

BIT: Bayesian Identification of Transcriptional Regulators from Epigenomics-Based Query Region Sets

Corresponding Author: Professor Xinlei Wang

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This manuscript presents BIT, a Bayesian hierarchical model designed to identify transcriptional regulators (TRs) from user-provided epigenomic regions. Through comprehensive simulation and application studies, the authors demonstrate that BIT consistently outperforms existing state-of-the-art methods, showcasing its robustness and accuracy. BIT quantifies the uncertainty for predicted TRs, and this is a significant difference and improvement compared to existing methods.

While acknowledging the inherent challenges posed by the dynamic and complex nature of TR regulation, such as cooperative binding and chromatin structure influences, the author mentioned their efforts to enhance computational efficiency, such as exploring variational inference, to handle the increasing volume of epigenomic data generated by modern high-throughput techniques.

Overall, BIT is a powerful tool for advancing our understanding of transcriptional regulation, particularly in the context of exploring single-cell and bulk ATAC-seq data. By addressing key challenges in TR binding prediction, BIT represents a step forward in the integration of Bayesian modeling with epigenomics research.

Following are some specific comments:

1. In the methods, the authors use the “naive” method to estimate μ (line 671). However, $\hat{\mu}$ as defined in line 671 seems to be the mean of “dataset-level” importance score, not the mean of “TR-level” importance score as the authors defined previously. The mean of TR-level importance score should be estimated as follows: (please see equation in the attached doc). If this is correct, the related results need to be corrected as well.
2. The author benchmarked the performance of BIT and other methods in identifying functionally essential TRs in Figure 4F. This benchmark experiments cover 9 cancer types, under different numbers of top ranked TRs. The message shown in this figure is not very clear. It might be better to summarize the results in one chart to make the conclusion more intuitive. Also, it is unclear what the meaning of comparing the p-values across different methods, if the p-values are not deemed significant anyway (e.g., > 0.5).
3. In Figure 5, the author focused on B cells in PBMC data, and among the other cell types, only Monocytes and Dendritic cells were run and shown in the supplementary data. How about other cell types in PBMC? To demonstrate the performance of BIT in this highly heterogeneous dataset, the author should run BIT on each existing cell types and present the results in an unbiased manner.
4. The authors compared top TRs identified by BIT and SnapATAC2 in figure 5D and highlighted the strength of BIT by pinpointing specific TRs out of other family members. To demonstrate BIT’s performance, the authors should also name true positive TRs identified by BIT but missed by SnapATAC2 Motif in the text. Also, it is unclear why only snapATAC2 is selected to compare with BIT but not other methods. Performance comparison with other existing methods is necessary.
5. Figure 5F seems to demonstrate that top TRs identified by BIT for liver malignant cells are prognostic biomarkers. However, some TRs are associated with better prognosis, while others are associated with worse prognosis. Considering these TRs are predicted to be functional in malignant cells, it is confusing and hard to understand why some oncogenic TRs

lead to better prognosis. The authors should explain these results thoroughly.

6. The author nicely demonstrates that BIT makes accurate predictions on perturbation data. Now that some single-cell perturbation data have been published and more TRs are included in those datasets, the author should test their method on these datasets and include the results.

7. Benchmarking results using other scATAC-seq methods (e.g. scBasset), should also be included to demonstrate the performance of BIT.

(Remarks on code availability)
Acceptable.

Reviewer #2

(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 #3

(Remarks to the Author)

The authors propose a novel method for inferring Transcriptional Regulators (TRs) from genomic or epigenomic regions (e.g., peaks). This approach relies on:

1. A comprehensive reference library of TR-ChIP datasets from the GTRD, comprising over 10,140 human and 5,681 mouse datasets, which eliminates the need for motif-based predictions.
2. A hierarchical model that accounts for both within- and between-dataset heterogeneity, enhancing robustness and mitigating biases caused by TRs with disproportionately large dataset representation. This method reduces over-fitting while preserving critical insights by integrating information across multiple datasets.
3. Probabilistic characterization of TRs regulation. Unlike frequentist techniques, this approach provides credible intervals on the matching/versus informative regions mean, offering a more nuanced and reliable analysis.

The paper is well-written, and the authors present extensive proof-of-concept results, including simulations, perturbation datasets, cancer-related data and single-cell experiments, to validate their assumptions and model. The proposed model is formally correct and well-suited to its context, providing straightforward and easily interpretable estimates for both expert and non-expert researchers.

Additionally, they have made significant efforts to provide both a GitHub package and a user-friendly web-based platform, enabling users to apply the method to their specific regions of interest.

However, I have several suggestions regarding the following aspects:

Reproducibility.

In several parts of the text, critical details necessary for reproducing the results are missing. For example, the authors do not specify the software and packages used for the analysis (e.g., R—what version?), or the release versions of the datasets (e.g., GTRD [see 550: add the link, or the name of the reference dataset, the release versions, ...], TCGA [see 551: did the authors use the online platform? What is the release?]), the version/released of the other methods employed (see [694]: what is the version of each package?). In general, a more detailed description of the data processing steps and tools used for analysis would be beneficial.

Although the authors provide scripts to reproduce the figures presented in the paper, these rely on pre-processed tables that are loaded and plotted, rather than starting from raw data. I recommend that the authors include complete scripts to process the analysis from raw data—at least for some examples—if they used the R package or other tools. This would greatly enhance reproducibility and transparency.

State-of-the-art methods.

Since the authors propose a new computational tool, it would be highly beneficial to expand the discussion of state-of-the-art methods in the introduction [lines 75–88]. This could include a detailed exploration of the technical aspects of other models and/or the reference libraries they rely on. Adding a comparison table that highlights the strengths and weaknesses of existing methods—or classes of methods (e.g., motif-based, region-based, distance-based)—would provide valuable context.

Such a table would allow readers to better understand the unique features of the proposed method, such as its ability to integrate multiple datasets or to identify TRs that are not characterized by specific motifs [lines 384–385]. This added context would also help readers appreciate the results of the study, particularly in comparison to other approaches.

Moreover, in comparing Figure 3F and Figure 4F, it would be insightful to report the top 10/50/100 ranked TRs for each method and compare them in terms of the similarity of common elements. This would reveal whether there are patterns among methods with similar models (or classes of models, such as motif-based or region-based approaches).

Fixed non-overlapping windows.

The authors use 1,000 bp fixed windows to bin the genome, but it is unclear why this specific window size was chosen. Given that transcription factor binding sites can be much smaller (~50 bp, depending on the TF), the impact of this choice should be addressed. Additionally, since the package allows users to select different window sizes (e.g., 200, 500, or 1,000 bp), it would be helpful to evaluate and discuss the implications of this flexibility on the results.

Computational time.

The authors briefly mention the issue of computational time in the discussion, noting that the web tool warns users that analyses can take approximately 50 minutes. However, it would be helpful to provide actual runtime measurements, particularly for the real-data examples, and to compare these with runtimes of other methods. This would give readers a clearer understanding of the computational efficiency of the proposed approach.

Credible intervals.

Although the authors emphasize the importance of credible intervals, these are not used in any of the provided results (e.g., Figure 3A). I recommend adding whiskers to the main figures. Additionally, in sections such as [87] and [110] (discussing "Uncertainty estimation" and "credible intervals"), the distinction between frequentist and Bayesian approaches may be challenging for non-expert readers to grasp. Rephrasing these sections to make them clearer and more accessible would improve the manuscript.

General clarifications in the manuscript:

[206] "Parameters in Results section": Are the authors referring to Figure 6A–B in the Methods section?

[206 and Figure 6A] Translating the parameter μ into terms of the proportion of matching versus over-informative elements would provide a more intuitive interpretation, helping to convey what a "real scenario" entails.

[247, 249] Parameters such as τ (controlling between-TR heterogeneity) and σ (noise levels) are critical concepts that would be nice to be highlighted in the method section.

[600] It would be valuable to discuss how the depth of each reference library influences the parameters of the distributions presented in Figure 6. For instance, does the variability primarily stem from the number of datasets included, or does sequencing depth/number of peaks also play a significant role in shaping the results?

[624] Despite the reference that gives the details of possible parameters, I would like to know how they decide for the Upper and Lower limits of the non-informative model.

[340]: Please report the rank of the other methods. Is there any difference between Mouse/Human?

[Figure 1A]: The vignette illustrating the first step is somewhat misleading. The "000" sequence visually corresponds to the actual input region, which should instead be represented as "111." Additionally, the color of the nucleosomes matches the "1" in the vector, which could cause confusion. I recommend revising the initial representation to align with the input vector for consistency and clarity.

[Figure 3C]: It would be helpful to depict the input regions (e.g., using rectangles) to visualize the overlap between the input and reference data more clearly, alongside the track. This enhancement would make the comparison more intuitive for readers.

[Figure 3E]: The meaning of the point colors is unclear. Adding a legend or explanation would help clarify this aspect.

(Remarks on code availability)

Github package.

I found the documentation on <https://bitbayesian-identification-of-transcriptional-regulators.readthedocs.io/en/latest/index.html> to be clear and comprehensive. However, the GitHub documentation contains several missing details. For instance, since the Zenodo repository is required to run BIT, I recommend including a direct reference to it in the GitHub tutorial. Similarly, the internal help files in R (e.g., for the BIT function) lack details about parameters such as burnin. Harmonizing the documentation across these platforms would greatly enhance usability.

- Since the method relies on a Gibbs sampler, it would be beneficial to allow users to set a random seed for reproducibility.
- The function appears to use all available CPU cores, by default. Providing recommendations for hardware requirements and enabling users to configure the number of cores would improve user experience and resource management.

Online tool

- File upload: Providing a downloadable example file as an alternative to the "click and run" option would help users better understand the required input format, even though the supported standards are well known. Additionally, the main text mentions the use of summit information when available—this is a critical detail that should be explicitly communicated to users within the tool.

- Retrieve key: Allow users to download a text file containing the automatically generated key along with a summary of the input parameters, so they can save and retrieve it later. Additionally, consider offering users the option to receive an email notification once the analysis is complete, as it can take more than an hour.

- Result download issue: I attempted to download the results using the following search keys, but it did not work: I) Search key: 00fe1dfd5098f8e89191ea7b0bba30ab, ii) Search key: 01520f69b0316ac16d8e4e8276447d33. Namely, the format of the file was not readable.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have fully addressed all concerns raised during the initial review. The manuscript has been significantly improved.

(Remarks on code availability)

Reviewer #2

(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 #3

(Remarks to the Author)

I have read the revised manuscript, and I appreciate the authors' thorough revisions. They have carefully addressed all my comments and provided detailed solutions to the points I raised. I believe the manuscript has improved significantly.

(Remarks on code availability)

I have examined the code, and both the online platform and the R package are well documented and fully functional.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license,

unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Authors' Responses to the Initial Review of

“BIT: Bayesian Identification of Transcriptional Regulators from Epigenomics-Based Query Region Sets”

Zeyu Lu^{1,3}, Lin Xu^{2,5*}, Xinlei Wang^{3,4*}

We sincerely thank the editors and reviewers for their time and effort in reviewing our BIT manuscript. In response to the feedback, we have extensively revised our manuscript. Specifically, we added one main table and six new supplementary figures to address the reviewers' concerns. We also updated five main figures and four supplementary files based on their suggestions. Furthermore, we expanded our comparison by incorporating two additional single-cell ATAC-seq analysis tools and included a new single-cell TR perturbation dataset for validation. Finally, we revised the entire manuscript to improve clarity and readability and added new references.

We hope that this revised version fully addresses the reviewers' concerns.

Reviewer 1

This manuscript presents BIT, a Bayesian hierarchical model designed to identify transcriptional regulators (TRs) from user-provided epigenomic regions. Through comprehensive simulation and application studies, the authors demonstrate that BIT consistently outperforms existing state-of-the-art methods, showcasing its robustness and accuracy. BIT quantifies the uncertainty for predicted TRs, and this is a significant difference and improvement compared to existing methods.

While acknowledging the inherent challenges posed by the dynamic and complex nature of TR regulation, such as cooperative binding and chromatin structure influences, the author mentioned their efforts to enhance computational efficiency, such as exploring variational inference, to handle the increasing volume of epigenomic data generated by modern high-throughput techniques.

Overall, BIT is a powerful tool for advancing our understanding of transcriptional regulation, particularly in the context of exploring single-cell and bulk ATAC-seq data. By addressing key challenges in TR binding prediction, BIT represents a step forward in the integration of Bayesian modeling with epigenomics research.

>>> We thank the reviewer for the valuable comments and suggestions, which have significantly helped us revise and improve our manuscript. To facilitate your review, we have included the relevant changes from the revised manuscript, along with updated and newly added figures, directly in this letter. We have addressed each of your concerns in a point-by-point manner, detailing how the revisions were included.

Following are some specific comments:

1. In the methods, the authors use the “naïve” method to estimate μ (line 671). However, $\hat{\mu}$ as defined in line 671 seems to be the mean of *dataset-level* importance score, not the mean of *TR-level* importance score as the authors defined previously. The mean of TR-level importance score should be estimated as follows: (please see equation in the attached doc). If this is correct, the related results need to be corrected as well.

>>> Thanks for pointing out that the original version of the naïve method was estimating the dataset-level average of the importance scores. Following your suggestion, we have revised our formula for estimating μ accordingly. We observed a decrease in the MSE of the naïve method after implementing the revised formula, indicating improved performance compared to the original version. Nonetheless, BIT continues to exhibit a consistently lower MSE, underscoring its superior accuracy in parameter estimation relative to the naïve methods. We have incorporated the following revision into the Methods section to reflect this change and updated **Fig. 2A**:

Page 38, Line 753 – Line 755: “To compare the performance of BIT in estimating model parameters, we used a naïve method as the baseline, where μ is estimated by

$$\hat{\mu} = \frac{\sum_{i=1}^M \frac{\sum_{j=1}^{J_i} \text{logit}\left(\frac{x_{ij}}{n_{ij}}\right)}{J_i} + \sum_{i'=1}^{M^c} \text{logit}\left(\frac{x_{i'1}}{n_{i'1}}\right)}{M + M^c};$$

”

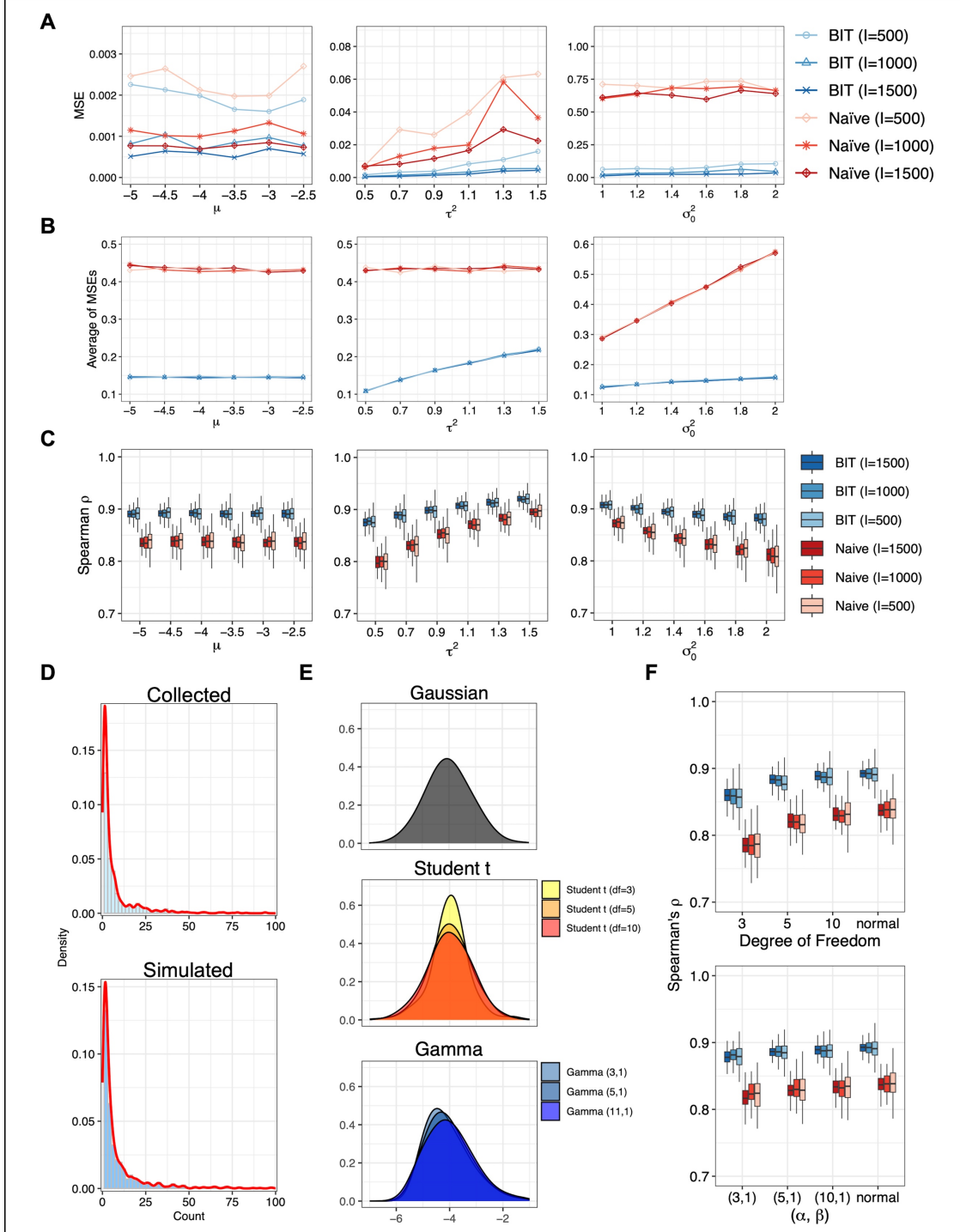


Figure 2. Simulation results under different parameter settings. **A.** Mean squared errors for estimation of global parameters using 100 replicate datasets in each simulation setting. **B.** The average of MSEs for estimation of importance parameters (averaged over different TRs) over 100 replicates. **C.** Boxplots of Spearman's rho correlation over 100 replicates between ranks from BIT or the baseline method and the true ranks of simulated TRs. **D.** The distributions of simulated vs. real numbers of ChIP-seq datasets of individual TRs. **E.** Data generated in our sensitivity analysis from t and gamma distributions, to examine whether heavy-tailedness and skewness would significantly affect the performance of BIT. **F.** Boxplots of Spearman correlation over 100 replicates under t and gamma distributions, with baseline under normal distribution.

2. The author benchmarked the performance of BIT and other methods in identifying functionally essential TRs in Figure 4F. This benchmark experiments cover 9 cancer types, under different numbers of top ranked TRs. The message shown in this figure is not very clear. It might be better to summarize the results in one chart to make the conclusion more intuitive. Also, it is unclear what the meaning of comparing the p-values across different methods, if the p-values are not deemed significant anyway (e.g., > 0.5).

>>> According to the reviewer's recommendation, we have revised Figure 4F to more clearly illustrate our findings. In the original figure, our aim was to show that the top-ranked TRs identified by BIT exhibit a higher overlap with functionally essential TRs, as determined by DepMap's CRISPR-Cas9 knockout screens (which measure cancer cell survival following depletion of a specific TR). A lower Chronos score indicates that a gene is more likely to be functionally essential for cancer cells. The original Figure 4F and the corresponding Fisher exact test compared how many functionally essential TRs each method included in its top-ranked TR list.

Following the reviewer's advice, we replaced it with a bar chart that directly shows the number of TRs within the top 10 and top 50 that overlap with functionally essential TRs identified by DepMap. As presented in the updated figure, BIT identifies the highest total number of functionally essential TRs among the top 10 and top 50 TRs separately, summed across nine cancer types.

We have revised **Fig. 4F** and updated the main text:

Page 21, Line 422 – Line 424: "... We observed that BIT identified the highest total number of functionally essential TRs, which was verified by DepMap in the top 10 and top 50 TRs separately, by summing the results from 9 cancer types (**Fig. 4F**)."

Page 39, Line 773 – Line 774: "...We then examined top-ranked TRs to determine how many were functionally essential in each cancer type."

[Figure Redacted]

Figure 4. BIT can identify transcriptional regulators from cancer-type specific accessible regions. **A.** 32 TRs identified by BIT from TCGA cancer-type-specific accessible regions validated by existing literature. **B.** BIT scores of top 10 TRs with upper and lower bound of 95% credible intervals. **C.** Jaccard index compares cancer-specific accessible regions between any two cancer types (upper triangle) and the sets of top TRs identified by BIT between any two cancer types (lower triangle). **D.** Cumulative number of BIT-identified top 10 and top 50 TRs with minimum Chronos scores of ≤ -2 to >0 . **E.** Boxplots showing the Pearson correlation coefficients between Chronos scores and BIT scores for the top 50 TRs identified by BIT across 209 cell lines from 9 cancer types. **F.** Numbers of functionally essential TRs identified in the top 10 and 50 TRs by BIT and state-of-the-art methods (BART, ChIP-Atlas, HOMER, i-cisTarget, WhichTF). **G.** The results of gene-ontology enrichment analysis of top 20 TRs identified by BIT from breast cancer.

3. In Figure 5, the author focused on B cells in PBMC data, and among the other cell types, only Monocytes and Dendritic cells were run and shown in the supplementary data. How about other cell types in PBMC? To demonstrate the performance of BIT in this highly heterogeneous dataset, the author should run BIT on each existing cell types and present the results in an unbiased manner.

>>> We totally agree on the necessity of systematically applying BIT to all existing cell types and presenting the results in an unbiased manner. In response, we have comprehensively revised Supplementary **Fig. S7** (original: **Fig. S5**) and **Fig. S8** (original: **Fig. S6**) to address these points.

For the PBMC scATAC-seq dataset, we applied BIT to the nine available cell types: B cells, monocytes, NK cells, CD4+ T cells, CD8+ T cells, dendritic cells, gdT cells, HSPCs, and MAIT cells. In the liver cancer scATAC-seq dataset, we analyzed seven different cell types: Malignant cells, T cells, tumor-associated macrophages (TAMs), B cells, neutrophils, cancer-associated fibroblasts (CAFs), and tumor endothelial cells (TECs). For each cell type, we now provide the top 10 TRs ranked by BIT, accompanied by Gene Ontology (GO) enrichment analysis of these top-ranked TRs. Results for B cells (PBMC dataset) and malignant cells (liver cancer dataset) are presented in the main text (**Fig. 5**) while results for the remaining cell types are detailed in **Fig. S7** and **Fig. S8**.

GO enrichment analysis for BIT-identified TRs across PBMC cell types revealed significant enrichment of biological processes directly linked to their core functions (e.g., B cell activation, T cell differentiation, and myeloid cell differentiation). In the liver cancer dataset, while cells also exhibited enrichment for lineage-specific biological processes (e.g., T cell differentiation), we observed additional context-specific pathways, such as "positive regulation of defense response" in TAMs and "lymphocyte activation in immune response" in B cells. These findings highlight the tumor microenvironment's influence on TR activity, where proximity to malignant cells may perturb canonical regulatory programs through microenvironmental crosstalk or oncogenic signaling.

In addition, we expanded the gene activity plots to demonstrate that top-ranked TRs exhibit cell type-specific activity patterns consistent with their functional roles. Due to space constraints, we cannot present the gene activity of all TRs in one figure. However, all raw data, visualization code, and computational workflows required to generate analogous plots for all the TRs are available in Zenodo and GitHub to ensure full transparency and reproducibility.

We have revised the main text (**Lines 527 – 538**) to reflect these changes. Finally, as suggested in other points, we have reorganized the results from benchmarking BIT against other tools including ArchR, SnapATAC2, and scBasset into a new **Supplementary Figure S6**. This standalone figure provides a detailed evaluation of performance from the four methods, which we will address in subsequent sections of this response.

Page 26, Line 527 – Line 538: “For an unbiased analysis, we present results for the eight remaining PBMC cell types (**Fig. S7**) and six remaining primary liver cancer cell types (**Fig. S8**) besides those for PBMC B cells and liver malignant cells. These results align with observations from **Fig. 5**. For example, GO enrichment analysis of BIT-identified TRs across PBMC cell types (**Fig. S7C**) revealed significant enrichment of biological processes directly linked to their core functions (e.g., T cell differentiation, myeloid cell differentiation). In the liver cancer dataset, we identified additional context-specific pathways (**Fig. S8C**), such as "positive regulation of defense response" in TAMs and "lymphocyte activation in immune response" in B cells. These findings highlight the tumor microenvironment’s influence on TR activity, where proximity to malignant cells may perturb canonical regulatory programs through microenvironmental crosstalk or oncogenic signaling. In addition, gene activity plots (**Fig. S7B & S8B**) demonstrate that top-ranked TRs exhibit cell type-specific activity patterns consistent with their functional roles.”

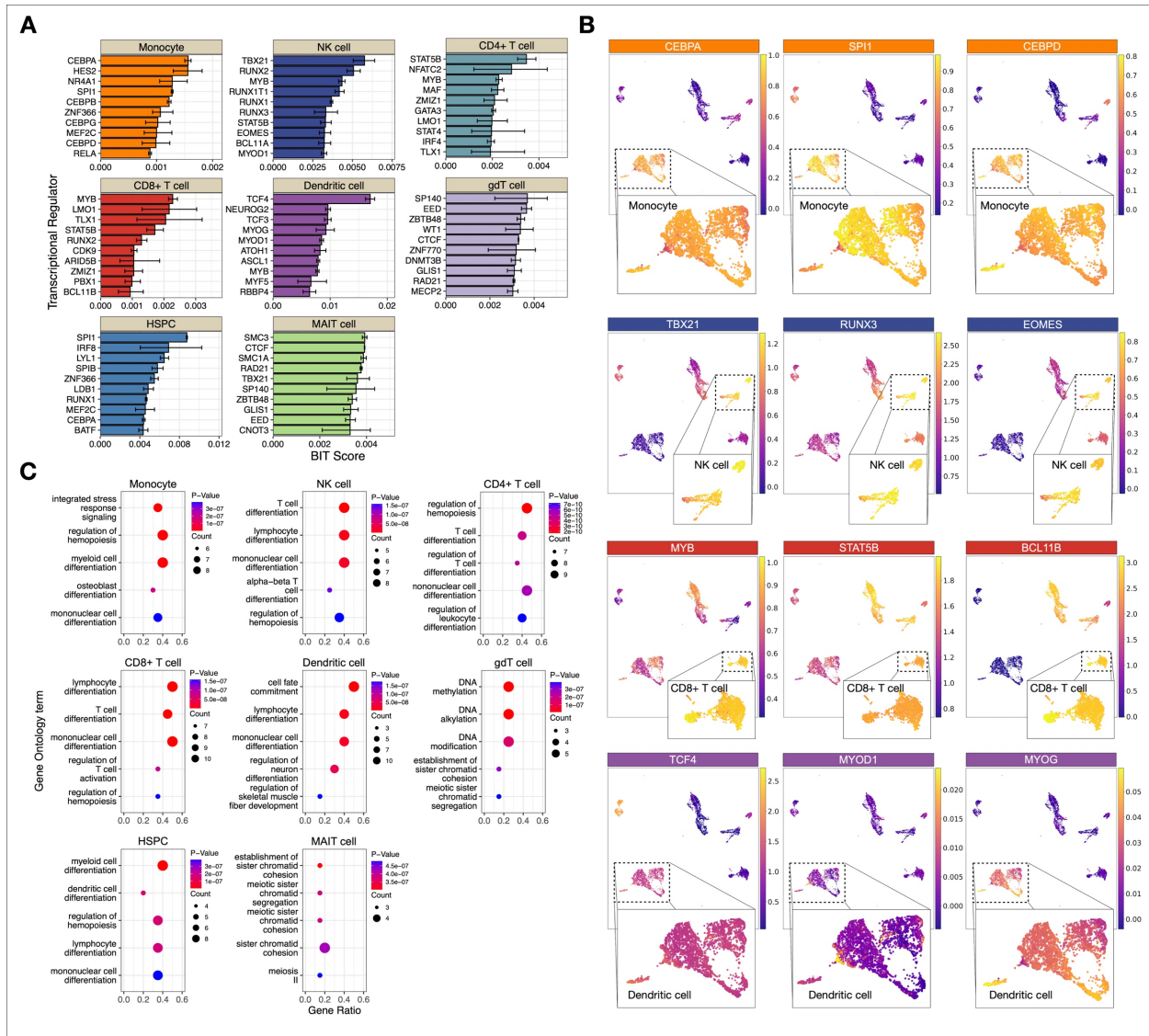


Figure S7. BIT can identify critical TRs in various cell types from PBMCs. **A.** Top 10 BIT-identified TRs of eight cell types (Monocyte, NK cell, CD4+ T cell, CD8+ T cell, Dendritic cell, gdT cell, HSPC, and MAIT cell) from PBMC single-cell ATAC-seq data. **B.** UMAP visualization of gene activity for CEBPA, SPI1, CEBPD, TBX21, RUNX3, EOMES, MYB, STAT5B, BCL11B, TCF4, MYOD1, and MYOG in PBMCs, each identified by BIT with high activity in one of the four largest cell types (Monocyte, NK cell, CD8+ T cell, and Dendritic cell). **C.** GO enrichment analysis results of BIT-identified top TRs in the eight cell types.

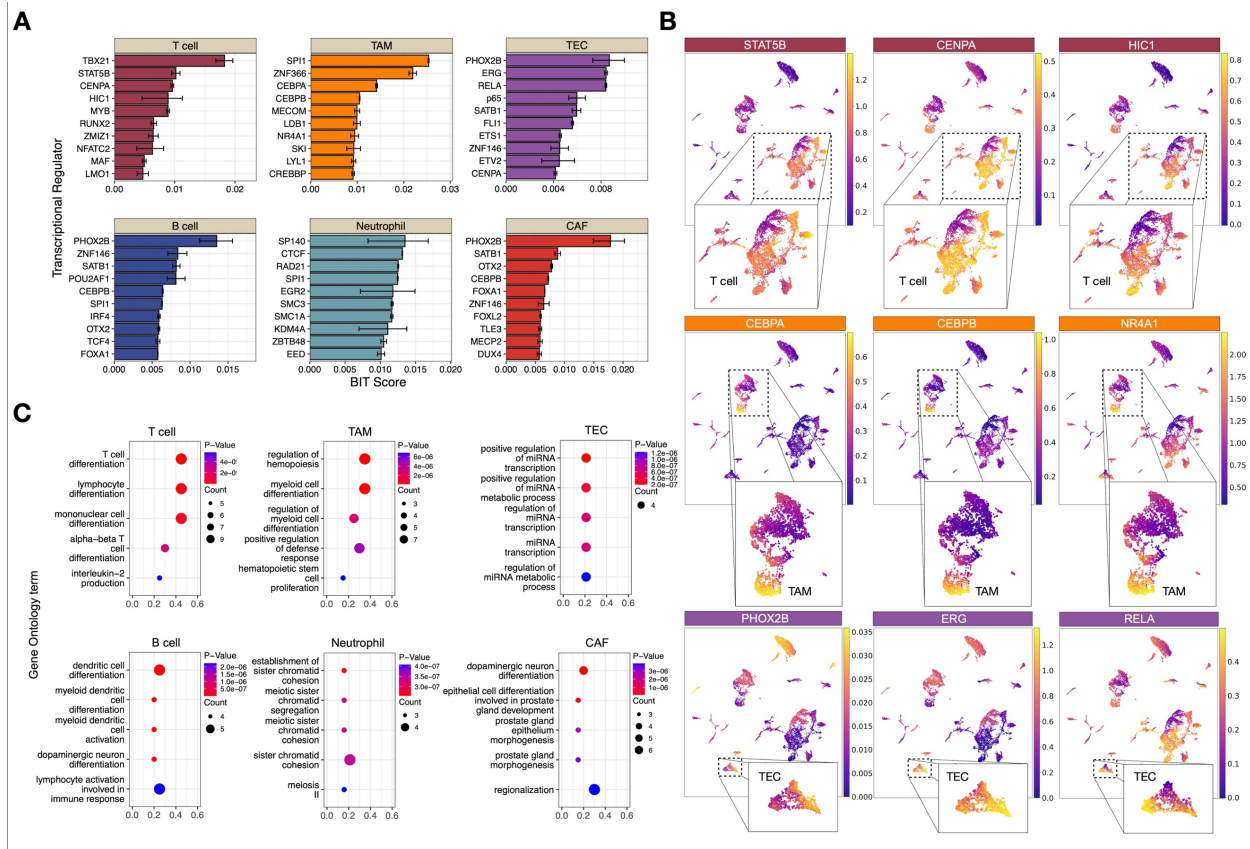


Figure S8. BIT can identify critical TRs in various cell types of liver cancer. **A.** Top 10 BIT-identified TRs of six cell types, including T cell, Tumor-Associated Macrophage (TAM), Tumor Endothelial Cells (TEC), B cell, Neutrophil, and Cancer-Associated Fibroblast (CAF), from liver cancer single-cell ATAC-seq data. **B.** UMAP visualization of gene activity for STAT5B, CENPA, HIC1, CEBPA, CEBPB, NR4A1, PHOX2B, ERG, and RELA in liver cancer cells, each identified by BIT with high activity in one of the three largest cell types (T cell, TAM, and TEC). **C.** GO enrichment results of BIT-identified top TRs in each of the six cell types.

4. The authors compared top TRs identified by BIT and SnapATAC2 in figure 5D and highlighted the strength of BIT by pinpointing specific TRs out of other family members. To demonstrate BIT's performance, the authors should also name true positive TRs identified by BIT but missed by SnapATAC2 Motif in the text. Also, it is unclear why only snapATAC2 is selected to compare with BIT but not other methods. Performance comparison with other existing methods is necessary.

>>> We thank the reviewer for the valuable comment on annotating true positive TRs. In response, we have rigorously annotated true positive TRs in both the main text (**Fig. 5D**) and the newly added Supplementary **Fig. S6A**. To define true positives, we leveraged the Human Protein Atlas (HPA), which aggregates single-cell RNA-seq data from public PBMC datasets to identify TRs with elevated expression or functional enrichment in six PBMC cell types (B cells, CD4+ T cells, CD8+ T cells, NK cells, monocytes, and dendritic cells) and liver tissue. HPA-verified TRs were cross-referenced with those identified by BIT and other benchmarking methods.

We also appreciate the reviewer's constructive suggestion to expand our comparison beyond SnapATAC2. In this revision, we have incorporated ArchR and scBasset, two additional tools capable of processing scATAC-seq data and inferring TR activity. This broader benchmarking is now consolidated in **Fig. 5G** and **Fig. S6**. We note that SnapATAC2, ArchR and scBasset offer multiple functions in scATAC-seq data analysis, one of which is to infer TR activity inference. In contrast, BIT is not a general-purpose single-cell sequencing analysis tool. Thus, our comparison focuses solely on inferred TRs.

To quantify accuracy, we evaluated the overlap of TRs identified by BIT, SnapATAC2, ArchR, and scBasset with HPA-verified TRs. For PBMCs, BIT demonstrated higher concordance with HPA-verified TRs across the six HPA-available cell types (**Fig. S6B**). In liver cancer, we revised **Fig. 5G** to include results from ArchR and scBasset, revealing that BIT retains higher specificity for functionally essential TRs in liver cancer cell lines.

We have revised the main text to reflect the updates:

Page 23, Line 450 – Line 457: “SnapATAC2 was applied to generate cell type-specific accessible regions. We then applied BIT to accessible regions from various cell types in the PBMC and liver cancer datasets and compared the results (**Tab. S5**) with those from SnapATAC2, ArchR, and scBasset. These tools integrate multiple steps for processing and analyzing scATAC-seq data, with motif enrichment analysis being a common approach for inferring TR activity in each annotated cell type. Since these tools are widely used with scATAC-seq analysis, we specifically evaluated BIT’s performance in comparison to their motif enrichment analysis functions from a practical perspective. The results were thoroughly validated by multiple approaches, as described below.”

Page 23, Line 461 – Line 467: “Notably, BIT successfully captured several TRs critical for B cell development. These include PAX5, EBF1, SPIB, IRF4, IKZF1, and BCL11A, which were previously reported for their roles in maintaining B cell identity and verified by HPA as lineage-specific TRs. In addition, POU2AF1, MEF2B, BCL6, BACH2, POU2F2, BATF, IKZF3, REL, and PBX1, though not present in the HPA-verified list, serve as pivotal regulators in B cell proliferation, differentiation, the germinal center transcriptional program, and immune response, as supported by the literature (**Tab. S4**).”

Page 25, Line 481 – Line 488: “For liver cancer malignant cells, six of the top 20 TRs (HNF4A, FOXA2, NR2F6, HNF1A, HLF, and AR) were verified as tissue-specific by HPA and reported in the literature. Furthermore, the top TRs identified from scATAC-seq (**Fig. 5B**) showed strong consistency with those from bulk ATAC-seq (**Fig. 4B**). Seven of the top 10 TRs identified from bulk data (SOX13, HNF4A, NFIL3, NR2F6, HNF4G, HLF, and NFIA) also appeared in the list from single-cell data.”

Page 25, Line 496 – Line 503: “Third, for PBMC B cells and liver malignant cells, we compared the top 20 TRs identified by BIT with those from motif enrichment analysis in ArchR, scBasset, and SnapATAC2. The four sets showed little to no overlap (**Fig. 5D**). A similar pattern was observed across all HPA-available cell types in the PBMC dataset (**Fig. S6A**). As mentioned in the introduction, the similarity of binding motifs can hinder the differentiation between TRs from the same protein family. This is evident in the top-ranked TRs identified using ArchR and SnapATAC2, where results of PBMC B cells include the POU and IRF families, and results of liver cancer malignant cells exhibit enrichment exclusively for the FOX and ZNF families.”

Page 26, Line 508 – Line 511: “We further compared the number of HPA-verified TRs among the top 20 for the six HPA-available cell types from PBMCs using these four approaches. BIT identified the most HPA-verified TRs in four cell types and had the highest total count overall (**Fig. S6B**).”

Page 26, Line 514 – Line 515: “...Conversely, the motif-enrichment-identified TRs displayed enrichment for less specific biological processes (**Fig. 5E**).”

Page 26, Line 524 – Line 526: “...In addition, CRISPR/Cas9 knockout experiments in liver cancer cell lines revealed that BIT identified more functionally essential TRs compared to TRs identified by ArchR, SnapATAC2, and scBasset, respectively (**Fig. 5G**).”

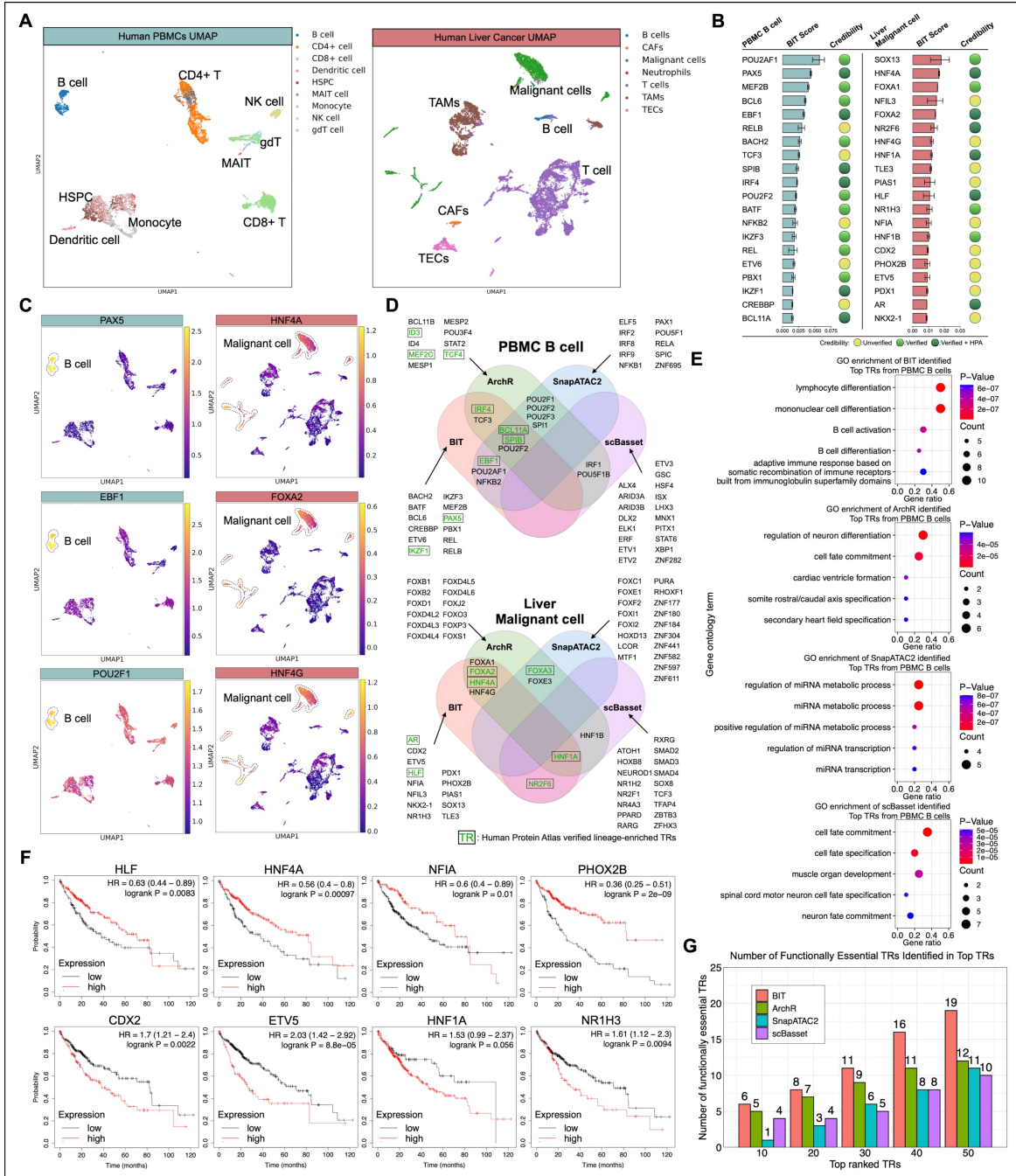


Figure 5. BIT can identify transcriptional regulators from cell-type-specific accessible regions. A. UMAP projection of human Peripheral blood mononuclear cells (PBMCs) and primary liver cancer cells. **B.** Top 20 TRs identified by BIT using PBMC B cell-specific and liver malignant cell-specific accessible regions, verified by existing literature and Human Protein Atlas (HPA). **C.** UMAP visualization of gene activity for PAX5, EBF1, and POU2F1 in PBMCs and HNF4A, FOXA2, and HNF4G in liver cancer samples, each identified by BIT with high activity in B or malignant cells. **D.** Venn plots of top 20 TRs identified by BIT, ArchR, SnapATAC2, and scBasset from PBMC B cells and liver malignant cells. **E.** Gene ontology enrichment results of top 20 TRs identified by BIT, ArchR, SnapATAC2, and scBasset from PBMC B cells. **F.** Kaplan-Meier Plots of HLF, HNF4A, NFIA, PHOX2B (potential tumor suppressors or differentiation regulators) and CDX2, ETV5, HNF1A, NR1H3 (potential oncogenes) in liver cancer samples. **G.** Numbers of BIT, ArchR, SnapATAC2, and scBasset identified top 10 to 50 TRs from liver malignant cells with minimum Chronos score ≤ 0.4 .

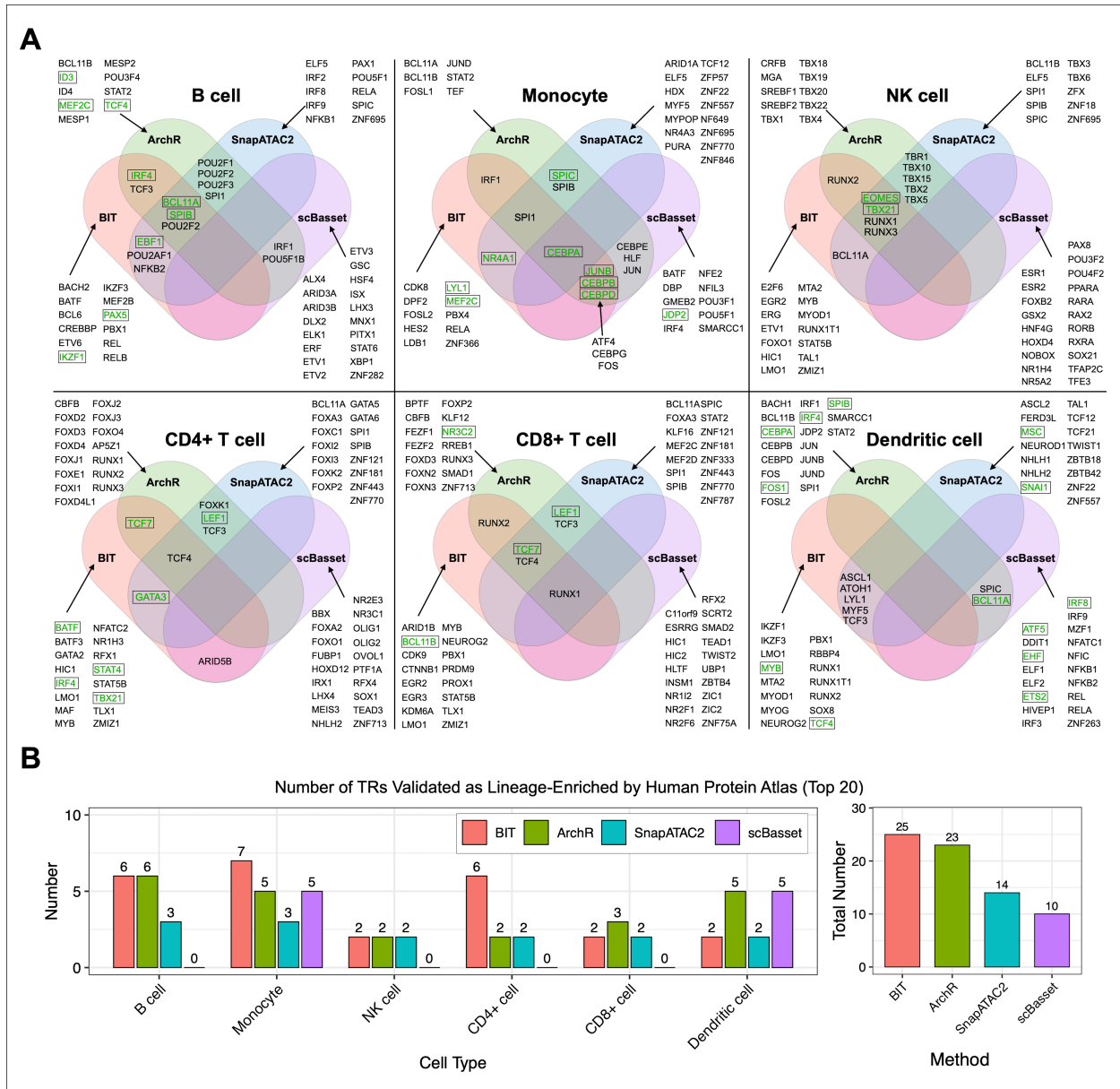


Figure S6. Comparative analysis of TR predictions across computational tools on PBMC single-cell ATAC-seq data using six cell types available in Human Protein Atlas (B cells, CD4+ T cells, CD8+ T cells, Monocytes, NK cells, and dendritic cells). A. Venn diagrams displaying the overlap among the top 20 TRs identified by BIT, ArchR, SnapATAC2, and scBasset. **B.** Number of TRs ranked to top 20 by BIT, ArchR, SnapATAC2, and scBasset that are validated as lineage-enriched TRs by the Human Protein Atlas.

5. Figure 5F seems to demonstrate that top TRs identified by BIT for liver malignant cells are prognostic biomarkers. However, some TRs are associated with better prognosis, while others are associated with worse prognosis. Considering these TRs are predicted to be functional in malignant cells, it is confusing and hard to understand why some oncogenic TRs lead to better prognosis. The authors should explain these results thoroughly.

>>> We appreciate the reviewer's insightful comment regarding the prognostic relevance of TRs identified by BIT in liver malignant cells. Prognostic biomarkers are not limited to TRs linked to worse outcomes; they also include TRs whose activity correlates with improved survival. While some TRs may act as oncogenic drivers, others may function as tumor suppressors or be involved in differentiation programs that are associated with a less aggressive tumor phenotype.

To clarify this, we have expanded our discussion in the main text to explicitly distinguish between TRs associated with poor prognosis (potential oncogenes) and those linked to better prognosis (potential tumor suppressors or differentiation regulators).

Page 26, Line 516 – Line 524: “Fifth, we explored the potential of BIT-identified TRs from liver malignant cells as prognostic biomarkers. Kaplan-Meier survival analysis revealed statistically significant associations between their expression levels and patient survival (**Fig. 5F**). While some TRs correlated with worse prognosis, others were linked to improved survival, suggesting diverse roles in tumor progression and differentiation. TRs correlated with better prognosis may reflect transcriptional programs that constrain malignant potential or promote tumor differentiation, while those linked to worse prognosis may drive tumor progression and survival under selective pressures. These findings highlight the functional relevance of BIT-identified TRs in shaping tumor behavior and their prognostic value in liver cancer.”

6. The author nicely demonstrates that BIT makes accurate predictions on perturbation data. Now that some single-cell perturbation data have been published and more TRs are included in those datasets, the author should test their method on these datasets and include the results.

>>> We sincerely thank the reviewer for pointing out the availability of single-cell perturbation data, which we have now incorporated into our study. Specifically, we included a single-cell perturbation dataset that integrated CRISPR with ATAC-seq to assess effects of knocking down 40 TRs on chromatin accessibility. This addition is detailed in the new **Fig. S4**.

We preprocessed the data using the original authors' scripts¹, which can be accessed through: (https://github.com/GreenleafLab/SpearATAC_MS_2021). We observed that GATA1 knockdown exhibited the strongest target perturbation. In contrast, the perturbation effects of the remaining TRs were weaker and difficult to distinguish (**Fig. S4A**), consistent with the finding from the original publication¹.

We next applied BIT to the differentially accessible regions (DARs) identified by comparing cells treated with single guide RNA (sgRNA) targeting one specific TR against control cells treated with non-targeting single guide RNA (sgNT). In the case of sgGATA1, BIT successfully ranked GATA1 as the top candidate (**Fig. S4B**). In addition, GO enrichment analysis, performed using the top-ranked TRs identified by BIT from sgGATA1-treated cells, revealed strong enrichment for terms associated with leukemia (**Fig. S4C**). To further benchmark the performance of BIT, we calculated the mean reciprocal rank for all 40 TRs and compared BIT's results with five alternative methods, demonstrating that BIT consistently outperformed the other approaches (**Fig. S4D** and **Fig. S4E**).

It is worth mentioning that most single-cell perturbation datasets are based on RNA sequencing. For example, scPerturb, a resource published in 2024 that comprises 44 single-cell perturbation studies, includes only three studies sequenced using ATAC-seq, with the dataset employed in our study being the most recent one. While BIT exhibits superior performance relative to other methods, we acknowledge that there is considerable room for improvement in the analysis of single-cell perturbation data derived from ATAC-seq, as this technology¹ continues to evolve rapidly.

We therefore added **Fig. S4** and revised the main text:

Page 17, Line 343 – Line 359: “In addition to bulk TR perturbation data, we leveraged a single-cell TR perturbation dataset⁶⁰ for validation, which screened 40 TRs with the CRISPR technique in K562 (a well-known leukemia cell line). Among them, GATA1 knockdown exhibited the strongest target perturbation, consistent with the finding from the original study⁶⁰. In contrast, the perturbation effects of the remaining TRs were weaker and difficult to distinguish from each other using UMAP projection (**Fig. S4A**). This may partly result from the essential role of GATA1 in erythrocyte differentiation and development. As anticipated, BIT accurately ranked GATA1 as the top TR when provided with DARs generated by comparing GATA1 knockdown cells with negative control-treated cells (**Fig. S4B**). In addition, GO enrichment analysis of the top-ranked TRs identified by BIT in GATA1 knockdown cells also reveals significant enrichment for terms related to leukemia (**Fig. S4C**).

Finally, to compare BIT with other state-of-the-art methods (BART, ChIP-Atlas, i-cisTarget, WhichTF, and HOMER), we calculated the mean reciprocal rank (MRR) for each method using five perturbed TRs in bulk datasets (**Fig. 3F**) and 40 TRs in the single-cell perturbation dataset (**Fig. S4D**). MRR is a commonly used metric to evaluate a method's ability to rank relevant terms at top positions⁶¹. Among the six methods, BIT achieved the highest MRR for both bulk and single-cell perturbation data. Additionally, BIT identified the most perturbed TRs within the top 10 and top 50 lists from the single-cell perturbations (**Fig. S4E**).”

1 Pierce, S. E., Granja, J. M. & Greenleaf, W. J. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat Commun* **12**, 2969 (2021). <https://doi.org/10.1038/s41467-021-23213-w>

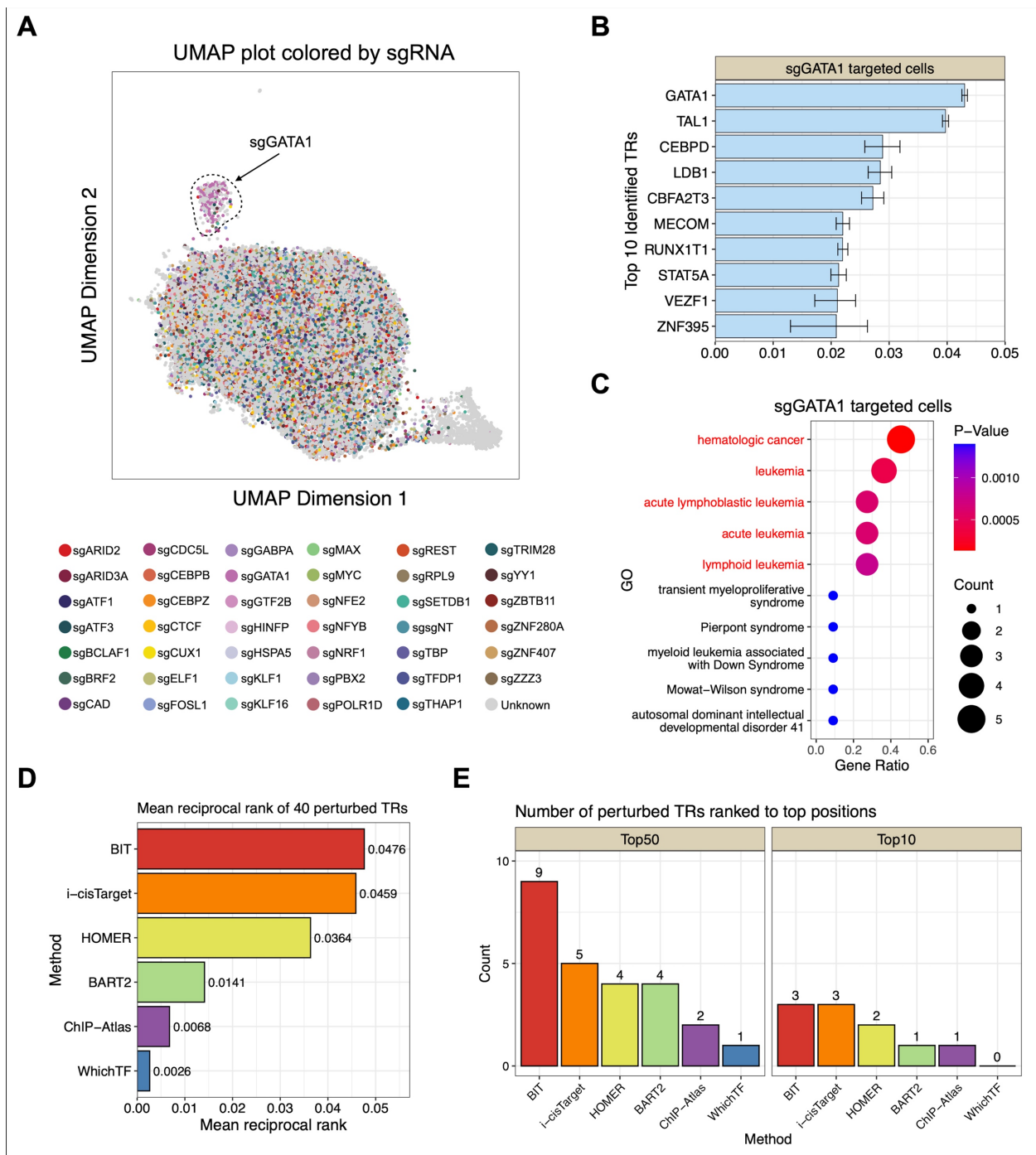


Figure S4. Validating BIT's performance on a single-cell TR perturbation screen dataset. A. UMAP projection of K562 cells, colored by sgRNA targeting status. **B.** Top 10 TRs identified by BIT in sgGATA1 targeted cells. **C.** Gene ontology enrichment results of top TRs identified by BIT in sgGATA1 targeted cells. **D.** Mean reciprocal rank (MRR) using 40 perturbed TRs to compare BIT and five state-of-the-art methods (BART, ChIP-Atlas, i-cisTarget, WhichTF, and HOMER). **E.** Number of perturbed TRs ranked within the top 10 or top 50 by the six methods.

7. Benchmarking results using other scATAC-seq methods (e.g. scBasset), should also be included to demonstrate the performance of BIT.

>>> Following your suggestion, we incorporated two additional scATAC-seq analysis tools, scBasset and ArchR, and compared their performance using CRISPR–Cas9-identified functionally essential TRs from liver cancer cell lines and enriched TRs in PBMCs verified by Human Protein Atlas. Our findings, presented in **Fig. 5** and **Fig. S6**, confirm that BIT consistently outperformed these three methods in ranking functionally essential TRs. Thank you!

Reviewer 2

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.

>>> We appreciate this reviewer' helpful comments to improve our manuscript. Thank you!

Reviewer 3

The authors propose a novel method for inferring Transcriptional Regulators (TRs) from genomic or epigenomic regions (e.g., peaks). This approach relies on:

1. A comprehensive reference library of TR-ChIP datasets from the GTRD, comprising over 10,140 human and 5,681 mouse datasets, which eliminates the need for motif-based predictions.
2. A hierarchical model that accounts for both within- and between-dataset heterogeneity, enhancing robustness and mitigating biases caused by TRs with disproportionately large dataset representation. This method reduces over-fitting while preserving critical insights by integrating information across multiple datasets.
3. Probabilistic characterization of TRs regulation. Unlike frequentist techniques, this approach provides credible intervals on the matching/versus informative regions mean, offering a more nuanced and reliable analysis.

The paper is well-written, and the authors present extensive proof-of-concept results, including simulations, perturbation datasets, cancer-related data and single-cell experiments, to validate their assumptions and model. The proposed model is formally correct and well-suited to its context, providing straightforward and easily interpretable estimates for both expert and non-expert researchers.

Additionally, they have made significant efforts to provide both a GitHub package and a user-friendly web-based platform, enabling users to apply the method to their specific regions of interest.

However, I have several suggestions regarding the following aspects.

>>> We are deeply grateful to the reviewer for the insightful comments and suggestions, which have greatly strengthened our manuscript and improved the accessibility of BIT. To streamline your evaluation, we have integrated all revisions directly within this response letter, including updated figures, newly added supplementary data, and key refinements.

Each concern has been thoroughly addressed in a point-by-point manner below, with detailed rationales for all modifications and direct references to revised text, figures, or analyses. In addition, we have updated the manual on readthedocs and updated the README file on GitHub to address coding concerns. The data on Zenodo has also been updated to reflect modifications in this revision.

1. Reproducibility.

In several parts of the text, critical details necessary for reproducing the results are missing. For example, the authors do not specify the software and packages used for the analysis (e.g., R—what version?), or the release versions of the datasets (e.g., GTRD [see 550: add the link, or the name of the reference dataset, the release versions, ...], TCGA [see 551: did the authors use the online platform? What is the release?]), the version/released of the other methods employed (see [694]: what is the version of each package?). In general, a more detailed description of the data processing steps and tools used for analysis would be beneficial.

>>> Following your suggestion, we added the version information of each package for the reproducibility in the main text. We also updated the release information of the GTRD, TCGA, HPA, and DepMap data used in this project, as well as the links to the datasets in the GitHub pages:

Page 31, Line 610 – Line 629: “**Bulk data:** In the TR perturbation-derived DAR analysis, raw ATAC-seq sequence reads were first cleaned using Trim-galore (v0.6.10). Read quality was checked by FastQC (v0.12.1). Paired-end reads were mapped to the human genome (hg38) or mouse genome (mm10) using bowtie2 (v2.5.2) applying parameters “bowtie2 --very-sensitive -X 2000”. SAMtools (v1.16.1) was used for the following steps: (i) convert SAM files to BAM files, (ii) remove unmapped reads and duplicates, (iii) generate index files for BAM files. Peaks were called using MACS2 (v2.2.9.1). BAM files, index files, and peak files were used for DAR analysis using

DiffBind (v3.10.1). Reference TR ChIP-seq datasets were retrieved from GTRD (v21.12), datasets with too low number peaks (<500) and individual peaks with low q-value (<10) were removed. Cancer-type-specific accessible regions were directly retrieved from TCGA.”

Single-cell data: The single-cell TR perturbation dataset was retrieved from the GEO archive. We obtained DARs for 40 perturbed TRs using the pipeline provided by the data authors (https://github.com/GreenleafLab/SpearATAC_MS_2021), by comparing cells treated with single guide RNA (sgRNA) targeting one specific TR against control cells treated with non-targeting single guide RNA (sgNT). 10K PBMCs and primary liver cancer scATAC-seq fragment files were retrieved from the 10XGenomics and GEO archive. The cell-type-specific accessible regions were generated from each dataset using SnapATAC2 (v2.6.0). When applicable, the cell type annotations were also retrieved and applied. For comparative benchmarking with established single-cell data analysis methods, in addition to SnapATAC2, we used ArchR (v1.0.2) and scBasset (<https://github.com/calico/scBasset>).”

Page 39, Line 775 – Line 777: “To compare with other state-of-the-art methods, the original implementation of BART (v2.0), HOMER (v4.11.1), and WhichTF (v0.2) were retrieved from <https://github.com/zanglab/bart2>, <http://homer.ucsd.edu/homer/>, and <https://bitbucket.org/bejerano/whichtf>.”

Page 39, Line 780 – Line 781: “We used the R package “clusterProfiler (v4.14.4)” to conduct gene ontology enrichment analysis.”

Page 40, Line 787 – Line 795:” Only public datasets were used in this study. The reference ChIP-seq datasets were retrieved from GTRD (v21.12, <http://gtrd.biouml.org:8888/downloads/current/>). HPA-verified TRs were retrieved from Human Protein Atlas (v24.0). TR perturbation experiment data were retrieved from GEO (GSE153237, GSE173416, GSE114102, GSE234331, and GSE168851). Cancer-type specific accessible regions were retrieved from TCGA (GDC 2019

v18.0, <https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>). CRISPR/Cas9 knockout screening data were retrieved from DepMap (<https://depmap.org/portal/>). scATAC-seq data were retrieved from 10XGenomics and GSE227265. We initialized a data repository on Zenodo (<https://zenodo.org/records/14231098>).”

Although the authors provide scripts to reproduce the figures presented in the paper, these rely on pre-processed tables that are loaded and plotted, rather than starting from raw data. I recommend that the authors include complete scripts to process the analysis from raw data—at least for some examples—if they used the R package or other tools. This would greatly enhance reproducibility and transparency.

>>> Thank you for pointing out the importance of providing scripts that process raw data for reproducibility. In response, we have uploaded the complete scripts used to process raw data into final results. To enhance transparency, we have added four example scripts on GitHub: (1) the knockout of Runx1 in mice thymocytes, (2) the newly-added single-cell TR perturbation dataset, (3) the cancer-type-specific accessible region dataset, and (4) the PBMC single-cell ATAC-seq dataset. These scripts demonstrate how to generate the final plots from raw data, ensuring greater reproducibility of our study.

2. State-of-the-art methods.

Since the authors propose a new computational tool, it would be highly beneficial to expand the discussion of state-of-the-art methods in the introduction [lines 75–88]. This could include a detailed exploration of the technical aspects of other models and/or the reference libraries they rely on. Adding a comparison table that highlights the strengths and weaknesses of existing methods—or classes of methods (e.g., motif-based, region-based, distance-based)—would provide valuable context.

Such a table would allow readers to better understand the unique features of the proposed method, such as its ability to integrate multiple datasets or to identify TRs that are not characterized by specific motifs [lines 384–385]. This added context would also help readers appreciate the results of the study, particularly in comparison to other approaches.

>>> Thank you for your valuable suggestion. We have added a table summarizing key differences between BIT and existing methods, all of which accept query region sets as input. This new **Table 1** highlights differences in reference libraries, technical aspects, and implementation, providing a clearer context for understanding BIT’s unique features.

Method	Implementation	Reference Library			Statistical Aspects			
		Type	TR Binding Activity	Data Source	Framework	Technique	Uncertainty Quantification	Information Pooling
BIT	R/Web portal	ChIP-seq-based	ChIP-seq binding sites represent the most accurate in vivo TR binding profiles.	GTRD	Bayesian	Bayesian hierarchical modeling	Bayesian framework inherently quantifies uncertainty.	Allow information sharing across TRs, mitigating overfitting.
BART2	Python/Web portal	ChIP-seq-based		CistromeDB	Frequentist	Wilcoxon rank sum test	Frequentist methods overlook variability in experiment replicability, potentially causing misinterpretation.	Analyzes each TR or dataset independently, lacking mechanisms for information sharing or overfitting mitigation.
ChIP-Atlas	Web portal	ChIP-seq-based		NCBI/DDBJ/ENA	Frequentist	Fisher exact test		
HOMER	Unix-style CLI	Motif-based	Motifs are computationally predicted, complicating differentiation of similar TRs.	Literature	Frequentist	Binomial test		
I-cisTarget	Web portal	Motif-based		ENCODE	Frequentist	Recovery analysis		
WhichTF	Python	Motif-based		JASPAR/TRANSFAC	Frequentist	Hypergeometric test		

Table 1. Comparison of state-of-the-art methods accepting query region sets.

Moreover, in comparing Figure 3F and Figure 4F, it would be insightful to report the top 10/50/100 ranked TRs for each method and compare them in terms of the similarity of common elements. This would reveal whether there are patterns among methods with similar models (or classes of models, such as motif-based or region-based approaches).

>>> We included Supplementary **Fig. S9** to evaluate Jaccard similarity indices of the top 10/50/100 ranked TRs identified by different methods and deposited the direct output files from these analyses on Zenodo under the "Outputs" folder. As illustrated in **Fig. S9**, ChIP-seq-based methods (BIT, BART2, and ChIP-Atlas) tend to exhibit relatively higher Jaccard similarity than the motif-based methods. This observation can be attributed, in part, to the overlapping reference databases used by these methods, as they all rely on publicly available ChIP-seq datasets. In

contrast, methods such as HOMER and i-cisTarget, which are primarily motif-based, and WhichTF, which incorporates motifs indirectly, demonstrated lower similarity due to their distinct preprocessing pipelines and enrichment analysis strategies.

We revised the manuscript to include a discussion of the similarity of top-ranked TRs across methods.

Page 29, Line 575 – Line 585

“Ranking similarity across methods

To assess the similarity of top-ranked TRs across methods, we computed Jaccard similarity indices of the top 10, 50, and 100 TRs identified by BIT, BART2, ChIP-Atlas, i-cisTarget, HOMER, and WhichTF in our studies of TR perturbation-derived DARs and cancer-type-specific accessible regions. Overall, Jaccard similarity among these approaches is generally low (**Fig. S9**). However, ChIP-seq-based methods (BIT, BART2, and ChIP-Atlas) exhibit relatively higher similarity compared to motif-based methods, likely due to their reliance on overlapping reference databases derived from publicly available TR ChIP-seq datasets. Since TR ChIP-seq provides the most accurate information for TR binding, this shared foundation contributes to the observed similarities. It is worth mentioning that the highest similarity indices often occur between BIT and another approach, suggesting BIT's strong ability to capture true signals.”

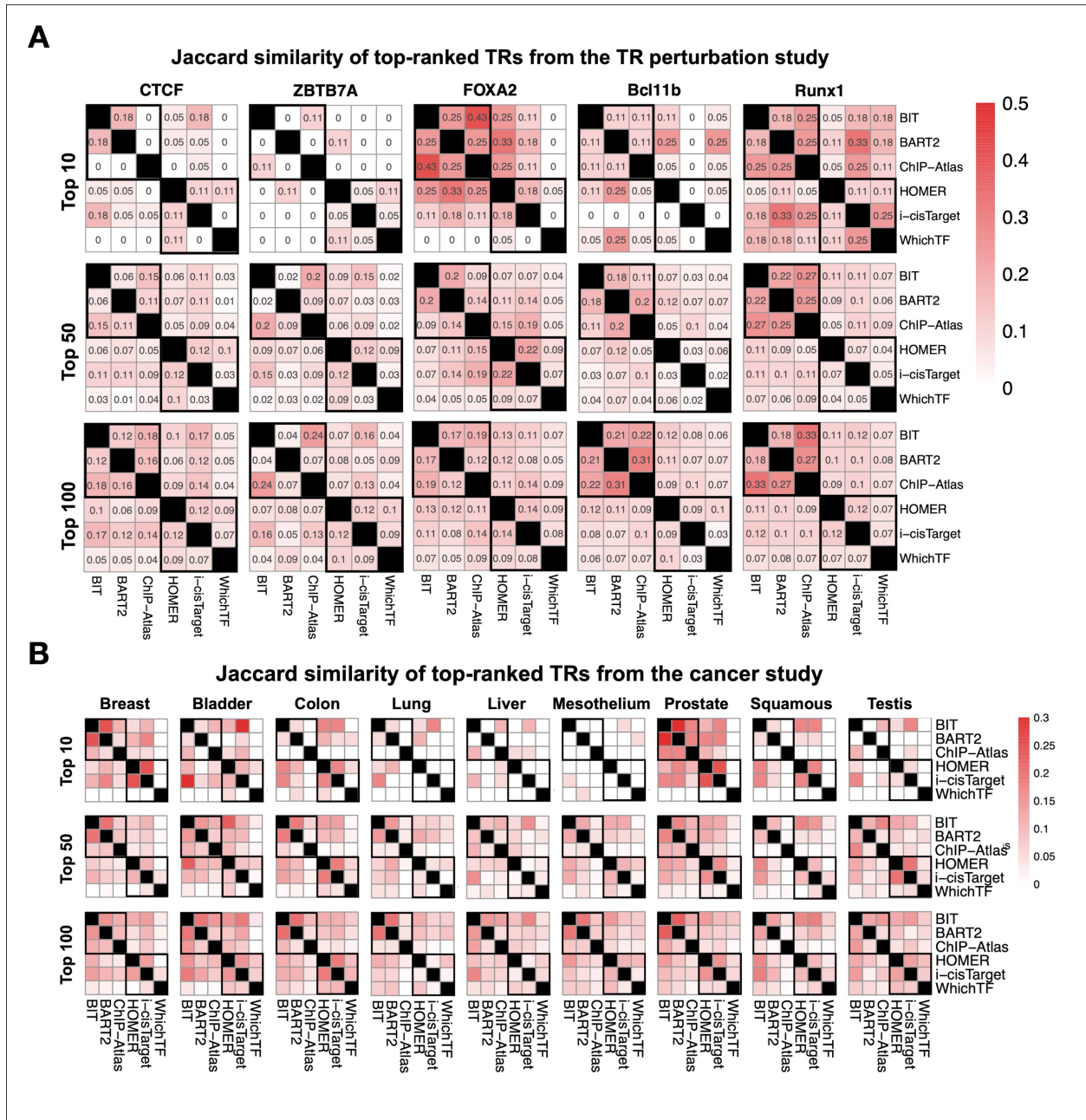


Figure S9. Jaccard similarity indices of top 10/50/100 ranked TRs identified by six methods (BIT, BART2, ChIP-Atlas, i-cisTarget, HOMER, and WhichTF). A. Heatmap of Jaccard similarity in the TR perturbation study. B. Heatmap of Jaccard similarity in the cancer study.

3. Fixed non-overlapping windows.

The authors use 1,000 bp fixed windows to bin the genome, but it is unclear why this specific window size was chosen. Given that transcription factor binding sites can be much smaller (~50 bp, depending on the TF), the impact of this choice should be addressed. Additionally, since the package allows users to select different window sizes (e.g., 200, 500, or 1,000 bp), it would be helpful to evaluate and discuss the implications of this flexibility on the results.

>>> Following your suggestion, we evaluated the impact of different window sizes (bin widths) using five bulk TR perturbation datasets, as shown in **Fig. S11**. The results show that our default 1kb window size performed best. However, in all five cases, BIT consistently ranked the perturbed TRs within the top 10, regardless of the window size.

We mention that a 1 kb window is a reasonable and practical choice. While individual TR binding sites are typically ~50 bps, the reference database was constructed using peak sets identified by widely used methods such as MACS, which generally produce peaks ranging from 200 to 500 bps. In addition, the input region sets in this study, primarily derived from chromatin accessibility profiling methods, also yield peaks within the 200-500 bps range. Using a 1 kb window ensures that both TR binding sites and accessible chromatin regions are adequately captured.

We have incorporated a discussion on window size in the main text.

Page 32, Line 643 – Line 646: “To assess the impact of bin width on BIT’s performance, we analyzed bulk TR perturbation datasets using bin widths of 200 bps and 500 bps. The results indicate that our default 1000 bps setting performed well, with no significant effect from varying bin width (**Fig. S11**).”

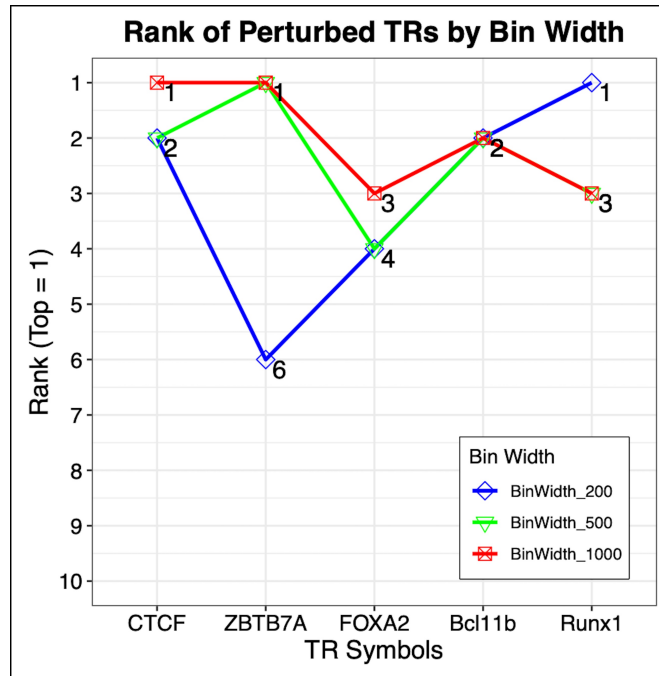


Figure S11. Rankings of five perturbed TRs determined by BIT using different bin widths (200, 500, 1000 bps) for processing both the reference ChIP-seq datasets and input region sets from the bulk TR perturbation data.

4. Computational time.

The authors briefly mention the issue of computational time in the discussion, noting that the web tool warns users that analyses can take approximately 50 minutes. However, it would be helpful to provide actual runtime measurements, particularly for the real-data examples, and to compare these with runtimes of other methods. This would give readers a clearer understanding of the computational efficiency of the proposed approach.

>>> As suggested, we evaluated the computational efficiency of BIT and compared it with other state-of-the-art methods. As shown in **Fig. S10A**, two primary factors influence BIT's runtime: the length of the Gibbs sampling chain and the reference database used. A longer chain (more iterations) increases computational time, as each iteration involves sequential sampling of parameters from their conditional posterior distributions. Meanwhile, a larger reference database requires sampling more parameters, leading to longer runtime. Consequently, BIT runs slower

when using the human reference library, which contains more TRs and datasets than the mouse reference library.

The default iteration count in BIT is set to 5000 to ensure MCMC convergence. However, as shown **Fig. S10B**, convergence was reached around 1000 iterations across all five datasets. To ensure a fair comparison with other methods, we measured BIT's runtime using 1000 iterations. For benchmarking, we ran BART, WhichTF, and HOMER in a local environment while for i-cisTarget and ChIP-Atlas, both of which lack local implementations, we recorded the elapsed time from submission to result retrieval. As illustrated in **Fig. S10C**, BIT's speed is slightly slower than BART but significantly faster than the other four methods.

Page 29, Line 587 – Line 595: “BIT implements a Gibbs sampling scheme to draw posterior samples for thousands of parameters defined within the model. As shown in **Fig. S10A**, its computational efficiency depends on two main factors: (1) the number of parameters and (2) the length of the MCMC chain. The number of parameters is determined by the available TRs and associated datasets in each reference library.

We compared the computational efficiency of BIT with the other methods using bulk TR perturbation data. For BIT, we measured the time to convergence (**Fig. S10B**), while for other methods, we recorded the time to obtain results. BIT's speed is slightly slower than BART but significantly faster than the other four methods (**Fig. S10C**).”

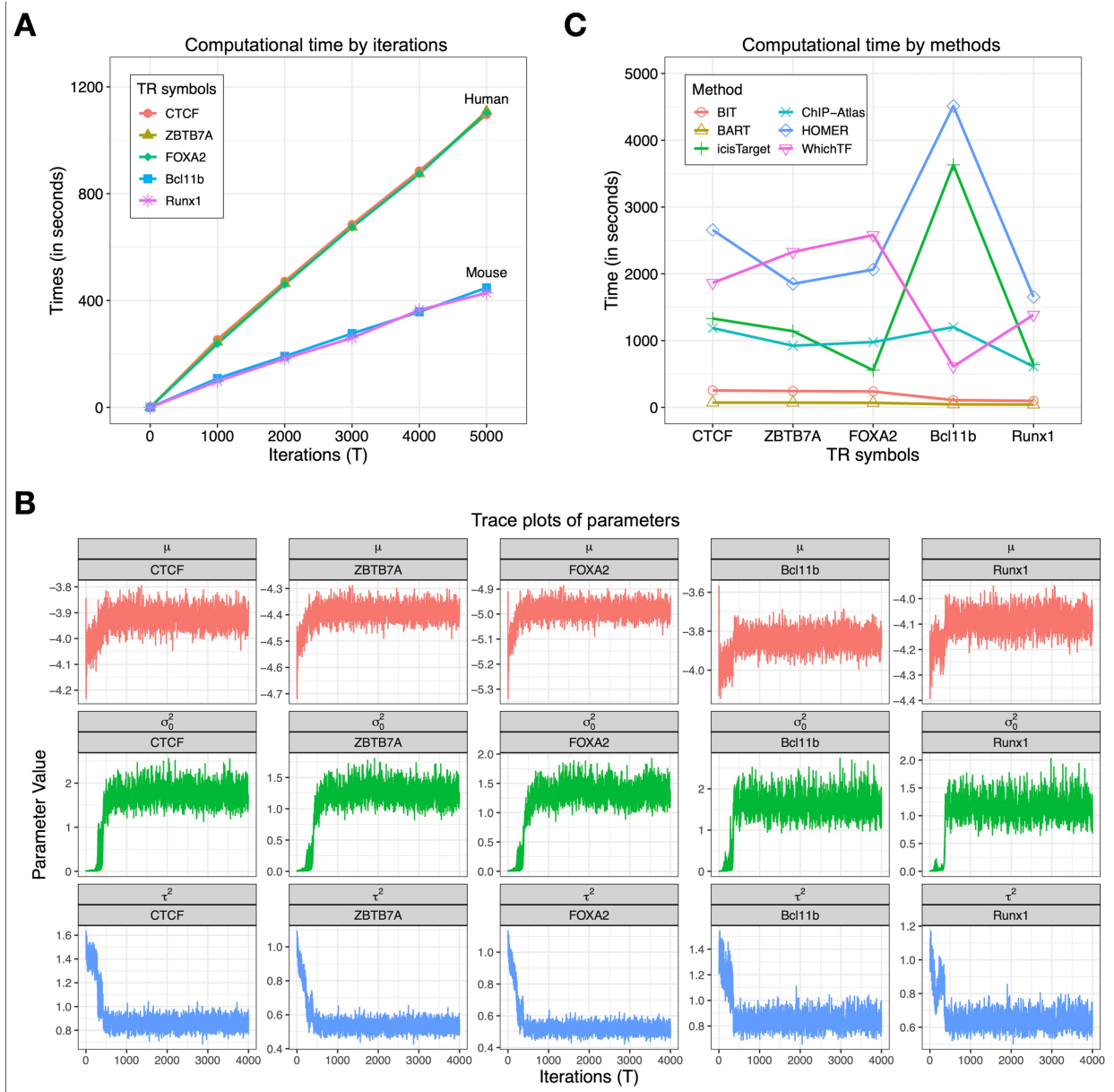


Figure S10. Comparative analysis of computational efficiency between BIT and state-of-the-art methods using bulk TR perturbation data. **A.** BIT's runtime across three human (CTCF, ZBTB7A, and FOXA2) and two mouse (Bcl11b and Runx1) TR perturbation datasets with different numbers of iterations for its Gibbs sampler. **B.** Trace plots for BIT's global parameters (μ , τ^2 , σ_0^2) across the five datasets, illustrating convergence behaviors. **C.** Runtime comparison for BIT, BART2, ChIP-Atlas, HOMER, *i-cisTarget*, and WhichTF across the five datasets.

5. Credible intervals.

Although the authors emphasize the importance of credible intervals, these are not used in any of the provided results (e.g., Figure 3A). I recommend adding whiskers to the main figures. Additionally, in sections such as [87] and [110] (discussing "Uncertainty estimation" and "credible intervals"), the distinction between frequentist and Bayesian approaches may be challenging for non-expert readers to grasp. Rephrasing these sections to make them clearer and more accessible would improve the manuscript.

>>> We added the credible intervals as whiskers to **Fig. 3A** as well as similar plots throughout the manuscript. We also extended the discussion of the benefits of Bayesian hierarchical modeling and emphasized how it differs from the frequentist-based approaches.

Page 5, Line 85 – Line 90: "...(iii) Most existing methods use frequentist approaches for measuring TR statistical significance, lacking coherent probabilistic interpretation or uncertainty estimation. For example, ChIP-Atlas³² ranks TRs using p-values from Fisher's exact tests, which assess significance under a fixed-parameter null hypothesis (e.g., odds ratio = 1) but fail to capture epistemic uncertainty, especially when experiments cannot be identically replicated. Given the dynamic nature of TR binding activity, the ability to quantify uncertainty is not just beneficial but essential."

Page 6, Line 112 – Line 115: "...(i) It enables straightforward quantification of estimation uncertainty by deriving 95% credible intervals from posterior sampling, representing the range within which the true value of an unknown quantity (e.g., a parameter of interest or a function of parameters) likely falls."

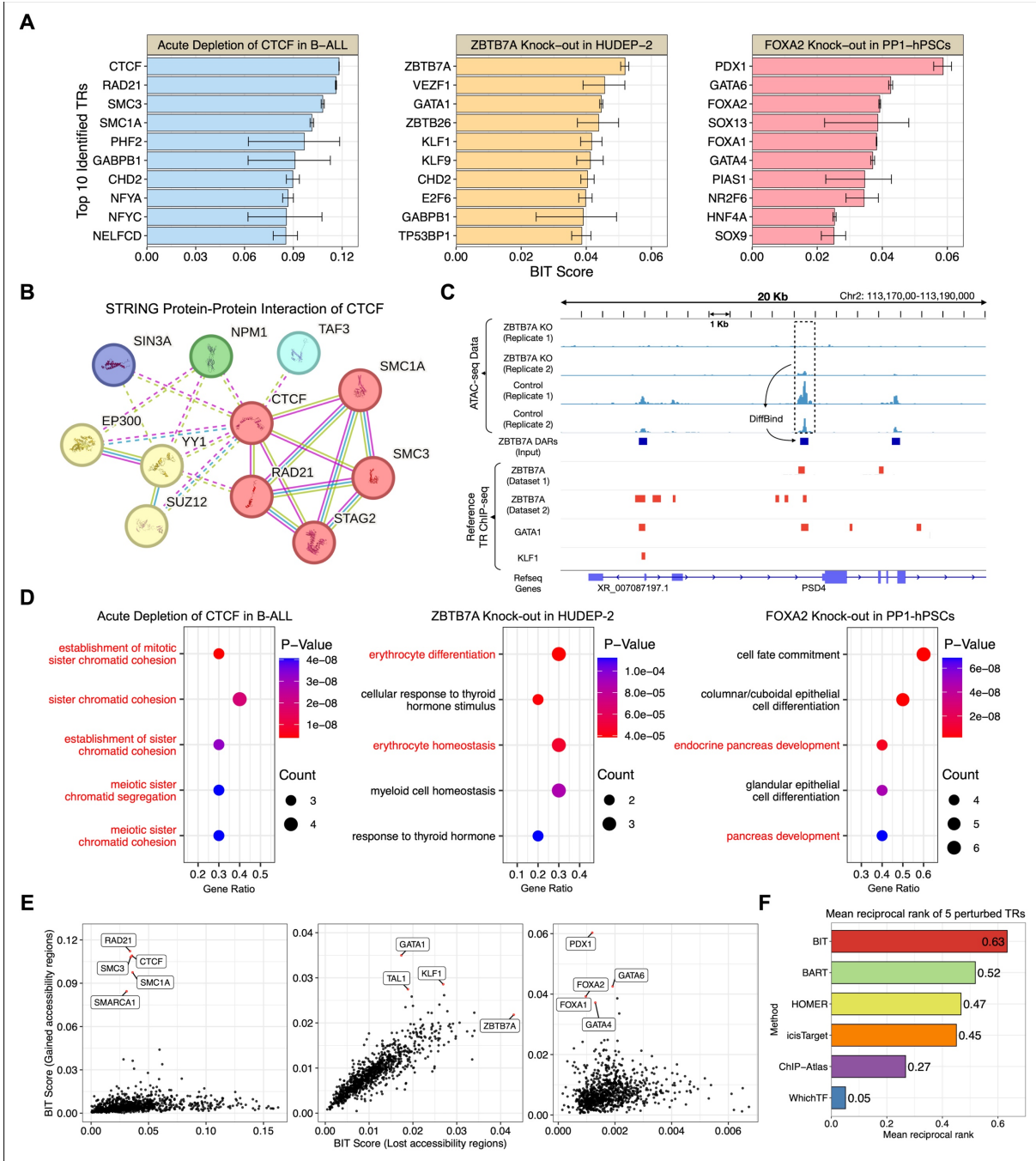


Figure 3. BIT can identify transcriptional regulators (TRs) from differentially accessible regions (DARs). **A.** The top 10 transcriptional regulators (TRs) identified by BIT using DARs from each of the three TR perturbation experiments. **B.** Protein-protein interaction of CTCF derived from STRING. **C.** BIT-identified TRs have binding patterns consistent with the differentially accessible regions observed in TR perturbation experiments. **D.** The top 5 terms ranked by gene ontology enrichment analysis of top TRs in the output list of three cases. **E.** The scatterplot of TRs by using BIT scores from gain accessible regions versus loss accessible regions. Each dot represents a TR. **F.** Mean reciprocal rank (MRR) using five perturbed TRs to compare BIT and five state-of-the-art methods (BART, ChIP-Atlas, i-cisTarget, WhichTF, and HOMER).

General clarifications in the manuscript:

[206] “Parameters in Results section”: Are the authors referring to Figure 6A–B in the Methods section?

>>> Yes. Thank you for pointing out the typo. Fixed now.

[206 and Figure 6A] Translating the parameter μ into terms of the proportion of matching versus over-informative elements would provide a more intuitive interpretation, helping to convey what a “real scenario” entails.

>>> We added the following sentence:

Page 10, Line 206 – Line 208: “...Thus, μ represents the global mean (logit-scale) proportion of matching across all TRs while τ^2 captures the between-TR variability.”

Thank you!

[247, 249] Parameters such as tau (controlling between-TR heterogeneity) and sigma (noise levels) are critical concepts that would be nice to be highlighter in the method section.

>>> As suggested, we have revised the manuscript to highlight the parameters τ^2 , σ_i^2 , and σ_0^2 :

Page 33, Line 678 – Line 681: “...We emphasize that τ^2 controls the TR heterogeneity, while each σ_i^2 governs the noise levels of the datasets associated with the same TR. Additionally, the σ_0^2 accounts for the noise levels for TRs with only a single dataset. These parameters are critical as they balance the estimates between TR-specific data and global population mean, reducing overfitting through regularization.”

[600] It would be valuable to discuss how the depth of each reference library influences the parameters of the distributions presented in Figure 6. For instance, does the variability primarily

stem from the number of datasets included, or does sequencing depth/number of peaks also play a significant role in shaping the results?

>>> We used GTRD as the main source for reference TR ChIP-seq data. The number of datasets and TRs covered depends on the data stored in the database. After performing the quality control to filter out datasets with too few peaks or low-quality peaks, we obtained 10,140 human TR ChIP-seq datasets covering 961 TRs and 5,681 mouse TR ChIP-seq datasets covering 652 TRs. The difference in the number of datasets and TRs covered affects the number of parameters in the model and computational efficiency, as shown in Figure S8A.

Following your suggestion, we also examined the influence of the following factors: (1) the number of datasets associated with each TR, and (2) the number of peaks in each dataset. Regarding the sequencing depth, we did not evaluate it as we lacked access to the raw sequence data. However, GTRD's manuscript² implies that their peak-calling process removes effects related to sequencing depth.

We first selected all 487 TRs with at least 3 datasets. To examine the effect of the number of datasets associated with each TR, we categorized these TRs into four groups of approximately equal size based on the quantiles of their dataset numbers. We then computed the log odds of matching for each dataset of each TR. We next computed the standard deviation of log odds associated with each TR, resulting in 487 standard deviation values categorized by the four groups. The results were visualized as boxplots for the four groups, shown in **Fig. S12A**. No significant differences were observed among the groups, indicating that the number of datasets associated with TRs appears not to significantly affect the results.

Next, we examined the number of peaks within each dataset. We re-categorized the 487 TRs into four groups of roughly equal size based on the average number of peaks in their datasets and replicated the analysis. The results in **Fig. S12B** show no significant patterns, again.

Page 33, Line 660 – Line 663: “It is worth mentioning that we also examined how factors, including the number of reference datasets available for a TR and number of peaks in a reference dataset, affect the within-TR variability of the log odds. No significant patterns were observed (Fig. S12).”

2 Kolmykov, S. et al. GTRD: an integrated view of transcription regulation. Nucleic Acids Res 49, D104-d111 (2021). <https://doi.org/10.1093/nar/gkaa1057>

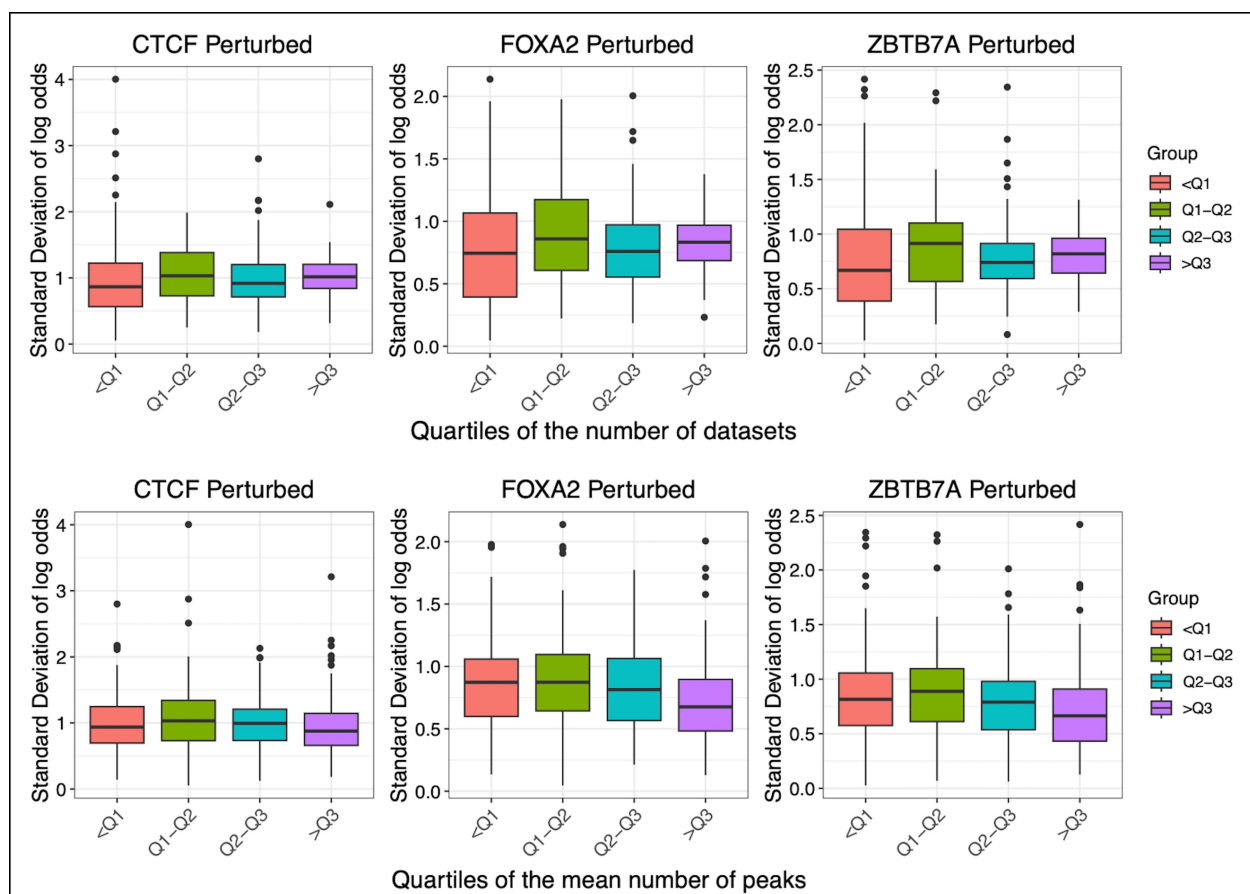


Figure S12. Assessing factors influencing variability by using ATAC-seq data from three bulk TR perturbation experiments. A. Variability in log odds of "matching" bins among "informative" bins for 487 TRs, stratified by the number of available ChIP-seq datasets per TR using the 1st, 2nd, 3rd quantiles. **B.** Variability in log odds of "matching" bins among "informative" bins for 487 TRs, stratified by the mean number of peaks per TR using the 1st, 2nd, 3rd quantiles.

[624] Despite the reference that gives the details of possible parameters, I would like to know how they decides for the Upper and Lower limits of the non-informative model.

>>> We determined the upper and lower limits of the non-informative prior distribution for the global mean μ through exploratory analysis. We selected a conservative range of -10 to 10, which corresponds to a proportion of matching cases among informative cases between 0.00005 and 0.99995. Since this defines the global mean, this range is sufficiently broad to encompass all plausible scenarios. We revised the manuscript to include these details:

Page 35, Line 705 – Line 709.” The global location parameter μ and all variance parameters including τ^2 , σ_0^2 , and $\{\sigma_i^2\}_{i=1}^M$ are assumed *a priori* independent. For μ , we consider the non-informative flat prior $\mu \sim \text{uniform}(L_\mu, U_\mu)$. The default values are $L_\mu = -10$ and $U_\mu = 10$, which correspond to a proportion of matching cases among informative cases between 0.00005 and 0.99995, providing a sufficiently wide coverage for all plausible values of μ suggested by data.”

[340]: Please report the rank of the other methods. Is there any difference between Mouse/Human?

>>> We have now included this into Supplementary **Tab. S1**, which presents the rankings of the perturbed TRs as reported by BIT and other methods. Note that methods such as i-cisTarget and WhichTF only provide rankings for TRs exceeding a certain threshold, which prevents the generation of a complete ranking list for all available TRs. For these cases, missing values were assigned a rank of $+\infty$ when calculating the mean reciprocal rank (MRR). To ensure transparency and reproducibility, we have uploaded the output files of the five methods that provide complete rankings for all candidate TRs to Zenodo.

We did not observe any notable differences between mouse and human results. Both genomes exhibited cases of strong and weak performance across different methods. However, BIT consistently demonstrated robust performance across all five cases examined.

[Figure 1A]: The vignette illustrating the first step is somewhat misleading. The "000" sequence

visually corresponds to the actual input region, which should instead be represented as "111."

Additionally, the color of the nucleosomes matches the "1" in the vector, which could cause confusion. I recommend revising the initial representation to align with the input vector for consistency and clarity.

>>> We apologize for the misleading representation in **Fig. 1A**, we have revised the vignette to ensure that the chromatin accessibility status aligns with the sequence representation. We also changed details such as the nucleosome color to improve clarity and avoid confusion.

[Figure Redacted]

Figure 1. Overview of BIT's framework. **A.** Transcription regulators (TRs) such as transcription factors, cofactors, and chromatin modulators mostly bind to accessible chromatin regions to regulate gene transcription. NGS techniques such as TR ChIP-seq and ATAC-seq can acquire information on TR binding sites and accessible regions, respectively. Input regions and collected TR ChIP-seq peaks are transformed into binary vectors, where the proportion of the "matching" cases among all "informative" cases serves as the response variable. **B.** BIT is a Bayesian hierarchical model, combined with Pólya-Gamma data augmentation, to integrate information from multiple TRs and multiple ChIP-seq datasets. **C.** A Gibbs sampler is used to draw posterior samples and estimate key model parameters. **D.** TR-level BIT scores are generated and ranked to output a TR ranking list.

[Figure 3C]: It would be helpful to depict the input regions (e.g., using rectangles) to visualize the overlap between the input and reference data more clearly, alongside the track. This enhancement would make the comparison more intuitive for readers.

>>> We appreciate the reviewer's suggestion. To enhance clarity, we have added the input regions (i.e., DARs) derived using DiffBind from ATAC-seq data to Figure 3C, illustrating how the input and reference data align more intuitively.

[Figure 3E]: The meaning of the point colors is unclear. Adding a legend or explanation would help clarify this aspect.

>>> Thank you for your feedback. We originally used a red-to-blue color scale to indicate whether a TR was ranked higher in gained or lost accessibility regions. In this revision, we removed the color scale to simplify the plot, as this distinction is already clear from the position of each TR in the two-dimensional space.

Github package.

I found the documentation on <https://bitbayesian-identification-of-transcriptional-regulators.readthedocs.io/en/latest/index.html> to be clear and comprehensive. However, the GitHub documentation contains several missing details. For instance, since the Zenodo repository is required to run BIT, I recommend including a direct reference to it in the GitHub tutorial. Similarly, the internal help files in R (e.g., for the BIT function) lack details about parameters such as burnin. Harmonizing the documentation across these platforms would greatly enhance usability.

>>> We thank this reviewer for the feedback on our GitHub package and documentation. As suggested, we have added a direct reference to the updated Zenodo repository in the GitHub documentation. We also revised the help files within the R package to include details of default parameters, such as the number of MCMC iterations, burn-in period, random seed, and other

relevant settings. Additionally, we updated our Read the Docs manual to reflect these modifications.

- Since the method relies on a Gibbs sampler, it would be beneficial to allow users to set a random seed for reproducibility.

>>> We have modified the main function of the BIT package to allow users to set a random seed for reproducibility. By default, the seed is set to 42. Thank you!

- The function appears to use all available CPU cores, by default. Providing recommendations for hardware requirements and enabling users to configure the number of cores would improve user experience and resource management.

>>> We have added detailed hardware requirements for running the BIT package in a local environment to the GitHub page. Since most of our testing was conducted on a high-performance computing (HPC) system, we have also included information about the HPC environment.

Our function does not explicitly specify the number of CPU cores, as Bayesian computation using MCMC with Gibbs sampling is inherently sequential, meaning that multiple CPU cores would not significantly improve computational speed.

Thank you for your feedback!

Online tool

- File upload: Providing a downloadable example file as an alternative to the “click and run” option would help users better understand the required input format, even though the supported standards are well known. Additionally, the main text mentions the use of submit information when available—this is a critical detail that should be explicitly communicated to users within the tool.

>>> Thank you for your valuable suggestion regarding our online portal. We have updated the

portal's homepage to include a downloadable example file to help users better understand the required input format. Furthermore, we have explicitly noted on the main page that if a column containing summit information is present in the uploaded file, the portal will use these summits instead of the region midpoint during the conversion process.

- Retrieve key: Allow users to download a text file containing the automatically generated key along with a summary of the input parameters, so they can save and retrieve it later. Additionally, consider offering users the option to receive an email notification once the analysis is complete, as it can take more than an hour.

>>> We appreciate your suggestions for improving the retrieval process. We have added an option that allows users to download a text file containing the automatically generated key and a summary of the input parameters. Additionally, we have implemented an automated email notification feature to alert users when the analysis is complete, which is especially helpful for longer runs exceeding one hour.

- Result download issue: I attempted to download the results using the following search keys, but it did not work: I) Search key: 00fe1dfd5098f8e89191ea7b0bba30ab, ii) Search key: 01520f69b0316ac16d8e4e8276447d33. Namely, the format of the file was not readable.

>>> We sincerely apologize for the inconvenience. The issue causing the downloaded file to be mislabeled as a .csv despite having an Excel (.xlsx) format has been resolved. If you have any additional concerns regarding the web portal or encounter any errors, please report them on our GitHub page. We appreciate your understanding and continued support.