Reviewer’s report:

The manuscript by Furey et al. compares performance of the software (AA-ALIGNER) in identifying allelic imbalance for quantitative short DNA read data (e.g. ChIP-seq) between the case where genotypes are known completely, the case where partial genotype information is available and the case where alleles in the population are known, but the specific genotype for the samples analyzed is unknown. This is a unique contribution to a field that is growing in importance and there is a lot of demand for new analytical approaches to deal with challenges specific to this type of data. The manuscript is focused on dealing with map bias (one source of bias in AI) when there is little or no information about genotype. Other challenges, for example different statistical approaches (e.g. FDR) and other sources of bias are not addressed, but are discussed as areas requiring further consideration. The major conclusion is that results from analyses using partial genotypes or using common alleles (and no specific genotypes) are not substantially worse than using the complete genotypes, which is promising. An additional interesting finding was that predicting genetic variants directly from the sequence data being analyzed gave very different results from using preexisting genotype information. Overall, this is a focused manuscript presenting a novel approach to missing genotype information, which also makes some very nice points with respect to strategies for dealing with map bias. However, there are some issues with clarity as outlined below and statistical caveats are a concern.

Major Compulsory Revisions

1) The finding that poor SNP calls are associated with false positives for imbalance is not novel (e.g. Heap et al. 2010; León-Novelo et al. 2014), but the high false positive rate found in this study is significant because it emphasizes the particular importance of this issue with respect to making calls from this type of data in the absence of other genotype information. However, the methodology for calling heterozygous sites from the sequence data needs to be clarified. Are these calls included in the part of the pipeline indicated by the box labeled 1 (Create Custom Reference Genome-described briefly in lines 132-134)? Could poor quality heterozygous site calls be remedied by adjusting the approach in the pipeline or are they a property of the type of sequence data analyzed? It was also unclear if these are false positives in the new variant calls or false negatives in the reference sequence?
2) The validation study is appropriate for determining if there is differential binding at a particular site, but it is incomplete with respect to testing the analytical approach. Only sites with inferred AI were tested, thus the frequency with which differences in protein binding would be detected when no imbalance is inferred is unknown. To conclude that AA-aligner is detecting AI correctly, it would need to be shown that protein binding is concordant with detecting/not detecting AI.

Minor Essential Revisions

3) Only mapping bias is substantially discussed in the background, other sources of bias are not. Other sources of bias could contribute to false positives, as potentially indicated by cases where AI is detected in the analysis, but no protein binding differences are detected. This should be discussed.

Discretionary Revisions and Comments

4) The bioinformatics overall are well described and executed, the statistical approach was not a focus of the paper and was very basic. While caveats are addressed in the discussion, I found the lack of consideration of statistical concerns distracting. It might be worthwhile to prominently point out that the software can be used with whatever statistical approach is most appropriate for the user.

5) Are the input control sequence data available? Perhaps there is some reason not to, but it seems to me that with appropriate filtering detection of AI in these data would be an appropriate negative control for the overall analysis. This would be similar to the approach used for peak calls themselves.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.