SOFTWARE TOOL ARTICLE

REVISED GAC: Gene Associations with Clinical, a web based application [version 4; referees: 2 approved, 1 approved with reservations]

Xinyan Zhang[1], Manali Rupji [1], Jeanne Kowalski[1,2]

[1]Winship Cancer Institute of Emory University, Atlanta, GA, 30322, USA
[2]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, 30322, USA

v4 **First published:** 03 Jul 2017, **6**:1039 (doi: 10.12688/f1000research.11840.1)
Second version: 26 Sep 2017, **6**:1039 (doi: 10.12688/f1000research.11840.2)
Third version: 02 Jan 2018, **6**:1039 (doi: 10.12688/f1000research.11840.3)
Latest published: 15 Feb 2018, **6**:1039 (doi: 10.12688/f1000research.11840.4)

Abstract

We present GAC, a shiny R based tool for interactive visualization of clinical associations based on high-dimensional data. The tool provides a web-based suite to perform supervised principal component analysis (SuperPC), an approach that uses both high-dimensional data, such as gene expression, combined with clinical data to infer clinical associations. We extended the approach to address binary outcomes, in addition to continuous and time-to-event data in our package, thereby increasing the use and flexibility of SuperPC. Additionally, the tool provides an interactive visualization for summarizing results based on a forest plot for both binary and time-to-event data. In summary, the GAC suite of tools provide a one stop shop for conducting statistical analysis to identify and visualize the association between a clinical outcome of interest and high-dimensional data types, such as genomic data. Our GAC package has been implemented in R and is available via http://shinygispa.winship.emory.edu/GAC/. The developmental repository is available at https://github.com/manalirupji/GAC.

This article is included in the RPackage gateway.

Open Peer Review

Referee Status: ✓ ✓ ?

	Invited Referees 1	2	3
REVISED version 4 published 15 Feb 2018			
UPDATE version 3 published 02 Jan 2018		✓ report	
UPDATE version 2 published 26 Sep 2017	✓ report	? report	? report
version 1 published 03 Jul 2017	? report		

1 **Shengjie Yang**, NorthShore University HealthSystem, USA

2 **Matthew N. McCall**, University of Rochester Medical Center, USA

3 **Cedric Simillion**, University of Bern, Switzerland

Discuss this article

Comments (0)

Corresponding author: Jeanne Kowalski (jeanne.kowalski@emory.edu)

Author roles: Zhang X: Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Rupji M**: Software, Writing – Original Draft Preparation, Writing – Review & Editing; **Kowalski J**: Conceptualization, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Zhang X, Rupji M and Kowalski J. **GAC: Gene Associations with Clinical, a web based application [version 4; referees: 2 approved, 1 approved with reservations]** *F1000Research* 2018, **6**:1039 (doi: 10.12688/f1000research.11840.4)

Grant information: Research reported in this publication was supported in part by the Biostatistics and Bioinformatics Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

First published: 03 Jul 2017, **6**:1039 (doi: 10.12688/f1000research.11840.1)

REVISED **Amendments from Version 3**

Based on the reviewer's comment for our paper, we have been requested to update Figure 1–Figure 3 according to the changes that were made in our previous version 3. Figure 1–Figure 3 have been updated to reflect the customized output formatting.

See referee reports

Introduction

Heterogeneity in terms of tumor characteristics, prognosis, and survival among cancer patients has remained a persistent problem. It has been well established that clinical factors alone are not sufficient to explain differences in prognosis. For example, based on clinical factors only, two tumor patients may have the same prognosis, but may not respond to the same treatment as the tumors may have a completely different molecular composition[1–5]. Despite the introduction of a tumor's genomic profile to explaindifferences in prognosis, there remains unexplained heterogeneity in tumor response to treatment. One factor potentially attributing to such unexplained differences may be due to inaccurate prognosis and prediction resulting from the analysis approach used to define prognostic markers of response.

For this purpose, Bair and Tibshirani[6] introduced a supervised principal component (SuperPC) method within the context of defining expression-based cancer subtypes of prognostic significance. The method uses both gene expression and clinical data for predicting patient prognosis. This approach was applied to several publicly available datasets that demonstrated it's ability to accurately predict the clinical outcome of interest based on a given gene expression profile. Since its inception, SuperPC has been introduced as a powerful tool for reducing dimensionality in selecting features (a.k.a., genes) of prognostic relevance in cancer.

Currently, the SuperPC method has been developed as an R package, 'superpc' but as it stands, it is unable to address the following:

1) clinical association based on a binary outcome (e.g. responders versus non-responders);
2) ease of use for clinicians and researchers with limited programming skills; and
3) a visual summary of results.

To address these limitations, we developed GAC: Gene Association with Clinical, an interactive, GUI-based web-based application for analysis of gene associations with various clinical outcomes of interest. We developed GAC based on the R packages 'shiny' and 'superpc'. Our GAC tool enables the user to perform a SuperPC analysis for three types of outcomes: time-to-event, continuous, and binary, and provides a summary of results using forest plots that may be readily exported into a file.

Methods

Supervised principal component analysis

SuperPC is a generalization of principal component analysis, which generates a linear combination of the features or variables of interest that capture the directions of largest variation in a dataset. Instead of using the whole dataset directly, SuperPC defines a list of genes based on their association with an outcome of interest. To select the list of genes, a univariate score for each gene is calculated and those features (a.k.a., genes) whose score exceeds a threshold are retained as input into a principal component analysis, based on the retained features. For details, refer to Bair and Tibshirani[6].

Time-to-event outcome

SuperPC for time-to-event was conducted using the 'superpc' package in R. Depending on the sample size of the original dataset; the researcher selects what proportion of the dataset to split into training and testing. The researcher can also specify how many numbers to test to check which the optimal threshold is. The number of folds for cross validation to determine the threshold also needs to be determined. There is also an option to run the analysis randomly, or upload fold IDs to replicate an analysis that was previously carried out. The association between the time-to-event outcome and the predicted principal component may be represented in a KM plot by dichotomizing the principal component using the median (Figure 1).

Continuous outcome

SuperPC for continuous outcomes is implemented using the '*superpc*' package in R, with the same options as time-to-event analysis. The predicted principal component is presented visually as continuous values through a scatter plot along with Pearson's correlation (Figure 2). The predicted principal component could also be presented as binary groups (cutoff at median) through a boxplot, with a t-test applied.

Binary outcome

In the 'superpc' R package developed by Bair and Tibshirani (2004), SuperPC analysis can be performed on both continuous and survival outcomes. We have extended this tool to include SuperPC for binary outcome (example 'responders' vs 'non-responders'). This extension follows a similar analysis workflow as the other two outcomes in that a list of genes is defined based on a univariate score to which a threshold is applied and the genes whose scores exceed the threshold are used as input into a principal component analysis. For modeling gene associations with binary outcomes a logistic regression has been implemented. The predicted principal component can be visualized as either a continuous variable through a box plot, with a t-test to summarize the statistical association (Figure 3), or as binary groups (cutoff at median) using a bar plot, with a chi-square test to summarize the statistical association between the predicted and the observed outcome.

Forest plot

A forest plot is a graphical display of point estimates of association widely used in meta-analysis. It has become popular for displaying the associations between clinical and genomic data. With our GAC tool, users have the option to generate a forest plot to display results (Figure 4).

Implementation

The GAC tool is written in R and tested using version 3.3.0. The interactive plots and data tables are made available using the shiny R package (www.rstudio.com/shiny).

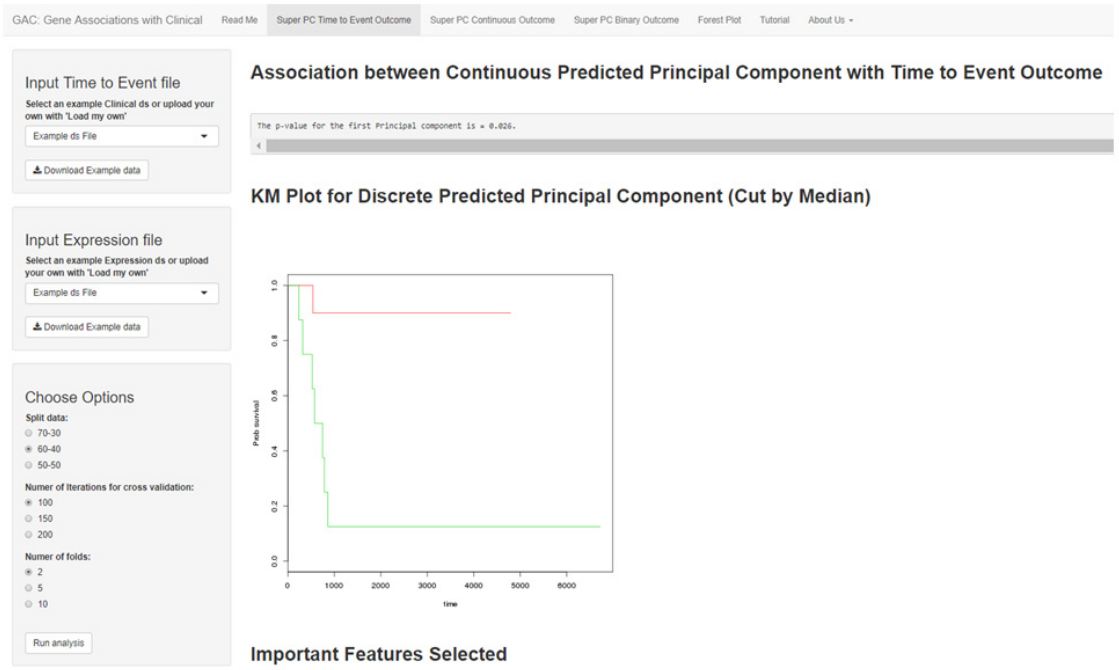

Figure 1. SuperPC time-to-event outcome. The interface shows an example SuperPC for time-to-event outcome. The left panel allows the user to select the various options such as data split, the optimal threshold, number of folds and a choice of generating new fold IDs or using a pre-existing set to replicate results. The right panel includes the results of the analysis. Use the 'Run Analysis' button (in left panel) to display results based on updated option(s). The KM plot displays the association of the outcome with the predicted principal component by the median. In addition, the univariate analysis regression scores and tables are also available for download.

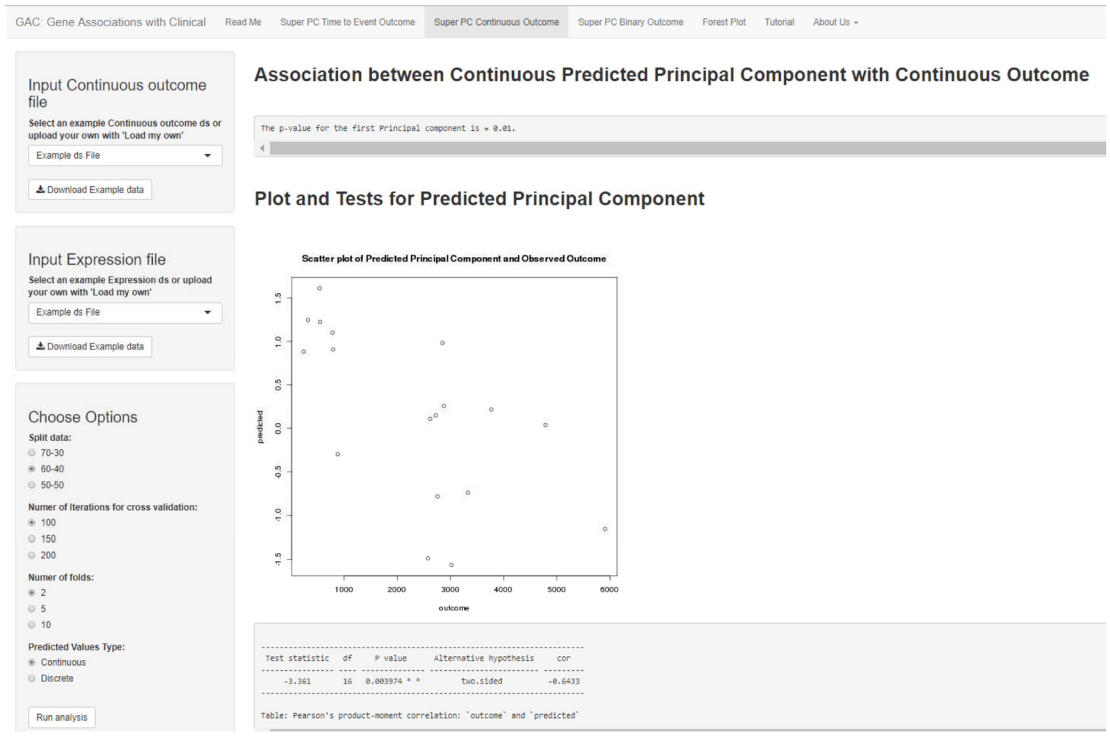

Figure 2. SuperPC continuous outcome. The interface shows an example SuperPC for continuous outcome. Similar to Figure 1, the user can choose the appropriate settings using the left panel and view the results of the analysis to the right. The association of the continuous outcome with the predicted principal component is summarized using a scatter plot (as seen). The user can alternatively choose to summarize these results through boxplots by dichotomizing the predicted principal component based on the median. As in Figure 1, the univariate regression scores and plots are available for download in the left panel.

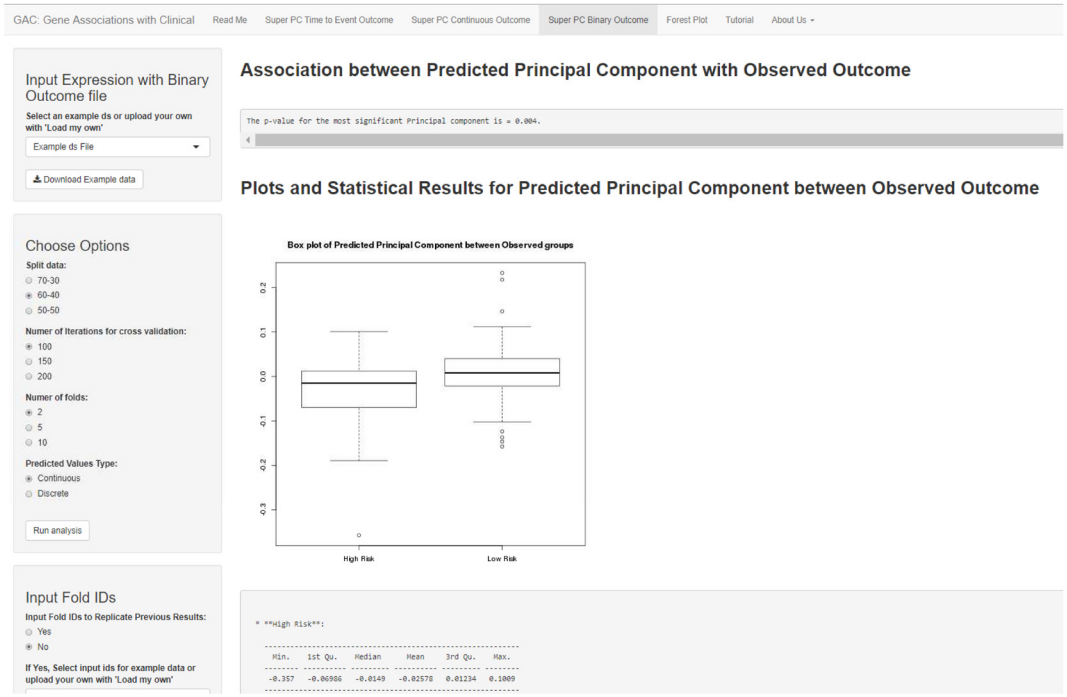

Figure 3. SuperPC binary outcome. The interface shows an example SuperPC for binary outcome. The user can opt for similar options as in the previous figures. Also, as in Figure 2, the user has a choice to display the continuous predicted principal component as a scatter plot, or divide it into binary discrete groups (using median cut-off) to represent the association through a barplot. Similar download options are available.

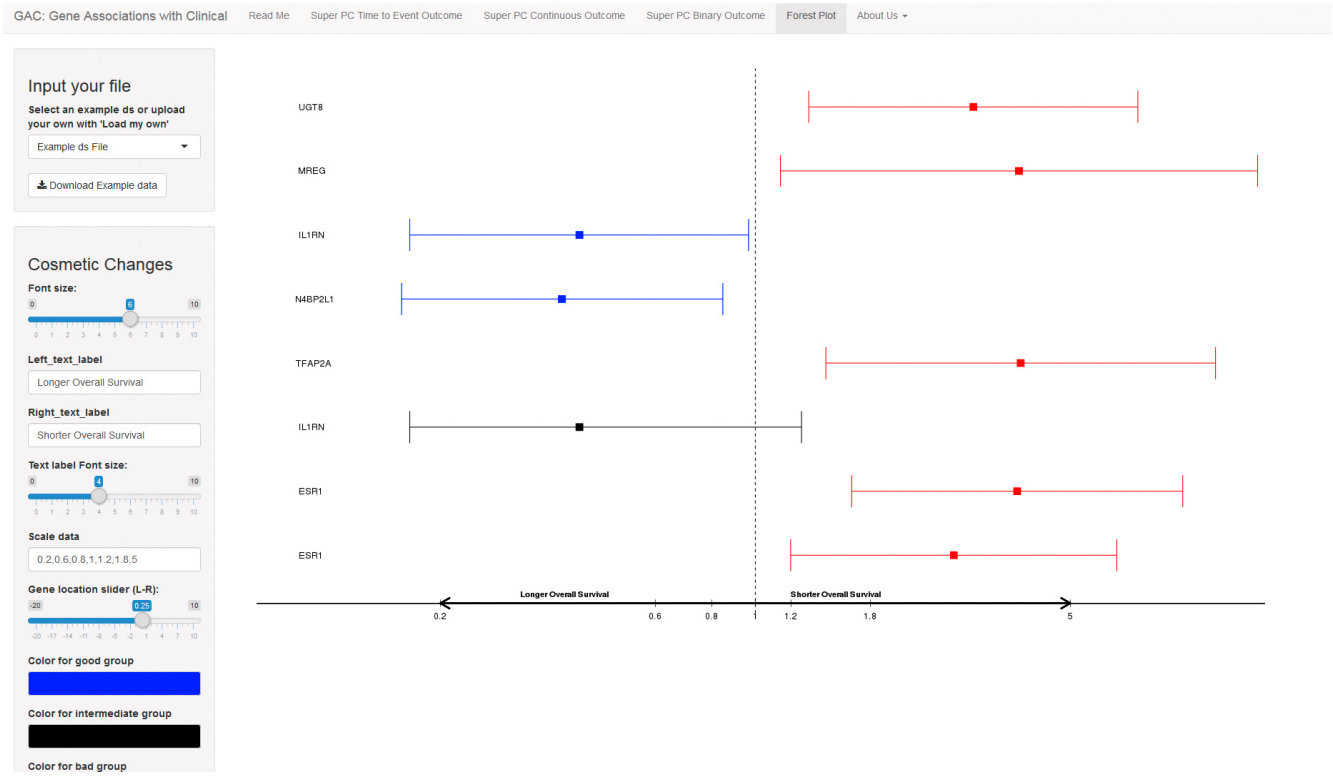

Figure 4. Forest plot. The interface shows an example forest plot. The left side comprises a user menu and the right includes the result plot. Users can upload their own summarized results with hazard ratio and confidence intervals for the survival outcome, or odds ratio and confidence intervals for the binary outcome. For graphical display, the researchers could choose to input different labels, font sizes and colors.

Operation
Using a windows 7 Enterprise SP1 PC with a 32.0 GB RAM and a 3.30 GHz Intel® Xeon® Processor E5 Family, the time-to-event and regression analysis with 45 patients from the TCGA BRCA dataset was completed in 2.54 s and 1.80 s, respectively. The binary outcome analysis using the CoMMPass data of 135 patients was completed in 65.05 s. Source code is available at https://doi.org/10.5281/zenodo.1064841.

Discussion
GAC is a suite of tools that allows the user to conduct statistical analysis to identify and visualize the association between clinical outcomes of interest and genomic data using an interactive application in R.

Data and software availability
For SuperPC time-to-event and regression analysis, we used TCGA BRCA RNASeqV2 gene expression and clinical data, downloaded from the TCGA data portal (now accessed at https://portal.gdc.cancer.gov/)[7]. The data included 380 differentially expressed genes when favorable (patients who did not die with at least 7 years of follow up) and unfavorable (patients who died 30 months post-treatment) outcomes from 45 patients were compared.

For SuperPC binary outcome analysis, CoMMPass IA9 RNASeq expression and clinical data was downloaded from the publicly available Multiple Myeloma Research Foundation (MMRF) database (https://research.themmrf.org/rp/download). Patient cytogenetics for outcome dichotomization was obtained from the MMRF data portal's 'Analysis Tools' section (https://research.themmrf.org/rp/explore/). 135 patients with clinical, gene expression and copy number data were classified as high risk based on cytogenetics and the remaining 343 as not high risk. Among these patients, the top 1450 most variable genes were used.

The developmental repository is available at https://github.com/manalirupji/GAC.

The example data used in this article is available in Zenodo. User uploaded data should be in the same format as the example data provided.

Archived source code as at the time of publication: https://doi.org/10.5281/zenodo.1064841[8]

Competing interests
No competing interests were disclosed.

Grant information
Research reported in this publication was supported in part by the Biostatistics and Bioinformatics Shared Resource of Winship Cancer Institute of Emory University and NIH/NCI under award number P30CA138292. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Alizadeh AA, Eisen MB, Davis RE, *et al.*: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature.* 2000; **403**(6769): 503–11.
PubMed Abstract | Publisher Full Text
2. Sørlie T, Perou CM, Tibshirani R, *et al.*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci U S A.* 2001; **98**(19): 10869–74.
PubMed Abstract | Publisher Full Text | Free Full Text
3. van 't Veer LJ, Dai H, van de Vijver MJ, *et al.*: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature.* 2002; **415**(6871): 530–6.
PubMed Abstract | Publisher Full Text
4. van de Vijver MJ, He YD, van't Veer LJ, *et al.*: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med.* 2002; **347**(25): 1999–2009.
PubMed Abstract | Publisher Full Text
5. Lapointe J, Li C, Higgins JP, *et al.*: **Gene expression profiling identifies clinically relevant subtypes of prostate cancer.** *Proc Natl Acad Sci U S A.* 2004; **101**(3): 811–6.
PubMed Abstract | Publisher Full Text | Free Full Text
6. Bair E, Tibshirani R: **Semi-supervised methods to predict patient survival from gene expression data.** *PLoS Biol.* 2004; **2**(4): E108.
PubMed Abstract | Publisher Full Text | Free Full Text
7. NCI and NHGRI: **The Cancer Genome Atlas (TCGA) Data Portal.** Accessed in December 2015.
Reference Source
8. manalirupji: **manalirupji/GAC: GAC v1.2.0.** *Zenodo.* 2017.
Data Source

Open Peer Review

Current Referee Status: ✔ ✔ ?

Version 3

Referee Report 08 January 2018

doi:10.5256/f1000research.14759.r29420

✔ **Matthew N. McCall**
Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA

I thank the authors for incorporating my suggestions into their work; however, the figures in the paper should be updated to reflect the changes in the shiny app, which were made in response to comment 2 from my initial review:

"2. The numeric output of the GAC shiny app is often unformatted R output. For example, on the Super PC Binary Outcome tab below the box plots. It would be better to provide formatted results (e.g. using knitr::kable). Also, the p-values (and other numeric outputs) should likely be rounded to a number of significant digits that conveys the precision of the values."

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 15 Feb 2018

Manali Rupji, Emory University, USA

We thank the referee for their suggestions which have lead to an improved paper. In the latest version, we have incorporated updates to Figures 1-3 to reflect the changes to output formatting made in version 3.

Competing Interests: No competing interests were disclosed.

Version 2

Referee Report 26 October 2017

doi:10.5256/f1000research.13806.r26969

Cedric Simillion
Interfaculty Bioinformatics Unit and SIB Swiss Institute of Bioinformatics, University of Bern, Berne, Switzerland

I agree with the previous referee that the authors have developed a nice user interface to the SuperPC algorithm, which should make its application accessible to many experimental scientists. I read the manuscript and I have little to no comments to make on its contents.

However, I have tried to use the tool myself on my own dataset and was unable to get it to work. I had formatted my data exactly as in supplied example input files, even with the exact same column names. After submitting my data, I only got an unhelpful error message "ERROR: An error has occurred. Check your logs or contact the app author for clarification." I could not see any logs or any other means of finding out why the program didn't work.

If the problem lies with the formatting of the input files, the authors need to provide some more documentation on the format of the input data, including column names, newline convention, etc.

Given that I presently cannot verify that the tool is working properly, I therefore cannot accept the manuscript for indexing.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 Dec 2017

Manali Rupji, Emory University, USA

We thank the reviewer for their thoughtful comments.

Below is our response to the reviewer's comments :

Thank you for pointing this problem out. As it turns out, there was indeed a bug with reading the input data that caused the problem described. This bug has been fixed in the new version of the tool available on the same link http://shinygispa.winship.emory.edu/GAC/.
Additionally, we have included a sentence about data formatting in the Data and software availability section: 'User uploaded data should be in the same format as the example data provided.' The tutorial has also updated to state the same.

The tool has also been updated to indicate errors in code directly as seen on the R console as a way to better inform on the possible nature of the error.

Competing Interests: No competing interests were disclosed.

Referee Report 23 October 2017

doi:10.5256/f1000research.13806.r26812

Matthew N. McCall
Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA

The authors describe a software package for interactive (via a shiny webapp) use of the superpc R package. This work will likely make the methods in the superpc package available to users who are unable to work in R.

My primary criticism of this work is that there does not seem to be a clear description of the methodology to which GAC provides GUI access. I understand that these methods were not developed by the authors of this paper; however, if the target audience are clinician scientists, it is doubtful that will have read the original paper[1].
A lay description of the method would greatly improve this manuscript.

I also have a few minor criticisms of the article:

1. The authors write, "Currently, prognosis and outcome predictions are solely based on clinical factors." This seems to dismiss a lot of recent work in this area. On a related note, the most recent paper cited is from 2004.

2. The numeric output of the GAC shiny app is often unformatted R output. For example, on the Super PC Binary Outcome tab below the box plots. It would be better to provide formatted results (e.g. using knitr::kable). Also, the p-values (and other numeric outputs) should likely be rounded to a number of significant digits that conveys the precision of the values.

3. There are several grammatical errors. For example:
   “The researcher can also specify how many numbers to test to check which the optimal threshold is.”

4. The superPC method is referred to in a variety of ways; it would be good to pick one of these:
   “Our GAC tool enables the user to perform a SuperPC analysis...”
   “... super pc analysis can be performed on both continuous …” “We have extended this tool to include super PC…”
   “Figure 2. SuperPC continuous outcome”

References

1. Bair E, Tibshirani R: Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol*. 2004; **2** (4): E108 PubMed Abstract | Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
No

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 Dec 2017

Manali Rupji, Emory University, USA

We thank the reviewer for their thoughtful comments that have led to both an improved paper and tool.

We have included in the introduction section a brief summary of the 'superpc' method.

Additionally, below, we respond individually to each comment.

1. The authors write, “Currently, prognosis and outcome predictions are solely based on clinical factors.” This seems to dismiss a lot of recent work in this area. On a related note, the most recent paper cited is from 2004.

We agree with the reviewer and have updated the introduction.Considering that the main method, supervised principal components analysis, was published in 2004, the other papers referenced earlier reflect the cited results of the method.

2. The numeric output of the GAC shiny app is often unformatted R output. For example, on the Super PC Binary Outcome tab below the box plots. It would be better to provide formatted results (e.g. using knitr::kable). Also, the p-values (and other numeric outputs) should likely be rounded to a number of significant digits that conveys the precision of the values.

We have modified the output in both the SuperPC binary and continuous outcome sections to provide formatted results using pander from the ‘knitr’ package. which also provides greater control on the display of results. The numeric output has also been modified as suggested by provided rounded results in the table.

3. There are several grammatical errors. For example: “The researcher can also specify how many numbers to test to check which the optimal threshold is.”

We have corrected the grammatical errors in the manuscript.

4. The superPC method is referred to in a variety of ways; it would be good to pick one of these:“Our GAC tool enables the user to perform a SuperPC analysis...”
“... super pc analysis can be performed on both continuous …” “We have extended this tool to include super PC…”
“Figure 2. SuperPC continuous outcome”

We thank the reviewer for pointing out this inconsistent nomenclature. When specifically referring to the R package, we use the name of the package as ‘superpc’ and when referring to the tool, method or software, ‘SuperPC’ is used throughout the manuscript.

Competing Interests: No competing interests were disclosed.

Referee Report 27 September 2017

doi:10.5256/f1000research.13806.r26394

Shengjie Yang
Program for Computational Genomics & Medicine, NorthShore University HealthSystem, Evanston, IL, USA

The authors have addressed my concerns appropriately in the revised version and this work is worthy of indexing.

Competing Interests: No competing interests were disclosed.

Referee Expertise: Clinical trial design

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Sep 2017

Manali Rupji, Emory University, USA

We thank the referee for taking the time to review our work, and thank them for the kind review.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 18 September 2017

doi:10.5256/f1000research.12794.r26010

Shengjie Yang
Program for Computational Genomics & Medicine, NorthShore University HealthSystem, Evanston, IL, USA

The authors did a nice job to develop a web tool called GAC, which is used to explore the association between clinical and genomic data. Essentially, GAC is a GUI of R package ‘superpc’. They not only provided a handy web tool but also published the source code to allow researchers to do some further expansion according to their own specific study objects. The manuscript is well written, but I have some comments:

1. Since the example data is from TCGA and this work aims to facilitate clinicians and researchers with limited programming skills, it is necessary to cite some literature to tell readers how to retrieve and process the TCGA Data, such as TCGA-Assembler (Nature methods, 2014), and it would be better if such tools can be embedded into GAC.

2. The authors stated that ‘superpc' is unable to address the limitation about clinical association based on a binary outcome in the Introduction. But in the Methods section, it said ‘SuperPC for binary outcome follows the same analysis workflow as the other two outcomes’. It confuses me and please explain the details.

3. I find that the calculation will be automatically triggered when changes any one option in the sidebar. That is not an user-friendly experience, especially it happens in the ’Super PC Binary Outcome’ module which can not return the result immediately. Think about that if the users want to change two default options, they have to adjust the first one, wait a few seconds and then adjust the second option, which is not efficient. So adding a ‘submit’ button to trigger the calculation is a better way.

4. Build a user manual page to explain the details of each function and use an example to demonstrate the usage.

Is the rationale for developing the new software tool clearly explained?
Partly

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 20 Sep 2017

Manali Rupji, Emory University, USA

We thank the reviewer for their thoughtful comments that have led to an improved paper. Below, we respond individually to each comment.

1. Since the example data is from TCGA and this work aims to facilitate clinicians and researchers with limited programming skills, it is necessary to cite some literature to tell readers how to retrieve and process the TCGA Data, such as TCGA-Assembler (Nature methods, 2014), and it would be better if such tools can be embedded into GAC.

   Various tools are available for extracting TCGA data (now available through GDC data portal). Data can be downloaded directly from the GDC data portal (https://gdc-portal.nci.nih.gov/) or by use of R packages. Some of the more recent packages include TCGA Biolinks (Colaprico *et. al*, 2015) which is available through Bioconductor and the TCGA Assembler (Zhu *et. al*, 2014) available as an R package. Users may refer to the detailed tutorials of these tools for instructions on extracting TCGA data types.

2. The authors stated that 'superpc' is unable to address the limitation about clinical association based on a binary outcome in the Introduction. But in the Methods section, it

said 'SuperPC for binary outcome follows the same analysis workflow as the other two outcomes'. It confuses me and please explain the details.

We have included a sentence in the methods section that clarifies how the superPC tool has been extended to include analysis of binary outcome data.

3. I find that the calculation will be automatically triggered when changes any one option in the sidebar. That is not a user-friendly experience, especially it happens in the 'Super PC Binary Outcome' module which cannot return the result immediately. Think about that if the users want to change two default options, they have to adjust the first one, wait a few seconds and then adjust the second option, which is not efficient. So adding a 'submit' button to trigger the calculation is a better way.

   A 'Run Analysis' submit button is now available under each analysis option that will update results based on user-specified options to multiple parameters, as opposed the previous version which triggered a calculation after each separate change.

4. Build a user manual page to explain the details of each function and use an example to demonstrate the usage.

   We have included a tutorial as a separate tab within the tool to demonstrate the usage of each function with example data.

References:

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot T, Malta TM, Pagnotta SM, Castiglioni I, Ceccarelli M, Bontempi G and Noushmehr H (2015). "TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data." *Nucleic Acids Research*. doi: 10.1093/nar/gkv1507, http://doi.org/10.1093/nar/gkv1507.

Genomic Data Commons (GDC) Data Portal: https://gdc-portal.nci.nih.gov/. NIH and NCI, accessed in August 2016.

Y. Zhu, P. Qiu, Y. Ji. TCGA-Assembler: Open-Source Software for Retrieving and Processing TCGA Data. Nature Methods. Vol. 11, No. 6, pp:599-600, 2014. | doi:10.1038/nmeth.2956

Competing Interests: No competing interests were disclosed.