Gmail by Google

Eduardo Gade Gusmão <eggduzao@gmail.com>

Fwd: Decision on Nature Methods submission NMETH-BC24264A

2 mensagens

Ivan Gesteira Costa Filho <ivan.costa@rwth-aachen.de> 9 de setembro de 2015 16:28
Para: Eduardo Gade Gusmão <eggduzao@gmail.com>, Martin Zenke <martin.zenke@rwth-aachen.de>

Dear all,

Here is the review. Only one of the reviewers still have (many) requests. We will need some further experiments (and some arguing with the referee).

More interestingly, the editor suggest to have an even larger paper format now (Analysis article).

"An Analysis article reports comprehensive comparative analyses of technologies, methods or reagents of key importance for a field of research, leading to important practical conclusions about their performances. Analysis articles may also report new analysis of existing large datasets that lead to a novel, exciting or arresting conclusion. The main text (excluding abstract, online Methods, references and figure legends) is approximately 3,000 words. The abstract is typically 100-150 words, unreferenced. Analyses have no more than 6 display items (figures and/or tables). An introduction (without heading) is followed by sections headed Results, Discussion and online Methods. The Results and online Methods should be divided by topical subheadings; the Discussion does not contain subheadings. References are limited to 50."

best,

Ivan

---------- Forwarded message ----------
From: <n.rusk@us.nature.com>
Date: 2015-09-09 15:52 GMT+02:00
Subject: Decision on Nature Methods submission NMETH-BC24264A
To: ivan.costa@rwth-aachen.de

4th Sep 2015

Dear Dr Gesteira Costa Filho,

Your Brief Communication entitled "Addressing DNase-seq cleavage bias and residence time on computational footprinting" has now been seen again by 2 referees, whose comments are attached. While they find your work of potential interest, one of them has raised serious concerns which in our view are sufficiently important that they preclude publication of the work in Nature Methods, at least in its present form.

Should further experimental data allow you to fully address these criticisms we would be willing to look at a revised manuscript (unless, of course, something similar has by then been accepted at Nature Methods or appeared elsewhere). In particular we ask you to address the questions of benchmarking against DHS datasets from different cell types and include a more detailed description of the strengths and weaknesses of each method. We realize that this exceeds the format of a Correspondence, your initial format of submission, and we are willing to consider the format of an Analysis, if all the reviewer's requests are addressed.

We hope you understand that until we have read the revised paper in its entirety we cannot promise that it will be sent back for peer-review.

If you are interested in revising this manuscript as discussed above for submission to Nature Methods in the future, please contact me to discuss your appeal. Otherwise, we hope that you find our referees' comments helpful when preparing your paper for resubmission elsewhere.

Best regards,

Nicole

Nicole Rusk, Ph.D.
Senior Editor
Nature Methods

Reviewers' Comments:

Reviewer #2:
Remarks to the Author:
This paper by Gusmao et al is a short contribution that developed out of an earlier letter to the editor. First of all, it is a strong point that now 14 footprinting methods are compared, so this turned into an all-inclusive analysis. It is also positive that they have now extended their HINT method to include bias correction inferred from naked DNA too, not only DHSs.

In general, I feel that this was a significant amount of work. However, given that this version is essentially a new submission, my comments and requests are more extensive than they would normally be at the stage of a revision. Most importantly, there are issues with the type of data and evaluation strategy.

Bias:
According to equation three, the authors estimate DNase sequence bias for each 6-mer by calculating the ratio of observed DNase cleavages to k-mer abundance genome-wide (in DHSs or whole genome). This approach is similar to used in He et al. and Yardimci et al. and is an established way of measuring bias. Both regular DNase-seq experiments and naked DNase-seq experiments are used to estimate bias, and the results are similar in terms of AUC.

The approach from estimating bias from DHS appears problematic. It is simply not "DNase bias", as this is purely driven by the enzyme and estimated on naked DNA. In supplementary figure 3, they show a correlation matrix of the bias values inferred from DHSs vs naked DNA, both for single and double cut protocols. While the naked DNA datasets cluster together, the DHS-inferred bias values show less agreement within each other. There are two such datasets that don't agree with anything else, for instance (HepG2 and K562, both UW). Since the underlying enzyme is the same, the bias values inferred for different datasets should in principle have a certain degree of agreement. The absence of this leads one to question the chosen strategy.

A final note on this issue is the argument that DHS-based bias estimation improves the footprint predictions for factors with GC-rich motifs. Since DHSs correspond to regulatory elements like promoters and enhancers, and since these elements are known to be GC-rich, it is of course expected that when the bias is estimated from DHSs, the k-mers that seem to be most associated with bias will be the GC-rich ones. This trend is clearly observed in supplementary figure 4. Note that here, the naked DNA datasets do not show this trend. So, the DHS-estimated bias correction is in fact to a certain extent a GC-bias correction. DNase cleavages in DHSs are driven by many different factors (nucleosome positioning, chromatin remodeling, GC content of promoter sites and TF binding events) and may correspond to a useful background model, but it is clearly *not* DNase bias. This should at the very least be clarified and contrasted to the approach taken by others.

Validation:

The title suggests that the study addresses cleavage bias; however, the focus lies on the discussion of predictive performance of various footprinting methods, and an elaboration of differences in different kinds of bias modeling or their impact is missing. It is better to replace the title with what the current version of the paper addresses.

The authors use AUC values at %10 FPR to compare methods. While this is a reasonable choice, related publications used full AUC values or those at 1% FPR to assess performance. Furthermore, CENTIPEDE, PIQ and FLR are turned into binary predictions (footprint vs. no footprint) before assessing performance. This is understandable as authors aim to compare both segmentation and site-centric methods; however using a singular AUC metric at a specific threshold may unfairly affect the performance of models that make non-binary predictions. Footprinting methods tend to result in high sensitivity at low FPR thresholds since the most obvious footprints tend to fall in high confidence ChIP-seq peaks; 10% FPR is too unrealistic and the authors should at least give the 1% values as well.

The most problematic point is that the authors use solely ChIP to evaluate footprints. Using ChIP-seq peaks as a sole gold standard is not adequate, since a subset of these miss footprints due to indirect binding (Neph

et al, Yardimci et al. Sung et al.), are known to contain artifacts (Teytelman et al. , Park et al.), and do not have the resolution to discover footprints at nucleotide resolution. Therefore, successful bias modeling need not result in better ChIP prediction accuracy, and this type of benchmarking has caused a lot of confusion in the field. The authors need to find a different way to address this; they may for instance follow the suggestion of Yardimci et al to evaluate presence or absence of footprints in DHS, for cell lines that express resp. do not express the relevant transcription factor. (Additional analyses on "differential peaks" has also been requested by the other reviewer.)

Some methods (Wellington, Cuellar-Partida prior) included in this paper are not affected by sequence bias in the way He et al. described (depletion of DNase-seq signal centered at TF binding site), or they combine many additional features in addition to the footprint (CENTIPEDE and the authors' own approach). Methods do not just separate into site-centric and footprint-centric, but also into footprint-specific and multi-feature, and to discern the contribution of bias and footprints, a more careful evaluation is needed. At the current stage, the comparison without any elaboration of differences in the binding sites they discover, or which features are informative, fails to add novel insight into the issue other than showing the success of multi-feature methods to predict ChIP-seq signals.

Minor points:

Residence Time:

The authors discuss TF residence times as discussed by Sung et al. to explain lack of footprints. This is a reasonable hypothesis for explaining poor performance of footprinting for certain nuclear receptors but not yet an established fact at this point. ["Sung et al.7 showed that short-lived TFs display a lower DNase I cleavage protection pattern". Sung et al. proposed this hypothesis, but did not prove it or test it extensively.]

The observations in the paper are largely a confirmation of the model inspired by the Sung et al. results, and are in line with the observations in Yardimci et al that bias modelling still may not lead to successful predictions across all factors. It is valuable to the footprinting field that observations about nuclear receptors are consolidated in an effort to dispel the idea that footprinting methods can be used for all DNA binding proteins, but the authors should be clear about their specific contribution.

Correction of Equation 3 in Supp. Material. (o * R) / (r * O)

Explanation of how OBS is calculated for each TF in Fig1A is missing.

There are some changes in Figure 1 that should be explained to the reviewers. Even though the datasets used to generate the two versions are almost exactly the same (with the exception that a few TFs were added in the current version), Centipede did not have a significant negative correlation with bias previously and now it does. Also, it seems to have much better accuracy now. The authors should explain this and make sure that they indeed get correct information from the compared predictors.

Reviewer #3:
Remarks to the Author:
The authors have addressed my concerns in the updated manuscript and have added protection score to address the concern on binding time.

** For Nature Publishing Group general information and news for authors, see http://npg.nature.com/authors.

**********END**********

--
Ivan G. Costa

www.costalab.org
IZKF Computational Biology Research Group
RWTH University Hospital Aachen

> doctoral stipends in computational engineering available in
> graduate school AICES - see http://www.aices.eu

Martin Zenke <martin.zenke@rwth-aachen.de> 9 de setembro de 2015 17:08
Para: Ivan Gesteira Costa Filho <ivan.costa@rwth-aachen.de>, Eduardo Gade Gusmão <eduardo.gusmao@rwth-aachen.de>

Dear Ivan and dear Eduardo,

well done and my congratulations !!!

I think it is great to get invited for the Analysis Article format and we should make an effort to address the points by Reviewer 2. Reading the comments they appear to me of being reasonable and doable, either by additional analysis or words. We are very close of getting this work published in a high profile journal.

It is also the career of a paper nowadays: starting from a Correspondence, then a Brief Communication and now Analysis Article format. Obviously, the Analysis Article format will generate maximal visibility and impact on the field and it is worth going for it. Therefore my congratulation again for receiving this invitation by the editor.

I suggest that you make a first draft of (i) what can be done and addressed by additional analysis and (ii) what should be addressed by words. I should be happy to discuss with you details next week, when I am back in Aachen.

We should also let the editor know that we go for a revised version (see her statement in the email "...If you are interested in revising this manuscript as discussed above for submission to Nature Methods in the future, please contact me to discuss your appeal.") and let her know an estimate of the time required.

Best regards

Martin

[Texto das mensagens anteriores oculto]

> [Texto das mensagens anteriores oculto]
> --
> Ivan G. Costa
>
> www.costalab.org
> IZKF Computational Biology Research Group
> RWTH University Hospital Aachen

Martin Zenke, PhD, Professor of Cell Biology and Chairman
Institute for Biomedical Engineering
Department of Cell Biology
Universitätsklinikum Aachen, RWTH
Pauwelsstrasse 30
52074 Aachen

Helmholtz Institute for Biomedical Engineering
Pauwelsstrasse 20
52074 Aachen
Germany

Voice +49-241-80 80759 (secretary)
+49-241-80 80760 (office)
Fax +49-241-80 82008

email: martin.zenke@rwth-aachen.de
eveline.mierau@rwth-aachen.de
andrea.offergeld@rwth-aachen.de
www: http://www.molcell.de