OVERALL COMMENT

~~Dear Dr. Cho,~~

We want to thank the reviewers for recognizing the importance of our study and offering insightful comments. We have significantly revised the manuscript to address their comments. In particular, we have performed a major overhaul of the pipeline and corresponding analyses. Per reviewers #2 and #3's suggestions to include the newer advances in the field of allele-specific variant detection, we now use a beta-binomial test to account for the overdispersion properties of RNA-seq and ChIP-seq datasets in order to call ASE and ASB variants. Additionally, we implemented a first-pass filter in which we compute the overdispersion parameter for each of the 1,280 ChIP-seq and RNA-seq datasets prior to the pipeline to identify and exclude those that are highly overdispersed. Per reviewer #1's suggestion, we have uniformly re-called the peaks for 83 ChIP-seq datasets with a common peak caller. The peak calling was performed using the personal genomes for each of the 14 individuals with matching ChIP-seq data. To further build AlleleDB as a resource, we developed novel formalisms to call allele-specific genes and genomic elements. We also included new analyses and figures to illustrate the advantages of having results from allele-specific analyses obtained from a large number of genomes.

The specific reviewers' comments are further addressed below.

Reviewer #1

Reviewer's comment	This manuscript by Chen, et al. entitled "Allele-specific binding and expression: a uniform survey over many individuals and assays". This study is an exploration of the effects of genomic variation on expression of one of alleles and on transcription factor binding using previously generated RNAseq and ChIPseq datasets. The importance of this topic is timely and potentially significant. The manuscript is written in a relatively clear manner. While there is much to recommend this manuscript, several areas and questions need to be addressed to assist the reader to better understand or accept the findings. The major issues include:
Response	*We thank the reviewer for acknowledging the importance and timeliness of our study and for his/her thorough examination of our manuscript. We have made significant changes to specifically address the technical and analytical aspects of our manuscript for the reader.*

HEAD + OUTLINE

Reviewer's comment	1) It's not clear why Bowtie1 (DNA aligner) was used for aligning RNA-seq reads. This significantly reduces the number of mappable reads, and also can miss the allele-specific splicing. The use of a RNA (spliced) aligner might better be used for RNA-seq mapping.
Response	*While we do agree with the reviewer that using a RNA (spliced) aligner improves the number of mappable reads (<16%), ~~however~~, we have performed some additional analyses and found that the resultant AS calls between using a*

THIS SAID

GRAM

Deleted: helpful

Deleted: W

Deleted: remove

Deleted: , PeakSeq

Deleted: *might*

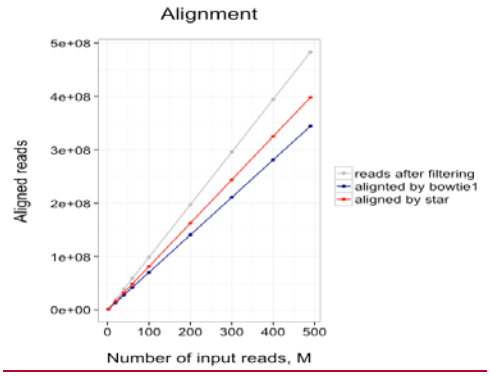
Deleted: *to demonstrate that the use of a RNA aligner increases the number of mappable reads by less than 16%.*

Move up

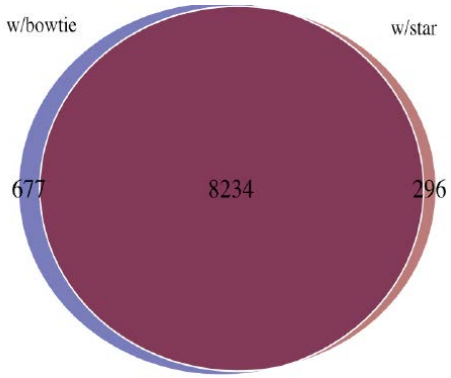
DNA and RNA aligner are very comparable (figure shown below). In panel a, we compare the number of mapped reads between using Bowtie1 and STAR at various coverage, sub-sampled from a RNA-seq dataset from NA12878 (the final data point is the entire dataset). On average the amount of increment in mappability by STAR is ≤ 16%. In panel b, we then detected ASE sites based on the entire dataset using Bowtie1 and STAR and found that the number of ASE calls between the two aligners are very similar. Thus, despite the increase in mappable reads, the effect of using a RNA aligner is relatively modest. ~~Moreover,~~ one of the main thrusts of our study is uniformity. Since a RNA aligner is not typically used in ChIP-seq alignment, we chose a consistent aligner in order to uniformly process both ChIP-seq and RNA-seq datasets.

[the figures will be updated shortly]

a.



b.



Deleted: *In addition, w*

Deleted: *to use*

Deleted: *re-*

Deleted: *the existing*

Deleted: *Hence, in light of these two points, we decided to use a DNA aligner for the alignment of both ChIP-seq and RNA-seq reads in our study.*

Deleted: ¶

Reviewer's comment	2) While calculating ASE for each of the SNV is straightforward, it's not clear how the ASEs of the genes are calculated. This would require combining ASE from multiple SNVs (and isoforms) of the same gene
Response	*We agree with the reviewer and indeed, the use of multiple SNVs of the same gene is the central premise of our enrichment analyses in determining the amount of allele-specific behavior for a gene. Perhaps our description in the method section was not sufficiently clear; we have clarified this in the 'Methods' section of the revised manuscript.*
Excerpt	*"Enrichment analyses were performed in two ways: 'collapsed' and 'expanded' (Figure 4b). In both cases, we aggregate ASB and ASE SNVs within a specific genomic element, such as a gene or an enhancer. We then use the Fisher's exact test to calculate the odds ratio and the hypergeometric p value…"*

Deleted: *totally*

Reviewer's comment	3) AS SNVs are counted if they fall within the called ChIPseq peaks. However, the peaks used were the ones used for each of the datasets studied with the exception of the McVicker's set. The lack of a uniform peak called for the calling of peaks will lead to significant variability due to the disparity in the results derived from various algorithms (i.e. some peaks cover more for the genome). In turn this can potentially inflate or diminish the number of sites evaluated. At least some evidence should be presented that the use of various peak callers will not significantly alter the number of variant with allele-specific phenotypes.
Response	*We agree with the reviewer that uniformity in peak calling is vital and have taken major steps in the revision to alleviate the concern. We addressed this by re-aligning reads from all the 276 ChIP-seq datasets. This is additionally performed in the context of each of the 14 pairs of personal haplotypes of each dataset, re-calling the peaks for each haplotype using a common peak-caller, PeakSeq, and then re-combining the peaks per dataset and for each personal diploid genome (haplotype pair).* WE MAKE THE UNI PEAKS AVAILABLE.
Excerpt	*"Peak regions are determined by first performing PeakSeq[60] for each of the personal haploid genome. Only a single read per strand per position is kept and duplicates removed. The fragment length is set to 200 bps. Peak calling is performed with default parameters and the final peak set for each transcription factor is identified at a false discovery rate of 5%. Finally, the coordinates of the peaks (based on the respective personal haploid genomes) are mapped to the reference genome and then finally being merged between the haploid genomes."*

Deleted: *strongly*

Reviewer's comment	4) While the study has as one of its strengths the development of a pipeline that can handle many (380) genomes with low coverage, it is unclear what the biological insights on this tour de force are other than the identification of 144K and 169K unvalidated ASEs and ASBs, respectively.
Response	*The reviewer's criticism is that he/she believes that the sole utility of detecting allelic events in many genomes is merely the identification of large numbers of ASE and ASB SNVs. We contend that it is precisely because of the identification of large numbers of ASE and ASB SNVs using multiple genomes, when appropriately processed, that more biological insights and uses can be developed.* *Our downstream analyses provide a window into some of these possibilities when many genomes are available. For instance, the enrichment analyses will not be feasible without a large number of ASE and ASB SNVs. It is important to appreciate that many SNVs are rare, thus the abundance and detection of rare allele-specific variants increase with many genomes. Previous studies mostly focus on a very small number of genomes. Hence, it is difficult to perform allele-specific analyses on rare variants from a single study. Yet, having large number of rare variants is important, especially when we want to define allele-specificity across a genomic region, as this requires combining information from multiple SNVs across the region, as already alluded to in the second comment of the same reviewer. In this case, further annotation and biological insights can then be provided for regions that seem to be more attuned to allele-specific behavior. In particular, these allelic regions form a useful genome annotation. Additionally, the aggregation of many genomes enables a more confident identification of common SNVs that have corroborating allele-specific evidence across multiple individuals, which in itself can serve as both a validation and a biological observation. Our study shows the value of combining many genomes across multiple studies in providing biological insights.* *In the revised manuscript, we have included a discussion of how having many genomes, when appropriately processed, can be useful, for instance, in enrichment analyses that aggregate rare variants in gene- or element-centric analyses. We have also included a new analysis to quantify the effects of having common allele-specific variants across multiple individuals. We now have two new figures: Figure 4, to illustrate the advantage of visualizing and having many genomes in validating common variants, and Figure 5, to capitalize on common variants in performing a population-aware enrichment analysis.*
Excerpt	*"An expanded population-aware approach emphasizes on common allele-specific variants found across multiple genomes to determine the allele-specificity of an element. An element is deemed more likely to be allele-specific if it is supported by more evidence of an allele-specific SNV occurring in*

Deleted: *It is important to appreciate that p*

Deleted: *The abundance and detection of allele-specific rare variants increases with many genomes and can be combined to provide the necessary statistical power to define allele-specificity across a genomic region, as already alluded to in the second comment of the same reviewer.*

Deleted: *In addition*

Deleted: *Thus i*

	multiple individuals. On the other hand, a collapsed approach treats each common and rare variant independently. An element that is deemed more allele-specific in this case, but not in the population-aware enrichment analysis, might mean that there are many more rare variants exhibiting allele-specific behavior."

Reviewer's comment	5) Minor point: No definition for CEU (Northern Europeans from Utah) RPB2, PAX5, etc.
Response	*All the definitions of the various human populations used for the 1000 Genomes Project were, in fact, already included in the original Methods section of the manuscript. They are intentionally omitted from the main text to enable readability. In the current revision, we have added a sentence referring the reader to the Methods section in the main text. We have also included the full names of the transcription factors (or short descriptions if no full names) such as RPB2 and PAX5, in the main text, at instances where we first mentioned them.*
Excerpt	*"The number of rare allele-specific SNVs (MAF ≤ 5%) is about two folds higher in the YRI than the other European sub-populations of comparable (CEU, FIN) or larger (TSI) population sizes (see Methods for full explanation of population abbreviations)."* *"Our visualization shows ASB loci from POL2 (RNA polymerase II largest subunit), RPB2 (RNA polymerase II second largest subunit) and MYC (also c-Myc, or v-myc avian myelocytomatosis viral oncogene homolog)…"*

Reviewer #2

Deleted: ¶
¶
Section Break (Next Page)

Reviewer's comment	This is an exceptionally naïve analysis of ASE and ASB patterns. The analysis to identify the ASE/ASB patterns is flawed, the statistical modeling is too basic, and the enrichment analysis is crude.
Response	*We thank the reviewer for the thorough examination of our manuscript. AlleleDB is, in fact, intended as a resource for ASB and ASE. Nonetheless, we have taken into account advances in the field and implemented more sophisticated changes to the statistical underpinnings of our pipeline. We have also included more rigorous analyses in our revised manuscript.*

Reviewer's comment	I have two concerns that, in my mind, are fatal flaws of the current analysis: First, mapping to a personal diploid genome indeed reduces the reference bias, but it does not eliminate the error associated with differences in mappability between the two alleles. In other words, the bias is gone, but the inflated variance due to mappability issues still persists. The only solution to date has been to map each allele separately and only retain reads that map uniquely at each allele, before the counting is done. This is a crucial aspect of the analysis presented in this paper and it must be addressed.
Response	*We agree with the reviewer that, in addition to building a personal diploid genome, mapping only unique reads to the individual haplotype or allele before the counting process is important. Our approach does encompass this and we have emphasized this point in the revised manuscript to better reflect this.*
Excerpt	*"Reads are aligned against each of the derived haploid genome (maternal/paternal genome for trio) using Bowtie 1.[56] When a read is aligned to the same locus, we only pick the alignment that map better to a haplotype. Otherwise, if a read is tied in alignment to both haplotypes, we discard the reads. No multi-mapping is allowed and only a maximum of 2 mismatches per alignment is permitted. "*

Reviewer's comment	Second, the ASE analysis was performed using a simple binomial test. This leads to a large number of falsely identified ASE patterns because of over dispersion in the data. Over dispersion in both RNA-seq and ChiP-seq data sets has been documented and commented on in a large number of papers. The correct analysis must use some strategy to estimate the over dispersion parameter and take it into account when testing for ASE.
Response	*While we thank the reviewer for his/her suggestion, we also note that many very recent publications have also used a binomial test in their detection of ASE and ASB SNVs. We list some of them here:* *Lappalainen, T. et al. (2013). Nature. 501(7468):506-11* *Kilpinen H. et al. (2013). Science. 342(6159):744-7* *Ding, Z. et al. (2014). PLoS Genet. 10(11):e1004798* *Dixon, JR. et al. (2015). Nature. 518(7539):331-6* *The GTEx Consortium. (2015). Science. 348(6235):648-60* *Nonetheless, we agree with and have taken to heart the reviewer's comment in order to provide a repository with 'cleaner' sets of ASE and ASB SNVs. As a result, we have significantly revamped our pipeline in terms of its statistical underpinnings and also re-processed all the 1,280 ChIP-seq and RNA-seq datasets. We now use a beta-binomial distribution to estimate the overdispersion behavior of each dataset and then use this as a filtering step to*

Deleted: *Reddy, T. et al. (2012). Genome Res. 22(5):860-9*¶

Deleted: ,

	exclude those that exhibit a greater overdispersion estimated from the allelic ratio distribution. As pointed out by the reviewer, overdispersion leads to a large number of false positives. This first step thus acts as a first pass in identifying datasets that are too overdispersed to start off with. Subsequently, we then use a beta-binomial test to estimate and account for overdispersion during our ASE and ASB detection. *The new Figure 1 and Methods section give a summary of our revised pipeline.*

Reviewer's comment	I have a few other major concerns: It is not entirely clear to me how the 'control SNVs' were defined. Are these simply cases where ASE was not detected? This seems a bit naïve to me; is the probability of including as a control a case where the null is rejected with a marginal p value is the same as a case where the null is rejected at, for example, P > 0.8? Also, I don't understand what it means to match the controls to the test cases by 'accessibility for statistical significance'. The terminology used is strange to me; is this a complicated way to say that you matched the power? If so, how was it done? If a cutoff for power was used, this would not result in true matching because the controls would probably be biases towards the lower threshold. More details on this analysis are needed.
Response	*We thank the reviewer for pointing out our definition of 'control SNVs'. Perhaps we were not sufficiently clear in our descriptions* *Our 'control SNVs' are not simply the complement of the allele-specific set nor are they defined by a single cutoff for power. We describe our control SNVs as "non-allele-specific accessible SNVs". Each set of accessible SNVs is defined by a read-depth cutoff. This cutoff is the minimum read depth that a heterozygous variant needs to possess to be able to even reach the significant p value threshold (accessible). Since the significant p value threshold is dataset-specific, there is a minimum read depth cutoff for each dataset such that each set of control SNVs generated is matched, in terms of the minimum number of reads, to each individual dataset's AS SNVs. In addition, our significant p values are also corrected using FDR estimated from an explicit computational simulation, which takes into account the number of reads at each heterozygous SNV when correcting the p values. This would also attenuate the bias towards having SNVs with lower read depth.* *We have provided more detailed explanations in both the main text and the Methods section to further clarify the definition.*
Excerpt	*"We define accessible SNVs as all heterozygous SNVs that exceed the minimum number of reads detectable statistically by the beta-binomial test for each dataset, including both allele-specific and non-allele-specific SNVs. This*

Deleted: *Previously in our manuscript, we described o*

Deleted: *as "a set of heterozygous, non-allele-specific but accessible SNVs." We also described accessible SNVs as heterozygous SNVs that in principle, can be detected as allele-specific, because they meet the minimum read depth requirement, which is computed relative to the size of the ChIP-seq or RNA-seq dataset. Thus*

Deleted: *, e*

Deleted: *control*

Deleted:

Deleted: *matched with the corresponding set of allele-specific SNVs for each dataset. Perhaps we were not sufficiently clear in our descriptions.*¶
¶
We have provided more detailed explanations in both the main text and the Methods section to further clarify how we obtain the 'control' SNVs.¶

	is an additional criterion imposed on top of the minimum threshold of 6 reads. The minimum number of reads thus varies with the pooled size (coverage) of the ChIP-seq or RNA-seq dataset. Thus, the accessible SNVs are dataset-specific; they are determined for each pooled ChIP-seq (grouped by individual and TF, not by study) or RNA-seq dataset (grouped by individual).… Considering an extreme allelic imbalance case where all the reads are found on one allele (all successes or all failures, i.e. allelic ratio is 0 or 1), this minimum N can be obtained from a table of expected two-tailed beta-binomial probability density function, such that accessible SNVs are all SNVs with number of reads, n = max(6,N). By considering only the cases with the largest effect size, we underestimate the number of accessible SNVs and this provides a conservative approximation of the statistical significance of the enrichment (or depletion). 'Control' SNVs are subsequently derived from accessible SNVs that are non-allele-specific, i.e. they are the set of accessible SNVs that has excluded the respective ASB or ASE SNVs for each dataset."

Formatted: Font: Italic

Deleted: …

Reviewer's comment	The heritability analysis (using a single trio...) is confusing to me. It is, in a sense, a corrupted version of what is typically considered heritability analysis. The comment that analysis was performed separately for single parent - child pairs in order to 'maximize statistics' is entirely unclear, and in general, the entire analysis seems ad hoc and does not result in what we typically consider a measure of heritability.
Response	*We thank for the reviewer for pointing out that this is not the measure of heritability of what we typically conceive. In our original manuscript, we did recognize that this is an* ***adapted*** *version of the conventional heritability analysis in population genetics. Even though we do not have a population of trios, there is still information about heritability that can be gleaned from a single trio. In fact, many other studies have also adapted their analyses in similar ways to show inheritance; we provide the citations for some of them:* *McDaniell, R. et al. (2010). Science. 328(5975):235-9* *Kasowski, M. et al. (2013). Science. 342(6159):750-2* *Kilpinen, H. et al. (2013). Science. 342(6159):744-7* *We have re-worded it to better reflect this.*
Excerpt	*"The conventional measure of 'heritability' allows the estimation of (additive) genetic contribution to a certain trait. The population genetics definition of 'heritability' in a parent-offspring setting is described by the slope, β, of a regression (Y=βX + α), with the dependent variable being the child's trait value (Y) and the independent variable (X) being the average trait values of the father and the mother ('midparent').[60] This is a population-based measure typically performed on a large set of trios for a particular trait (e.g. height) and β is not necessarily bound between 0 and 1… Given that we have only a*

	single trio, we adapt the typical definition of 'heritability' to quantify allele-specific inheritance for each TF."

Reviewer's comment	The analysis of functional annotation of the identified SNVs implicitly makes the assumption that these are causal variants. This is not the case, especially for the ASE, where the typed SNV is most likely in LD with the causal regulatory locus. As has been previously shown, the causal assumption is a poor one when ASB is considered as well.
Response	*We are in total agreement with the reviewer that these AS SNVs are not necessarily causal and have never intended to imply causality in our writing. We will re-word the manuscript to better reflect this.*
Excerpt	*"…it is important to note that the AS SNVs detected are not causal. The resultant allelic difference in gene expression and binding can be due to another undetected causal variant that has a strong linkage disequilibrium with the detected variant or, it could be due to a group of variants that act collectively to give the resultant allelic expression or binding.[37]"*

Reviewer's comment	Minor comments: From the intro: "AS variants can be detected regardless of their population allele frequencies." - This is actually not true in practice. ASE in intermediate frequency alleles are still easier to detect in the entire population because one can estimate the over dispersion parameter more precisely.
Response	*The sentence was not meant to refer to the ease in detection, but rather the range, of allele frequencies that can be detected in allele-specific variants. We will re-word the manuscript to better reflect our intention.*
Excerpt	*"Using each allele in a diploid genome as a perfectly matched control for the other allele, allele-specific variants can be detected even at low population allele frequencies."*

Reviewer's comment	I applaud the author's computational competence, but is the sentence, in the Results, on the amount of CPU time needed for the analysis really adds to the narrative? I think such details should be reported in the methods section.
Response	*We agree with the reviewer and have moved the sentence from the main text to the Methods section.*

Reviewer #3

Reviewer's comment	This manuscript provides analysis of allele-specific binding (ASB) and expression (ASE) data for many individuals and assays. The authors compile this information in a database and further focus on describing the properties of transcription factors and genes which are enriched or depleted in ASB and ASE. I have the following major comments:
Response	*We thank the reviewer for the thorough examination of the manuscript.*

Reviewer's comment	1) The methods do not take into account the known statistical challenges of calling ASB or ASE and current advances in this area. The authors simply resort to a simple binomial test setting a minimum depth of 6 reads. It should be obvious that even with an FDR of 5 or 10% that low depth sites are going to be enriched in significant sites. They should plot depth by percentage of significant sites as a sanity check.
Response	*We agree with the reviewer's comment that using a binomial test, even with multiple hypothesis correction, significant sites will be enriched with low read depth sites. Hence, we have built a more sophisticated statistical model using a beta-binomial test to account for overdispersion. As mentioned also in response to reviewer #2's comment, we have now re-processed all the 1,280 ChIP-seq and RNA-seq datasets by first estimating the overdispersion parameter of each dataset. We then filter out those that exhibit a greater overdispersion (overdispersion parameter, $\rho > 0.3$ for ChIP-seq datasets and $\rho > 0.125$ for RNA-seq datasets) estimated from the allelic ratio distribution. Finally, we use a beta-binomial test to estimate and account for overdispersion during our ASE and ASB detection. The new Figure 1 and Methods section give a summary of our revised pipeline.* *Additionally, we provide a new supplementary figure (Supplementary Figure 2) to show that the percentage of AS SNVs called by our pipeline, with respect to the accessible SNVs is consistent across various read depths.*

Reviewer's comment	2) Many of the datasets they use have ASB and ASE already called on them. The advantage of their approach to existing data is not compared. How different are there ASB/ASE calls to gEUVADIS or McVicker et al? One would expect that even if these are not online, the methods should minimally be compared as there will be an expectation that the AlleleDB database would yield similar quality results.
Response	*As also noted by the first reviewer, it is important to appreciate that there is a fair amount of heterogeneity in the parameters and tools used in other studies, for instance the peak callers, aligners, detection strategies and reference*

Deleted: ¶
¶

genomes. Hence, there will naturally be great disparities and variability when comparing AlleleDB with results from the other studies. In fact, AlleleDB is motivated by the need to harmonize and uniformly reprocess all the datasets for allele-specific detection instead of simply combining the results from these various studies. We have included a new supplementary table (Supplementary Table 1) to show some of the differences in methods used across various studies.

In response to the reviewer, we also did attempt to compare our AS SNVs and those called, if any, in the studies that are used in AlleleDB. Lalonde et al. does not call allele-specific variants. We were unable to find SNV lists for the study by Kasowski et al. The ASB and ASE analyses performed by the ENCODE consortium were done with our pipeline but using the original binomial test. The foci of Montgomery et al., Pickrell et. al, the gEUVADIS consortium and McVicker et al. were not AS analyses. They primarily utilized ASE variants for refinement of their detection of eQTLs or candidate regulatory variants. Unfortunately, they did not provide any ASE SNV lists online for comparison. The only study we were able to make at least a minimal comparison was the study by Kilpinen et al., who provided the per-individual mean number for ASE, ASB and accessible sites called in their RNA-seq and some ChIP-seq datasets. Their analyses were for eight unrelated CEU individuals.

	Kilpinen			*AlleleDB*		
	ACC	*ASB/ASE*	*Proportion*	*ACC*	*ASB/ASE*	*Proportion*
POL2	*9254*	*525*	*0.06*	*2,817*	*199*	*0.07*
PU.1	*930*	*154*	*0.17*	*83*	*3*	*0.04*
ASE	*4061*	*190*	*0.05*	*1,266*	*14,923*	*0.08*

We note that while the proportions of accessible SNVs being AS SNVs are very similar, the mean number of SNVs called per person is different.

Reviewer's comment	3) How is AS inheritance at binding sites not a universal phenomenon of TFs like MYC or RPB2? This seems like a pretty bold assertion. Isn't it more likely that there is something wrong with your method for these sites? Low read depth, poor antibody efficiencies, non-specificity of binding profiles, etc. Why make a biological claim before you have exhausted technical sources of error.
Response	*We agree with the reviewer and have removed the assertion from the revised manuscript.*

Deleted: *with*

Deleted: ¶ ¶

Deleted: *between*

Formatted: Font: 12 pt

Formatted Table

Formatted: Centered

Reviewer's comment	4) The phrasing of the paper suggests that ASE sites are actually causal. For instance, the relative numbers of sites with ASE and ASB are compared. What does this even mean? How are these even directly comparable? An ASE effect suggests the genes are imbalanced. This imbalance could be due to multiple causal ASB events. Furthermore, the ASE site is not causal. It is only indicating the potential presence of a causal regulatory variant.
Response	*We are in total agreement with the reviewer that these sites are not necessarily causal and have never intended to imply causality. We have re-worded the manuscript to make this point explicit.* *The AS sites are used as markers for potential allele-specificity in the genome. By visualizing ASE and ASB SNVs side by side, we had meant to provide some context and possibly set the stage for some biological insights. We have included a new figure, Figure 4, to show how concurrent visualization of ASB and ASE SNVs in a population of individuals for the gene ZNF331 can potentially provide some biological insights into the allele-specific properties of the gene and its specific sub-regions (exons).*
Excerpt	*"However, even with lower number of false positives, it is important to note that the AS SNVs detected are not causal. The resultant allelic difference in gene expression and binding can be due to another undetected causal variant that has a strong linkage disequilibrium with the detected variant or, it could be due to a group of variants that act collectively to give the resultant allelic expression or binding.[37]"*

Reviewer's comment	5) The authors don't seem to understand why a gene would be depleted in allele-specific behavior. Is there expectation that allele-specific behavior should influence all genes equally? Furthermore, I worry that depth might be more deterministic of which genes are enriched or depleted.
Response	*We agree with the reviewer that we do not expect allele-specific behavior to influence all genes, or even sub-regions within a single gene, equally. Indeed, genomic regions with lower read depth will not have enough power to detect AS behavior. However, for regions with sufficient read depth, our implementation of an explicit FDR simulation takes into account the respective read depth of each heterozygous SNV in order to determine whether a SNV is allele-specific or not. Additionally, to determine whether a gene (or genomic element) is enriched or depleted in allele-specific behavior, we incorporated the use of 'control' SNVs that are well-matched in power to the allele-specific SNVs.*

Deleted: ¶
¶

Formatted: Font: Bold

Formatted: Font: Italic

Reviewer's comment	6) Why would ASB be under less selective constraint that ASE SNVs? This probably only has to do with the background of being in a gene versus being in a non-coding region. Again ASE SNVs are not causal, so what is selection acting on. Figure 4 makes no sense. Beyond this, I don't even see a difference between the ASB +/- sites at low frequencies.
Response	*We had intended to report the rare variant load in ASE and ASB sites only as a general observation that might be suggestive, and not indicative, of less natural selection in ASE sites. We have re-worked the main text to better reflect this.* *We agree with the reviewer that it is reasonable that the background of being in a gene versus a non-coding region can contribute to the higher enrichment of rare variants in ASE SNVs than ASB SNVs. Hence, we have removed the sentence that makes this ASE-ASB comparison.* *Furthermore, inspired by the reviewer's comment, we have constrained our analysis to only the coding DNA sequences and the transcription factor binding motifs and re-calculate the rare variant loads in ASE, accessible non-ASE, ASB and accessible non-ASB sites. Indeed, we do not find any significant enrichment of ASB sites, compared to accessible non-ASB sites.*
Excerpt	*"Our results in Figure 6 show a statistically significant lower enrichment of rare variants in ASE SNVs as compared to non-ASE SNVs (Fisher's exact test odds ratio=0.2, p<2.2e-16) but statistically insignificant higher enrichment of rare variants in non-ASB SNVs than ASB SNVs (Fisher's exact test odds ratio=1.4, p=0.08). This observation seems to suggest that ASE variants may be under weaker selection than non-ASE variants, which can be a result of accommodating varying levels of gene expression across individuals."*

Reviewer's comment	7) Do the authors have any insight into how well their calls replicate and then their replication at various depths.
Response	*In response to the reviewer's suggestion, we have included two new analyses and corresponding figures as supplementary materials.* *The first analysis investigates the replication of AS calls between two technical replicates by comparing their AS calls. We obtained two equally-sized subsets of 200M ('M' denotes 'million of reads'), which are randomly sampled from a pooled RNA-seq dataset of NA12878. We show that the calls are almost identical, demonstrating that our calls replicate very well (Supplementary Figure 3).* *The second analysis investigates the replication of AS calls at increasing read depths randomly subsampled from the same pooled RNA-seq dataset of*

DID WE PUT IN DISC.

Deleted: *We had intended to report the rare variant load in ASE and ASB sites only as a general observation that might be suggestive, and not indicative, of less natural selection in ASE sites. We have re-worked the main text to better reflect this.*¶
¶

Deleted: . ¶
¶
The first one investigates how well our pipeline recapitulates, by comparing the detected AS SNVs from two biological replicates from the ENCODE RNA-seq datasets of the same cell line, NA12878. We observe that the number of AS SNVs detected is highly dependent not only on the size of the dataset (number of reads), but also on how overdispersed the dataset is. Between two biological replicates, when the one is more overdispersed, it typically gives a higher number of AS sites, even when the other replicate is of a higher read depth.¶
¶
The second one investigates the reproducibility of the AS variant calls from pseudo-datasets of various read depths taken from the same ENCODE RNA-seq NA12878 dataset. The majority of the AS sites overlap at each read depth, showing how well the calls replicate. [need to check on these…]

	NA12878, namely subsets of 100M, 200M, 300M, 400M and 490M. Each subsample is a direct subset of the other larger subsamples, e.g. 100M is a direct subset of all the other sets. We then re-align each pool of reads to NA12878 personal genome and re-call the AS sites for each subset. Supplementary figure 4 shows that the majority of the AS sites (>80%) are consistent in all 5 subsets and the 3 larger subsets (300M, 400M and 490M) share most of their AS sites as well.