

BBSI Proposal Summer 2008

Incorporating cluster analysis into BioBike language

Andrew Whittam
Washington & Jefferson College
Information Technology Leadership
Biology

Introduction

The microarray has become the standard way to obtain a holistic view of gene expression in a cell. Only with the advancements in technology has the ability to collect and interpret the massive amounts of data produced by microarrays become possible. When approaching such large dataset there are many varying techniques to filter results and help draw conclusions. From the isolating of those genes whose expression had varied significantly to linking genes whose expressions seemed to be related, different tools allow for different observations. Clustering is the use of algorithms to organize genes according to similarity in pattern of gene expression. In “Cluster analysis and display of genome-wide expression patterns” by Eisen, Spellman, Brown, and Botstein the use of clustering to interpret microarray results is described. This paper is often cited in many microarray articles and their results. Combining these results with other microarray techniques along with the ability to directly use the results in other tests would be beneficial to scientists. The capability to take the results and began direct computation work on their genomic sequence or upstream factors allows for the exploring and interpreting of linkages suggested by clustering. BioBike is a visual programming language that caters to non programmers and allows them the ability to produce programs. BioBike uses both original programming and already existing

protocols in concert to allow users to tailor their own programs and experiments to meet their needs. Incorporation of cluster analysis into BioBike would allow for more freedom with results along with increasing ability to further investigate similarities in clustered genes.

Method

Integration of cluster analysis into BioBike will allow for the users to better interpret results and further explore them. BioBike is built upon the computer languages of BioLisp and Lisp. We decided to build a cluster analysis program from the ground up in Lisp as well. Tailoring a program specifically to BioBike will allow for faster results along with less software complications. Using microarray data already stored on BioBike the program will cluster genes and return the results to the user in variable form allowing for instant incorporation into framework for new programs.

Results

This summer I will be working on incorporating cluster analysis into BioBike and making sure it works seamlessly within BioBike. Along with this programming I will also be working to expand the microarray database of BioBike by downloading and formatting microarray results. I will also be working on the creation of a utility to allow users to upload their own microarrays to the BioBike environment.

References

Michael B. Eisen, Paul T. Spellman, Patrick O. Brown, and David Botstein

Cluster analysis and display of genome-wide expression patterns

PNAS, Dec 1998; 95: 14863 - 14868.

<http://ramsites.net/~biobike/> - BioBike Website