Reviewer’s report

Title: ViSEAGO: a Bioconductor package for clustering biological functions using Gene Ontology and semantic similarity

Version: 0 Date: 13 Mar 2019

Reviewer: Patryk Orzechowski

Reviewer's report:

The paper describes a software called ViSEAGO, which represents a tool for analysis of gene enrichment and visualization of GO terms. ViSEAGO takes advantage of multiple R packages, including topGO, GOSemSim and others.

Although I highly acknowledge time and effort invested in development of the software, there are a couple of concerns I would like to raise.

My major concern comes from the difficulties to understand what kind of novelty is associated with the software.

* Is it automated annotations acquisition? I believe this may be simply accomplished by keeping software updated and also reporting the research with sessionInfo() for the reproducibility.

* Is it the tool for conversions between ENTREZ, Ensembl and UniProt? I believe this isn't new and could be realized with simple mapping between annotations, or using ClusterProfiler.

* Is it a new visualization? Cluster map of GO terms? Not novel either, cluster maps were created a couple of years ago:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4406664/figure/fig04/

* Is it the workflow or combination of tools? There are multiple workflows already existing in the literature, for example:


https://f1000research.com/articles/5-1384/v1
* Is it a feature that automatically gathers multiple dataset related to the same organism and allows to analyze them simultaneously? If so, this raises many different questions on the right choice of the datasets for the analysis, which wasn't addressed in the paper.

* Is it the advantage of analyzing multiple dataset within the same workflow, Well, this can be done and was done in Bioconductor, e.g. https://bioconductor.org/packages/release/data/experiment/html/curatedBreastData.html

Thus, my question is what makes ViSEAGO different from existing tools? What does ViSEAGO offer that other tools don't? Is it a combination of different tools? I am not sure if I saw the answer within the text and I believe this needs to be clear to the reader.

Secondly, although the authors claim that the compared tools don't offer end-to-end workflow, a couple of workflows have already been proposed, either as a part of Bioconductor, or independently (see above).

How is ViSEAGO workflow different from the existing workflows in the literature? I haven't seen any mention about the existing workflows.

What makes ViSEAGO workflow unique? I believe the novelty aspect needs to be far more emphasized in the paper.

My third major concern is on integration of ViSEAGO with Bioconductor. I have noticed in previous commits on GitLab that the software was intended to be proposed as a part of this open source software for Bioinformatics (Commit adb17bd4). What are the reasons that it eventually wasn't integrated?

I would be far more convinced with ViSEAGO, and thus with this paper, if I reviewed a Bioconductor package, not a standalone software. Each package in Bioconductor goes through a strict review process and requires code quality standards and thorough documentation. Otherwise, although ViSEAGO is well documented and has examples on GitLab, I can't be certain of the quality and thoroughness of documentation. I can't also be sure that the software doesn't double functionalities that have already been implemented in Bioconductor.

My fourth point concerns presented applications on genomic data. I believe the analyses presented in the main paper should be accompanied by a source code in supplementary materials to this paper.
It can also be easier for potential users if a tutorial in markup language with fragments of code was added to the repository.

For the fifth, I have found the visualizations of GO terms with clusters very illegible and unconvincing (see Fig. 3 and 4).

Please either make a font smaller, consider a cut-off on the number of GO-terms that may be visualized or wrap a couple of GO terms forming the cluster together and substitute the figure with the representative of the cluster.

Leaving overlapping text doesn't make any sense, as no one is able to read it.

My minor comment is on a very poor quality of Figure 1. Please send figure in other format (eps?) and in a better resolution.

Summarizing, I would highly recommend that the authors reconsider integrating the software with Bioconductor. Please also pay more attention to differentiating in the paper the software and existing tools and workflows in the literature.

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?
4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes