Reviewer #1 (Remarks to the Author):

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:

[[ ==MG-note==> ]]thank reviewer

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: We have some preliminary analyses, demonstrating that the use of spliced aligners does not increase the number of mappable reads by more than 20%. Thus, the effect of using a splice aligner might be modest. We can include these analyses in the revised manuscript.

[[ ==MG-note==> ]]consistency betw chip & rna

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 how allele-specific a gene is in ASE and ASB. Perhaps our description in the method section was not sufficiently clear; we will clarify the text in the revised manuscript.

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 we will address this in the revised manuscript.[[ ==MG-note==> ]] didn’t we do this?

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, 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, enrichment analyses will not be feasible without a large number of ASE and ASB SNVs. It is important to appreciate that previous studies mostly focus on a very small number of genomes. 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. In this case, further annotation and biological insights can be provided for regions that seem to be more attuned to allele-specific behavior. In addition, 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 a validation and a biological observation.

We will better address this concern in our revised manuscript.

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 are in fact already included in the Methods section. They are intentionally omitted from the main text to enable readability. We will include them, along with the definitions of the transcription factors (if any) such as RPB2 and PAX5, in the main text of the revised manuscript where we first mention them.

Reviewer #2 (Remarks to the Author):

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: AlleleDB is, in fact, intended as a resource for ASB and ASE. Nonetheless, we will include more rigorous analyses and make the statistical modeling more sophisticated in our revision.

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 will emphasize in the revised manuscript to better reflect this.

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: Many very recent publications have also used a simple binomial test in their detection of ASE and ASB SNVs (we list some of them below). However, we agree with the reviewer’s comment and will build a more sophisticated statistical model to accommodate overdispersion. [[ ==MG-note==> ]] more on this

McDaniell, R. *et al.* (2010). *Science*. 328(5975):235-9

Montgomery, S. *et al.* (2011). *PLoS Genet*. Jul;7(7):e1002144

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

Lappalainen, T*. et al.* (2013). *Nature*. 501(7468):506-11

Ding, Z. *et al.* (2014). *PLoS Genet.* 10(11):e1004798

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: In our manuscript, we describe our ‘control SNVs’ as a set of heterozygous, non-allele-specific but ‘accessible’ SNVs. We also describe ‘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, each set of control SNVs is matched with the corresponding set of allele-specific SNVs for each dataset.

We will provide more detailed explanation to further clarify how we obtain the ‘control’ SNVs.

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: In our manuscript, we did recognize that this is an adapted version of the conventional heritability analysis in population genetics, since we do not have a population of trios. In fact, many other studies have adapted the analysis in similar ways to show inheritance; we provide the citations for some of them below.

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

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 causal and have never intended to imply causality in our writing. We will re-word the manuscript to better reflect this.

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.

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 will move the sentence from the main text to the Methods section.

Reviewer #3 (Remarks to the Author):

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:

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: As mentioned, we agree with the reviewer’s comment and will build a more sophisticated statistical model to accommodate overdispersion.

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 genomes. Hence, there will naturally be great disparities and variability when comparing AlleleDB with results with 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.

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 will remove the assertion from the revised manuscript.

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 causal and have never intended to imply causality. 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 will re-work the manuscript to better reflect this.

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: On the contrary, our (null) expectation is a ‘balanced’ (biallelic) behavior in expression or binding. Hence, a statistically significant depletion of allele-specific behavior in a genomic element can be interpreted as more ‘balanced’. We had mentioned FHIT as an example of such a more balanced behavior, but could not understand the significance in relation to its role in cancer. To prevent further confusion, we will exclude citing the example in our revised manuscript. [[ ==MG-note==> ]]shouldn’t we discuss depth issue ?

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 agree with the reviewer that a substantial contribution to the comparison of selective constraints based on the enrichment of rare variants between ASE and ASB might be due to the background of being in a gene versus a non-coding region respectively. We will remove the only sentence making the ASE-ASB comparison. In addition, to attenuate such a background effect, we will constrain our analysis to only the coding DNA sequence and the transcription factor binding motifs in the revised manuscript.

The main aim of this analysis has been to investigate how selective constraints affect allele-specific and non-allele-specific sites for expression and binding. Again, we are not implying any causality. If natural selection is acting equally on non-causal ASE and non-ASE SNVs, we would not expect any significant difference between them. The fact that we are observing a significant difference indicates that a considerable proportion of ASE sites is directing the selection. In this case, the ASE SNVs seemed to experience less selective constraints than the non-ASE SNVs, suggesting that these regulatory sites can accommodate more variability [[ ==MG-note==> ]] reword this interpretation

7) Do the authors have any insight into how well their calls replicate and then their replication at various depths.

Response: We can provide some analyses to address the concerns.