Reviewer’s report

Title: PROGgeneV2: Enhancements on the existing database.

Version: 4  Date: 10 October 2014

Reviewer: Bülent Arman Aksoy

Reviewer’s report:

# Major Compulsory Revisions

1) PROGgeneV2 allows users upload their own data to the server, which is great for users who want to simply use this tool on a custom data set; but I do not think IP address based constraining of the data is a solid security implementation. Since the tool is aimed for use by clinically relevant researchers, it is safe to assume that data sets uploaded to this service should be highly protected. A majority of the institutions utilize firewalls for outgoing access and use proxy servers when providing internet access to its users. This means that IP addresses are commonly shared within a working environment, hence it is not safe to assume that same IP always means same user. I strongly suggest authors to implement a better solution for restricting access to the data, either through a user registration facility or by uniquely identifying a data set and allowing user to access to the data only when this ID is known. The former is better, but most services also adopt the latter for practicality.

2) The following directory exposes all the IP addresses that ran an analysis on the web site: http://watson.compbio.iupui.edu/chirayu/proggene/database/output/

Based on the IP address based restriction, this creates a minor security vulnerability and I suggest the authors to disable listing feature of this folder.

# Minor Essential Revisions

1) When user uploads data, the web site does not show a message confirming the successful upload operation. The only way to check whether the data gets uploaded is to go back to the analysis and see the data set name. This might be confusing for some users.

2) The authors provide a table listing all studies supported by the web service, but this information is not immediately available on the web site. I suggest the authors to provide this table with links to the relevant data sources on the web page, therefore making it easier for users to get to the up-to-date list.

3) v1 selects all matching studies by default when running a query, whereas the v2 lists all matching studies deselected by default. I suggest the authors to allow users to select/deselect all studies in the second step of the analysis for easier use.

# Discretionary Revisions

The results subsection in the abstract can be improved by:
- Instead of saying "almost 2 folds", simply putting the actual numbers of studies in v1 and v2
- Fewer uses of "users can now"
- Fixing the ambiguity due to 'and' usage: "... ratio of expression of two genes and curated and published gene signatures ...."

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests