RESPONSE TO REVIEWERS FOR "IDENTIFICATION OF ENRICHED REGIONS IN CHIP-SEQ EXPERIMENTS USING A MAPPABILITY CORRECTED MULTISCALE SIGNAL PROCESSING FRAMEWORK"

RESPONSE LETTER

-- Ref1 – General Remarks --

Reviewer Comment	This manuscript presents an improved method for identifying significantly enriched regions from ChIP-seq experiments, called MUSIC. The current methods for analysis of ChIP-seq data suffer from 2 major limitations. First, none of them account for repetitive DNA sequence or other regions that can affect ChIP-seq peak calling due to mappability issues. Second most of the available programs struggle with accurately calling peaks for broadly distributed histones, such as H3K36me3 and H3K4me1. The authors present compelling evidence to show that MUSIC offers a solution to both of these issues, and is a significant improvement over most previous methods. The paper is well written and the data are presented clearly and the conclusions are well supported by the results. I only have minor comments, which can be addressed in a revised version. Overall, I'm very pleased with this manuscript and I look forward to using MUSIC as my preferred ChIP-seq analysis package.
Author Response	We thank the reviewer for the constructive comments. We address the comments point by point below.
Excerpt From Revised Manuscript	

-- Ref1.1 – Smoothed Signal Track Outputs --

Reviewer Comment	MUSIC outputs a list of peaks and corresponding enrichment scores. This is fine, but it would be very helpful if the program also outputted a track of the processed (smoothed) data that could be loaded onto a Browser, so that the degree of smoothing could be visualized. A "before" and "after" version of these tracks would be particularly useful.
Author Response	We agree with the referee that visualization of the data is an important aspect of assessing the results and would make MUSIC much more useful. For this, we added an option to MUSIC to dump the smoothed signal tracks (for each track) in bedGraph format with the output ER's. These can easily be converted to bigwig file format for viewing the files locally or

	uploading to other genome browsers like UCSC Genome Browser. We updated the manuscript to present that the smoothed tracks are saved in bedGraph files.
Excerpt From Revised Manuscript	

-- Ref1.2 – Troughs in the Signal --

Reviewer Comment	Second, I'm struggling a bit to assess how much of the smoothing removes details in a given signal that can sometimes be informative. For example, when ChIP-seq data is plotted in aggregate, it's clear that transcription factors often bind in the "trough" of a bimodal histone-peak, corresponding to the nucleosome free region. Does MUSIC smooth this out? How often are such regions called as a single histone-peak with MUSIC? Do the other programs (MACS, etc) tend to split these regions into 2 called peaks? A comparative analysis would be helpful.
Author Response	The referee raised a very important point. We agree that the troughs can be very important since they may mark the nucleosome free regions where the regulatory factors (like TFs) can interact with DNA. The identification of these troughs are especially hard since the decrease in signal can be related to either real decrease in histone modification levels, or simply a decrease in the mappability. Our inspection of the ChIP-Seq signal profiles, however, shows that the dips in the signal is very frequently caused by the decrease in mappability. In fact, it is very hard to distinguish between the non-mappable troughs and mappable troughs. Therefore, MUSIC currently merges these regions together because it aims at identifying the ER as a complete unit, for example, for H3K36me3 marks, the whole gene body. The other peak callers do not generally merge these regions and tend to oversegment the signals. In order to quantify the amount of decrease in the signal with respect to mappability versus the real nucleosome-free regions, we concentrated on H3K36me3 signal and identified the regions that were identified in ERs identified by MUSIC not identified in MACS ERs. Then we performed aggregation of the multi-mappability signal to assess if there is significant increase in multi-mappability signal compared to a set of control regions and observed that there is a very significant difference in the mappability (See Supplementary Figure S6). This result suggests that the regions that MUSIC merges (but MACS does not) are significantly enriched in regions that have low mappability. However, since we believe that this is a very important point for analysis of punctate histone marks like H3K27ac and H3K4me3,

	we also added a functionality to punctate peak calling mode of MUSIC to identify the smallest dip in the signal in each ER that has good mappability (at the level of exonic mappability) requirement in each identified ER. We believe this is a valuable addition to the functionalities of MUSIC. We added a paragraph in the paper summarizing our point above.
Excerpt From Revised Manuscript	

-- Ref 2 – General Remarks --

Reviewer Comment	Harmanci et al. present a new algorithm called MUSIC to identify enriched regions in the ChIP-seq experiments. MUSIC aims to correct the systematic noise introduced by non-uniform read mappability and devices a smoothing strategy to merge fragmented enriched regions in ChIP-seq experiments. Furthermore, they applied MUSIC at multiple length scales to automatically consider both the narrow and broad peaks. They compared the performance of MUSIC with several peak-finding algorithms on H3K36me3. Using RNA-seq signals as a gold standard, they showed that MUSIC achieved better F-measures than the existing methods. In particular, they investigated the RNA ploymerase II binding ChIP-seq data and showed distinct expressions of genes with different length scale of binding peaks. For a computational method, it provides some new features such as smoothing peaks using read mappability and considering multiple length scales. The major concern is that the performance assessment is not as thorough as it can be. Some details of the parameters set should be provided as well.
Author Response	We thank the referee for constructive comments. The referee's main concern is that MUSIC is not compared to other methods in terms of more punctate events such as H3K4me3 and TF's. We updated the benchmark section with the requested comparisons in terms of addition of methods and more datasets using comparison with different metrics. We first would like to note that we chose the other methods in the benchmark because they all have broad ER identification modes implemented in them. We address these concerns below point by point. We also would like to emphasize an important point about MUSIC. MUSIC is specifically designed to process the ChIP-Seq signal using a multiscale approach and the multiscale approach is utilized best for the signals that show ERs at large spectrum of

	length scales. [[Punctate events are not affected much by the non-uniform mappability and multiscale comparisons]]
Excerpt From Revised Manuscript	

-- Ref 2.1 – ChIP/Input Normalization Factor Computation --

Reviewer Comment	In page 4, "The MUSIC computes a scaling factor using linear regression between the ChIP and control signal profiles. The slope of the regression is used as normalization factor for control." It is unclear how this regression was done. A brief explanation would be helpful for readers to understand how this was done.
Author Response	We agree with the reviewer that this is an important point in the paper that needs to be clarified. For this, we added Section 4.1 in Methods Section to explain the computation of input normalization factor in full detail.
Excerpt From Revised Manuscript	

-- Ref 2.2 – Parameter Selection --

Reviewer Comment	In page 4, how are the parameters of l(start) and l(end) determined? Also, how are the default values of gamma and tau determined? In Methods, it is noted that these parameters are set by trial and error. What is this "trial and error" procedure? How to judge what parameter values perform better? Is there any general guidance of choosing the values? Does the choice impact the results?
Author Response	We agree with the reviewer that the selection of parameters is an essential part of MUSIC workflow and should be clarified. For selecting l(begin) and l(end) for broad marks (H3K36me3, H3K27me3), we utilize the fact a median filter of length l removes all the features of length smaller than l/2 within it (See Section 4.11 and Supplementary Figure S3). Given the distribution of gene-gene distances and gene lengths, we used this fact to minimize over-merging and over-segmentation of the signal while detecting gene-wide ERs. For punctate marks like H3K4me3 the enrichments are expected at scales of at most several kbs thus we set l(end) to 2000bp's. For these we set l(start) to be 100bp's so as not to miss any small ERs. For most transcription factors, there is almost no concept of multiscale processing since the binding is assumed to happen at certain point. For TFs, we use l(start)=100bps and l(end)=200bps. tau is estimated (for all the modes) as the threshold that satisfies 5% false positive rate under the null model that the reads are distributed with Poisson distribution with mean estimated from the

	1 megabase windows over the genome (See Section 4.6). Thus it is not a free parameter. gamma is the threshold on the smoothing statistic that is introduced to avoid overmerging of the ERs by oversmoothing of the signal in the decomposition. In principle, this oversmoothing test is a proxy for a statistical test that would compare the distribution of signal in the regions in smaller scales that get merged regions in the higher scales and determine if there is significant shift in the signal levels: We expect that as the signal is smoothed, it will diffuse out and become smaller. We realized, however, that this would be computationally too costly and implemented the test with thresholding the simple test statistic presented in Section 4.5. For illustrating how different gamma values change the smoothing levels, we plotted the distribution of p-values of regions with respect to the smoothing statistics for each SSER in a large scale decomposition (See Supp Fig. 5). Following these, we decided gamma=4 (where we capture around 90% of the SSERs) is a reasonable value for thresholding the smoothing statistic. For selecting sigma, interscale multiplicative factor, we evaluated different values and observed that above 2, MUSIC starts missing too many SSERs. Since sigma has to be greater than 1, we chose 1.5 as a good value for sigma. We updated the Section 4.11 to more clearly explain the selection procedure for these parameters.
Excerpt From Revised Manuscript	

-- Ref 2.3 –Comparison to TFs and DHSs--

Reviewer Comment	When evaluating the performance of MUSIC, the authors selected H3K36me3 and used RNA-seq signals as the gold standard. Clearly MUSIC outperformed the other methods. This is not completely unexpected because MUSIC tends to identify long enriched regions. What about a comparison on signals with narrow peaks of TFs and DHS? There are many TF ChIP-seq available and their motifs are also known. It would be interesting to see whether MUSIC recovers peaks of these TFs containing the motifs.
Author Response	The referee brought up an important point. We added a new benchmarking section to the manuscript (Sections 2.2.2 and 2.2.3) for comparing the methods with respect to their accuracies for TFs and we also included H3K4me3 in the comparisons.
Excerpt From Revised Manuscript	

-- Ref 2.4 –Zinba, F-seq, DFilter--

Reviewer Comment	There are several recently developed methods that should be included for comparison, such as Zinba, F-seq and DFilter. These methods also provide flexibility of detecting peaks at different length.
Author Response	We thank the reviewer for pointing out these methods. We added the mentioned methods (ZINBA, F-Seq, and DFilter) in our ER identification comparisons, updated the results, and highlighted the manuscript. We ran all the tools (including other programs) in the broad ER identification mode. We added one section to the manuscript (Section 4.13) on the details of the options used to run the other programs in the benchmarking. The results show that MUSIC performs favorably compared to other methods for broad marks.
Excerpt From Revised Manuscript	

-- Ref 2.5 –H3K4me3 and TF Comparisons--

Reviewer Comment	F-measure is only one simple statistics and a better performance on H3K36me3 alone is not sufficient to demonstrate that MUSIC is superior to other methods. There are additional criteria for performance comparison such as comparing active promoters overlapping with H3K4me3 peaks called by the methods, percentage of peaks located within 50bp of motifs for TFs et al.. A thorough comparison can be found in the DFilter and Zinbe papers.
Author Response	We thank the reviewer for the suggested comparisons. We updated the benchmark section with comparisons of H3K4me3 and (enrichment of active TSS'es around identified H3K4me3 peaks) and TF experiments (enrichment of motif around 150 bp as it was used in ZINBA paper and several previous papers) of the identified peak summit). The results suggest that MUSIC performs favorably for H3K4me3 peaks and comparable with the best performing methods for motif enrichment.
Excerpt From Revised Manuscript	

-- Ref 2.6 –Study by Knijnenburg *et al.*--

Reviewer Comment	There was a recent paper published by Knijnenburg *et al.* Nature Methods, 11, 689-694, 2014 that provides a multiscale representation of genomic signals. Can the authors comment on that study and compare the multiscale features of MUSIC with Knijnenburg study?
Author Response	We thank the reviewer for pointing this relevant paper, which was published very close to our initial submission. The *Knijnenburg study* utilizes a Gaussian filtering based multiscale decomposition to summarize and visualize the genomic signals. Although the

	authors present a pruning approach for identification of peaks, MSR is not designed primarily for identification of ERs, thus we cannot directly compare MSR with MUSIC, i.e.: ".. Consistent with our principal goal of investigating the ability of the MSR method to enable multiscale comparisons between heterogeneous genomic signals (rather than the specific data reduction step of peak-calling), we used the 'unpruned' MSR for all subsequent analyses. ..." There is no discussion on selection of scales and we are afraid that the pruning methodology proposed in the study will cause extensive overmerging of the ER. We also would like to point out that multiscale feature detection is a very broad area of research in electrical engineering literature. The Gaussian scale space approach utilized by *Knijnenburg et al* is a linear approach and MUSIC utilizes a novel median filtering based non-linear approach. We believe there is significant difference in two decomposition methodologies. More specifically in terms of methodology, MUSIC is different from *Knijnenburg et al*'s approach in three basic aspects: First, MUSIC performs the multi-mappability correction before performing the multiscale decomposition. We observed that this increases accuracy of identified ERs significantly as shown in Figure 3e. *Knijnenburg et al* considers mappability for building a background in the enrichment step and not in the decomposition step and does not take into account the mappability effects while building the decomposition. But the assumption is that the genomics signal is never smaller than the mappability signal, which is definitely not correct for the datasets that we evaluated. When Gaussian decomposition is utilized, the lowly mappable regions will see high decrease in signal levels and this will seriously distort the tree based segmentation using the decomposition. MUSIC, however, takes this into account in the mappability correction stage. In order for the pruning and Gaussian decomposition to work, Gaussian decomposition to work, there are assumptions on how fine the scale space is sampled, which corresponds to the parameter sigma in MUSIC. In addition, MSR can be built only from the 50 scale decomposition of each chromosome which ensures that all chromosome becomes on node in the segmentation. In MUSIC, however, there is no constraint on the scale levels and the user can change this parameter freely but which is not possible in MSR building. Similar arguments can be made for l(begin) and l(end).

	Finally, MUSIC uses a median filtering based non-linear multiscale decomposition (unlike the linear Gaussian filter), which is shown to be much more robust with respect to the impulse noise (refer to the manuscript) introduced by the lowly-mappable regions in the genome and which has better edge preserving capability in the smoothing. This should allow MUSIC to recover the peaks more accurately [\cite]. Please note that non-linear multiscale feature detection is an active area of research in electrical engineering. We updated and highlighted the manuscript to reflect these arguments.
Excerpt From Revised Manuscript	