nature portfolio

Corresponding author(s): Xinlei Wang, Lin Xu

Last updated by author(s): Apr 21, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Not applicable
Data analysis	We have used following softwares for data analysis: Trim-Galore(https://github.com/FelixKrueger/TrimGalore), FastQC(v0.12.1), Bowtie2(v2.5.2), SAMtools(v1.16.1), MACS2(v2.2.9.1), DiffBind(v3.10.1), SnapATAC2(v2.6.0), ArchR (v1.0.2), scBasset (https://github.com/calico/scBasset), clusterProfiler(v4.14.4), BART(https://github.com/zanglab/bart2), ChIP-Atlas(https://chip-atlas.org), HOMER(v4.11), icisTarget(https://gbiomed.kuleuven.be/apps/lcb/i-cisTarget/), WhichTF(https://bitbucket.org/bejerano/whichtf/src/master/), BIT(https://github.com/ZeyuL01/BIT), KMPlotter(https://kmplot.com/analysis/)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Only public datasets were used in this study. The reference ChIP-seq datasets are available at GTRD (v21.12) [http://gtrd.biouml.org:8888/downloads/current/]. HPA-verified TRs lists are available at Human Protein Atlas (v24.0) [https://www.proteinatlas.org/]. The TR perturbation experiment data are available at GEO under the accession code GSE15323740 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE153237], GSE17341641 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE173416], GSE11410242 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114102], GSE23433159 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234331], and GSE16885160 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168851]. Cancer-type specific accessible regions are available at TCGA63 (GDC 2019 v18.0) [https://gdc.cancer.gov/about-data/publications/ATACseq-AWG]. CRISPR/Cas9 knockout screening data are available at DepMap64 [https://depmap.org/portal/]. PBMCs scATAC-seq data is available at 10XGenomics [https://support.10xgenomics.com/single-cell-multiome-atac-gex/datasets/1.0.0/pbmc_granulocyte_sorted_10k]. Liver cancer scATAC-seq data is available at GEO under accession code GSE22726589 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE227265]. Processed data is available at Zenodo [https://zenodo.org/records/14231098]. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The ATAC-seq and scATAC-seq datasets were decided to include a variety of experimental conditions, tissues and data types, resulting in ~20 datasets covering three different case studies with each study has more than 3 datasets from different conditions, tissues, or cell types. This broad inclusion was intended to enable an objective evaluation of BIT performance. When evaluating the individual computational methods, we did not generate results for multiple runs for the same dataset, as results generated by these methods show minimal or no variation. The BIT analysis of epigenomics-based region sets from ATAC-seq or scATAC-seq is about identifying the critical transcription regulators, which does not involve the comparison across samples. Therefore, the sample size of one for each ATAC-seq or scATAC-seq data is sufficient. No statistical method was used to predetermine sample size.
Data exclusions	Data exclusion criteria were established prior to analysis to ensure the integrity and quality of results. Datasets were excluded if they exhibited missing values beyond an acceptable threshold, significant outliers that could not be justified biologically or technically, or if they failed to meet predefined quality control benchmarks. Data exclusions were documented in method section.
Replication	When benchmarking individual computational methods, we did not report data from multiple runs as these methods generate results with minimal or no variation. For the proposed method all conclusions reported in this study are reproducible across at least three independent runs with different random seeds.
Randomization	In this study, there was no allocation of participants/samples/organisms into experimental groups. Instead, all computational methods were

Randomization	benchmarked uniformly across all datasets. Consequently, randomization was not a relevant consideration for our experimental design. The experiments were not randomized.
Blinding	Given that all computational methods in our study were benchmarked uniformly across all datasets within each case, blinding was neither necessary nor applicable to our experimental design. The Investigators were not blinded to allocation during experiments and outcome assessment

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable