Yale University

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

Aug 22, 2017

Dear Orli,

I am writing to submit our manuscript, titled "An integrative ENCODE resource for cancer genomics" for the exclusive consideration of publication in Nature.

In this submission, we attempted to change the paper substantially based on the comments you had from the initial reading. In particular, we have tried hard to make this paper more clearly be an ENCODE resource that is useful for people in cancer genomics, but not a research paper about findings in cancer genomics. We hope that our revisions to the text make this clear, and we appreciate your suggesting these beforehand. We also point out that it is always very hard writing a resource paper where one feels in balancing the potential research results that illustrate the value of the resource, from just describing the resource, and we hope we have struck the right balance here.

This work is intended to serve as a resource for disease research, in complement to the main ENCODE encyclopedia resource. The main contribution of this paper is deep annotation of cancer genomes using advanced and novel assays like STARR-seq, Hi-C, RAMPAGE, and eCLIP, together with a broad spectrum of traditional assays including hundreds of ChIP-seq experiments. We were able to use these data to:

1. Build an accurate background mutation rate model through massive data integration
2. Define compact annotation for maximized statistical power in mutation burden analysis.
3. Construct various transcription factor (TF) and RNA binding protein (RBP) regulatory networks from binding profiles of hundreds of ChIP-seq and eCLIP experiments.

As an illustration of these resources, we combined it with large cancer cohort data to

1. Identify mutationally burdened regions in multiple cancer types
2. Identify key regulators that drive tumor-to-normal differential gene expression
3. Measure direct regulation changes in the transition from normal to tumor cells
4. Prioritize key regulators, functional elements, and functional variants, and then validate these prioritizations using small-scale functional assays.

We make available all the results as flat text files on through the ENCODE project portal - encodec.encodeproject.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 below a list of our referees. These are broken into those more associated with basic genomics and those more associated with cancer genomics. Many of the cancer genomics referee suggestions come from PCAWG.

The one referee that we'd like to exclude is Peter Campbell who we believe has a conflict of interest.

Various people from PCAWG & other cancer projects

- Gad Getz (gadgetz@broadinstitute.org, the Broad Institute)
- Li Ding (lding@genome.wustl.edu, Washington University School of Medicine)
- Jan Kobel (korbel@embl.de, EBI)
- Lincoln Stein (lincoln.stein@oicr.on.ca, OICR)
- Han Liang (hliang1@mdanderson.org, MD Anderson)

Various people from basic genomics

- Matthieu Lupien (mlupien@uhnres.utoronto.ca, University of Toronto)
- Daniel Geschwind (dhg@mednet.ucla.edu, UCLA)
- Aleksandar Milosavljevic (amilosav@bcm.edu, Baylor College of Medicine)
- Peter Park (peter_park@harvard.edu, Harvard)
- Jay Shendure (shendure@uw.edu, UW)
- Charles Lee (Charles.Lee@jax.org, JAX)
- Marc Vidal (marc_vidal@DFCI.harvard.edu, DFCI)
- Kai Wang (kw2701@cumc.columbia.edu, Columbia University)

Many thanks,
Mark Gerstein