a very summarized form makes for a steep learning curve. A better page design would also help with program usability. I would personally vote for a basic versus advanced user double menu. A 'step-by-step' guided tutorial

would be very useful. In its present state, I would not define the site as suitable for the faint of heart or for the absolute beginner in clustering. With some improvements, however, it will become a very useful and easy-to-use research and didactic tool.

## Table of links

[EBI expression profiler](#)

## References

1. [EBI expression profiler](#).