BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 5 2013, pages 666–668
doi:10.1093/bioinformatics/bts708

Databases and ontologies

Advance Access publication January 6, 2013

RamiGO: an R/Bioconductor package providing an AmiGO Visualize interface

Markus S. Schröder[1,2,*], Daniel Gusenleitner[2,3], John Quackenbush[2,4,5], Aedín C. Culhane[2,5,†] and Benjamin Haibe-Kains[6,†]

[1]School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland, [2]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA, [3]Department of Bioinformatics, Boston University, Boston, MA 02215, USA, [4]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA, [5]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02215, USA and [6]Bioinformatics and Computational Genomics Laboratory, Institut de recherches cliniques de Montréal, Montreal, Quebec H2W 1R7, Canada

Associate Editor: Martin Bishop

ABSTRACT

Summary: The R/Bioconductor package RamiGO is an R interface to AmiGO that enables visualization of Gene Ontology (GO) trees. Given a list of GO terms, RamiGO uses the AmiGO visualize API to import Graphviz-DOT format files into R, and export these either as images (SVG, PNG) or into Cytoscape for extended network analyses. RamiGO provides easy customization of annotation, highlighting of specific GO terms, colouring of terms by *P*-value or export of a simplified summary GO tree. We illustrate RamiGO functionalities in a genome-wide gene set analysis of prognostic genes in breast cancer.

Availability and implementation: RamiGO is provided in R/Bioconductor, is open source under the Artistic-2.0 License and is available with a user manual containing installation, operating instructions and tutorials. It requires R version 2.15.0 or higher. URL: http://bioconductor.org/packages/release/bioc/html/RamiGO.html

Contact: markus.schroeder@ucdconnect.ie

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

Received on August 24, 2012; revised on December 10, 2012; accepted on December 12, 2012

1 INTRODUCTION

The Gene Ontology (GO) is a controlled vocabulary of gene annotation that was developed to provide consistent functional classification for genes across species; GO is also widely used in gene set enrichment analyses (Gene Ontology Consortium, 2012). It is organized as a directed acyclic graph with top-level ontologies molecular function, biological process and cellular component. Although several web-based or standalone tools are available for visualization of lists of GO terms as GO tree structures, these are not easily accessible through R/Bioconductor, where one has to either rebuild the GO tree using R packages such as GO.db or GOstats, or copy and paste the GO terms of interest into an external web service such as AmiGO Visualize (Carbon *et al.*, 2009, http://amigo.geneontology.org) to display the GO tree.

The free open source web application AmiGO provides users with access to ontology and gene annotation data. Visualization of the queried ontology data is provided by AmiGO Visualize, which uses Graphviz libraries to create GO trees. Graphviz is a collection of software for viewing and manipulating abstract graphs (Gansner and North, 2000). We developed RamiGO to provide R functions that connect directly with the AmiGO Visualize API and retrieve GO trees in various formats. The most common format being PNG or SVG image file, but a file representation of the GO tree in the Graphviz-DOT format is also possible. RamiGO provides a parser for the Graphviz-DOT format that returns an R S4 object, called *AmigoDot*, that includes (i) the tree as a graph object, (ii) the adjacency matrix of the tree, (iii) the annotation of the nodes, (iv) the relation between the nodes and (v) a list of the leaves of the tree. The GO tree is displayed with the set of input GO IDs and all parents of those GO IDs to the root of each GO category where relations between the nodes are represented using the same colour palette as AmiGO Visualize; green, red, black, blue and light blue represent 'positively regulates', 'negatively regulates', 'regulates', 'is a' and 'part of', respectively. In addition using RamiGO, one can display and interactively modify GO tree colours, annotations and relationships between nodes in Cytoscape directly from R.

2 USING THE RAMIGO PACKAGE

The main functions within the RamiGO package are as follows:

```
getAmigoTree(goIDs,color,filename,picType)
readAmigoDot(object,filename)
AmigoDot.to.Cyto(object)
```

which retrieve a GO tree from AmiGO Visualize in the preferred format type, read Graphviz-DOT format files and communicate with Cytoscape, respectively. The AmigoDot.to.Cyto() function displays the tree from an AmigoDot object in Cytoscape using the RCytoscape package. RCytoscape requires the Cytoscape plugin CytoscapeRPC, and the installation of this is described

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint Last Authors.

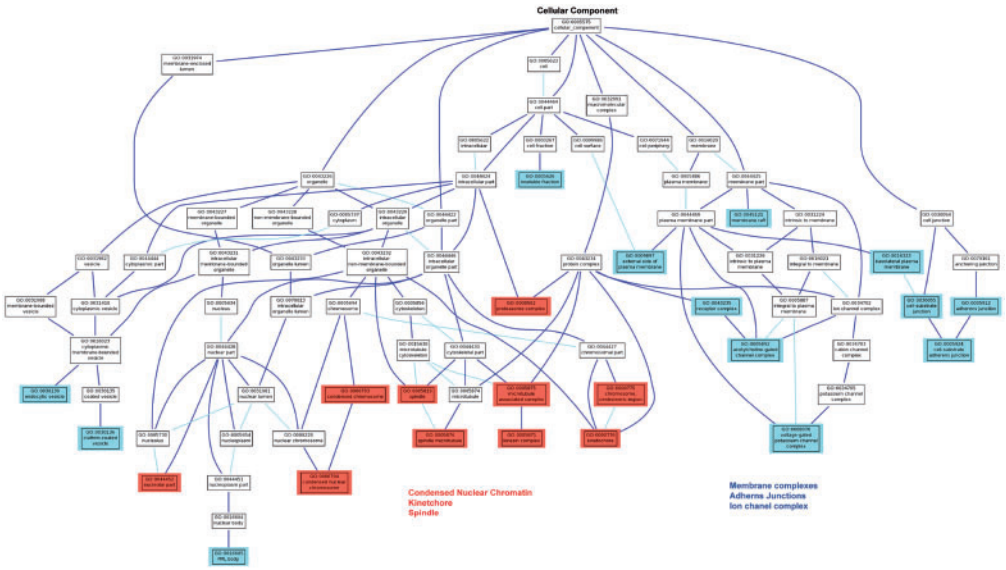

Fig. 1. Subset of GO terms (cellular component) reported from the pre-ranked GSEA analysis of the prognostic value of genes in the breastCancerVDX data with an FDR <0.15. Red nodes represent GO terms with a positive NES, which show high expression of nuclear genes in high risk patients, whereas the blue nodes represent GO terms with a negative NES, which show loss of expression of membrane genes in low risk patients

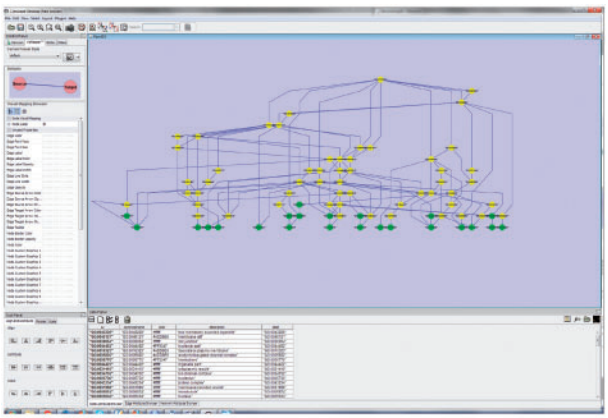

Fig. 2. Cytoscape representation of the GO tree hierarchy from Figure 1

in more detail in the RCytoscape manual. Additional functions to convert the AmigoDot S4 objects to graphAM and graphNEL formats are also provided (AmigoDot.to.graphAM, AmigoDot.to.graphNEL).

3 AN EXAMPLE

To showcase an analysis pipeline (Supplementary Material) that integrates the RamiGO package, we used the Bioconductor data package breastCancerVDX, which includes Affymetrix Human U133A microarray profiles from 344 breast cancer patients (Wang *et al.*, 2005; Minn *et al.*, 2007). We ranked the genes based on their prognostic value by computing the concordance index to estimate the association of each gene with distant metastasis-free survival using the package survcomp (Schröder *et al.*, 2011). This ranked list was then used to run a pre-ranked Gene Set Enrichment Analysis (GSEA) (Subramanian *et al.*, 2005), using the C5 (GO) subset of the Molecular Signature Database (v3.0) gene sets (http://broadinstitute.org/gsea/msigdb/collections.jsp). Figure 1 shows a subset of the GO cellular component gene sets that were reported with a false discovery rate <0.15. Using the customized colouring option in RamiGO, the nodes for gene sets with a positive Normalized Enrichment Score (NES) are red and nodes for gene sets with a negative NES are blue. The clustering of red and blue nodes is clearly visible and would have been missed when only looking at the GO gene set terms alone.

Figure 1 shows the PNG image file of the tree from AmiGO Visualize. By changing the picType parameter of the getAmigoTree() function, we can also specify SVG and DOT for the requested file type or it can be viewed in Cytoscape by exporting AmigoDot object using AmigoDot.to.Cyto (Fig. 2).

4 CONCLUSION

The R/Bioconductor package RamiGO provides an easy-to-use R interface to the AmiGO Visualize web server. It provides a simple and elegant way to retrieve Graphviz trees that display hundreds of GO IDs at once and efficiently study clusters or subcomponents of the GO tree in graph form. RamiGO provides functions to convert a GO tree into different formats and display it in Cytoscape without leaving the R environment. RamiGO is therefore a perfect companion to GSEA and GO analyses in R, as it helps one better analyse and interpret the long, and sometimes complex, lists of GO identifiers that these analyses produce.

Funding: This work was supported by the National Human Genome Research Institute (1P50 HG004233 to M.S.S.); Fulbright Commission for Educational Exchange to (B.H.-K.); US National Institutes of Health (#1U01CA151118-01A1 to J.Q., R01 LM010129-01 to B.H.-K. and J.Q.); Claudia Adams Barr Program in Innovative Basic Cancer Research (A.C.C. and

J.Q.); Career Development grant from DFCI Breast Cancer SPORE: CA089393, Dana-Farber Cancer Institute Women's Cancer Program (to A.C.C).

Conflict of Interest: none declared.

REFERENCES

Carbon,S. *et al.* (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics*, **25**, 288–289.

Gene Ontology Consortium. (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res*., Database Issue, D559–D564.

Gansner,E.R. and North,S.C. (2000) An open graph visualization system and its applications to software engineering. *Software Practice and Experience*, **30**, 1203–1233.

Minn,AJ. *et al*. (2007) Lung metastasis genes couple breast tumor size and metastatic spread. *Proc. Natl Acad. Sci. USA*, **104**, 6740–6745.

Schröder,M.S. *et al.* (2011) survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*, **27**, 3206–3208.

Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.

Wang,Y. *et al.* (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*, **365**, 671–679.