

GEPAS: towards new challenges in gene expression data analysis.

Herrero, Javier¹, Vaquerizas, J. M., Al-Sharour, F., Conde, L., Mateos, Á., Santoyo, J., Díaz-Uriarte, R., Dopazo, J.

Bioinformatics Unit, Biotechnology Programme, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro 3, E-28028 Madrid (Spain)

¹Present address: EnsEMBL Team, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

We present recent developments in the GEPAS, a web-based pipeline for microarray gene expression profile analysis. GEPAS stands for Gene Expression Profile Analysis Suite (<http://gepas.bioinfo.cnio.es>). It is composed of several interconnected modules which include tools for preprocessing, clustering, classification problems as well as finding genes differentially expressed among distinct classes of experiments, or correlated to diverse clinical outcomes.

GEPAS has expanded its scope by adding a module for normalising raw cDNA microarrays. On the other side of the pipeline, GEPAS is moving towards answering biological questions. A new tool is available for building predictors. In addition to this, the availability of accurate genomic data and the recent implementation of CGH-arrays, has opened up the possibility of mapping expression and genomic data on the chromosomes.

Due to the amount of data generated in a single experiment, data mining is one of the key part in the analysis of the results. Different initiatives, such as Gene Ontology (GO) consortium, pathways databases, protein functional motifs, etc. provide curated annotations for genes. FatiGO and FatiWISE, using such information and being well integrated with the rest of the modules, helps understanding the results of the other analyses from a biological point of view.