Oral presentation

Visualization of large microarray experiments with space maps

Nils Gehlenborg*[1,2] and Alvis Brazma[1]

Address: [1]European Bioinformatics Institute, Cambridge, CB10 1SD, UK and [2]Graduate School of Life Sciences, University of Cambridge, Cambridge, CB2 1RX, UK

E-mail: Nils Gehlenborg* - nils@ebi.ac.uk
*Corresponding author

Background

Heatmaps and profile plots are effective techniques to visualize expression profiles of several hundred genes across a few dozen samples. However, these techniques do not scale to data sets with expression profiles that have been measured across several hundred samples or even thousands of samples. Our motivation to find a solution to this scaling problem is based on the observation that with increasingly mature and affordable microarray platforms, the number of studies in ArrayExpress [1] including hundreds of samples has been increasing steadily over the years.

Methods

We have developed the glyph-based Space Maps visualization technique that is conceptually similar to Value and Relation Displays [2]. The technique comprises two steps: (1) Generation of glyphs to represent gene expression profiles and (2) arrangement of the glyphs to reflect relationships between genes. Both steps support the integration of biological knowledge into the visualization, for instance in form of ontologies that describe hierarchical relationships among the conditions in the data. We also use hierarchical organization of samples and aggregation of expression levels to summarize expression values of groups of samples, which enables the user to reduce the amount of data shown on each glyph. Similar to treemaps [3], this

construction makes it possible to start out with an overview of the data and then view details on demand.

Results

We have applied the Space Maps visualization to a data set with 5,372 samples (Margus Lukk, personal communication). This data set has been constructed from a large collection of publicly available gene expression data sets and a problem-specific hierarchy on the samples is available. We selected the 1,000 most variable genes from this data set and visualized this subset with our technique (Figure 1). The arrangement of the glyphs represents an overview of the global patterns in the data, such as clusters and outliers. Furthermore, the visualization provides insight into local patterns in the gene expression profiles. Since global patterns arise directly from local patterns we were able to explain several of the clusters and outliers and assign meaningful labels to them.

Conclusion

The Space Maps visualization technique is a novel approach to visualization of gene expression data that facilitates the visualization of expression profiles of genes with hundreds or thousands of samples without loss of context information. A major strength of this technique is that it allows a tightly coupled exploration of local and
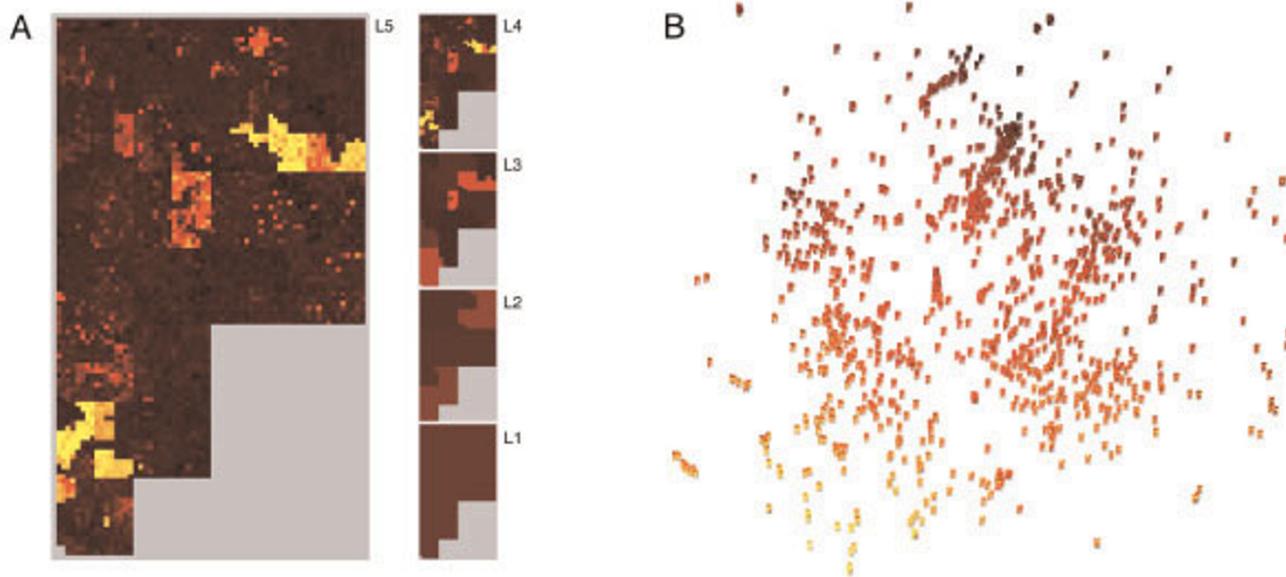
Figure 1
Space Maps visualization of 1,000 genes with 5,372 samples. (A) An expression profile at five levels of the hierarchy. Level L1 corresponds to the root and Level L5 corresponds to the leafs of the hierarchy. The information-content of the glyph increases as the levels increase. (B) A non-linear projection [4] of 1,000 expression profiles into 2D space. It is possible to make out global patterns such as clusters and outliers. Local patterns in the expression profiles can be identified as well, for instance in the lower left corner.

global patterns, which makes hypothesis generation more efficient than with traditional techniques.

References
1. Parkinson H, *et al*: **ArrayExpress update – from an archive of functional genomics experiments to the atlas of gene expression.** *Nucleic Acids Res* 2007, **37(Database issue):** D868–D872.
2. Yang J, *et al*: **Value and Relation Display – Interactive visual exploration of large data sets with hundreds of dimensions.** *IEEE Transactions on Visualization and Computer Graphics* 2007, **13:**494–507.
3. Johnson B and Shneiderman B: **Tree maps: A space-filling approach to the visualization of hierarchical information structures.** *Proceedings of the 2nd International IEEE Visualization Conference* 1991, 284–291.
4. Venna J and Kaski S: **Non-linear dimensionality reduction as information retrieval.** *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007)* 2007, 568–575.