Gmail by Google

Eduardo Gade Gusmão <eggduzao@gmail.com>

Fwd: Decision on Nature Methods submission NMETH-C24264

2 mensagens

Ivan Gesteira Costa Filho <ivan.costa@rwth-aachen.de> 8 de maio de 2015 06:47
Para: Eduardo Gade Gusmão <eggduzao@gmail.com>, Martin Zenke <martin.zenke@rwth-aachen.de>

Dear Eduardo and Martin,

There are many requests, but it is definitely worth a try.

best,

Ivan

---------- Forwarded message ----------
From: <n.rusk@us.nature.com>
Date: 2015-05-08 2:57 GMT+02:00
Subject: Decision on Nature Methods submission NMETH-C24264
To: ivan.costa@rwth-aachen.de

6th May 2015

Dear Dr Gesteira Costa Filho,

Your Correspondence, "Correction of DNase-seq cleavage bias impacts on quality of footprinting", has now been seen by 3 referees. As you will see from their comments (below), although the referees find your work of considerable potential interest, they have raised a number of concerns, to which we would like to see your response before we reach a final decision on publication.

We would therefore like to invite you to revise your manuscript to address these concerns, which we envisage will require the inclusion of more data. Please ensure that the revised version is as concise as possible. We shall return to the issue of format and presentation in the event that the manuscript is accepted for publication.

When revising your paper:

* include a point-by-point response to our referees and to any editorial suggestions

* please underline any additions to the text or areas with other significant changes to facilitate review of the revised manuscript

* ensure it complies with our general format requirements as set out in our guide to authors at www.nature.com/naturemethods

* complete the statistics and methods manuscript checklist described below

* resubmit all the necessary files electronically by using the link below to access your home page

http://mts-nmeth.nature.com/cgi-bin/main.plex?el=A2M7CaV2A2BBkb4J3A9ftdbzbX8iISp2QtDwYjLuYOgZ

We are trying to improve the quality and transparency of statistics and methods reporting in our papers. Therefore, when revising your manuscript, please fill out the checklist (available at: http://www.nature.com/nmeth/pdf/sm_checklist.pdf) to confirm that all relevant elements have been included in the manuscript. (If you have already filled out this checklist for a previous version of the manuscript, please only fill in the information relevant to any new experiments.) Please submit this checklist, and any previous version of the checklist, with your revised manuscript. It will be available to referees to aid in their evaluation if the paper goes back to the referees for re-review. If you have any questions about the checklist, please see http://www.nature.com/authors/policies/availability.html or contact me.

To further increase transparency, we would encourage you to provide, in tabular form, the data underlying the graphical representations used in your figures. This is in addition to our well-established data-deposition policy for specific types of experiments and large datasets. For readers, the source data will be made accessible directly from the figure legend. Spreadsheets can be submitted in .xls, .xlsx or .csv formats. Only one (1) file per figure is permitted: thus if there is a multi-paneled figure the source data for each panel should be clearly labeled in the csv/Excel file; alternately the data for a figure can be included in multiple, clearly labeled sheets in an Excel file. As with Supplementary Information, files sizes of up to 30 MB would be permitted; however it is anticipated that the vast majority of source data files will be much smaller than this. When submitting source data files with your manuscript please select the Source Data file type and use the Title field in the File Description tab to indicate which figure the source data pertains to.

We hope to receive your revised paper within four weeks. If you cannot send it within this time, please let us know. In this event, we will still be happy to reconsider your paper at a later date so long as nothing similar has been accepted for publication at Nature Methods or published elsewhere in the meantime. Should you miss the four-week deadline without contacting us and your paper is eventually published, the received date may be that of the revised, not the original, version.

[Nature Methods strongly encourages authors of papers reporting proteomic information to share their full data sets via a suitable public repository and to provide the accession number in the final version of the manuscript.]

We look forward to hearing from you soon.

Best regards,

Nicole

Nicole Rusk, Ph.D.
Senior Editor
Nature Methods

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The current manuscript by Gusmao et al. presents an interesting comparison of several DNase-Seq computational footprinting methods and

investigates how DNase-Seq cleavage bias affects their performance and accuracy. The manuscript is well written and easy to understand. The authors have evaluated footprinting methods with the respect to correlation with cleavage bias and made the following observations:

1. The authors question conclusions by He et al. regarding inferred cleavage patterns of many TFs representing intrinsic DNase biases rather than TF footprints. The authors argue such conclusions were made since He et al. did not include methods that have better performance than FS metric.
2. In their analysis He et al. used FS metric, which the Gusmao et al. demonstrate is the worst out of eight methods in terms of performance.
3. The best performer is HINT, which according to the authors is better at distinguishing footprints of nearly all TFs compared to the tag count (TC) metric. Overall, four methods outperform TC metric.

The analyses by Gusmao et al. are informative with the respect to comparison of computational footprinting methods, which was not systematically performed by He et al. This is an important clarification and is of general interest to the field. However, I am not convinced that results from Gusmao et al. disagree with conclusions of He et al. Rather they expand some of the He et al. analyses by including additional methods.

He et al. evaluated the effects of differential footprinting for predicting differences in TF binding between conditions and demonstrated high performance of this metric compared to footprinting score. It would be informative to see Gusmao and co-authors evaluate HINT and other methods for similar analyses.

General comments

1. I disagree with the claim by Gusmao et al. that He et al. claim that "the simplest method for detection of active binding sites possible, outperforms computational footprinting". The authors only made observations for the methods that were evaluated.
2. The authors demonstrate improved footprinting of six factors after DNase bias correction. However, no results are shown for AR and GR, which was given as a main example of uninformative footprints in He et al. Can the authors provide such analyses? Furthermore, it would be useful to compare corrected footprints for between He et al. and Gusmao et al. on the same set of 6+ factors (Fig. 5, SI).
3. It is currently difficult to evaluate the results for a specific TF from the graph (Fig. 3, SI). Can the authors summarize the results of their AUC vs OBS analyses in a spreadsheet wtr to individual TFs and footprinting methods, so that the performance of methods can be evaluated for individual TFs?
4. Can the authors comment on He et al.'s evaluation of 0500 and 0458 motifs? Do Gusmao et al. agree that these footprints are artifacts of DNaseI?

Reviewer #2:
Remarks to the Author:
This short note aims to demonstrate that contrary to He et al, appropriate computational approaches can successfully identify transcription factor footprints in DNase-seq data. To this end, the authors apply a method developed by them and compare it to some other, simpler computational tools.

There are several major concerns why this format may not be the best venue for this particular study.

1) The authors are overstating the implications from He et al. The

main point of that study was to raise concerns that simple methods that do not account for background, including some that had been used in high profile publications, could lead to false positive footprints (they do not suggest that no method would ever be able to identify footprints). In my opinion, this was shown unequivocally.

More importantly, they authors have missed several studies that have already addressed this very question, published as independent papers rather than a short note.

a) Hager lab, Mol Cell 2014 -- these authors develop explicit models for background and suggest (and show with a few examples) that many factors would not lead to discernable footprints. Specifically, depth of a footprint would be attributable to residence time, while the shape would be entirely due to background.

b) Crawford and Ohler labs, NAR 2014 -- this study came out almost at the same time and led to somewhat different conclusions. It contains a mixture model with a background component, and quantifies how well the actual footprint, rather than overall accessibility, would predict bound sites.

c) Gifford lab (Nat.Biotech 2014) with the method "PIQ": This is also a site-centric method where footprints are learned from ChIP and DNase signal magnitude and shape in a TF-specific manner. [This method might be less relevant since a 400bp window is considered instead of immediate motif vicinity and there are no claims/correction regarding bias, but it would certainly be relevant to see how it compares.

d) Ott lab (NAR 2013) with the method "Wellington" (implemented in pyDNase): This is a segmentation-based method that takes advantage of the strand-specific imbalance of DNase cuts, resulting from size selection in the double-hit protocol. So this would only be applicable to U Wash data analyses, b ut has been shown to perform well in several comparisons.

2) There is an issue with the estimation of the bias. Rather than explicitly using the bias parameters estimated on dechromatinized DNA by Crawford and Ohler for the Duke protocol, or earlier by Stamatoyannopoulos for the U Wash protocol, this is done using aligned reads inside DHSs. This is of course open chromatin, but not equivalent to naked DNA. This problem is reflected in supplementary figure 1, which shows the correlations of these 6-mer bias values in different datasets and protocols. Crawford and Ohler showed that the bias in Wash vs Duke protocols is positively correlation (0.74), whereas here this appears to be much lower or not detectable. Cleavage is not just be influenced by the sequence but also what is really unbound and accessible.

-) The authors do not include the nuclear receptors in their studies (with a note that they are not represented in major PWM repositories). This is the "poster child" for factors that likely do not leave discernible footprints, and they need to include those.

- In their analysis they find that some of the methods are not significantly influenced by bias. These are methods like Cuellar and Centipede that take advantage of smoothed signals in windows significantly larger than the motif itself (150-200bp). This is probably the reason for their relative immunity to the bias problem, and I don't find this very surprising. Therefore it is unclear to me what this finding contributes to the field precisely. It is certainly not enough to rule out the necessity for bias corrections. In fact the authors incorporate bias correction in their own method (HINT-BC). These broad differences between methods that use larger regions vs just sites is not clearly explained and phrased and should be revised.

Minor concerns:

- The labels in figure 1A appear to be wrong, they don't agree with the text. (Figure 1A is the same as supplementary figure 3, this seems a bit redundant)

- The method, HINT, appears like a promising approach. With the size constraints imposed by the format, there is too little information how this is trained and whether and how it might differ to the original publication.

- Some results and comparisons would benefit from a more in-depth discussion. For instance, the original Cuellar Partida is a prior derived from tag counts, and the comparably worse performance seems to result from a combination with FIMO (PWM scores). Similarly, "Neph" is compared against FS score (which is pretty much the same without optimized flank scores) but performs much better. Insights into the flank optimization (which in the original Neph appears to be done manually/ad hoc) would certainly help others in the field.

Overall, given that the main question has been addressed by others, the authors may fare best to write up a more in depth computational study that focuses on the method -- and insights how different approaches work or do not work -- rather than the background issue.

Reviewer #3:
Remarks to the Author:
The He paper outlined the problem of footprinting TF from DNase-seq experiments and showed how different TF are affect at different levels. Sung et al 2014 further defined the problem by examining factor dynamics in detecting footprints.

In this correspondence the authors address the overall bias across all factors, but do not show how there method corrects for the bias for specific factors which have short binding times.

For example, one of the key finding in He et al was that AR and p53 can not be accurately detected by DNase-seq footprinting. Does HINT-BS accurately detect AR or p53 footprints? With out this included it is hard to understand if correcting for bias will make a difference.

Details of the confidentiality and pre-publicity policy may be found
here http://www.nature.com/authors/policies/confidentiality.html

Privacy Policy | Update Profile

--
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

Ivan Gesteira Costa Filho <ivan.costa@rwth-aachen.de> 8 de maio de 2015 06:55
Para: Eduardo Gade Gusmão <eggduzao@gmail.com>

oi dudu,

vou ler logo a literatura relacionada. Nao demora nao. Temos soh 1 mes
para a revisao.

abs,
[Texto das mensagens anteriores oculto]