BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 17 2012, pages 2281–2282
doi:10.1093/bioinformatics/bts428

Systems biology

Advance Access publication July 4, 2012

Voronto: mapper for expression data to ontologies

Rodrigo Santamaría[1,*] and Philippe Pierre[2,3,4]

[1]Departmento de Informática y Automática, Universidad de Salamanca, 37008 Salamanca, Spain, [2]Centre d'Immunologie de Marseille-Luminy, UM2, Aix-Marseille Universite, 13288 Marseille, France, [3]Institut National de la Santé et de la Recherche Médicale (INSERM), U1104, 13288 Marseille, France and [4]Centre National de la Recherche Scientifique (CNRS), UMR7280, 13288 Marseille, France

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Several techniques analyze gene expression data from the point of view of biological ontologies. These methods focus on statistical and numerical analyses, but there is an additional need for fast, global and interactive visualizations of gene expression data in the context of biological ontologies.

Results: Voronto addresses these needs with an easy to use integration of custom gene expression data with custom or public ontologies. In order to do that, it implements Voronoi diagrams where ontologies are mapped with gene expression, providing a quick hierarchy browsing and integration with ontology's resources.

Availability: The tool is available at http://vis.usal.es/voronto. Source code available at https:/github.com/rodrigoSantamaria/Voronto.

Contact: rodri@usal.es

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

Received on May 8, 2012; revised on May 8, 2012; accepted on June 29, 2012

1 INTRODUCTION

Function and relation-focused techniques are gaining interest as part of expression data analyses. The growing of existing biological knowledge and its consolidation into structured ontologies (Ashburner *et al*., 2000; Kanehisa, 2000; Matthews *et al*., 2009) make it possible to integrate such information into the analysis process. The results of these techniques are usually lists of enriched terms, which focus on statistical significance and other numerical information, disregarding non-enriched terms, hierarchical relationships and expression levels. It is therefore hard, if not impossible, to get a picture of expression data in the context of the whole ontology at hand, and to easily inspect it. Here, we present Voronto, a tool that integrates expression data and biological ontologies, allowing the analyst to explore the whole ontology and detect changes on expression patterns inside the ontology.

2 APPROACH

Voronoi diagrams determine, for a set of points, the areas which are closer to each point, dividing the plane into adjacent polygons. If we consider each ontology term as a point, and recursively apply the techique to hierarchy levels, we obtain a treemap-like Voronoi visualization of the hierarchy. This technique was originally described by Balzer *et al*. (2005) and has been adapted to taxonomies (Horn *et al*., 2009) and to ontologies by Bernhardt *et al*. (2009). However, Bernhard *et al*.'s adaptation lacks the degree of interaction and immediacy to make it usable for exploratory analysis. Also, it is not directly connected to related ontology resources, which can be very helpful for detailed investigation. Voronto takes these former approaches as the basis to build a tool for exploration of both expression data and hierarchical ontologies, and its integration with ontology sources. We also take care on information visualization and visual analysis concepts to make it a highly interactive, easy to use, pattern discovery tool.

3 METHODS

Voronto uses a recursive algorithm similar to the one described by Balzer *et al*. Each step builds the tessellation for a hierarchy level. During each step, several iterations are performed in order to adjust location and sizes of the Voronoi cells to the ontology term sizes (i.e. number of annotated genes) by means of an additively weighted Voronoi tessellation [adapted from Takashi Ohyama's implementation (http://www.nirarebakun.com/voro/epwvoro.html)]. Each iteration also optimize cell shapes through a centroidal Voronoi tessellation approach (Lloyd's method).

Voronto includes parsed, up to date versions of KEGG, GO and Reactome ontologies, and mappings to annotated genes for most organisms. It also accepts custom ontolgies in a simple xml format. Gene expression data must be provided by the user, in a tab matrix format, including gene identifiers that match with the ontology identifiers for annotated genes. Voronto then maps gene expression to the ontology terms (the expression of a term corresponds to the average expression of every annotated gene for the term that is present in the user's expression data) and computes the tessellation.

The resulting visualization (Fig. 1) represents an ontology level, so each term is represented by a polygon (cell) colored with its expression. Terms with a common ancestor are represented by adjacent cells, and surrounded by a wider line. The user can go up or down in the hierarchy by using the up/down arrow keys. He can also browse the experimental conditions by using the left/right keys. To avoid losing context, a profile with the hovered term's expression level for each experimental condition is also shown.

Term cells can be clicked for two effects. Double-click opens the available web resources (AmiGO entries, Reactome pathway graphs or KEGG colored pathway images). Alt-click shows a heatmap with the expression profile of every gene in the term. Genes in the heatmap are double-clickable to open their GenBank entry. Several other options, such as term labeling, quantile expression scaling, gene or term searching, ontology areas expansion, etc. are also available.

*To whom correspondence should be addressed.

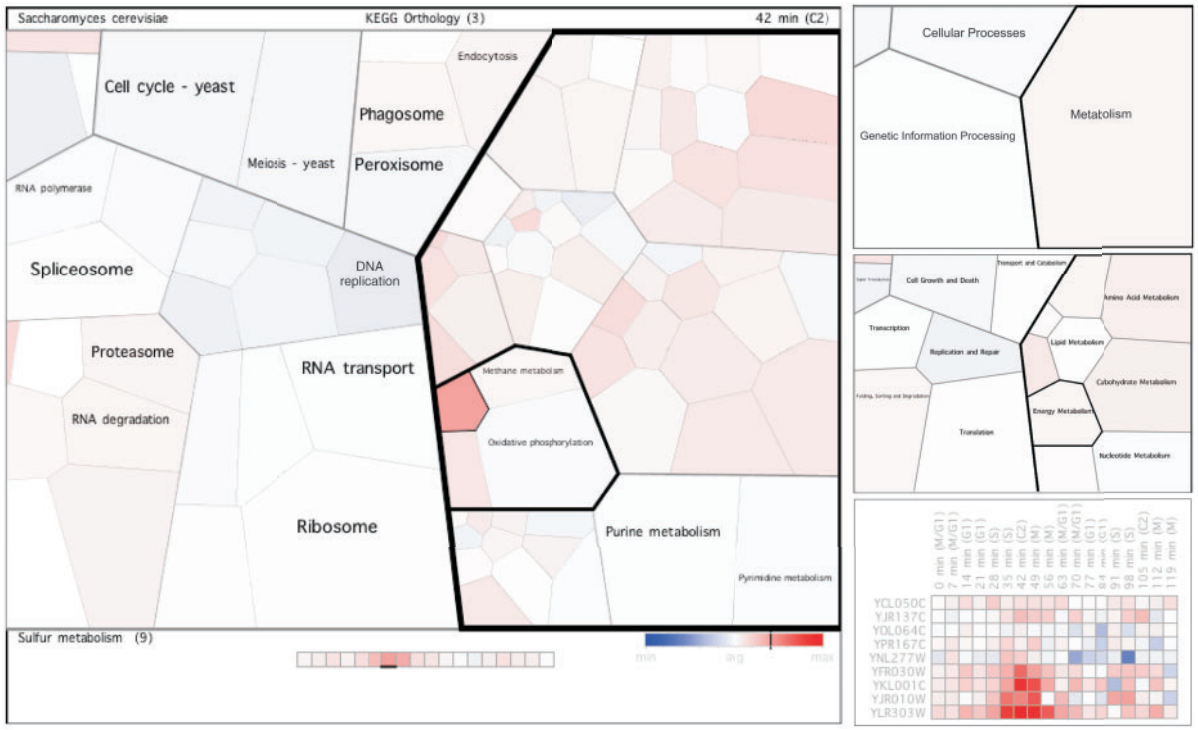

Fig. 1. KEGG orthology mapped with yeast expression along the cell cycle (data available at Voronto's web page). (Left) Visualization of the third hierarchy level for expression at 42′ (phase C2). The term 'Sulfur metabolism', with nine mapped genes, is highlighted, and its profile bar for the whole time series shown at the bottom. (Right, top and center) First and second levels of the hierarchy, 'Sulfur metabolism' belongs to 'Metabolism' and 'Energy metabolism'. (Right, bottom) Heatmap with 'Sulfur metabolism' annotatd genes, showing the overexpression at 42′ is mostly due to three of them.

4 DISCUSSION

The tool has been tested with several expression data, from different organisms and experimental contexts (see Supplementary Information for some use cases). Biologist's testing of the tool suggests that it is useful on a top-down approach: start by the highest levels of the ontology and narrow down the search to differentially expressed lower terms, helped by the profile bar and the gene heatmap. It has also been used with success in bottom-up approaches, starting with genes or terms known to be relevant for the experiment and searching them in the ontology to determine biological relations or expression patterns.

As an example, in Figure 1 we can easily identify 'Sulfur metabolism' as highly expressed, and at the same time, we can check other terms (e.g. 'RNA transport', with no significant expression, or ''Peroxisome'', slightly underexpressed) or other time points for 'Sulfur metabolism' (42′ is a peak around C2 phase, with no significant expression earlier or later on). From there we can go up on the hierarchy to check for more generic behaviors or narrow down to identify the genes involved in 'Sulfur metabolism'.

5 CONCLUSION

Voronto unburdens the analyst from mapping ontologies to gene expression and offers a clean, general view of the hierarchy. It favors exploration in the context of the ontology's resources and the experimental conditions, complementing functional enrichment analysis, common in systems biology studies. These characteristics make Voronto a powerful, yet simple and fast, way to study expression data in the context of ontologies.

ACKNOWLEDGEMENTS

The authors thank Jonathan Ewbank, Luca Beltrame and Manuel Santos for tool testing and biological feedback.

Funding: This work is funded by the European Commission (EU Framework Programme 7 Collaborative Project SYBARIS, Grant Agreement Number 242220).

Conflict of Interest: none declared.

REFERENCES

Ashburner,M. *et al.* (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.

Balzer,M. *et al.* (2005) Voronoi treemaps for the visualization of software metrics. In *Proceedings of the ACM Symposium on soft visualization*, ACM, New York, NY, USA, pp. 165–172.

Bernhardt,J. *et al.* (2009) Visualizing Gene Expression Data via Voronoi Treemaps. In *International Symposium on Voronoi Diagrams*, IEEE Computer Society, Copenhagen, Denmark, pp. 233–241.

Horn,M.S. *et al.* (2009) Visualizing Biodiversity with Voronoi Treemaps. In *International Symposium on Voronoi Diagrams*, IEEE Computer Society, Copenhagen, Denmark, pp. 265–270.

Kanehisa,M. (2000) Kyoto encyclopedia of genes and genomes. *Nucleic Acid Res.*, **27**, 29–34.

Matthews,L. *et al.* (2009) Reactome knowledgebase of human biological pathways and processes. *Nucleic Acid Res.*, **37** (**Supp. 1**), D619–D622.