RESPONSE TO REVIEWERS FOR "ALLELE-SPECIFIC BINDING AND EXPRESSION: A UNIFORM SURVEY OVER THE 1000-GENOMES-PROJECT INDIVIDUALS"

RESPONSE LETTER

Reviewer #1

-- Ref1 – General positive comment --

Reviewer Comment	This reviewer did not have formal comments to the authors as s/he found the revised paper to be satisfactory and endorses publication.
Author Response	We thank the reviewer for his/her thorough examination of our manuscript and endorsing our paper for publication.

Reviewer #2

-- Ref2.1 – General comment --

Reviewer Comment	The authors did not adequately address my two major concerns.
Author Response	We thank the reviewer for the thorough examination of our manuscript. We have provided additional analyses and responses.

-- Ref2.2 – mapping to the personal diploid genome --

Reviewer Comment	My first comment was that mapping bias should be addressed. The authors replied by explaining that they excluded reads that map to more than one location. This is indeed a standard step in more alignment. Yet, the challenge when looking for ASE is not standard. Different alleles may have different mapping probabilities and this must be taken into account. Failing to do so results in a high number of falsely identified ASE. I must admit that it is a bit concerning to me that the authors interpreted my comment as a question regarding their standard alignment approach. In my mind, it points to a deep lack of familiarity with the ASE literature.
Author Response	We agree with the reviewer that allelic mapping bias can be an issue, and it has first been mentioned in Degner *et al.* [1]. We performed additional analyses to show that this only affects a small proportion of our results, demonstrating that our approach is conservative and also alleviates this type of allelic bias. [1] Degner *et al.* (2009) *Bioinformatics.* 25(24),
Excerpt From	

Deleted: . ¶

Formatted: Font color: Auto

Deleted: .

Formatted: Font: Italic, Font color: Auto

Formatted: Font color: Auto

Deleted:] discussed two examples of mapping bias: the reference bias and allelic mapping bias, both of which can be resolved by the use of the personal genomes. The latter is accounted by the construction of two reference genomes, where both reference and alternate alleles are properly phased and represented; the former has been shown previously [2]. Van de Geijn *et al.* [3] very recently provided another example of allelic bias.

Formatted: Font color: Auto

Formatted: Font color: Auto

Deleted: ¶
[2] Rozowsky *et al.* (2011) *Mol Syst Biol.* 7(522)¶
[3] Van de Geijn *et al.* (2014) *bioRxiv*. doi: 10.1101/011221

Formatted: Font color: Red

Revised Manuscript	

-- Ref2.3 – Over-dispersion –

Reviewer Comment	My second major concern was regarding the binomial test to identify ASE. The authors begin their response by citing other papers that used such a test. I am not sure what it the argument presented here, especially since the authors proceed by acknowledging over-dispersion in their data. So, yes, other paper got it wrong in the past, but this is hardly a reason to perpetuate this mistake. As for their revised approach, estimating a global over-dispersion parameter is not effective. Removing some loci because of 'too much' over-dispersion is ad hoc and was not justified. But more importantly, there are at least 3 published methods now to identify ASE using models that estimate site-specific over-dispersion, account for mapping bias, and report p values based on permutation. Why not use one of those published methods?
Author Response	While we thank the reviewer for his/her comment, the purpose of the references is not to make any claims on the 'correctness' of the methods, but to point to the broader reality that there is currently a diversity of methods in the field, where there is no firm consensus on the 'right' approach. The fact that these publications are recent and peer-reviewed at influential journals indicates the plurality of the methods accepted by the community, each with their own advantages and limitations. For example, van de Geijn *et al.* presented a software that perform alignment to the human reference genome, accounts for allelic bias and allele-specific detection using the beta-binomial test to account for over-dispersion. However, it is not able to take into account indels and larger structural variants, which can be accommodated by the construction of personal genomes. In particular, we have utilized our approach in the 1000 Genomes Structural Variant group, whose manuscript has recently been peer-reviewed and accepted by *Nature.* Also, our revised approach estimates over-dispersion at two levels. An over-dispersion is estimated for each individual dataset to remove *entire datasets* that are deemed too over-dispersed and might result in higher number of false positives. After which, for each individual (and each transcription factor, TF, for ChIP-seq experiments), we pool the datasets and estimate the over-dispersion and apply this estimation to the beta-binomial test for each site in that individual (or TF). While the latter step have been employed extensively also in many recent software that detects allele-specific expression [1-5], we point out that our two-step serial procedure is novel and homogenizes the pooling by

Deleted: A global

Deleted: Hence,

Deleted: detection of

Deleted: variants using

Deleted: beta-binomial test is performed in a site-specific manner.

	removing datasets that are too over-dispersed in the first place. Perhaps we were not sufficiently clear, we have amended the manuscript to better reflect this.
Excerpt From Revised Manuscript	

Deleted:

Reviewer #3

-- Ref3.1 – General positive comment --

Reviewer Comment	The manuscript is much improved and the authors have sufficiently addressed the majority of my concerns. I have the following minor comments:
Author Response	We thank the reviewer for the thorough examination of the manuscript and we are pleased that the reviewer finds our improved manuscript satisfactory.

-- Ref3.2 – Include additional references --

Reviewer Comment	1) Imprinting discussion should reference recent imprinting paper from GTEx. Lappalainen in Genome Research. 2) Heritability analyses of ASE should reference Li, AJHG, 2014.
Author Response	We have included the references in the respective sections of the manuscript.
Excerpt From Revised Manuscript	Please refer to the 'Discussion' section and also the 'Results' section under "ASB and ASE Inheritance analyses using CEU trio". "It could also be a result of other epigenetic effects such as genomic imprinting where no variants are causal.[35]", where reference 35 is by the GTEx consortium and Baran *et al.* published in *Genome Research*. "The CEU trio is a well-studied family and with multiple ChIP-seq studies performed on different TFs. Previous studies have also presented allele-specific inheritance.[10,15,21]", where reference 21 is by Li *et al.* published in *American Journal of Human Genetics*.