Author’s response to reviews

Title: ClickGene: an open cloud-based platform for big pan-cancer data genome-wide association study, visualization and exploration

Authors:
Jia-Hao Bi (jh_bi@tju.edu.cn)
Yi-Fan Tong (tongyfan@gmail.com)
Zhe-Wei Qiu (zw_qiu@tju.edu.cn)
Xing-Feng Yang (nerosoft@outlook.com)
John Minna (john.minna@utsouthwestern.edu)
Adi Gazdar (adi.gazdar@utsouthwestern.edu)
Song Kai (ksong@tju.edu.cn)

Version: 2 Date: 22 May 2019

Author’s response to reviews:

From Reviewer!

The authors left unaddressed my first two comments:

- Regarding comment 1, following my recommendation the authors no longer claim their results are inconsistent with the current literature. However, they still use the mean, which in their GBM example is totally dominated by a few outliers. I do not see any reason for using the mean instead of the median.

Answer: We used the median instead of mean to do the Mountain plots this time. So sorry for the troubles.

- Regarding comment 2, the authors did not perform any statistical analysis of the significance of the small differences they find among the DTW scores of different tumor types. I do not see any reason for not doing such an analysis.
Answer: We used Bootstrapping test to test the DTW score between tumor samples and the corresponding non-malignant samples. The p-value is calculated as following.

To test whether the inconsistent between the Mountain curves of two different groups of samples are caused by random fluctuations, Bootstrapping based significance test is provided to evaluate the significance of the difference between these two kinds of samples.

For CNV genome datasets of a group with m samples and of another group with n samples (the DTW score between them is noted as DTWtn), the steps for Bootstrapping test are as follows:

1) Mix all samples into one group;
2) Randomly rearrange the samples of the mixed group;
3) Select the first m samples as group1 and the rest as group2;
4) Calculate the DTW score between group1 and group2;
5) Repeat step 2-5 10,000 times and get the DTW1…10000 scores
6) Sort DTWtn and all 10000 DTW scores together, if DTWtn ranked as the ith, then the corresponding p-value is i/10000.

If p-value<0.05, the DTW score between these two groups of samples are not considered to be caused by random fluctuations.

Even though more repeat times of Bootstrapping may be better, but the running time for it for 10,000 is already about 159s per chromosome now (CPU: Intel Core i7 s960X, 6 cores 12 threads, 4.2GHz, DDR3 1600MHz). For the RPC server we are renting isn’t as good as the computer we mentioned above, therefore the response time should be longer than 3 mins. Considering about the response time for the end-users and the sample size (normally smaller than 1000), we believe 10,000 times Bootstrapping is acceptable. We are working on making the online response time of it shorter.

Thank you for your kind correction and comments which made our manuscript much better. We really appreciate it!