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 at present a diversity of methods in the community. For example, while the GTEx consortium [1] did attempt to correct for allelic mapping bias, 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, Ding *et al.* [2] performed their alignment on the human reference genome and allele-specific detection using binomial tests, but did *not* correct for allelic mapping bias. Given the plurality of current approaches, the fact that the reviewer is insisting on his/her points of view suggests his/her prejudice for a particular 'right' approach, where there is no firm consensus.

In our endeavor to mine 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 [3]. 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. Specifically, the personal genome construction is especially useful in structural variant analyses since it is able to incorporate indels and structural variants; the other methods are only limited to single nucleotide variants. The manuscript from

Deleted: because

Deleted: incorporating

Deleted: Building

the Structural Variant Group of the 1000 Genomes Project consortium has just been recently peer-reviewed and accepted by *Nature*.

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 Degner *et al* [4] and van de Geijn *et al* [5].

We agree that allele-specific analyses are challenging. Therefore, there is a plethora of approaches, with corresponding pros and cons, developed to address various concerns. Reviewer #2 has made some reasonable suggestions, thus we have made significant efforts in trying to incorporate his and *all* the reviewers' comments. While we have also tried to add new analyses and responses in this round of review to address specifically reviewer #2's concerns, we fear an insistence on his/her single approach in performing allele-specific detection when there are multiple ways. Nonetheless, we are encouraged by the other reviewers' endorsements of our current manuscript and indeed, strongly believe that our approach and resource 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] Djebali *et al.* (2012). *Nature*. 489(7414):101-8
[4] Degner *et al.* (2009). *Bioinformatics*. 25(24):3207-12
[5] van de Geijn *et al.* (2015). *bioRxiv*. doi: http://dx.doi.org/10.1101/011221

We list a number of suitable reviewers for the paper:

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

Deleted: van de Geijn *et al* [4]. 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*.

Deleted: Some of reviewer #2's

Deleted: were reasonable

Deleted: incorporating,

Deleted: [4

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