nature portfolio

Peer Review File

Robust footprinting with sample-specific Tn5 bias correction for bulk and single cell ATAC-seq

Corresponding Author: Professor Nancy Zhang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

A version of this paper was originally rejected for publication by Nature Communications, however that decision was reconsidered after appeal by the authors.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In this study by Lin et al., the authors present TraceBIND, a computational framework that corrects sample-specific Tn5 bias. They apply this framework to the aging rat kidney, demonstrating age-associated dynamic regulatory changes, linking footprint activity to epigenetic drift.
At its heart, this study is a quality control study. The authors have identified a potential solution for one form of bias (Tn5) inherent in scATAC-seq data that could impact batch correction. A credible argument is made that this form of bias is more impactful between datasets (as opposed to within) and would result in imperfections in footprinting / sub-peak calling in aged samples. TraceBIND could be useful for integrating large datasets with 100's of samples across a wind range of ages in order to improve footprint informed feature calls.

Critiques:
1. The degree of improvement is significant, but not large. TraceBIND appears to cut down on false positives (Fig 3E) while maintaining discriminatory potential (Fig 3i). However, the overall improvement over existing approaches seems modest in figures 3, 4 and 5. There are many of Tn5 Bias corrections approaches, it is hard to gauge the relative impact of this one. I am moderately enthusiastic as I thought it was creative to use mitochondrial insertions as an estimate for the naked DNA.
2. The Cross-sample variation sections sometimes lack multi-sample summary comparisons across all possible samples. The authors pick examples and it is unknown whether the examples are representative of the whole merged datatset. Presumably, there are more than 2 cell types, more than 2 samples in the same study, etc.
a. Figure 1A - It is unclear which samples fall into which category (same sample, same study, different labs, etc), other than the 4 comparisons which are labeled.
b. Figure 1f-l – The samples selected are unclear. Further, it is unclear whether these correlations apply across all comparisons that can be similarly categorized for the 37 samples.
c. Figure 3b – Again, one sample comparison is given here and there is a 6% difference in correlation.
3. Can the authors demonstrate that cross-sample variability may be due to differences in experimental factors such as Tn5 concentration and the duration of exposure in their datasets? This seems feasible as most data were derived from the authors and their collaborators' labs.
4. This study focuses mainly on post-processing. It would be helpful to provide practical recommendations or evaluation of pre-analytic and experimental parameters.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
ATAC-seq data is widely available and have often been used to study TF binding events, combined with TF motifs. Default analysis workflows use ATAC-seq peaks, which have limited resolution. TF footprint analysis can pinpoint the exact location

of TF binding and thus can be particular useful to study TF regulation. TraceBIND used novel approach to leverage mitochondrial DNA as an in-sample negative control for Tn5 bias prediction. In contrast to the recent PRINT method, that assumes the same Tn5 bias for all samples, TraceBIND does sample-specific correction for Tn5 bias. The improved sensitivity and specificity of TraceBIND has been validated in diverse datasets.

I have a few questions, mainly for clarifications.

1. First, TraceBIND builds on PRINT and fine-tune it using sample-specific mitochondrial DNA data. How to judge whether there is sufficient data for finetuning? e.g., compared with the amount of data used for initial training of PRINT? Different cell types have different number of copies of mitochondrial DNAs. Does this mean that finetuning is more effective for some cell types than other cell types?

2. TraceBIND performs sample-specific finetuning using the insertion counts from mtDNA. Though from the description in the main text, it is not clear what kind of finetuning was done. It has been clarified in the method section, but it would be helpful to mention it in the main text as well.

3. The lower half of panel 2a needs more clarifications, either in figure caption or in figure legend. For example, in the bottom left of panel 2a, the -log10 p-value panel and effect size panels shows -lgo10 p-value and effect size by different colors? Then what are the corresponding values for different colors? What does "Poisson mixture point process with inhomogeneous background rate" refers to? I did not find it in the Methods Section. I assume the lower half of panel 2a is about computing a scan statistic. If so, it might be better to separate it out as a one panel.

4. The comparison of false positives in Figure 3d and 3e is impressive. I appreciate the explanation of PWM methods and why they have limited performance. Similar brief summaries of TOBIAS and HINT-ATAC would help readers to better understand the results. Are they supervised tools?

5. the GitHub https://github.com/lyx-lin/TraceBIND is not available.

Minor concerns

Figure 6b shows a nice monotone trend. Though the plot is dominated by outliers. A log transformation of effect size may help increase the readability of the figure.

Line 587 in Methods section, the definition of N_{t}^{w, L, flanking} is not clear. what is " the window of size L of (t – w/2, t + w/2)"? In the following equation, "Nbinomial" means negative binomial distribution? Here \pi be the probability that the Tn5 insertions are located in the center window? if so, should the probability of S being Nbinomial(N, \pi) rather than Nbinomial(N, 1 - \pi)?

(Remarks on code availability)
It seems the GitHub is not public?

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)
During my review, the provided link is invalid.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I found that the authors satisfactorily addressed my critiques.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
the author have addressed all my concerns. congratulations for this nice work!

(Remarks on code availability)
I read the documentation and briefly check the organization of the codes. all look good.

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer 1:

In this study by Lin et al., the authors present TraceBIND, a computational framework that corrects sample-specific Tn5 bias. They apply this framework to the aging rat kidney, demonstrating age-associated dynamic regulatory changes, linking footprint activity to epigenetic drift.
At its heart, this study is a quality control study. The authors have identified a potential solution for one form of bias (Tn5) inherent in scATAC-seq data that could impact batch correction. A credible argument is made that this form of bias is more impactful between datasets (as opposed to within) and would result in imperfections in footprinting / sub-peak calling in aged samples. TraceBIND could be useful for integrating large datasets with 100's of samples across a wind range of ages in order to improve footprint informed feature calls.

We thank the reviewer for the thoughtful summary of our study and for the comments regarding the utility of TraceBIND, particularly in large-scale single-cell ATAC-seq datasets. We also appreciate the recognition that cross-sample Tn5 bias can propagate to affect footprint-informed feature calls.

While we agree that addressing Tn5 bias has important implications in quality control, we respectively clarify that the central contribution of this work extends beyond QC. Existing footprinting methods typically assume a globally shared Tn5 bias profile derived from naked DNA controls. Our analyses demonstrate that Tn5 cleavage bias varies substantially across samples, even when nominally generated under the same protocol. When this variability is not modeled, it leads to inflated false positive footprint calls, particularly in cross-sample comparisons.

Moreover, TraceBIND is designed to integrate with downstream single-cell analysis frameworks. By refining peaks into high-confidence sub-peak footprint features, TraceBIND improves transcription factor activity inference (via chromVAR) and enhances cell-state and age prediction (via Epitrace), demonstrating practical benefits beyond bias correction itself.

We also present a base-resolution comparison between TraceBIND footprints and attribution peaks from a cell type-specific DNA foundation model (Decima). This integration enables a direct comparison between sequence-derived regulatory predictions and experimentally inferred in vivo occupancy, highlighting both concordance and complementary regulatory information. To our knowledge, this is among the first systematic nucleotide-level integrations of footprinting with a DNA foundation model.

We therefore view TraceBIND as a methodological advance in statistical modeling of chromatin accessibility data that enables high-confidence, cross-sample footprint analysis and facilitates integration with emerging single-cell and sequence-to-function modeling frameworks, rather than solely as a quality-control refinement.

1. *The degree of improvement is significant, but not large. TraceBIND appears to cut down on false positives (Fig 3E) while maintaining discriminatory potential (Fig 3i). However, the overall improvement over existing approaches seems modest in figures 3, 4 and 5. There are many of Tn5 Bias corrections approaches, it is hard to gauge the relative impact of this one. I am moderately enthusiastic as I thought it was creative to use mitochondrial insertions as an estimate for the naked DNA.*

We thank the reviewer for the thoughtful assessment and for recognizing the creativity of leveraging mitochondrial insertions as an empirical proxy for naked DNA. We respectfully clarify that although intermediate bias-correction metrics may appear modest in isolation, these translate into 20–60-fold reductions in false positive rates relative to existing approaches while maintaining comparable sensitivity (Supplementary Figure. 7). This resolves the specificity barrier for applying footprinting analysis to single cell ATAC sequencing data sets.

We believe that the mis-reading of our results was due to the way it was presented. We focused too much on intermediary statistics and did not clarify the appropriate final benchmark. We have revised the manuscript to more clearly present these end-user benchmarks and to emphasize the practical implications of improved specificity for large-scale bulk and single-cell ATAC-seq studies. In the new Figures 3d–g, we present a comprehensive benchmarking of footprinting methods, evaluating the balance between sensitivity and specificity across multiple data types and assays.

Specifically, in Figures 3e and 3f, we benchmarked TraceBIND against existing methods using public naked DNA and ChIP-seq datasets, and observed an approximately 25% improvement over the next-best method in the balance between specificity and sensitivity as measured by area under the ROC curve. Importantly, this improvement is driven primarily by a marked reduction in false positives rather than inflation of sensitivity. In the DNase-seq benchmark, TraceBIND recovers approximately 30% more validated footprints than the next-best approach, demonstrating that the improvement generalizes beyond a single assay type.

Indeed there are many Tn5 bias correction methods. However, TraceBIND is distinct in its use of mitochondrial insertions as an empirical estimate of naked DNA, rather than relying solely on sequence-based or in silico models. TraceBIND also contains multiple technical innovations that allow rigorous type 1 error control while maintaining high power. These advances enables reliable genome-wide footprinting at a level of specificity not achieved by prior methods, which makes this type of analysis viable for large-scale genomic studies.

2. *The Cross-sample variation sections sometimes lack multi-sample summary comparisons across all possible samples. The authors pick examples, and it is unknown whether the examples are representative of the whole merged dataset. Presumably, there are more than 2 cell types, more than 2 samples in the same study, etc.*

   a. Figure 1A - It is unclear which samples fall into which category (same sample, same study, different labs, etc), other than the 4 comparisons which are labeled.

   b. Figure 1f-I – The samples selected are unclear. Further, it is unclear whether these correlations apply across all comparisons that can be similarly categorized for the 37 samples.

c. Figure 3b – Again, one sample comparison is given here and there is a 6% difference in correlation.

We thank the reviewer for highlighting the importance of demonstrating that the reported patterns are representative across the full dataset rather than based on selected examples. We agree and have revised the manuscript accordingly.

a. We have added a color bar to Figure 1a to indicate laboratory origin for each sample. We also removed the "different studies same lab" category, as the public datasets generated from the same lab we used may include samples generated by the same technician, making this category potentially misleading. As shown in the revised figure, laboratory origin alone does not fully explain the correlation structure, highlighting the complexity of batch effects and reinforcing the need for sample-specific bias correction.

b. We have provided correlations across *all* pairwise comparisons among the 37 samples in Supplementary Figures 3 and 4, demonstrating that the trends shown in Figure 1f-i are observed in different datasets.

c. For Figure 3b, we have expanded the benchmarking to include all samples rather than a single illustrative comparison and have added an additional bias-correction method for completeness (Fig. 3b and Supplementary Fig. 7g). Across the full dataset, sample-specific fine-tuning consistently improves correlation between predicted and observed Tn5 bias relative to global models, indicating that the improvement is systematic rather than example-specific.

3. *Can the authors demonstrate that cross-sample variability may be due to differences in experimental factors such as Tn5 concentration and the duration of exposure in their datasets? This seems feasible as most data were derived from the authors and their collaborators' labs.*

We appreciate this insightful suggestion. Identifying the experimental factors underlying cross-sample variability in Tn5 bias, such as differences in Tn5 concentration or exposure duration would be highly informative.

However, the datasets used in this study were not generated with the explicit goal of characterizing Tn5 bias, and they all followed the standard 10x single cell ATAC-seq protocol with the same kit. Thus, on paper, they all tried to follow the same procedure. However, as illustrated in Figure 1, even when technicians from the same lab follow the same protocol, slight variations between experimental runs can not be controlled and batch effect is inevitable.

Thus, while experimental factors such as enzyme concentration, reaction kinetics, or handling differences may plausibly contribute to this variability, these covariates were not explicitly controlled or recorded in a manner that would allow formal attribution. Importantly, the central objective of this work is not to identify the causal sources of bias

variation, but to demonstrate that such variability exists and materially impacts footprinting specificity. These observations motivate a sample-specific correction strategy that does not rely on known covariates, but instead leverages embedded negative controls (mitochondrial insertions) to empirically estimate bias within each dataset.

To clarify this, we revised the paper to state the following:

> *The relationship between Tn5 bias and GC content also differed across samples in a nonlinear and non-monotonic manner (Supplementary Fig. 2l), further indicating that these differences cannot be explained by simple compositional effects. We speculate that subtle experimental factors, such as Tn5 concentration, reaction kinetics, or enzyme handling, may contribute to these shifts in cleavage preference[27]. Notably, even among datasets generated using the same 10x Chromium Single Cell ATAC reagents and protocol, we observed substantial variation in mtDNA-derived bias profiles. Together, these findings demonstrate that Tn5 cleavage bias exhibits pronounced, sample-specific variability that is not captured by protocol metadata and cannot be assumed to be constant across experiments.*

4. *This study focuses mainly on post-processing. It would be helpful to provide practical recommendations or evaluation of pre-analytic and experimental parameters.*

We thank the reviewer for this thoughtful suggestion. While TraceBIND is primarily a computational framework applied after sequencing, our findings have important implications for experimental design and preprocessing as well.

First, we demonstrate that mitochondrial reads, commonly discarded during ATAC-seq preprocessing, provide a valuable in-sample reference for estimating Tn5 bias and assessing batch effects. We therefore recommend retaining mitochondrial reads during preprocessing, as they enable sample-specific bias assessment and empirical false discovery control without requiring additional experimental controls.

Second, our analyses show that substantial cross-sample variability in Tn5 bias can arise even when standardized protocols (e.g., the 10x Genomics single-cell ATAC-seq kit) are followed. This suggests that experimental reproducibility cannot be assumed at the level of enzymatic cleavage preference, and that bias evaluation should be treated as a routine quality-control step in ATAC-seq workflows.

More broadly, TraceBIND provides a framework for diagnosing and correcting bias variability at the sample level, enabling more reliable downstream footprinting and transcription factor activity analyses. We have clarified these practical recommendations in the revised manuscript to better articulate how our findings inform both preprocessing decisions and downstream regulatory analysis strategies (see the Results section corresponding to Figure 1 and the Discussion).

Reviewer 2:

ATAC-seq data is widely available and has often been used to study TF binding events, combined with TF motifs. Default analysis workflows use ATAC-seq peaks, which have limited resolution. TF footprint analysis can pinpoint the exact location of TF binding and thus can be particularly useful to study TF regulation. TraceBIND used a novel approach to leverage mitochondrial DNA as an in-sample negative control for Tn5 bias prediction. In contrast to the recent PRINT method, that assumes the same Tn5 bias for all samples, TraceBIND does sample-specific correction for Tn5 bias. The improved sensitivity and specificity of TraceBIND has been validated in diverse datasets.

We thank the reviewer for this thoughtful and accurate summary of our work. We agree that while ATAC-seq peak-based workflows are widely used, their resolution is inherently limited, and footprinting provides a more precise view of transcription factor occupancy at base-pair resolution.

As noted, a key distinction of TraceBIND is its use of mitochondrial DNA insertions as an in-sample negative control to estimate sample-specific Tn5 bias. In contrast to approaches that assume a globally shared bias model, our analyses demonstrate that Tn5 cleavage bias varies substantially across samples and that modeling this variation is essential for controlling false positives in footprint detection.

We further appreciate the reviewer's recognition that the improvements in sensitivity and specificity are supported across multiple validation settings, including naked DNA controls, matched ChIP-seq/CUT&RUN data, DNase-seq footprints, and perturbation experiments. We believe these orthogonal validations strengthen the case that sample-specific bias modeling materially improves the reliability of footprint-based regulatory inference.

1. *First, TraceBIND builds on PRINT and fine-tune it using sample-specific mitochondrial DNA data. How to judge whether there is sufficient data for finetuning? e.g., compared with the amount of data used for initial training of PRINT? Different cell types have different number of copies of mitochondrial DNAs. Does this mean that finetuning is more effective for some cell types than other cell types?*

    We thank the reviewer for this important question regarding data sufficiency and robustness of the fine-tuning procedure.

    TraceBIND builds on PRINT by fine-tuning the model using sample-specific mitochondrial DNA insertions as an empirical proxy for naked DNA. In our datasets, mitochondrial insertions are substantially more abundant than the naked DNA insertions used in the released PRINT resources. The released PRINT naked DNA data have a mean insertion count of 4.2 per position, whereas mitochondrial DNA insertions in our datasets show substantially higher coverage (median 399.5 and ranging from 36.9 to 996.9, Supplementary Table 3). This substantially higher coverage provides sufficient signal for stable fine-tuning of the pretrained model.

    We agree that mitochondrial copy number varies across cell types, which could in principle affect fine-tuning effectiveness if performed at the individual cell-type level.

Indeed, as shown in Supplementary Fig. 6, some cell types have limited mitochondrial coverage.To address this, we perform fine-tuning at the sample level rather than the individual cell-type level, pooling mitochondrial insertions across all cells within a sample to improve robustness.

To formally evaluate whether sample-level Tn5 bias can represent cell type specific-bias, we fitted a Negative Binomial model to quantify the dispersion of Tn5 bias across cell types within the same sample, using the sample-level bias as the mean (Methods). Supplementary Fig. 5 shows that within-sample dispersions are <0.1, as compared to cross-sample dispersions which are orders of magnitude higher. This indicates that Tn5 bias is highly consistent across cell types within a sample, and that sample-level fine-tuning provides an accurate and robust estimate for all constituent cell types.

Together, these analyses demonstrate that mitochondrial coverage is sufficient for reliable fine-tuning, and that variability in mitochondrial copy number across cell types does not materially affect bias estimation.

2. *TraceBIND performs sample-specific finetuning using the insertion counts from mtDNA. Though from the description in the main text, it is not clear what kind of finetuning was done. It has been clarified in the method section, but it would be helpful to mention it in the main text as well.*

   We thank the reviewer for this helpful suggestion. We have revised the main text to more clearly describe the fine-tuning procedure. Specifically, we now state that TraceBIND initializes from pretrained PRINT weights and performs sample-specific fine-tuning using mitochondrial insertion counts, while freezing the second convolutional layer to preserve generalizable mid-level sequence features. The early and late layers are allowed to adapt to sample-specific cleavage patterns and output scaling.

   This clarification ensures that readers can understand the fine-tuning strategy directly from the main text, without needing to consult the Methods section.

3. *The lower half of panel 2a needs more clarifications, either in figure caption or in figure legend. For example, in the bottom left of panel 2a, the -log10 p-value panel and effect size panels shows -log10 p-value and effect size by different colors? Then what are the corresponding values for different colors? What does "Poisson mixture point process with inhomogeneous background rate" refers to? I did not find it in the Methods Section. I assume the lower half of panel 2a is about computing a scan statistic. If so, it might be better to separate it out as a one panel.*

   Thank you for the suggestion! We have revised the figure and caption to improve transparency and consistency with the Methods section.

Specifically, we added explicit color legends for the −log10 p-value and effect size panels to clarify the values represented by different colors. We also unified our language regarding the model: We replaced the description of a “Poisson mixture point process with an inhomogeneous background rate” with “negative binomial”, used throughout the Methods section, as essentially that is what it is. In addition, we separated the lower half of panel 2a into a standalone panel to more clearly illustrate the scan statistic computation.

4. *The comparison of false positives in Figure 3d and 3e is impressive. I appreciate the explanation of PWM methods and why they have limited performance. Similar brief summaries of TOBIAS and HINT-ATAC would help readers to better understand the results. Are they supervised tools?*

   We have added concise summaries of PRINT, TOBIAS, and HINT-ATAC to provide readers with better intuition for their underlying assumptions and to explain why these methods exhibit higher false positive rates in our benchmarks. In short, PRINT and HINT-ATAC are supervised models trained on publicly available samples, while TOBIAS is an unsupervised method whose footprint scores depend strongly on local chromatin accessibility.

5. *the GitHub https://github.com/lyx-lin/TraceBIND is not available.*

   Thank you for pointing this out. The GitHub repository is now public and accessible. We have verified that all code required to reproduce the analyses in this study is available.

Minor concerns

6. *Figure 6b shows a nice monotone trend. Though the plot is dominated by outliers. A log transformation of effect size may help increase the readability of the figure.*

   We thank the reviewer for this helpful suggestion. The effect size is defined in the Methods as the log ratio of insertion counts between center and flanking regions, adjusted by predicted preference. To further improve visualization and reduce the influence of extreme values, we applied an additional log transformation for display purposes. The monotonic trend remains significant after this transformation and is shown in Supplementary Fig. 9. We believe this revised visualization improves readability while preserving the underlying biological signal.

7. *Line 587 in Methods section, the definition of N_{t}^{w, L, flanking} is not clear. what is " the window of size L of (t – w/2, t + w/2)"? In the following equation, "Nbinomial" means negative binomial distribution? Here \pi be the probability that the Tn5 insertions are located in the center window? if so, should the probability of S being Nbinomial(N, \pi) rather than Nbinomial(N, 1 - \pi)?*

We thank the reviewer for carefully examining the statistical notation. We have revised the Methods section to clarify the definition of *N_{t}^{w, L, flanking}.* Specifically, we now explicitly define the window of size *L* as the interval (*t* – *w*/2 - L, *t* -+ *w*/2 - 1) or (*t* + *w*/2 + 1, *t* + *w*/2 + L), and clarify how the flanking region is constructed relative to the central window. We also clarify that “NBinomial” refers to the negative binomial distribution. In addition, we corrected a typographical error in the probability parameter: the model uses *\pi* as defined in the text, and the previous appearance of *1 - \pi* was a notation error. We confirm that this correction affects only the presentation in the manuscript and does not impact the implementation or any results reported in the study.

Reviewer 3 (Remarks on code availability):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

1. During my review, the provided link is invalid.

   Thank you for bringing this to our attention. The GitHub repository has now been made public and is fully accessible at the provided link. We have verified that all code and documentation required to reproduce the analyses are available.