Discovery Park

Bindley Bioscience Center

Interactive Analysis of 'Omics analyte Expression Data

Mingwu Zhang, Qi Ouyang, Alan Stephenson, Michael D Kane, David E Salt, Sunil Prabhakar, Charles Buck, Xiang Zhang
Purdue University, West Lafayette, IN 47907

Overview

Systems biology [1, 2] aims to understand biological systems on a comprehensive scale, such that the components that make up the whole are connected to one another and work through dependent interactions. Two essential components are featured in systems biology: powerful tools for data acquisition and computational bioinformatics. For the later, a growing list of data analysis and data modeling methods have been developed leveraging the disciplines of computer science, engineering, statistics, and mathematics.

Molecular correlations [3] and time lapse study of molecular expression are crucial to establishing interdependent connections in systems biology. Presented is a novel interactive visual data mining application, SysNet, that provides an interactive environment for the analysis of 'omics high data volume biological systems.

System Implementation

- SysNet has been developed in Microsoft Visual Studio .Net using Visual C++.
- Microsoft Access database was selected to store analyte expression information including expression data, sample meta information and analyte characteristics.
- There are two major functionalities in the current version of SysNet:
  - One function is interactive analysis of analyte correlations in a single "SysData", which is a data structure that contains analyte expression information generated from various 'omics studies including genomics, proteomics, metabolomics, ionomics, etc.
  - The other function is interactive analysis of time lapse data such that analytes measured in multiple SysData are aligned and displayed in a single environment. These functions were developed as two distinct forms, even through they share multiple analysis and visualization routines.
  - For both functionalities, SysNet enables user to interactively change the contents of analyte information. The related information is then automatically updated in the graphic display.
- SysNet visualizes 'omics expression data as a two-dimensional network supporting circular and heatmap layout. In the circular layout, analytes are represented as nodes located on circles. The inter-analyte correlations are represented as links or edges between nodes.

Experiment

Arabidposis thaliana plants were seeded and stratified for 3 days at 4 ºC and allowed to grow for 5 weeks at 19-22 ºC under 90 uEm-2s-1 of fluorescent light. The growth medium was Sunshine Mix LB2 which has been spiked with As, Ca, Li, Ni, Pb and Se. Plants were watered twice per week with quarter-strength type 2 Hoaglands in which the normal iron was replaced with 0.5 to 30 uM Fe-HBED mixed with equimolar amount of iron(III) nitrate and brought to pH 6.0 with 4 M KOH. Elmental analysis was perfomred with an inductively coupled plasma-mass spectrometry (ICP-MS) for Li, B, Na, Mg, P, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Mo, and Cd.

Results

The SysNet program has been utilized to analyze elemental profile changes with increasing concentration of iron in growth medium. Following are some screenshots of utilizing SysNet analyzing the ionomics data.

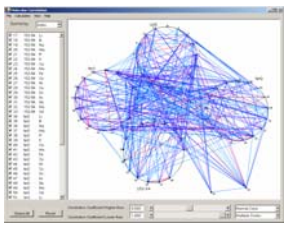

Figure 1. Interactive display of analyte correlation in multiple circles with the normal color schema. The left panel displays analyte information and the right panel displays analyte correlation. Each circle represents analytes measured in a bio-domain. The user can re-arrange the correlation map by simply selecting a circle or node and moving the computer mouse. Col0 is a wild-type plant whereas Ler2, fpt2 and 152-54 are mutants.

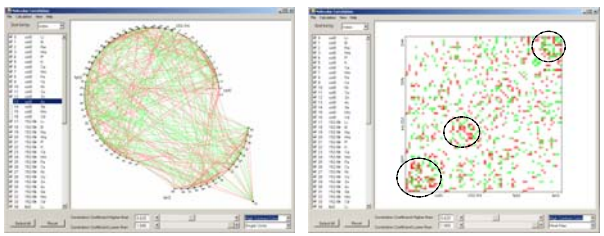

Figure 2. Graphic display of analyte correlation in (a) one circle and (b) heat map with high contrast color schema. High contrast color schema is used to show the correlation directions. Green indicates positive correlation while red indicates negative correlation. All analytes contained in a SysData are displayed in one circle (a), with breaks in the circle representing the divisions between the different bio-domains; col0, ler2, fpt2 and 152-54, in this case. All analytes belonging to the same bio-domain are displayed in the same arc. Each arc and analyte node can be re-arranged for easy visualization. Clicking on each node in (a) shows figure 3 while clicking on the edge in (a) or a node in (b) shows figure 4.

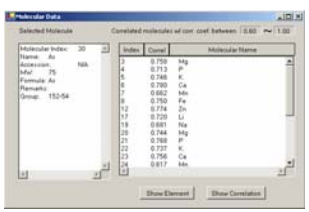

Figure 3. List of analytes that correlates with the interest analyte. The left panel shows the interested analyte. All analytes correlated with the interested analytes are displayed in the right panel.

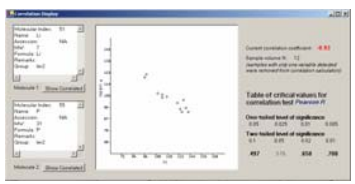

Figure 4. Correlation between two elements (Li in ler2 and P in ler2). The information for two correlating analytes is displayed in the two list boxes on the left. The middle graph shows the expression data of the two analytes. Each data point represents peak intensity of the two analytes in the same sample (elements in this ionomics example). The right panel displays statistical information for the correlation to assess significance.

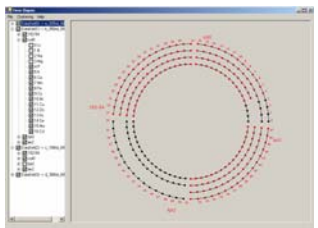

Figure 5. Visualization of time lapse data. All analytes are aligned in concentric circles based on analyte name. The left panel lists experimental information. The user can focus on a single analyte to investigate its behavior in multiple SysData. By clicking a node on the graph of the time lapse window, a four panel 'evolution window' will appear that displays the response of that analyte in each SysData (see figure 6).

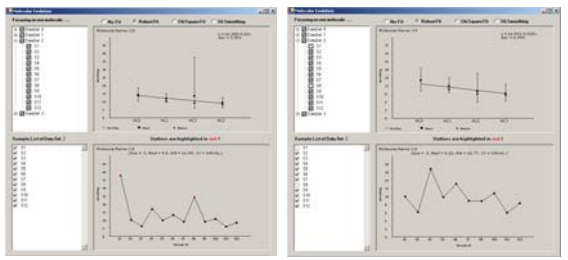

Figure 6. Display of analyte behavior in a time course. SysData 2 is the active dataset in the display window. The behavior of the analyte of interest in each sample of the active SysData is displayed in the lower right panel. a) the analyte response in sample 1 and 8 were detected as outliers and highlighted in red; b) the outliers in the SysData 2 have been removed by unchecking S1 and S8 in the left panel [4].

Conclusions

- SysNet was developed using an interactive visual data mining approach to enable experimental scientists performing data mining.
- The SysNet program has been utilized to analyze elemental profile changes with increasing concentration of iron in growth medium.
- Current version of SysNet supports analyte correlation analysis and time lapse study.

References

- Hood L: **Systems biology: integrating technology, biology, and computation**. Mechanisms of Ageing and Development 2003, 124:9-16.
- Weckwerth W: **Metabolomics in system biology**. Annu Rev Plant Biol 2003, 54:669-689.
- Clish CB, Davidov E, Oresic M, Plasterer T, Lavine G, Londo TR, Meys M, Snell P, Stochaj W, Adourian A, Zhang X, Morel N, Neumann E, Verheij E, Vogels JTWE, Havekes LM, Afeyan N, Regnier FE, Greef J, Naylor S: **Integrative biological analysis of the APOE*3-Leiden transgenic mouse**. Omics: A Journal of Integrative Biology 2004, 8:3--13.
- Sheskin DJ: **Handbook of Parametric and Nonparametric Statistical Procedures**. 2nd edition Chapman & Hall/CRC, Washington, DC, 2000.