Systems biology

Caleydo: connecting pathways and gene expression

Marc Streit[1,*], Alexander Lex[1], Michael Kalkusch[1], Kurt Zatloukal[2]
and Dieter Schmalstieg[1]

[1]Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16
and [2]Institute of Pathology, Medical University Graz, Auenbruggerplatz 25, 8010 Graz, Austria

ABSTRACT

Summary: Understanding the relationships between pathways and the altered expression of their components in disease conditions can be addressed in a visual data analysis process. Caleydo uses novel visualization techniques to support life science experts in their analysis of gene expression data in the context of pathways and functions of individual genes. Pathways and gene expression visualizations are placed in a 3D scene where selected entities (i.e. genes) are visually connected. This allows Caleydo to seamlessly integrate interactive gene expression visualization with cross-database pathway exploration.

Availability: The Caleydo visualization framework is freely available on www.caleydo.org for non-commercial use. It runs on Windows and Linux and requires a 3D capable graphics card.

Contact: caleydo@icg.tugraz.at; streit@icg.tugraz.at

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Pathways have proven to be fundamental in understanding the role of genes in biological processes. Pathway databases such as KEGG (Kanehisa *et al.*, 2008) and BioCarta are well known and heavily used. Pathways model relations in a specific functional context. One approach to show interconnections is to visualize the whole network. However, working with such large graphs containing thousands of nodes is tedious. Furthermore, automatic layouting, as used in such representations, often does not produce desirable results. Finally, an automatic layout does not make use of the innate human ability to recognize pictures and patterns. Therefore, a system using existing, hand-crafted pathway layouts is preferable (Streit *et al.*, 2008). Caleydo focuses on the visualization of interconnections between pathway graphs and the roles of genes in different pathways.

The second central goal of Caleydo is to facilitate combined analysis of pathways and regulation data. To achieve this, in-place augmentation on the node of KEGG pathways has been proposed (e.g. Lindroos and Andersson, 2002). However, the number of distinguishable experiments per node is limited by the node's tiny size. Multi-mappings, where one node corresponds to more than one set of expression experiments, cannot be visualized at all. Furthermore, pathways which do not use rectangular shapes

to represent nodes, like BioCarta, are not compatible with this technique at all.

Our contribution enables seamless interaction in a visual environment across pathway databases (currently, KEGG and BioCarta), and the simultaneous display of linked gene expression profiles. Additionally, our system provides visual filtering methods which allow the user to discover properties in data intuitively.

2 APPLICATION OVERVIEW

Caleydo uses two types of views: pathways (currently, *Homo sapiens* models only, *Mus musculus* in the near future) and gene expression visualizations. By treating each pathway as a separate view, the problem of relating multiple pathways can be addressed by the multiple view technique known in the information visualization community. When showing relations between pathways, the straightforward method would be to place them side by side and highlight identical nodes. However, the cognitive load of identifying related items in different views is high. As a remedy, we use visual links (Collins and Carpendale, 2007), which are lines between selected entities, to connect items directly. This visual linking is used in Caleydo to identify relations between different pathways and to relate gene expression data to pathways. Visual linking is superior to traditional gene expression augmentation of nodes because it is not limited to the small area on top of the nodes. It also works equally well for any shape of node and allows true multi-mapping.

In addition to linking, the user can opt to use augmentation of the nodes. Previous work tries to simultaneously map many values on a tiny node. In Caleydo, the node is mapped with one single color, which corresponds to one single experiment. The experiments can be selected interactively, thus allowing a user to view differences between experiments for a whole pathway by selecting each experiment in turn.

The limitations of the available screen space and the nature of the visual links suggest an arrangement of views in 3D space. Therefore, we have placed the views in a square bucket-like setup, with views at the bottom, on the walls and on the rim of the bucket (Fig. 1). The view at the bucket's bottom is undistorted and large text is readable. To explore this view in detail, the user zooms into the bucket. At this level, all the information in this view is clearly recognizable. The walls of the bucket present contextual information. Pathways related to the center view or gene expression visualizations can be placed there.

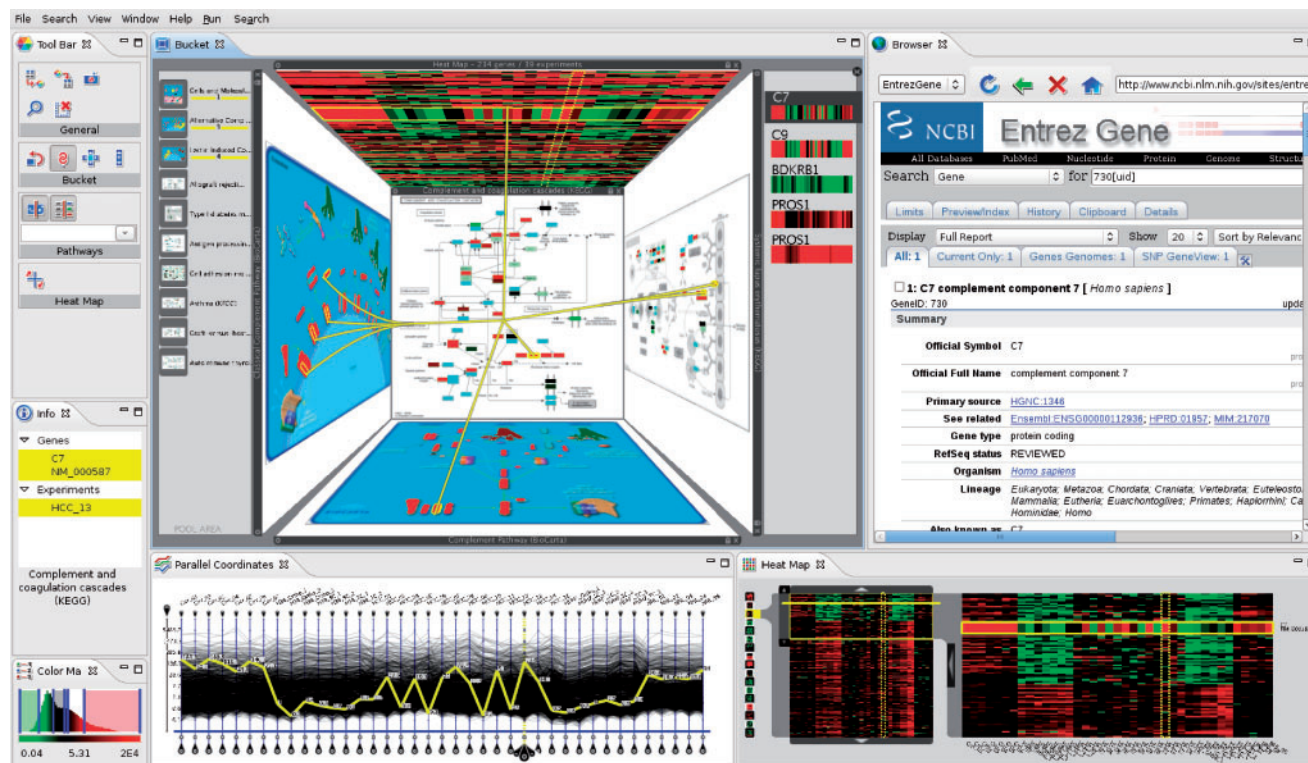*To whom correspondence should be addressed.

Fig. 1. Caleydo with the bucket setup, a hierarchical heat map, parallel coordinates and a linked browser. The visual links connect the selected gene *C7* in different pathways with their expression values of 40 microarray experiments. Further related pathways are placed on the left rim of the bucket. The separate parallel coordinate browser and hierarchical heat map also show the gene *C7*.

Multiple view applications usually do not deal with a large number of views, but there are more than 500 pathways in KEGG and BioCarta combined. Caleydo uses two methods to cope with this large database. First, pathways can be accessed by a search feature. More importantly, Caleydo handles loading of related pathways dynamically. By right-clicking on a gene, all pathways that also contain this gene are displayed. The user can pick interesting ones and load them into the bucket.

Caleydo also features sophisticated gene expression visualization tools such as a parallel coordinates browser, which can be used to filter a dataset of interesting genes, and a hierarchical heat map (Fig. 1). Furthermore, several clustering algorithms can help identify interesting patterns. Results of a filtering process can then be explored in the context of pathways in detail. Whenever an entity is selected, a summary is displayed in the info area, while an integrated web browser loads detailed information from KEGG, BioCarta, EntrezGene or PubMed. The actual gene expression values are shown in a linked tabular data viewer.

3 CONCLUSIONS

Caleydo facilitates a deeper understanding of the biological effects of altered gene expression in disease conditions across pathway boundaries. The integrated gene expression analysis tools enable scientists to derive hypotheses about potential biological consequences of altered gene expression in an interactive manner. The system allows them to give particular consideration to the fact that several gene products may exert multiple functions in different pathways. The concept presented may also be applied to different molecular datasets such as gene copy number variations, protein expression or metabolite concentrations, thus making it a widely applicable data analysis tool.

Conflict of Interest: none declared.

REFERENCES

Collins,C. and Carpendale,S. (2007) VisLink: revealing relationships amongst visualizations. *IEEE TVCG*, **13**, 1192–1199.

Kanehisa,M. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, 480–484.

Lindroos,H. and Andersson,S.G.E. (2002) Visualizing metabolic pathways: comparative genomics and expression analysis. *Proc. IEEE*, **90**, 1793–1802.

Streit,M. *et al.* (2008) Navigation and exploration of interconnected pathways. *Comput. Graph. Forum*, **27**, 951–958.