----------------------- REVIEW 1 ---------------------

PAPER: 203

TITLE: Pathway-based visualization of cross-platform microarray datasets

AUTHORS: Clemens Wrzodek, Johannes Eichner and Andreas Zell

In this manuscript, Wrzodek et al describe a pathway-based visualization of microarray datasets, able to simultaneously display mRNA and miRNA expression data, protein modifications, and DNA methylation - all in a pathway context. I believe the software described here can be a valuable tool for people working with these types of data. Nevertheless, I do have some questions and remarks regarding the text as submitted.

[1] It would be nice if the authors would include Pathline (Meyer et al, 2010) into their overview of tools available. Although InCroMAP does differ from Pathline, there is significant overlap in functionality which would certainly warrant discussion.

[2] When underlaying the KEGG picture with the raw KEGG pathway (which is a graph), how do the authors preserve consistency of overlap? Does this mean that the GraphML representation also stores the actual x- and y-coordinates on the screen? Although the approach followed does help in showing e.g. cellular structure, it does put constraints on what other additional information can be displayed (see [3]). It might have been more useful to loosen the constraint on position and find alternative ways to represent the additional information available in the original pathway map.

[3] Related to [2]. The KEGG pathway pictures are optimized for display and fixed. This means that adding additional information (e.g. miRNA and protein data) breaks this optimal placement of nodes. The result can therefore become very crowded (e.g. Fig1d, and placement of miR-30a in Fig1f).

[4] For displaying DNA methylation, the tool adds a box in front of the box representing a gene. A bar in that box from the middle to the left represents hypomethylation, while one to the right represents hypermethylation. I wonder if another visual encoding would have been more effective here; it is difficult to distinguish (and definitely impossible to identify pre-attentively) promoters that are slightly up-regulated from those that are slightly down-regulated. Also: what does it mean if the complete right side of a box is black? In other words: what is 100% here?

[5] "Visualization of protein expression data". This is confusing. It's either transcript expression, protein translation, or protein modification. From the rest of the text, it seems to be about protein modification.

[6] Is any interactivity available or planned for this tool?

[7] Are the authors planning on utilizing other sources of pathways as well, like e.g. BioCarta?

----------------------- REVIEW 2 ---------------------

PAPER: 203

TITLE: Pathway-based visualization of cross-platform microarray datasets

AUTHORS: Clemens Wrzodek, Johannes Eichner and Andreas Zell

In this paper Wrzodek et al make use of KEGG pathways to display microarray data

from mRNA, micro RNA (miRNA), and methylation array data. What makes this paper

attractive is not the ability to map mRNA expression data onto KEGG pathways

(this feature has been available from KEGG for many years), but rather the fact

that miRNAs, are not part of KEGG pathways (their targets are). The display of

methylation data is also challenging as this needs to be represented in the form

of hypo-to-hyper methylation of the promoter region of the gene. Wrzodek et al.

approach the display of all these types of data in a very attractive way so that

as they say at the end of their paper using their software InCroMAP it is

possible to see that: "the hypomethylation of a promoter may cause the upregulation of a

miRNA which may in turn downregulate the corresponding target mRNA, whereupon a

connected pathway protein could change its expression or activation state.

Such complex cascades of effects, which involve multiple levels of gene

regulation can be deduced from the pathway images generated by the here presented

visualization method." The approach used by the authors is attractive. However,

the method appears limited to the use of KEGG pathways using KEGGtranslator to

convert the data to a GraphML format. It would be good to see the ability to

convert other pathway data resource, eg. Reactome and Biocarta to GraphML as

well. Overall this is very appealing software and idea, but in this age of

massively parrallel sequencing, the use of mRNA, miRNA and methylation data has

diminished for many users.

----------------------- REVIEW 3 ---------------------

PAPER: 203

TITLE: Pathway-based visualization of cross-platform microarray datasets

AUTHORS: Clemens Wrzodek, Johannes Eichner and Andreas Zell

The manuscript presents a set of methods for enhancing pathway visualizations to include hetrogeneous data, including gene expression, gene methylation, protein expression, and miRNA abundance. These methods are implemented in a tool (InCroMAP), which I verified does work with the datasets provided. Overall, I am not convinced that the individual proposed visualization methods are particularly novel - e.g., the method for visualizing miRNA and mRNA are quite standard. The methods for visualizing protein expression and gene methylation seem somewhat more novel, but I also find them somewhat problematic (e.g., the left/right scheme for methylation clashes with neighbours).

I feel the result of adding these different data types using the proposed methods leads to an overly complex visualization that requires significant cognitive load before it can be interpreted by the end-user to gain insight from these data. I believe that this visual complexity could be significantly improved by careful application of some basic principles from the literature and practice of the data visualization community (e.g., Colin Ware, Tamara Munzner, Edward Tufte, etc.). For example, the Cerebral pathway visualization from Tamara Munzner combines a similarly impressive amount of heterogeneous data, but remains relatively uncluttered and easy to interprete.

Although I think the visualization methods could and should eventually be improved, to my knowledge InCroMAP is one of only very few systems currently available that can manage this particular set of heterogeneous, high-throughput data.

However, prior to acceptance the authors should address the following point: Since one of the key aspects of the proposed visualization is to distinguish information specific for genes (e.g., methylation) from information about proteins, it should use the accepted nomenclature for case-variant and particularly for italics/normal font when referring to gene and protein names - e.g., the ”Egfr” gene can be referred to simply as EGFR (in italics), and the protein names or identifiers written without italics. See <http://en.wikipedia.org/wiki/Gene_nomenclature>

Minor issues:

(1) Frequent use of incorrect quotation marks, e.g., running in the wrong direction (”Cell Cycle” instead of "Cell Cycle"). In addition, quotation marks are often used when not needed (e.g., "Smc3"). Finally, quotation of single words or very short phrases should normally be done using single rather than double quotation marks, which are typically used for quoting text from other authors.

(2) Syntax & typos in Fig 2 legend: Change 'darker' to 'more saturated' in "White indicates a fold-change of zero and darker colors correspond to stronger differential expression.". Change "In out example" to "In our example".

(3) Fig. 2 legend should explicitly mention the significance of the yellow highlighting used to indicate Smc3, Trp53, and Ttk.