Yale University

Bass Building, Rm 432A
266 Whitney Avenue
New Haven, CT 06520-8114
Telephone: 203 432 6105
pi@gersteinlab.org

Jan 2nd, 2018

Dear Editor,

I am writing to submit our manuscript, titled "RADAR: An integrative framework for variant annotation and prioritization in post-transcriptional regulome for RNA binding proteins" for the exclusive consideration of publication in Genome Biology.

This paper is submitted to as a ENCODE3 companion paper focusing on variant impact in post-transcriptional regulations. It incorporates the entire eCLIP, shRNA RNA-Seq, and RNA Bind-n-Seq experiments in ENCODE3. Two other papers in the current submission package, which have just got positive reviews back from Nature, also used this dataset with different focuses.

Many methods have previously been developed on this topic, but most of them only focused on transcriptional-level regulation. However, post-transcriptional regulation – such as splicing, cleavage, polyadenylation, editing, localization, stability, and translation – precisely controls the fate of cells, and its dysregulation has been proven to be disease-causal. The post-transcription regulome covers a much larger region than one may think (1.5 times the size of the whole exome) and it demonstrates larger than expected cross-population and cross-species conservations. Hence it is necessary to develop computational tools to annotate and prioritize the post-transcription regulome.

Our RADAR framework contains two components. It uses a baseline score to measure a variant's baseline impact through the use of a pre-built data context and also has a tissue-specific score to highlight specific disease-related variants. Specifically, we incorporated the following features for the baseline score:

- *RBP regulome information:* 318 eCLIP experiments for 112 RBPs in ENCODE
- *Conservation information:*
  - Cross-population sequence conservation inferred by SNPs from 1000 Genomes
  - Cross-species sequence conservation from Gerp score
  - RNA structure conservation from Evofold
- *Network information:* 472 shRNA RNA-Seq experiments
- *Motif information:* 76 Bind-n-Seq experiments and *de novo* motifs directly from peaks

For the tissue-specific score, RADAR incorporates user-specific inputs, such as differential expression/mutational profiles or prior knowledge of genes to further highlight disease-relevant variants. Results on somatic and germline variants demonstrate that RADAR can successfully pinpoint intronic, splicing-disruptive variants in key genes such as TP53, which cannot be fully detected by current methods. We believe that RADAR can serve as a useful tool to annotate and

prioritize the post-transcriptional regulome for RBPs, which has not been covered by most of the current variant functional impact interpretation tools. We make available all the results as flat text files on our website (radar.gersteinlab.org).

We appreciate you taking the time to review and respond to our manuscript. Please address all correspondence concerning this manuscript to pi@gersteinlab.org. We attach a list of our suggested referees. These are broken into those more associated with basic genomics and those more associated with disease studies.

Referees that we would like to exclude are Nuria Lopez-Bigas and Jay Shendure, who we believe have a conflict of interest.

We recommend the following experts with a strong background in basic genomics and expertise in RNA regulation as our potential referees.

- *Experts in RNA field*
  - Zhi John Lu (lulab@biomed.tsinghua.edu.cn, Tsinghua University)
  - Grace Xiao (gxxiao@ucla.edu, UCLA)
  - Ouyang Zhengqing (zhengqing.ouyang@jax.org, the Jackson Lab)
  - Jingyi Jessica Li (jli@stat.ucla.edu, UCLA)

- *Experts in integrative genomics*
  - Steven E. Brenner (brenner@compbio.berkeley.edu, UC Berkeley)
  - Chongzhi Zang (zang@virginia.edu, University of Virginia)
  - Chao Cheng (chao.cheng@dartmouth.edu, Dartmouth Colledge)
  - Xianghong Jasmine Zhou (xjzhou@mednet.ucla.edu, UCLA)

Many thanks,
Mark Gerstein