Yale University

Bass Building, Rm 432A
266 Whitney Avenue
PO Box 208114
New Haven, CT 06520-8114

203 432 6105
360 838 7861 (fax)
mark@gersteinlab.org

27th August 2015

Nature Communications
75, Varick Street
Fl 9, New York
NY, 10013-1917
USA

Dear Dr. Cho,

Thank you for the invitation to revise and resubmit the manuscript. We are heartened that reviewers #1 and #3 find our responses satisfactory and have endorsed our manuscript for publication in *Nature Communications*. However, we are rather surprised by reviewer #2's comments.

The publications that we cited in our responses are a selection of the most current work performed by authorities in the field and peer-reviewed by colleagues in the community. The main point we are trying make is not to show the 'correctness' of these methods, but to point to the broader reality that there is currently a diversity of methods in the community, where there is no firm consensus on the 'correct' methodology. For example, while the GTEx consortium [1] did attempt to correct for allelic mapping bias by filtering out susceptible sites, they performed their alignment on the human reference genome and allele-specific detection using binomial tests, not accounting for over-dispersion. On the other hand, while Ding *et al.* [2] similarly performed their alignment on the human reference genome and allele-specific detection using binomial tests, they did *not* correct for allelic mapping bias. To further underscore the plurality of current approaches, Kasowski *et al.* [3] did *not* correct for allelic mapping bias, and they performed their alignment using personal genomes and allele-specific detection using binomial tests, also not accounting for over-dispersion. However, the fact that the reviewer is insisting on his points of view suggests his prejudice for a particular 'right' approach.

In our endeavor to use the wealth of existing datasets, we have come to appreciate and acknowledge this diversity, and thus have advocated for the need to uniformly process these datasets. Our allele-specific detection approach has already been extensively discussed and ultimately utilized in the ENCODE, Epigenomics Roadmap and 1000 Genomes Project consortia. The ENCODE consortium has utilized an earlier version of our approach in its 2012 publication [4]. It is also currently being utilized by the Epigenomics Roadmap consortium in their allele-specific analyses. Moreover, our approach is used in the analyses of the 1000 Genomes Project Structural Variants group because the personal genome construction is

especially useful in incorporating indels and structural variants; the other methods are only limited to single nucleotide variants. Furthermore, building a personal genome not only reduces the reference bias as mentioned by reviewer #2, but, as we show in our new analyses and responses, it is also less affected by the type of allelic bias that was highlighted in van de Guijn *et al*. The manuscript from the Structural Variant Group of the 1000 Genomes Project consortium, which included analyses from our approach, has just been recently peer-reviewed and accepted by *Nature*. Thus, we are quite taken aback by reviewer #2's comments.

We agree that allele-specific analyses are challenging. Therefore, there is a plethora of approaches, with corresponding pros and cons, developed to attempt to address these concerns. We have made significant efforts to incorporate *all* the reviewers' comments. We are encouraged by two of the reviewers' endorsements of our work. But, given the general acceptance of our approach, and the myriad of methods observed in multiple influential and recent publications, we believe that reviewer #2's insistence that we are 'wrong', and his allusion to a community-wide agreement of a single approach, are unfounded. Nonetheless, we have added new analyses and responses to address specifically reviewer #2's concerns. We strongly believe that our approach will generate considerable interest in the community. Hence, we do hope to seek your understanding and do take into consideration this cover letter when making your decision.

Yours sincerely,

Mark Gerstein
Co-chair of 1000 Genomes Project Consortium Functional Interpretation Group and Member of the 1000 Genomes Project Consortium Structural Variation Group
Albert L. Williams Professor of Biomedical Informatics, Molecular Biophysics & Biochemistry, and Computer Science,
Co-director of the Yale Program in Computational Biology and Bioinformatics

[1] The GTEx Consortium (2015). *Science*. 348(6235):648-60
[2] Ding, Z. *et al.* (2014). PLoS Genet. 10(11):e1004798
[3] Kasowski *et al.* (2013). 342(6159):750-2
[4] Djebali *et al.* (2012). 489(7414):101-8

We list a number of suitable reviewers for the paper:

Professor Aleksandar Milosavljevic
Baylor College of Medicine, Texas, USA
amilosav@bcm.edu

Professor Tom Gingeras
Cold Spring Harbor Laboratory, New York, USA
gingeras@cshl.edu

Professor Roderic Guigo
Centre for Genomic Regulation, Barcelona, Spain
roderic.guigo@crg.cat

Professor Zhiping Weng
University of Massachusetts Medical School, Massachusetts, USA
zhiping.weng@umassmed.edu

Dr. Paul Bertone
EMBL-EBI, Cambridge, United Kingdoms
bertone@ebi.ac.uk

Professor Chris Mason
Weill Cornell Medical College, New York, USA
chm2042@med.cornell.edu

Due to conflict of interests, we would like to request that our manuscript not be reviewed by:

Professor Tuuli Lappalainen
New York Genome Center, New York, USA
tlappalainen@nygenome.org

Professor Emmanouil Dermitzakis
University of Geneva, Geneva, Switzerland
emmanouil.dermitzakis@unige.ch

Professor Jonathan Pritchard
Stanford University, California, USA
pritch@stanford.edu

Professor Lior Pachter
University of California at Berkeley, California, USA
lpachter@math.berkeley.edu